METAGENOMIC SEQUENCING DIRECTLY FROM BLOOD CULTURE AS A TOOL FOR CLINICAL DIAGNOSIS



Kumeren Nadaraj Govender

Linacre and Balliol College

Nuffield Department of Medicine

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity 2023

SUMMARY

Metagenomic sequencing directly from blood culture as a tool for clinical diagnosis

Dr. Kumeren Govender, Linacre/Balliol College, Nuffield Department of Medicine

A thesis submitted for the degree of Doctor of Philosophy, Trinity 2023.

Metagenomic sequencing is heralded as a transformative tool for diagnostic clinical microbiology, offering expedited species detection and antimicrobial resistance (AMR) prediction. This thesis introduces a real-time sequencing pipeline direct-from-blood cultures, aiming to replace traditional culture-based diagnostics and antimicrobial susceptibility testing (AST) for suspected bloodstream infections.

I assessed DNA extraction kits and human DNA depletion methods for optimal direct-from-blood culture sequencing. To address potential sequencing errors and taxonomic misclassifications, I systematically benchmarked prevalent classification tools, databases, and sequencing platforms using a combination of simulated and real-world datasets. I processed 271 liquid blood culture samples (211 culture-positive and 62 culture-negative) in real-time, and applied a threshold-based and machine learning approach for species classification.

For species, the initial diagnostic sensitivity was 97% (95 CI: 94-99%, 210/216) and specificity of 94% (90-97%, 193/205) in culture positive samples. After adjusting for plausible false positive samples, the pipeline had a sensitivity of 100% (98-100%, 229/229) and specificity of 100% (98-100%, 205/205). Additionally, 13 polymicrobial and 5

monomicrobial infections that were overlooked by standard routine procedures were identified. In the first ten minutes of sequencing, using re-trained filter thresholds, the sensitivity and specificity was 98% (95-99%, 224/229) and 100% (98-100%, 205/205) respectively, taking a total time of 3hrs30min to species result, nearly half the time (i.e., 10hrs33min) taken by the routine microbiology laboratory. Specificity in culture negative samples, after accounting for a single plausible infection detected, was 100% (94-100%, 61/61) across all timepoints.

For AMR prediction, ResFinder after 24 hours, using raw sequencing reads, had the highest overall sensitivity and specificity of 91% (87-94%, 204/224 drug-organism combinations) and 90% (88-91%, 923/1029) respectively across the top 10 clinically relevant species, while raw reads after two hours of sequencing resulted in a performance of 88% (82-92%, 196/224) and 93% (91-94%, 953/1029), respectively. Notably, this approach allowed for comprehensive AMR detection within 5hrs30min from positive blood culture, outpacing conventional microbiology methods by 22 hours for comparable samples. This study indicates that clinical metagenomics may be integrated into routine diagnostics, allowing for quicker, unbiased pathogen detection, including unculturable and overlooked polymicrobial organisms. This method not only promises prompt antimicrobial resistance predictions but also could bolster patient outcomes. Moreover, it introduces the possibility of bacterial strain typing, vital for infection prevention and oversight.

TABLE OF CONTENTS

1 PREFACE XI

1.1 ACKNOWLEDGEMENTS XI

1.2 FUNDING FOR ALL CHAPTERS XIV

1.3 ETHICS STATEMENT FOR ALL CHAPTERS XIV

1.4 DECLARATION AND ATTRIBUTIONS XIV

1.4.1 Chapter 3 attributions xiv

1.4.2 Chapter 4 attributions xv

1.4.3 Chapter 5 attributions xv

1.4.4 Chapter 6 attributions xv

1.4.5 Chapter 7 attributions xv

1.5 PUBLICATIONS ARISING XVI

1.5.1 Directly from this thesis xvi

1.5.2 Other publications during the course of the thesis xvi

1.6 LIST OF TABLES XVII

1.7 LIST OF FIGURES XXI

1.8 LIST OF ABBREVIATIONS AND ACRONYMS XXVII

1.9 LIST OF APPENDICES XXIX

2 BACKGROUND 1

2.1 INTRODUCTION 1

2.2 THE TYPICAL DIAGNOSTIC WORKFLOW FOR BLOOD CULTURES IN ROUTINE MICROBIOLOGY LABORATORIES 4

2.2.1 Sampling *5*

2.2.2 Specimen reception *5*

2.2.3 Laboratory techniques for blood culture *5*

2.2.4 Isolation and identification *7*

2.2.5 Antimicrobial Susceptibility Testing (AST) *9*

2.3 ADVANCES IN MOLECULAR APPROACHES FOR PATHOGEN DETECTION AND ANTIBIOTIC RESISTANCE 14

2.3.1 Evolution and Impact of Polymerase Chain Reaction (PCR) Techniques *15*

2.3.2 Variants and Innovations in PCR Techniques *16*

2.3.3 The Rise of Isothermal Amplification Methods in Rapid Diagnosis *17*

2.3.4 Application of Syndromic Panels in Clinical Diagnosis: The Case of BioFire *19*

2.3.5 Emergence of Other Technologies: The Case of DNA Microarrays and Beyond *21*

2.3.6 Conclusion *22*

2.4 METAGENOMICS APPROACHES 23

2.4.1 Sequencing technologies *24*

2.4.2 Library preparation *26*

2.4.3 Detecting species with metagenomics *28*

2.4.4 Detecting Genotypic Resistance with Metagenomics *30*

2.4.5 Databases *32*

2.4.6 Bioinformatic approaches *34*

2.4.7 Nucleic Acid extraction *36*

2.4.8 Quality controls *38*

2.4.9 Contamination *39*

2.4.10 Outbreak investigations *41*

2.4.11 Cost and time of metagenomic workflows *42*

2.4.12 Regulation as a medical device *44*

2.4.13 Largest metagenomic studies to date *45*

2.5 CONCLUSION 46

3 META-ANALYSIS AND SYSTEMATIC REVIEW 49

3.1 INTRODUCTION 49

3.2 METHOD 51

3.2.1 Search strategy and selection criteria *51*

3.2.2 Meta-analysis *53*

3.3 RESULTS 54

3.3.1 Clinical samples and range of pathogens *57*

3.3.2 Methodology: nucleic acid extraction and sequencing library preparation *57*

3.3.3 Methodology: controls *58*

3.3.4 Sequencing technologies *58*

3.3.5 Bioinformatic approaches *58*

3.3.6 Assessing true species presence versus contamination *60*

3.3.7 Performance: species identification *62*

3.3.8 Performance: antimicrobial resistance prediction *68*

3.3.9 Time of complete process *69*

3.3.10 Cost *71*

3.4 DISCUSSION 71

4 OPTIMIZATION OF DNA EXTRACTION AND SEQUENCING 100

4.1 INTRODUCTION 101

4.2 RESEARCH QUESTION 1 103

4.2.1 Methods *103*

4.2.2 Results *106*

4.2.3 Discussion *115*

4.3 RESEARCH QUESTION 2 117

4.3.1 Methodology *117*

4.3.2 Results *118*

4.3.3 Discussion *129*

4.3.4 Limitations *131*

5 BENCHMARKING TAXONOMIC CLASSIFIERS 132

5.1 BACKGROUND 133

5.2 METHODS 135

5.2.1 Simulated data *135*

5.2.2 Taxonomic classifiers and sequence databases *140*

5.2.3 Real-world data *140*

5.2.4 Statistical analysis *141*

5.3 RESULTS 143

5.3.1 Determinants of taxonomic classification accuracy *143*

5.3.2 Best performing combinations *148*

5.3.3 Choice of sequencer *152*

5.3.4 Impact of read length. *154*

5.3.5 Common misclassifications *155*

5.3.6 Relationship between species misclassification and nucleotide identity *160*

5.3.7 Comparison with real-world data *163*

5.4 DISCUSSION 165

6 SEQUENCING BLOOD CULTURE CLINICAL SAMPLES AND PREDICTING SPECIES ABUNDANCE 209

6.1 METHODS 214

6.1.1 Sample collection and processing *216*

6.1.2 DNA extraction from blood culture *217*

6.1.3 Library preparation and Nanopore sequencing *218*

6.1.4 Bioinformatic analysis *218*

6.2 RESULTS: METHOD 1 – OPTIMISED THRESHOLD APPROACH (PLASMID SEQUENCES REMOVED) 224

6.3 RESULTS: METHOD 2 – OPTIMISED THRESHOLD APPROACH (WITH MAPPING) 227

6.3.1 Assessing false negative results for plausibility in Method 2 (i.e., potential inaccurate sequencing results or inaccurate MALDI-TOF results) *230*

6.3.2 Assessing false positives in Method 2 by maximizing sensitivity (i.e., potential contaminants or potential missed infections by the routine laboratory) *235*

6.3.3 Assessing false positive results for plausibility: *237*

6.3.4 Performance after accounting for plausible false positive and false negative results *242*

6.3.5 Performance with even lower threshold cut-offs (after accounting for plausible false positives and false negatives) *244*

6.4 CULTURE NEGATIVE SAMPLES 247

6.5 METHOD 3 - RANDOM FOREST APPROACH 249

6.5.1 Random forest model 1: Not accounting for previously identified plausible false negative and false positive species results. *249*

6.5.2 Random forest model 2: Accounting for previously identified plausible false negatives and false positives species in the training and testing dataset. *252*

6.6 SAMPLES UNIDENTIFIABLE BY THE ROUTINE LAB IN POSITIVE BLOOD CULTURES 255

6.7 SEQUENCING DEPTH AND COVERAGE 257

6.8 CONTROLS 257

6.8.1 Negative control: No spike *257*

6.8.2 Negative control: Thermus thermophilus spike *257*

6.8.3 Negative control: Increasing Thermus thermophilus spike levels *260*

6.8.4 Positive control *262*

6.8.5 Internal spike control present in positive controls *264*

6.8.6 Internal spike control present in culture positive and culture negative clinical samples *266*

6.9 LIMIT OF DETECTION 269

6.10 BARCODE CONTAMINATION 269

6.10.1 Mis-assigned reads *270*

6.10.2 Cross-assigned reads *273*

6.10.3 Cross-assigned contamination by off-target barcode *276*

6.10.4 Probability of assigning a correct read *279*

6.11 TIME TO SPECIATION (ROUTINE LABORATORY VS METAGENOMIC APPROACH) 280

6.11.1 Time to speciation by the routine laboratory *280*

6.11.2 Time to speciation by the metagenomic pipeline *280*

6.12 CFU COUNTS, DNA EXTRACTION YIELD AND TIME TO POSITIVITY. 285

6.12.1 Does positive blood culture bottle weight (a proxy for inoculated blood volume) affect CFU count? *285*

6.12.2 Does blood culture bottle weight weight (a proxy for inoculated blood volume) affect incubation time i.e., blood culture time to positivity? *287*

6.12.3 Is CFU count associated with blood culture time to positivity? *289*

6.12.4 Does CFU count relate to DNA extraction yield? *290*

6.12.5 Does DNA extraction yield relate to the number of bases sequenced? *292*

6.12.6 Does available nanopore sequencing pores relate to the number of bases sequenced? *293*

6.13 COST PER SAMPLE 296

6.14 RESULTS SUMMARY 299

6.15 DISCUSSION 302

6.15.1 Performance 303

6.15.2 Quality control 306

6.15.3 CFU counts, DNA extraction yield and time to positivity. 307

6.15.4 Cost 308

6.15.5 Limitations of this study 310

6.15.6 Challenges and future direction 311

APPENDIX 1 320

APPENDIX 2 322

7 METAGENOMIC RESISTANCE PREDICTION IN BLOOD STREAM INFECTIONS 324

7.1.1 Introduction 324

7.2 METHODS 329

7.3 RESULTS 333

7.3.1 Study characteristics 333

7.3.2 Performance in monomicrobial infections 337

7.4 DISCUSSION 368

8 CONCLUSION AND FUTURE WORK 394

9 REFERENCES 404

1 PREFACE

1.1 Acknowledgements

Over the course of the past four years, I have been blessed with a most enjoyable experience. I attribute this, in no small part, to the fulfilling nature of engaging with a challenging, and intriguing research problem. Yet, equally significant was the overwhelming support and encouragement that I received from a community of remarkable individuals who stood by me throughout this journey. Their steadfast assistance, particularly during the trying times wrought by the COVID-19 pandemic, has been invaluable and deeply appreciated.

This thesis would not have been possible without the invaluable guidance, encouragement, and support of my supervisor Professor David Eyre. His unwavering commitment to excellence, deep knowledge and insightful feedback have been crucial to shaping the work presented here. In addition to David's impressive academic aptitude, I must acknowledge his kind and supportive nature, he not only motivated me to achieve my academic goals but also genuinely cared about my well-being as a person.

I am also grateful to Ali Vaughan for her constant support who has always ensured that I have had all the necessary resources to carry out my laboratory experiments. Dr. Teresa Street has been a beacon of light in the laboratory helping me to design and carry-out crucial experiments, as well as James Kavanagh, who had taught me, as a naive medic, laboratory techniques for the first time. I also thank Dr. Sam Lipworth, Dr. Nicole Stoesser, and Dr. Kate Dingle for their endless willingness to always discuss and debate ideas to improve upon my work as well as Professor Derrick Crook who has provided oversight of my thesis for which I am grateful.

I am also immensely appreciative for my friends who have been a constant source of support and encouragement throughout my academic journey. To Michael Chen, my partner in "pun-derful" adventures and fellow patagoon, always armed with a joke that hits the spot. To Lise du Buisson, my partner-in-crime at those formal Oxford dinners. Where there was laughter echoing a tad too loudly, rest assured, it was from our table. To Nabeela Kajee, an unwavering pillar of strength, whose persuasive charm scored us the best unpaid seats at the Wimbledon semi-finals. To Alexander Lachapelle, the maestro who could turn a baguette crisis into a strategic symposium. And to Bianca Bauer, with whom I attended so many White Tie balls that they started to feel like our regular hangouts. Thanks to the South African crew: Rahul Chaudhary, the schoolmate always armed with a timely joke; Winston Padayachee, our grounding voice of reason; and Riedwaan Fakier, a caring soul who doubles as both a formidable athlete and an ever-reliable snack supplier. Their support and infectious positivity have made this journey not only academically rewarding but also an enjoyable one.

To my Tawney Street friends and housemates Momo Komatsu, Maike Meier, Johny Bulled and Verena Sarrazin, thanks for turning my challenges into shared journeys with comforting meals, steadfast encouragement, and always lending a listening ear.

Last but certainly not least, I must express my deepest gratitude to my family and especially my parents, Nadaraj and Sarojini Govender, who have always been my pillars of strength and have unwaveringly supported me in my endeavours. Thank you for instilling in me the values of hard work, perseverance, and resilience. I am forever grateful for your sacrifices and firm belief in me. Thank you to my sister, Kemintha Govender, who has always stood by my side and to my grandfather, Poolingam Perumal Govender, who has been a constant source of guidance. Finally, I would like to dedicate this dissertation to my late grandmother, Ulaganiagie Govender, who had succumbed to sepsis. My hope is that the work undertaken in this thesis leads to better bloodstream diagnostics in the future to prevent such fatalities. Although she is no longer with us, her everlasting support and encouragement will continue to inspire me in all my future endeavours.

1.2 Funding for all chapters

I am forever indebted to the Rhodes Trust for funding my academic and living expenses during the course of my thesis, and especially for their continued support in making up for the time lost over the COVID-19 pandemic. This thesis was additionally supported by the National Institute for Health Research (NIHR) Biomedical Research Centre, Oxford, UK.

1.3 Ethics Statement for all Chapters

This study was done with Research Ethics Committee approval (REC reference 17/LO/1420).

1.4 Declaration and attributions

I, Kumeren Govender, designed and conducted all the analyses presented in this thesis with the support of my supervisor David Eyre without whom this thesis certainly would not have been possible. A detailed description of work undertaken, and assistance obtained is set out below.

1.4.1 Chapter 3 attributions

I designed and conducted a systematic review and meta-analysis under the supervision of Professor David Eyre. Dr. Teresa Street contributed by providing feedback on laboratory aspects, while Dr. Nicholas Sanderson provided feedback on bioinformatic aspects. Furthermore, Dr. Teresa Street served as an additional independent reviewer of studies.

1.4.2 Chapter 4 attributions

I designed and conducted experiments to optimise DNA extraction and human DNA depletion directly from blood, under the supervision of Professor David Eyre and Dr. Teresa Street. Dr. Teresa Street taught me PCR and saponin-use techniques which I have incorporated into the experimental design. James Kavanagh had taught me how to sequence using the Oxford Nanopore platform.

1.4.3 Chapter 5 attributions

I designed and conducted an experiment to benchmark classification tools, sequencers and read lengths, as well as simulate sequencing error profiles, under the supervision of Professor David Eyre. Dr. Sam Lipworth provided real world Illumina sequencing data for benchmarking.

1.4.4 Chapter 6 attributions

I designed and conducted an experiment to predict infecting species using clinical metagenomics directly from blood cultures, under the supervision of Professor David Eyre. Dr. Teresa Street assisted in laboratory methodology design. I designed all bioinformatic pipelines involved in this study.

1.4.5 Chapter 7 attributions

I designed and conducted an experiment to predict antimicrobial susceptibilities using clinical metagenomics directly from blood cultures, under the supervision of Professor David Eyre. Dr. Nicole Stoesser assisted by providing feedback on resistance prediction methodology.

1.5 Publications arising

1.5.1 Directly from this thesis

1. Govender, K. N., Street, T. L., Sanderson, N. D. & Eyre, D. W. Metagenomic sequencing as a pathogen-agnostic clinical diagnostic tool for infectious diseases: a systematic review and meta-analysis of diagnostic test accuracy studies. *J Clin Microbiol* **59**, e02916-20 (2021).

2. Govender, K. N. & Eyre, D. W. Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications. *Microb Genom* **8**, 000886 (2022).

1.5.2 Other publications during the course of the thesis

1. Govender, K.N. Precision pandemic preparedness: improving diagnostics with metagenomics. *J Clin Microbiol* ***59***, pp.e02146-20 (2021).

2. Kukull, W.A., Govender, K.N. & Bowen, J., 2022. *Oxford Textbook of Global Public Health*, 7 edn, Oxford Textbooks in Public Health (Neurological diseases, epidemiology, and public health), https://doi.org/10.1093/med/9780198816805.003.0068.

3. Moodley, Y., Govender, K., van Wyk, J., Reddy, S., Ning, Y., Wexner, S., Stopforth, L., Bhadree, S., Naidoo, V., Kader, S. & Cheddie, S. Predictors of treatment refusal in patients with colorectal cancer: A systematic review. *Semin. Oncol* **49**, 456-464 (2023).

1.6 List of Tables

TABLE 1. SEARCH STRATEGY. 51

TABLE 2. INCLUSION AND EXCLUSION CRITERIA FOR THIS SYSTEMATIC REVIEW. 52

TABLE 3. CHARACTERISTICS OF STUDIES INCLUDED IN THE META-ANALYSIS. 56

TABLE 4. QUANTITATIVE PCR PRIMERS AND PROBES. 105

TABLE 5. TOTAL DNA YIELD BY EXTRACTION KIT (N=120). 107

TABLE 6. TOTAL DNA YIELD BY PRE-TREATMENT METHOD (N=28). 119

TABLE 7. LOGISTIC REGRESSION MODEL FOR PERFORMANCE OF SPECIES ABUNDANCE CLASSIFICATION BY SEQUENCER, CLASSIFIER, DATABASE AND BACTERIAL SPECIES. 145

TABLE 8. LOGISTIC REGRESSION MODEL FOR PERFORMANCE OF GENUS ABUNDANCE CLASSIFICATION BY SEQUENCER, CLASSIFIER, DATABASE AND BACTERIAL SPECIES. 146

TABLE 9. PREDICTED ABUNDANCE CLASSIFICATION FOR *STAPHYLOCOCCUS AUREUS* AND *ESCHERICHIA COLI*. 147

TABLE 10. RECOMMENDED THRESHOLD ABUNDANCE TO CALL ADDITIONAL TRUE SPECIES. 158

TABLE 11. ANI BETWEEN TRUE SPECIES AND MISCLASSIFIED SPECIES AGAINST PERCENTAGE OF READS MISCLASSIFIED AS A SPECIFIC SPECIES. 162

TABLE 12. METHOD 1 APPLIED TO THE TRAINING DATASET: READ AND PERCENTAGE CUT-OFF MAXIMISING THE YOUDEN'S INDEX. 226

TABLE 13. METHOD 1 APPLIED TO THE TRAINING, TESTING AND FULL DATASET: READ AND PERCENTAGE CUT-OFF MAXIMISING THE YOUDEN'S INDEX. 226

TABLE 14. METHOD 2 APPLIED TO THE TRAINING DATASET: READ AND PERCENTAGE CUT-OFF MAXIMISING THE YOUDEN'S INDEX. 229

TABLE 15. METHOD 2 APPLIED TO THE TRAINING, TESTING AND FULL DATASET: READ AND PERCENTAGE CUT-OFF MAXIMISING THE YOUDEN'S INDEX. 229

TABLE 16. ASSESSING FALSE NEGATIVE SPECIES RESULTS FOR PLAUSIBILITY (I.E., POTENTIAL INACCURATE MALDI-TOF RESULTS). 233

TABLE 17. METHOD 2 APPLIED TO THE ENTIRE DATASET: READ AND PERCENTAGE CUT-OFF MAXIMISING FOR SENSITIVITY (RATHER THAN FOR YOUDEN'S INDEX). 236

TABLE 18. METHOD 2 APPLIED TO THE TRAINING, TESTING AND FULL DATASET: READ AND PERCENTAGE CUT-OFF MAXIMISING SENSITIVITY AND THEN SPECIFICITY. 236

TABLE 19. ASSESSING FALSE POSITIVE SPECIES RESULTS FOR PLAUSIBILITY 240

TABLE 20. METHOD 2 APPLIED TO THE TRAINING DATASET AFTER ADJUSTING FOR PLAUSIBLE INFECTIONS: READ AND PERCENTAGE CUT-OFF MAXIMISING FOR YOUDEN'S INDEX. 243

TABLE 21. METHOD 2 APPLIED TO THE TRAINING, TESTING AND FULL DATASET AFTER ADJUSTING FOR PLAUSIBLE INFECTIONS: READ AND PERCENTAGE CUT-OFF MAXIMISING FOR YOUDEN'S INDEX. 243

TABLE 22. METHOD 2 APPLIED TO THE ENTIRE DATASET AFTER ADJUSTING FOR PLAUSIBLE INFECTIONS: READ AND PERCENTAGE CUT-OFF WITH LOWER THRESHOLD CUT-OFFS. 245

TABLE 23. ASSESSING FALSE POSITIVE SPECIES RESULTS FOR PLAUSIBILITY WITH LOWER THRESHOLD CUT-OFFS. 246

TABLE 24. FALSE POSITIVE RESULTS IN RANDOM FOREST MODEL 1 TESTING DATASET (EXCLUDING FOUR FALSE POSITIVE RESULTS DUE TO MALDI INACCURACY). 250

TABLE 25. ADDITIONAL INFECTIONS IDENTIFIED IN THE THRESHOLD-BASED METHOD BUT NOT IN THE RANDOM FOREST MODEL 1 TESTING DATASET. 250

TABLE 26. SPECIES PERFORMANCE RESULTS ACROSS ALL MODELS. 254

TABLE 27. SEQUENCING RESULTS FOR FASTIDIOUS UNIDENTIFIABLE CULTURE POSITIVE RESULTS BY THE ROUTINE LABORATORY. 256

TABLE 28. DYNAMIC FILTERING THRESHOLDS DERIVED FROM THE TRAINING DATASET AT GIVEN INTERVALS FROM SEQUENCING START. 284

TABLE 29. ITEMIZED COSTS PER SINGLE METAGENOMIC SAMPLE IN THIS STUDY (BASED ON 6 SAMPLE MULTIPLEXING). 297

TABLE 30. PREDICTED ITEMIZED COSTS PER SINGLE METAGENOMIC SAMPLE (BASED ON 24, 96 SAMPLE MULTIPLEXING). 298

TABLE 31. BREAKDOWN OF CULTURE POSITIVE SAMPLES AND RESULTS 301

TABLE 32. PHENOTYPIC RESISTANCE AND SENSITIVITY BY SPECIES. 334

TABLE 33. PHENOTYPIC RESISTANCE AND SENSITIVITY BY DRUG. 335

TABLE 34. PHENOTYPIC RESISTANCE AND SENSITIVITY BY SPECIES-DRUG COMBINATION. 336

TABLE 35. GENOTYPIC PERFORMANCE OF TOP 10 CLINICALLY RELEVANT SPECIES BY TOOL AFTER GENOME ASSEMBLY. 339

TABLE 36. GENOTYPIC PERFORMANCE OF TOP 10 CLINICALLY RELEVANT SPECIES USING RAW READS. 340

TABLE 37. GENOTYPIC PERFORMANCE OF TOP 10 CLINICALLY RELEVANT SPECIES USING RAW READS AT EARLIER TIME INTERVALS. 340

TABLE 38. GENES IDENTIFIED IN OTHER TOOLS TO IMPROVE VME CALL IN RESFINDER. 363

TABLE 39. ORGANISM-DRUG COMBINATIONS WHERE NO TOOL PREDICTED RESISTANCE. 365

TABLE 40. PHENOTYPIC SENSITIVITY AND SPECIFICITY FOR POLYMICROBIAL ISOLATES. 367

1.7 List of Figures

FIGURE 1. METAGENOMIC FLOW DIAGRAM. 50

FIGURE 2. FLOW CHART SHOWING STUDY SELECTION. 55

FIGURE 3. METAGENOMIC BIOINFORMATIC FLOW DIAGRAM. 59

FIGURE 4. COMMON APPROACHES TO CONTAMINATION CONTROL. 61

FIGURE 5. QUADAS-2 TOOL FOR THE EVALUATION OF PRIMARY RESEARCH ARTICLES. 62

FIGURE 6. QUADAS-2 QUESTIONS. RISK OF BIAS IS JUDGED AS "LOW", "HIGH", OR "UNCLEAR". 63

FIGURE 7. TRAFFIC LIGHT PLOT OF STUDIES INCLUDED. 64

FIGURE 8. POOLED SENSITIVITY OF STUDIES GROUPED BY SAMPLE TYPE. 66

FIGURE 9. POOLED SPECIFICITY OF STUDIES GROUPED BY SAMPLE TYPE. 67

FIGURE 10. POOLED ANTIMICROBIAL RESISTANCE PREDICTION PERFORMANCE. 69

FIGURE 11. TIME TAKEN FOR SAMPLE PROCESSING GROUPED BY SAMPLE TYPE. 70

FIGURE 12. TOTAL DNA BY EXTRACTION KIT (NG). 106

FIGURE 13. HUMAN AND BACTERIAL COPY NUMBER PER 2µL GROUPED BY EXTRACTION KIT. 108

FIGURE 14. HUMAN AND BACTERIAL COPY NUMBER PER 2µL GROUPED BY DEPLETION METHOD AND GROUPED BY EXTRACTION KIT. 110

FIGURE 15. RATIO OF BACTERIAL COPY NUMBERS PER 2µL TO HUMAN COPY NUMBERS PER 2µL BY EXTRACTION KIT AND DEPLETION METHOD. 111

FIGURE 16. DISPLAYS BACTERIAL AND HUMAN COPY NUMBER PER 2µL IN CONTROLS GROUPED BY EXTRACTION KIT AND DEPLETION METHOD. 112

FIGURE 17. DISPLAYS BACTERIAL AND HUMAN COPY NUMBER PER 2µL IN CONTROLS GROUPED BY EXTRACTION KIT AND GROUPED BY DEPLETION METHOD. 113

FIGURE 18. DISPLAYS HUMAN AND BACTERIAL COPIES PER 2µL BY WATER CONTROLS, NEGATIVE CULTURE-MEDIUM CONTROL AND POSITIVE CULTURE. 114

FIGURE 19. TOTAL DNA EXTRACTED FROM NO TREATMENT VS. SAPONIN TREATMENT (NG). 119

FIGURE 20. BACTERIAL, HUMAN, VIRAL AND TOTAL BASES GROUPED BY DEPLETION METHOD. 120

FIGURE 21. BACTERIAL SEQUENCING DEPTH. 121

FIGURE 22. HISTOGRAM OF AVERAGE READ LENGTH DATA PER SAMPLE WITH DENSITY PLOT. 122

FIGURE 23. TOTAL BASES SEQUENCED PER BARCODE FOR EACH FLOW CELL. 124

FIGURE 24. KRAKEN2 RESULTS FOR METAGENOMIC SAMPLES. 127

FIGURE 25. KRAKEN2 RESULTS FOR SIMULATED SAMPLES FROM NCBI. 128

FIGURE 26. EXTRACTED DNA CONCENTRATION (µG/ML) PLOTTED BESIDE BACTERIAL BASES SEQUENCED. 129

FIGURE 27. ILLUMINA SEQUENCING READ DISTRIBUTION. 138

FIGURE 28. NANOPORE SEQUENCING READ DISTRIBUTION. 139

FIGURE 29. PERCENTAGE OF READS CLASSIFIED CORRECTLY OR INCORRECTLY ACROSS CLASSIFIERS FOR FIVE IMPORTANT BLOOD STREAM PATHOGENS. 149

FIGURE 30. THE PERCENTAGE OF READS WITH THE CORRECT AND INCORRECT ASSIGNED SPECIES AND GENUS, AND THE PERCENTAGE OF READS UNCLASSIFIED ARE SHOWN. 150

FIGURE 31. PERCENTAGE OF READS CLASSIFIED FROM SIMULATED HUMAN READS USING ILLUMINA SEQUENCING. 151

FIGURE 32. PERCENTAGE OF READS CLASSIFIED CORRECTLY OR INCORRECTLY BY SEQUENCING DEVICE AND ERROR PROFILE. 153

FIGURE 33. READ LENGTH VS. PROPORTION OF READS CORRECTLY CLASSIFIED. 155

FIGURE 34. ANI BETWEEN TRUE SPECIES AND MISCLASSIFIED SPECIES AGAINST PERCENTAGE OF READS MISCLASSIFIED AS A SPECIFIC SPECIES. 161

FIGURE 35. PERCENTAGE OF READS CLASSIFIED CORRECTLY OR INCORRECTLY IN ILLUMINA SIMULATED DATA AND REAL-WORLD DATA. 164

FIGURE 36. TRADITIONAL MICROBIOLOGY VS. METAGENOMIC PIPELINE. 215

FIGURE 37. BIOINFORMATIC PIPELINE: SPECIATION 219

FIGURE 38. FILTERING CALIBRATION HEAT MAP FROM TRAINING DATASET IN METHOD 1. 225

FIGURE 39. FILTERING CALIBRATION HEAT MAP FROM TRAINING DATASET IN METHOD 2. 228

FIGURE 40. BIOMERIEUX API (ANALYTICAL PROFILE INDEX) 20E TESTING FOR *KLEBSIELLA PNEUMONIAE* (SAMPLE 60) VS *KLEBSIELLA QUASIPNEUMONIAE* (SAMPLE 46 & 49) 234

FIGURE 41. PRE-FILTERED CULTURE NEGATIVE CLINICAL SAMPLES (FILTER THRESHOLDS NOT APPLIED). 248

FIGURE 42. ROC CURVE OF TEST DATASET WITHOUT ADJUSTING FOR PLAUSIBILITY IN BOTH TRAINING AND TEST DATASETS. 251

FIGURE 43. RANDOM FOREST MODEL 1 VARIABLE IMPORTANCE PLOT. 251

FIGURE 44. RANDOM FOREST MODEL 2 VARIABLE IMPORTANCE PLOT. 253

FIGURE 45. PRE-FILTERED NEGATIVE CONTROL WITH NO SPIKE (NO FILTER THRESHOLDS APPLIED). 258

FIGURE 46. UNFILTERED AND FILTERED SPIKED NEGATIVE CONTROL. 259

FIGURE 47. INCREASING LEVELS OF SPIKED *THERMUS THERMOPHILUS (TTH)* CONCENTRATION IMPROVES CONTAMINATION CONTROL. 261

FIGURE 48. UNFILTERED AND FILTERED POSITIVE CONTROLS. 263

FIGURE 49. UNFILTERED AND FILTERED POSITIVE CONTROLS. 265

FIGURE 50. INTERNAL SPIKE CONTROL SHOWN IN CULTURE-POSITIVE CLINICAL SAMPLES. 267

FIGURE 51. INTERNAL SPIKE CONTROL SHOWN IN CULTURE-NEGATIVE CLINICAL SAMPLES. 268

FIGURE 52. PROBABILITY OF READS MIS-ASSIGNED TO OTHER BARCODES PER RUN. 272

FIGURE 53. PROBABILITY OF CROSS-ASSIGNMENT IN CULTURE POSITIVE SAMPLES (ESTIMATED USING *ARCANOBACTERIUM HAEMOLYTICUM*). 274

FIGURE 54. PROBABILITY OF CROSS-ASSIGNMENT IN CULTURE NEGATIVE SAMPLES (ESTIMATED USING *ARCANOBACTERIUM HAEMOLYTICUM*). 275

FIGURE 55. MEAN PROPORTION OF OFF TARGET READS BY BARCODE. 277

FIGURE 56. CROSS-ASSIGNED *ARCANOBACTERIUM HAEMOLYTICUM* READS SHOWN ACROSS OFF-TARGET BARCODES. 278

FIGURE 57. SPECIES PERFORMANCE OVER SEQUENCING TIME (FILTERING THRESHOLDS BASED ON TRAINING DATASET AT THE END OF SEQUENCING). 282

FIGURE 58. SPECIES PERFORMANCE OVER SEQUENCING TIME (DYNAMIC FILTERING THRESHOLDS BASED ON TRAINING DATASET AT GIVEN TIME POINTS FROM SEQUENCING START). 283

FIGURE 59. SCATTER PLOT OF BLOOD CULTURE BOTTLE WEIGHT VS. CFU COUNT. 286

FIGURE 60. SCATTER PLOT OF BLOOD CULTURE INCUBATION TIME VS. BLOOD CULTURE BOTTLE WEIGHT. 288

FIGURE 61. SCATTER PLOT OF BLOOD CULTURE INCUBATION TIME VS. BLOOD CULTURE BOTTLE WEIGHT FOR *E. COLI* AND *S. EPIDERMIDIS*. 288

FIGURE 62. SCATTER PLOT OF CFU COUNT VS. TIME TO POSITIVE BLOOD CULTURE. 290

FIGURE 63. SCATTER PLOT OF CFU COUNT VS. DNA EXTRACTION YIELD. 292

FIGURE 64. SCATTER PLOT OF DNA EXTRACTION YIELD VS. NUMBER OF BASES SEQUENCED. 293

FIGURE 65. SCATTER PLOT OF AVAILABLE NANOPORES AT SEQUENCING START VS. BASES GENERATED. 295

FIGURE 66. GENOTYPIC PERFORMANCE BY SPECIES ACROSS TOOLS. 343

FIGURE 67. GENOTYPIC PERFORMANCE BY DRUG CLASS. 346

FIGURE 68. GENOTYPIC PERFORMANCE BY SPECIFIC DRUG (1/2). 347

FIGURE 69. GENOTYPIC PERFORMANCE BY SPECIFIC DRUG (2/2). 348

FIGURE 70. GENOTYPIC PERFORMANCE FOR SENSITIVE ONLY DRUG/DRUG CLASS. 349

FIGURE 71. UPSET PLOT SHOWING VARIOUS GROUPS WITHIN RESISTANT PHENOTYPE CALLS. 352

FIGURE 72. UPSET PLOT SHOWING VARIOUS GROUPS WITHIN SENSITIVE PHENOTYPES. 353

FIGURE 73. GENES IMPLICATED IN TRUE GENOTYPIC CALLS (RESFINDER WITH FASTQ READS AT 24HR). 356

FIGURE 74. GENES IMPLICATED IN MAJOR ERROR GENOTYPIC CALLS (RESFINDER WITH FASTQ READS AT 24HR). 359

FIGURE 75. SPECIFIC *BLA* GENES IMPLICATED IN FALSE AND TRUE POSITIVE GENOTYPIC CALLS (1/2). 360

FIGURE 76. SPECIFIC *BLA* GENES IMPLICATED IN FALSE AND TRUE POSITIVE GENOTYPIC CALLS (2/2). 361

1.8 List of Supplementary Tables

SUPPLEMENTARY TABLE 1. DETAILED LABORATORY METHODOLOGY. 77

SUPPLEMENTARY TABLE 2. BIOINFORMATIC DETAILS. 81

SUPPLEMENTARY TABLE 3. PERFORMANCE OF EACH STUDY. 84

SUPPLEMENTARY TABLE 4. QUADAS-2 RESULTS. 86

SUPPLEMENTARY TABLE 5. OTHER META-ANALYSIS MEASURES. 88

SUPPLEMENTARY TABLE 6. DETAILED THRESHOLD ABUNDANCE RECOMMENDED TO CALL ADDITIONAL TRUE SPECIES AND GENUS. 89

SUPPLEMENTARY TABLE 7. PREDICTED ABUNDANCE CLASSIFICATION FOR ALL INCLUDED SPECIES. 170

SUPPLEMENTARY TABLE 8. DETAILED THRESHOLD ABUNDANCE RECOMMENDED TO CALL ADDITIONAL TRUE SPECIES AND GENUS. 180

SUPPLEMENTARY TABLE 9. REFERENCE LIST OF GENOMES RANDOMLY SELECTED FROM NCBI TO PERFORM SIMULATION. 193

SUPPLEMENTARY TABLE 10. BIOINFORMATIC CONTAMINANTS (SAMPLES WITH MEDIAN >1% SHOWN FOR BREVITY). 315

SUPPLEMENTARY TABLE 11. PHENOTYPIC RESISTANCE AND SENSITIVITY BY SPECIES-DRUG COMBINATION. 375

SUPPLEMENTARY TABLE 12. ANTIBIOTIC CLASSES AND SPECIFIC DRUGS. 379

SUPPLEMENTARY TABLE 13. PHENOTYPE PANEL WHERE A GENETIC RESISTANCE CALL COULD BE MADE. 381

1.9 List of Abbreviations and Acronyms

AMR- Antimicrobial resistance

ARDB- Antibiotic Resistance Genes Database

ARG-ANNOT- Antibiotic Resistance Gene Annotation

AST- Antimicrobial susceptibility testing

BLAST- Basic Local Alignment Search Tool

BSI- Bloodstream infection

CAMI- Critical Assessment of Metagenome Interpretation

CARD- Comprehensive Antibiotic Resistance Database

CDC- Centre for Disease Control and Prevention

CLIA- Clinical Laboratory Improvement Amendments

CSF- Cerebrospinal fluid

ECDC- The European Centre for Disease Prevention and Control

FDA- Food and Drug Administration

HGT- Horizontal gene transfer

LAMP- Loop-mediated isothermal amplification

MALDI-TOF- Matrix-assisted laser desorption/ionization-time of flight

NGS- Next-generation sequencing

PCR- Polymerase chain reaction

VRE- Vancomycin-resistant enterococci

WGS- Whole-genome sequencing

1.10 List of Appendices

APPENDIX 1 320

APPENDIX 2 322

"The writing of the hurried mind does not compare to the musings of the reflective."

-Nabeela Kajee

2 BACKGROUND

2.1 Introduction

The term metagenomics first appeared in 1998 which referenced the idea that a collection of genes could be sequenced in a single sample without the need for isolation or lab cultivation of a specific species.[1] Driven by advances in sequencing technology, the study of metagenomics primarily encompasses infectious disease diagnostics, microbiome analysis, transcriptomics and oncological applications.[2] In infectious diseases, metagenomics has the potential to be a disruptive technology that revolutionises diagnostics, in place of traditional culture-based microbiology which has remained fundamentally unchanged over decades.

Traditionally most diagnostic microbiology depends on culture to identify the causative organism in an infection and to define antimicrobial susceptibilities. However, this process can take several days, or in the case of *Mycobacteria* and other slow growing

organisms several weeks. While waiting for results patients are treated empirically, which may result in unnecessarily broad-spectrum treatment in some, and ineffective treatment in others. Additionally, culture is imperfectly sensitive, for example, many patients with a serious infection have negative blood cultures, and culture may be impaired by prior antimicrobial exposure.

Molecular diagnostics based on detection of antigens, antibodies or nucleic acid are increasingly used to complement culture-based diagnosis. For example, multiplex-PCR syndromic panels are able to identify common pathogens and antimicrobial resistance determinants within a few hours. However, these panels can only identify a restricted number of pathogens and resistance determinants and are therefore typically bespoke for a given clinical syndrome.

Emerging real-time sequencing technologies might offer an unbiased approach to identify any pathogens and known resistance determinants present within hours. Studies have demonstrated that metagenomic pipelines can provide faster diagnoses than conventional culture-based methods which may hold crucial for critically unwell patients.[3] Metagenomics can yield diagnoses in challenging cases where no conventional infectious sources can be found, as in the case of a 29 year-old man in San Francisco who was incorrectly treated for tuberculous meningitis multiple times but found to have *Taenia solium* infection from metagenomic sequencing.[4] In contrast to culture-based techniques, obtaining results for a wide range of second- and third-line drugs requires minimal additional effort if sufficient sequence data for a given species are available. Where sufficient knowledge bases exist prediction of antimicrobial resistance

is potentially possible and detection of mixed infections is also possible, where a subpopulation of resistant strains can be detected.[5]

In the future, clinical metagenomics, potentially together with pathogen and host transcriptomic data, could provide precision infectious disease diagnostics. However, current metagenomic pipelines are largely research-focused, with further work required to improve performance and provide assurance on diagnostic accuracy, and to develop platforms for routine use. Many challenges exist for the implementation of metagenomics in clinical laboratories. These challenges specifically include sample preparation involving human DNA depletion, unbiased DNA extraction methods often with limited pathogen DNA present, sequencing errors, bioinformatic challenges from computational power to reliable databases, contamination, and quality control mechanisms. Furthermore, meeting regulatory standards is a complex endeavour with regular updates to the workflow or databases.

This chapter examines the traditional workflow in a diagnostic microbiology laboratory, with specific attention to blood culture processing, a cornerstone in detecting bloodstream infections. Concurrently, it delves into the landscape of novel molecular methods. These emerging techniques promise to bring transformative changes to clinical diagnostics by providing improved sensitivity and swift results, including those typically challenging to culture. This segues into the exploration of metagenomics, which holds potential for a more comprehensive understanding of microbial diversity within a clinical sample, shedding light on unculturable, rare, or novel organisms. However, this technique comes with its own suite of challenges and limitations, which are thoroughly scrutinized within this chapter. As a whole, this chapter encapsulates the current and

prospective future of diagnostic clinical microbiology (with a focus on blood culture), emphasizing the potential role of metagenomics as a culture-independent tool for infectious disease diagnostics.

2.2 The typical diagnostic workflow for blood cultures in routine microbiology laboratories

The diagnostic workflow in routine microbiology is a critical process, and my research is specifically focused on the blood culture workflow, a key component of this system. The purpose of a blood culture is to diagnose systemic infections, including bacteraemia, infective endocarditis, and conditions associated with unexplained fever (pyrexia of unknown origin). Mortality due to bacteraemia is related to the type of organism isolated and the nature of any underlying disease. Therefore, early positive results from a blood culture can provide valuable diagnostic information on which appropriate antimicrobial treatment can be used.

In the United Kingdom, the UK Health Security Agency (UKHSA) (previously known as Public Health England) sets out the standards for routine microbiology investigations. The "Investigation of Blood Cultures (for organisms other than *Mycobacterium* species)" document provide guidance, recommendations, and standardisation on which routine microbiology laboratories should adhere to; this includes guidance from sampling techniques to use of automated diagnostic equipment.[6]

2.2.1 Sampling

Sampling of blood for culture should be carried out in an aseptic manner, using United Kingdom Accreditation Service (UKAS) marked bottles only (previously CE). Ideally blood cultures should be taken before the start of antimicrobial therapy and as soon as possible after a spike of fever. The recommended specimen volume is 8-10ml in both the adult aerobic and anaerobic culture bottled, while 1-3ml is recommended when using the paediatric bottle. Users should not exceed or underfill bottles as this leads to less accurate results.

2.2.2 Specimen reception

Inoculated bottles should be incubated as soon as possible, within a maximum of four hours, due to their clinical significance. Usually, blood culture bottles are received as a pair; one aerobic and one anaerobic making a blood culture set. In certain cases, more than one set may be collected from a patient, and in the case of neonates and children a single paediatric bottle is routine.

2.2.3 Laboratory techniques for blood culture

Most clinical institutions employ automated, continuously monitored blood culture devices to incubate specimens delivered to the laboratory. These systems, manufactured by several different companies including BACTEC™ by BD (Becton, Dickinson and Company), BacT/ALERT® by bioMérieux, and VersaTREK™ by Thermo Fisher Scientific, and are renowned for their similar performance characteristics.[7] They are designed to maintain blood culture bottles at an optimal temperature for a user-specified period, signalling audibly and/or visually upon detecting microbial growth.

The medium used in each automated blood culture system varies based on user selection, with bottles typically housing a proprietary concoction of culture medium, an anticoagulant, and frequently, resins or charcoal mixtures. These additional ingredients are crucial for neutralizing the effects of antimicrobials and other potentially toxic compounds that could inhibit the growth of the microorganisms. Generally, the selection of medium combinations is done strategically to augment the recovery of the broadest range of microorganisms. Such combinations often consist of aerobic and anaerobic formulations. In certain situations, the use of a medium infused with reagents specifically designed for the cultivation of *Mycobacteria* and/or yeasts may be warranted. Studies have consistently shown that media fortified with antimicrobial binding or absorbing agents (like resins or charcoal compounds) have a clear edge in terms of recovering microorganisms, particularly *Staphylococci* and yeasts.[8]

Standard protocol for automated, continuous-monitoring systems dictates a routine incubation period of five days for blood cultures. This duration has been empirically proven to be adequate for the detection of the majority of pathogens, including fastidious bacteria from the *Haemophilus*, *Actinobacillus*, *Cardiobacterium*, *Eikinella*, and *Kingella* (HACEK) group.[9] Prolonged incubation past five days usually only serves to increase the recovery of contaminants rather than true pathogens. However, longer incubation times may be necessary under specific circumstances such as suspected dimorphic fungaemia or bacteraemia triggered by *Legionella*, *Brucella*, *Bartonella*, or *Nocardia* species. Cultures specifically testing for *Mycobacterium spp.* should be incubated for an extended period of four weeks to ensure accurate results.

2.2.4 Isolation and identification

2.2.4.1 Gram stain and microscopy

When blood cultures signal the presence of microbial growth, the next step typically involves performing a Gram stain. This is a basic test that allows for the preliminary classification of the bacteria based on their reaction to the staining technique. This initial information can guide immediate empirical antibiotic treatment decisions. For instance, should the Gram stain result be positive, which indicates a thick peptidoglycan cell wall, antibiotics such as flucloxacillin or vancomycin, typically effective against Gram-positive bacteria, might be deployed. Conversely, in the case of a Gram-negative result, which signifies a thin peptidoglycan cell wall covered by an outer membrane, antibiotics such as co-amoxiclav or gentamycin, often effective against Gram-negative bacteria, might be administered. Furthermore, Gram staining and microscopy helps guide the choice of an appropriate culture medium for further growth and testing. For instance, suspected *Staphylococci* will have a coagulase test, while suspected yeast might be further grown on sabouraud agar.

2.2.4.2 Sub-culture and species identification

Following the detection of microbial growth and the Gram stain procedure, subculturing on Petri dishes is the subsequent stage in the diagnostic workflow. This process involves transferring a specimen from the positive blood culture onto additional culture media to promote further microbial proliferation. It is noteworthy that high-risk samples necessitate handling within a Containment Level 3 facility, to minimize potential biohazards. Such samples are typically not dispatched for identification via Matrix-Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) mass spectrometry or Phoenix BD automated microbiology system. Instead, extensive subculturing is conducted, followed by manual identification methods.

The pathway for interpreting and conducting further examinations from positive cultures bifurcates into two routes. The first route involves cultures that follow a 6-hour incubation protocol, wherein the microorganisms are allowed to grow and multiply under controlled conditions for this specific duration. The second route encompasses isolates necessitating additional incubation beyond the initial 6 hours. This extended incubation is generally mandated for certain specimens, such as Gram-positive rods, *Candida* species, or instances when there is mixed microbial growth. Once sufficient growth is achieved, samples are then prepared for further identification processes, such as MALDI-TOF mass spectrometry.

MALDI-TOF mass spectrometry is a pivotal tool in modern microbiology, offering a rapid, precise and cost-effective solution for microbial identification. The principle of this technique involves ionizing microbial proteins using a laser and measuring their time of flight to determine their mass-to-charge ratios. These ratios are then compiled to generate a unique spectral profile, often referred to as a "fingerprint". This profile is compared to a reference database for the identification of the microorganism. The advantage of MALDI-TOF lies in its speed, accuracy and cost; a single test can generate results within minutes, drastically reducing the turnaround time compared to traditional culture-based identification methods.

Despite the advantages of MALDI-TOF, some laboratories, especially those in resource-limited settings, may rely on more conventional methods such as the Analytical Profile Index (API) test (bioMe'rieux, Craponne, France). API is a biochemical testing system that helps in the identification of bacterial species based on their metabolic activities.

An API test panel consists of multiple miniaturized biochemical tests and a numerical code based on the test results is compared with the database for identification. While API tests may be more time-consuming than MALDI-TOF and are not as broadly applicable, they still provide a valuable tool in the microbiologist's arsenal. Particularly, they may be of use when dealing with isolates that are not represented in the MALDI-TOF database, thus complementing the capabilities of mass spectrometry in microbial identification.

2.2.5 Antimicrobial Susceptibility Testing (AST)

2.2.5.1 The problem of growing AMR

The World Health Organisation in 2019 declared antimicrobial resistance as one of the top 10 global public health threats facing humanity.[10] AMR occurs when bacteria, fungus, and other microorganisms develop resistance to antibiotics, antifungals, or other antimicrobial drugs. In clinical settings, antibiotic resistance reduces available treatment options and increases morbidity and mortality compared with those caused by susceptible bacterial infections.[11] This leads to persistent infections which are difficult to treat, and may lead to life-threatening, multidrug-resistant infections. In fact, antibiotic resistance is seen to nearly all antimicrobials, including many last-resort treatments.[12] Resistance in first-line antimicrobials to bacteria alone affects more than 2 million people in the United States, resulting in an estimated US$20 billion cost to the health system and $35 billion cost in lost productivity annually.[13] From European data collected in 2015, it was estimated that there are 670,000 infections with resistant bacteria, accounting for an estimated 33,000 attributable deaths and loss of 870,000 disability adjusted life years (DALYs).[14] The increasing prevalence of AMR is limiting the number of antibiotics that can be used to successfully treat infections, while the Centre for Disease Control and Prevention (CDC) and The European Centre for Disease

Prevention and Control (ECDC) have estimated that up to 30-50% of all antimicrobials prescribed to human patients are unnecessary, promoting further development and spread of resistance.[15,16]

AST is a widely used method to determine antibiotic resistance to inform treatment plans in clinical settings. Methods used for AST are often phenotypic, which rely on observable characteristics of a bacterial strain, such as whether it can grow at particular concentrations of the antibiotic being tested, or can also be genotypic where methods detect the presence of resistance genes or mutations conferring resistance in a bacterial isolate.

2.2.5.2 Phenotypic AST testing on Isolates from Positive Blood Cultures

For AST testing, isolated bacteria are directly challenged to test for AMR. In broth dilution AST, bacteria are added to decreasing antimicrobial concentrations to find the point at which bacterial growth is inhibited at the lowest concentration of an antibiotic agent, known as the minimum inhibitory concentration. Solid media techniques such as the disk diffusion test or gradient diffusion strips can also be used to measure the clearance of bacteria from a source of the antimicrobial agent, known as the zones of inhibition. Further details on culture-based phenotypic AST are available elsewhere.[17] Standards developed by the Clinical and Laboratory Standards Institute (CLSI) or the European Committee on AST (EUCAST) are used to convert minimum inhibitory concentrations or zones of inhibition measurements to categorical resistance interpretations.

Several automated systems have been developed to conduct phenotypic AST in clinical settings which include the BD Phoenix[18], VITEK[19], ADAGIO[20], Accelerate Pheno[21] among others. These systems have many advantages including standardised chronometers, on-going monitoring systems and sensitive optical instruments for determining susceptibility results, increasing the consistency of results, and reducing active bench-time for clinical microbiologist. Although used as the standard in many microbiology laboratories, these systems are prohibitively expensive to set up in lowered resourced settings.

Other resistance determination methods currently being developed include fluorescence in situ hybridization (FISH), and microfluidics-based techniques which are nascent in development. Such rapid AST is ideal for slow-growing organisms such as *Mycobacterium tuberculosis*, for which conventional AST can take weeks.[22]

Phenotypic AST is useful for providing actionable results; however, it is low throughput. For every sample, a trained clinical microbiologist needs to set up and read the AST results. This is limiting as not every centre has trained clinical microbiologists especially in less well-resourced countries. It usually takes a minimum of 24hrs for growth of bacterial isolates, and a further 18-24hrs to obtain phenotypic AST results. Classic manual methods are time-consuming with (semi) automated broth microdilution systems (e.g. Phoenix) reducing incubation and hands-on times to 6–12 h. The high degree of standardization in a quantitative manner is an advantage. However, such systems have the limitations, in that they require viable bacteria, the number of bacterial species and antibiotics for which interpretative criteria exist is finite, and the costs especially set-up and maintenance costs. The development of faster AST thus far has not paralleled

technological advances in identification, limiting the potential impact of early pathogen identification.[23] Current phenotypic methods have low resolution in examining resistance determinants and provide less information about resistance gene epidemiology than whole-genome or metagenomic sequencing.

For instance, certain AMR genes are expressed at low levels resulting in *in vitro* phenotypes below clinical breakpoints for resistance and other deviations in detection may occur as well among leading hospital pathogens. An example is vancomycin-resistant enterococci (VRE) carrying *vanB* genes expressing low-level (inducible or not) resistance or carrying *vanA* genes but not expressing clinical resistance at all ('vancomycin-variable enterococci', [VVE]).[24] *VanB*-VRE types have been linked to vancomycin-susceptible outbreak isolates. Both *VanB*-VRE and VVE are increasingly reported across Europe, often not detected properly by routine phenotypic screening methods and automated systems, leading to pitfalls in routine VRE detection and an unnoticed transmission of *van* genes. Furthermore, in enterococci we have observed hospital strains that carry acquired and mobile resistance genes (*optrA, cfr, poxtA*) expressing 'borderline' or 'low-level' resistance to linezolid, meaning the MICs alternate between 4 (S, susceptible) and 8 (R, resistant) mg/l and thus along the clinical breakpoint of >4 mg/l.[25] The clinical consequences of these diagnostic uncertainties are unknown since large scale studies do not exist. Despite the vast diagnostic arsenal for carbapenemase-producing *Enterobacterales*, they can still evade common diagnostic screening systems. Particularly, *OXA-48* and some VIM producers are hardly detectable by common automated AST systems.[26] This is being increasingly associated with difficult to control outbreaks and the spread of strains or plasmids carrying *blaOXA-48*- or similar *blaOXA* genes like *blaOXA-244* that might express variable levels of carbapenem insusceptibilities.[27] An additional diagnostically challenging situation may

result from Enterobacterales with colistin resistance, due to mobile *mcr* genes. Various studies reported *mcr-1*-carrying isolates showing MICs of 2 mg/l, which is below the clinical breakpoint (>2 mg/l) and should be considered in targeted screening approaches requiring further genotyping confirmation.[28]

There are additional barriers in phenotypic AST prediction including the need for specific arrangements such as adding detergents to prevent surface attachment of lipoglycopeptides like dalbavancin, adding Ca^{2+} to measure daptomycin susceptibilities, providing fresh medium and a low oxygen concentration to determine tigecycline susceptibilities, and cefiderocol as a novel siderophore cephalosporin requires iron-depleted media for correct AST determination to mention some of the nuances.

The need for more comprehensive and quicker AST results has boosted the technical advancement for high-throughput technologies to become better developed into automated AST prediction. Where transcriptomics- and proteomics-based technologies are very much in the nascent phase, genomics-based (including metagenomics-based) prediction of AST are much more advanced.[29] It is likely that phenotypic AST continues to remain the gold standard for guiding patient treatment. However, pathways incorporating partly genotypic methods may be increasingly seen.

2.3 Advances in Molecular Approaches for Pathogen Detection and Antibiotic Resistance

The historical challenge in clinical settings has been the swift and accurate identification of bacterial pathogens and their antimicrobial resistance. Traditional methods, while reliable, can be slow and labour-intensive. The introduction of PCR in the 1980s revolutionized the field by allowing faster, direct detection from samples, and reducing laboratory workers' exposure risk. The evolution of molecular techniques, including real-time PCR, multiplex PCR, and microarray technologies, has further transformed bacterial infection diagnostics. These methods allow simultaneous detection of multiple pathogens and have improved sensitivity and specificity. Bacterial resistance is often genetically encoded by several mechanisms including the overexpression or duplication of existing genes, point mutations or the acquisition of entirely new genes via horizontal gene transfer from the same or different species. Advancements in DNA technology have facilitated antimicrobial resistance gene identification and characterization in genomes enabling antimicrobial resistance prediction and other complex applications such as surveillance or transmission. Genotypic AST methods include PCR, DNA microarray, loop-mediated isothermal amplification (LAMP) and next-generation sequencing based methods. These methods are effectively direct methods that reduce the need for bacterial cultures and long incubation requirements. However, understanding their strengths and limitations remains key for effective clinical application.

2.3.1 Evolution and Impact of Polymerase Chain Reaction (PCR) Techniques

The introduction of the Polymerase Chain Reaction technique in 1985 was a pivotal moment in the realm of molecular biology, fundamentally changing the landscape of pathogen detection. Devised by Kary Mullis, PCR was designed to amplify minuscule amounts of DNA, converting it into quantities sufficient for detailed analysis.[30] Utilizing short DNA fragments known as primers that are complementary to the target sequence, and a DNA polymerase enzyme that can withstand high temperatures, PCR allows for the exponential amplification of a specific DNA sequence.[31,32] PCR works by cycles of denaturation, annealing of the primers, and elongation of the primers by a thermostable DNA polymerase in a compatible buffer containing nucleotides, ions, and other substances. Each cycle of amplification doubles the target DNA molecule.

Applied to infectious disease diagnostics, the process involves collecting a patient sample, extracting and purifying total DNA or RNA (which includes genetic material of potential pathogens), and designing specific DNA primers that correspond to the pathogen's genome. The PCR process involves creating a master mix of necessary reagents, and through a cycle of heating and cooling, known as thermocycling, the targeted DNA or RNA region is amplified. This amplified sequence is then analysed; its presence confirms infection. The amplified target can be confirmed for the presence of resistance genes through multiple technics including electrophoresis, restriction fragment-length polymorphism, southern blotting, single-strand conformation polymorphism, DNA fingerprinting, molecular beacons, and other DNA sequencing analysis methods. Additionally, PCR can provide insights into the pathogen's genotype, strain, and antimicrobial resistance genes, facilitating personalized treatment strategies.

Variations of PCR, like real-time and multiplex PCR, have enhanced its speed, specificity, and efficiency, allowing simultaneous detection of multiple pathogens and resistance genes. Thus, PCR has significantly advanced infectious disease diagnostics and is one of the most reliable, rapid, and sensitive molecular tools for quantification of bacterial genes.

2.3.2 Variants and Innovations in PCR Techniques

Since its inception, PCR has undergone significant refinements and has been adapted into a range of specialized techniques tailored to diverse applications. These include methods like real-time PCR (qPCR), reverse transcription PCR (RT-PCR), multiplex PCR, nested PCR, and digital PCR (dPCR), each bringing unique advantages to the study of infectious diseases.

Real-time PCR, or qPCR, adds a layer of quantification to the conventional PCR process. It uses fluorescent markers to visualize the amplification of the target sequence in real time. This approach enables quantification of the initial concentration of the target, offering valuable insights into disease load or bacterial count in infection studies, and has significantly streamlined diagnostics due to its rapid turnaround time and high sensitivity. Reverse transcription PCR is another variant that has proved invaluable, particularly in the study of RNA viruses like SARS-CoV-2. RT-PCR first involves the synthesis of complementary DNA (cDNA) from an RNA template, a process known as reverse transcription. This cDNA can then be amplified using standard PCR, enabling the detection and study of RNA-based organisms.

Multiplex PCR has revolutionized the capacity for broad-spectrum detection by allowing simultaneous amplification of multiple target sequences in a single reaction. This has made it possible to screen for multiple pathogens or multiple resistance genes in a single test, providing a comprehensive snapshot of infection in a time-efficient manner. Nested PCR, involving two rounds of PCR using different primers, has greatly improved the sensitivity and specificity of PCR by reducing the chance of non-specific amplification. This technique is particularly useful in situations where the target DNA concentration is very low or is mixed with a large amount of non-target DNA. Lastly, digital PCR (dPCR) represents a more recent innovation that provides absolute quantification of DNA or RNA molecules without the need for a standard curve. This is achieved by partitioning the PCR reaction into thousands or even millions of individual reactions, with each containing either one or no target sequence. dPCR is exceptionally useful for detecting low-abundance targets, validating next-generation sequencing data, and performing precise quantification of gene expression levels.

In summary, the evolution of PCR techniques has greatly expanded our ability to detect, quantify, and study pathogenic organisms. The continued development of these technologies, and their increasing integration into the clinical laboratory setting, is expected to further improve infectious disease diagnostics and our understanding of microbial pathogenesis.

2.3.3 The Rise of Isothermal Amplification Methods in Rapid Diagnosis

Isothermal amplification methods have gained increasing importance as rapid and accurate diagnostics for species and antimicrobial resistance profiles. These methods

rely on the amplification of nucleic acids at a constant temperature, circumventing the need for precise temperature cycling in traditional PCR. In LAMP, the gene of interest is amplified at a constant temperature of 60–65 °C using a *Bst* DNA polymerase instead of *Taq* polymerase because of strong strand displacement activity.[33] The simplicity, speed, and cost-effectiveness of these methods make them particularly suitable for resource-limited settings and point-of-care testing. Several techniques fall under the umbrella of isothermal amplification, including Loop-mediated Isothermal Amplification (LAMP), Nucleic Acid Sequence-Based Amplification (NASBA), and Rolling Circle Amplification (RCA). Among these, LAMP has gained considerable attention for its robustness, rapidity, and high sensitivity and specificity.

2.3.3.1 Species Identification

LAMP has been extensively used for species-level identification of numerous pathogens, including bacteria, viruses, and parasites. For example, LAMP assays have been developed to rapidly identify *Escherichia coli*, *Staphylococcus aureus*, and *Mycobacterium tuberculosis* with high accuracy.[34,35] The method's ability to detect even low levels of pathogen DNA quickly makes it a valuable tool for early diagnosis and treatment.

2.3.3.2 AMR Detection

Beyond species identification, isothermal amplification techniques have been employed in detecting AMR. For instance, LAMP has been utilized to identify *methicillin-resistant Staphylococcus aureus* (MRSA), a major cause of hospital-acquired infections.[35] Furthermore, LAMP has been used to detect resistance genes in pathogens like the carbapenem-resistance genes in *Enterobacteriaceae* and the *mcr-1* colistin resistance

gene in *Escherichia coli*.[36] Though isothermal amplification methods offer many benefits, they are not without limitations. Potential challenges include the risk of non-specific amplification and primer-dimer formation. Despite these challenges, the simplicity, speed, and cost-effectiveness of these methods continue to drive their adoption and innovation in the realm of infectious disease diagnostics.

2.3.4 Application of Syndromic Panels in Clinical Diagnosis: The Case of BioFire

Syndromic panels have become pivotal in the evolution of clinical diagnostics, providing a comprehensive and swift approach to identify infectious pathogens. Various diagnostic companies across the globe have developed these multiplex PCR technologies to facilitate high throughput analysis. These include the Verigene System by Luminex Corporation (Austin, Texas, USA), the ePlex System by GenMark Diagnostics (Carlsbad, California, USA), and the QIAstat-Dx System by QIAGEN (Hilden, Germany) among others.[37] Among these, the FilmArray system developed by BioFire (Salt Lake City, Utah, USA), a subsidiary of bioMérieux (Marcy-l'Étoile, France), has made a significant impact in the field.[37] My discussion here particularly focuses on BioFire's technology and its application using the blood panel.

The BioFire's FilmArray system merges several steps—sample preparation, amplification, detection, and analysis—into a single, automated process that delivers results in about an hour. The core of the process lies in the FilmArray pouches, which house all the reagents necessary for the multiplex PCR. The testing procedure involves sample injection, nucleic acid purification, reverse transcription for RNA targets, a first-

stage PCR, and a second-stage, nested multiplex PCR. This latter step allows for the simultaneous detection of a multitude of targets, contributing to the high throughput and rapid turnaround times. The BioFire Blood Culture Identification (BCID) Panel can identify a comprehensive range of pathogens—both bacterial and fungal—and antimicrobial resistance genes directly from positive blood cultures in about an hour.

The incorporation of syndromic panels into clinical diagnostics presents considerable advantages. Primarily, these panels offer rapid diagnostics, a critical factor in the efficient management of infectious diseases. Additionally, their ability to detect an extensive array of potential pathogens in one single assay significantly enhances patient outcomes. A recent study conducted across intensive care units in the United Arab Emirates demonstrated the impact of this approach: the 30-day mortality rate fell by almost 10% when routine methods were supplemented with the BioFire Blood Panel.[38] Syndromic panels further offer an automated process, which drastically reduces hands-on time, thereby decreasing the potential for human error. Nonetheless, it's crucial to recognize the limitations of syndromic panels. Their use can be expensive, particularly if employed as a first-line test. Furthermore, while some panels can detect certain antimicrobial resistance markers, they cannot be a replacement for susceptibility testing due to the limited array of gene markers they can identify. Future research and cost-effectiveness studies are needed to ascertain the best ways to integrate these panels into diagnostic algorithms.

2.3.5 Emergence of Other Technologies: The Case of DNA Microarrays and Beyond

Over the years, new methods continue to emerge, expanding the landscape of molecular diagnostics. One such approach is the use of DNA microarrays, also known as gene chips or biochips. DNA microarrays represent a significant leap in high-throughput diagnostics. They are miniaturized platforms that allow simultaneous detection and analysis of thousands of genetic markers in a single assay.[39] Each microarray chip contains a grid of microscopic DNA spots, with each spot hosting a unique DNA sequence that complements a gene or nucleotide sequence of interest. When a fluorescently labelled sample of DNA or RNA is exposed to the chip, it binds - or hybridizes - to the complementary DNA spots. The pattern of hybridization can then be scanned and analyzed to identify which sequences are present in the sample.

This technology can help detect the presence of a wide array of pathogens or their antimicrobial resistance genes in a single patient sample, thereby offering comprehensive diagnostic information that can be crucial for targeted therapy.[40] For instance, the determination of rifampin and isoniazid resistance (*rpoB, katG* and *inhA*) in *Mycobacterium tuberculosis* has been demonstrated successfully through DNA microarrays and chips.[41,42] While DNA microarrays hold significant promise, they are not without their limitations. They can be costly, complex to use, and require advanced equipment and expert interpretation of results.

Beyond DNA microarrays, the field continues to advance with the development of technologies like CRISPR technology for disease detection. Platforms like CRISPR-Cas13 and CRISPR-Cas12 are being harnessed to create diagnostic tools that are highly specific, rapid, and potentially usable in point-of-care settings.[43] While these

technologies are still in their relative infancy and have challenges to overcome, their potential to transform the diagnosis and management of infectious diseases is undeniable.

2.3.6 Conclusion

The past decades have witnessed an impressive evolution in the field of molecular diagnostics for infectious diseases. Traditional culture-based methods, while still essential in many aspects of diagnostics, have been significantly complemented, and in some cases, supplanted by rapid, sensitive, and specific molecular techniques. PCR and its variants, isothermal amplification methods, syndromic panels such as BioFire, DNA microarrays, and more recent technologies like CRISPR-based diagnostics, have significantly enhanced our ability to identify pathogens and their antibiotic resistance profiles. However, while molecular techniques have ushered in a new era of infectious disease diagnosis, the quest for a perfect diagnostic tool continues. The next step in this evolution seems to be the emergence and establishment of metagenomics or Next-Generation Sequencing (NGS) technologies. These technologies offer an unprecedented level of genetic detail, metagenomics can provide comprehensive insights into the identity, diversity, and resistance profiles of pathogens, thereby pushing the boundaries of molecular diagnostics. In the following section, I delve into the principles, advantages, applications, and prospects of metagenomic technologies in the field of infectious disease diagnostics.

2.4 Metagenomics approaches

The advent of metagenomics signifies a pivotal shift in the field of infectious disease diagnostics. It propels us beyond traditional culture or PCR-based methods and into an era of direct and comprehensive genomic analysis of all microorganisms present within a clinical specimen, such as blood culture, sputum, or cerebrospinal fluid. Metagenomics, by sequencing millions of DNA fragments in parallel, offers an unprecedented overview of the entire microbial landscape within a sample. This capability renders metagenomics not merely a tool for identifying known pathogens, but also a powerful discovery platform capable of detecting novel, unexpected, or rare infectious agents without any prior knowledge of their genetic composition. Moreover, metagenomics transcends mere identification of pathogens and delves into a more detailed analysis of the pathogen genome. It can decipher the presence of virulence factors and antimicrobial resistance genes, thus providing invaluable insights into the pathogenicity and resistance profiles of the organisms. This capacity has the potential to guide clinicians in making more informed therapeutic decisions and allows researchers and epidemiologists to better track the trends of antimicrobial resistance.

However, metagenomics is not without its complexities. The process requires meticulous sample preparation, DNA extraction, library construction, and sophisticated computational resources for data analysis. Even small variations in these steps can introduce biases and affect the sensitivity and specificity of the results. Given these intricacies, it becomes apparent that while metagenomics is a powerful tool, its effective integration into clinical microbiology requires a deep understanding of its components and potential pitfalls. In the subsequent sections, we will delve deeper into the specific

components that are integral for optimizing the performance of metagenomic approaches in infectious disease diagnostics.

2.4.1 Sequencing technologies

Remarkable progress has been made in the development of sequencing technologies since the completion of the Human Genome Project in 2003 which has led to a 50 000-fold decline in sequencing costs per megabase and led to an increased number and diversity of genomes sequenced from microbes to humans. However, for sequencing technology to be useful clinically it will require rapid turn-around times with high accuracy. Here I discuss sequencers used in current metagenomic literature and their specific application in infectious diseases. For information on other sequencing platforms, a review is available by Goodwin S *et al.*[44]

Illumina technology is the most common sequencing technology found in the metagenomic literature which offers high single read accuracy of 99.9% however it is limited by a minimum time per run of at least 4-24 hours before any results are available. Multiple Illumina sequencing platforms are available (iSeq, MiniSeq, MiSeq, NextSeq, HiSeq and NovaSeq) that cater for different applications, throughput, and sequencing requirements. This technology is based on clonal amplification and sequencing by synthesis chemistry. During the process DNA bases are identified and simultaneously incorporated into a nucleic acid chain. As each base is added to the growing strand, it emits a unique fluorescent signal which is used to determine the order of the DNA sequence. Multiplexing is also permitted however may introduce barcode index switching errors of 0.1–2% especially with patterned flow cell systems which may result

in higher false results. If used in a clinical setting, samples would all have to be collected in large batches before sequencing to optimise cost and additional samples received would not be able to be added on for many hours until the platform has completed the initial run.

The ION Torrent offered by Thermo Fisher Scientific (Waltham, MA) uses a semiconductor to detect the release of H+ ions in a microwell when flooded with sequential single species of deoxyribonucleotide triphosphate (A, C, G, T) during the sequencing by synthesis process. It offers a single read accuracy of 99.6% and is fast with a 2-7 hours run time making it a worthy candidate for clinical use. However, its major drawback is the poor accuracy in homopolymer regions which lack single-base accuracy in measuring homopolymers larger than 6–8 bp.[44]

Oxford Nanopore Technologies (Oxford, United Kingdom) offers portable and benchtop sequencing devices (Flongle, MinION, GridION, PromethION). Unlike most other sequencers, it does not monitor by DNA hybridisations of nucleotides. Instead, DNA is passed through a protein pore while a current is applied. As DNA enters the pore though the action of a secondary motor protein, a voltage blockade occurs that modulates current changes across the pore. These characteristic shifts in voltage are used to determine the DNA sequence. Oxford Nanopore Technologies provide real-time results but are limited to a single read accuracy of ±92-97% which can generally be compensated for by using longer read lengths and a higher depth of sequencing. Multiplexing up to 12-96 samples is possible on each flow cell. The flow cells can also be washed, stored, and reused allowing clinical samples to be potentially processed on demand with reduced cost. Furthermore, Nanopore sequencing does not require prior PCR amplification if the

sample input is >400ng which may reduce sample-to-sequencing duration for selected samples.

It is yet unclear which sequencing platform will emerge as the ideal device for metagenomic sequencing. Successful platforms will need to be scalable, flexible for different batch sizes, and offer stable and predictable performance. Current real-time sequencing devices such the Oxford Nanopore technology yield rapid data, which may be a significant development for infection diagnostics, but are limited by relatively high per read error rates which can be problematic especially if only low coverage depth is achieved of the pathogen genome. Although improvements in error rates have occurred during the writing of this thesis.

2.4.2 Library preparation

Library preparation is the process of preparing extracted DNA or RNA to be compatible for the sequencing device to be used. This process may grossly vary between different types of technologies and subtly within the same type of technology depending on the specific need and context of the sequencing run. Assessing the quantity and quality of nucleic acids and performing clean up steps, for example, magnetic DNA purification before and after library preparation removes fragmented, low-purity and degraded nucleic acids and yields high-quality sequencing data. However, library preparations also introduce a degree of representation biases which are influenced by factors, for example, by high or low GC content, the number of PCR cycles, thermocycling conditions and buffers.[45,46] In just 1μg of DNA nearly 5 x 10^{14} base pairs are present, while in contrast a sequencing run on Illumina HISeq 2500 processes about 10^{11} base

pairs, approximately 0.02% of the original content.[47] Hence, library preparations subsample the original nucleic acid content as the amount of biological DNA data is much larger than what can be practically sequenced even by the latest high-throughput sequencers possibly introducing some bias.[47] However, metagenomic sequencing is probably the most unbiased approach to library preparation as the original DNA is randomly subsampled, providing a uniform library that covers all genomic regions based on their prevalence, at least conditional on the relative efficiency of DNA extraction.[47]

In general, most platforms follow the following steps for library preparation with the exception of Illumina's Nextera prep: (1) fragmentation, (2) end-repair, (3) phosphorylation of the 5′ prime ends, (4) A-tailing of the 3′ ends to facilitate ligation to sequencing adapters, (5) ligation of adapters, and (6) cycles of PCR s to enrich for products that has adapters ligated to both ends. Illumina's Nextera tagmentation technology uses a transposase enzyme which simultaneously fragments and inserts adapter sequences into dsDNA which has several advantages including reduced sample handling and preparation time.[46] In the Ion Torrent workflow, blunt-end ligations are added to different adapter sequences. Detailed information on library preparation methods can be found in a review by S. Head *et al.*[46]

Illumina sequencing using the Nextera XT DNA Library Preparation Kit has been commonly used in metagenomic applications, with an input DNA requirement of 1ng. Standard Oxford Nanopore methods require 400ng of DNA, while a combined PCR-amplification library preparation can be used requiring only 10ng of DNA. Choice of library preparation has varied in literature and is based on suitability for purpose of study taking into consideration trade-offs among improving sequencing quality, reducing

preparation time, maximising throughput, and utilising low input DNA. While the optimal library preparation would uniformly cover genomic samples and eliminate bias, competition has driven the quality and performance up and the price and complexity of use down.

2.4.3 Detecting species with metagenomics

Koch's postulates, developed in the 19th century by German physician Robert Koch, represent a cornerstone of microbiology, guiding the establishment of causal relationships between specific microbes and diseases. Central to these postulates is the culture of pathogens, wherein a suspected pathogen is isolated from a diseased host and grown in a controlled laboratory setting. However, these postulates present major limitations in the contemporary understanding of pathogens, particularly those that cannot be conventionally cultured. A reframed set of criteria that is more attuned to modern methodologies, such as metagenomics, has been proposed by Fredericks and Relman.[48] This reconsideration maintains the core intent of determining causal relationships but modifies the methodology to accommodate advancements in technology and our understanding of microbial ecology. It acknowledges the role of molecular methods for detecting a specific microorganism directly from the site of the disease, circumventing the necessity for culturing the pathogen. If metagenomics is shown to be diagnostically accurate, it becomes vital to understand the clinical implications of positive results, especially when all other conventional assays, including culture-based ones, return negative results. In this context, an adapted perspective on Koch's postulates could prove instrumental in deciphering the correlation between metagenomic detections and their clinical significance, thereby fortifying the role of this

technology in infectious disease diagnosis. This will necessitate large-scale interventional and longitudinal studies to establish the implications for patient outcomes.

2.4.3.1 Species detection performance

Metagenomic sequencing has proven to be a powerful tool in identifying a wide array of pathogens in various clinical scenarios, often outperforming conventional diagnostic methods. Studies to date have utilised metagenomic sequencing as a tool for pathogen-agnostic diagnosis of clinical infection on a variety of clinical samples including from blood, urine, respiratory samples, cerebrospinal fluid, orthopaedic sonication and synovial fluid, intraocular fluid, and cardiac valvular tissue. Performance in these studies have varied which may reflect the number of organisms present in each sample type and the relative abundance of human cells and any other bacteria present. Among CSF studies, Wilson, M *et al.* report 55% sensitivity and 98% specificity, possibly reflecting a greater effort to obtain a diagnosis in the reference standard used for comparison including culture, multiplex molecular diagnostics and obtaining a variety of samples including tissue biopsy and abscess fluid.[49] Furthermore, conservative preestablished bioinformatic thresholds were used which resulted in negative results even though species-specific reads might have been used to identify the appropriate organism at lower thresholds. Among studies on blood, Somasekar *et al.* reported a 96% sensitivity and 100% specificity, while a study from Rossoff *et al.* reported a 92% sensitivity and 64% specificity.[50,51] Studies with low sample sizes generally performed worse, possibly reflecting the most limited experience developed in these studies.[52–55] Among orthopaedic studies, Street *et al.* reported a 88% sensitivity and 88% specificity, while another study by Ruppé E *et al.* reported lower sensitivity likely arising from low DNA extraction yields and frequent freeze/thaw cycles of samples reducing DNA stability.[56,57]

CSF studies performed better compared to blood and orthopaedic studies in part as preestablished conservative thresholds were used, trading off sensitivity for higher specificity.[49,58] Specificity performance varied substantially across orthopaedic studies due to non-standardised thresholds for determining a negative result in the presence of high contamination and low DNA extraction yields, resulting in many organisms of unknown clinical significance. [56,57] A comprehensive analysis of species performance metrics can be found in the subsequent chapter, featuring a meta-analysis of various studies that have utilized metagenomic sequencing.

2.4.4 Detecting Genotypic Resistance with Metagenomics

Developments in whole genome sequencing and bioinformatic tools have made using this approach feasible for AST.[59] Third generation systems such as Illumina MiniSeq, Pacific Biosciences PacBio, or Oxford Nanopore MiniON/PromethION can provide high coverage at great speed. It has been well demonstrated that for *Mycobacterium tuberculosis* WGS can provide a rapid and comprehensive view of the genotype of the organism, and thus enable reliable prediction of resistance, including to second line drugs, within a clinically relevant timeframe.[60] Similar studies demonstrate this for other species such as *Staphylococcus aureus, Escherichia coli* and *Klebsiella pneumoniae* among others.[61,62]

Predicting resistance from metagenomic samples is a challenging feat as meta-assemblies can contain overlapping elements leading to complex assembly structures that make it challenging to identify true resistant genes that can be interpreted in the right context. Although most tools require assembly to identify resistant genes, tools

have been developed using k-mer alignment to allow for direct and fast processing of reads.[63] Despite the number of studies concerned with species identification, there have been limited species agnostic studies.

Genomic AST prediction works well for well-known antibiotic classes with a defined genotype, for instance, those with single-nucleotide polymorphism gene presence. However, the genetic mechanism and profile for the resistance of all bacteria is not yet defined. Furthermore, many novel AMR phenotypes, some to last resort antibiotics, may lack a distinct and clear genetic correlate making genetic predictions less accurate. Hence, further research is required to understand these resistance mechanisms and complex gene interactions to fully harness the potential of genotyping. A comprehensive analysis of AMR performance metrics can be found in the subsequent chapter, featuring a meta-analysis of various studies that have utilized metagenomic sequencing.

2.4.4.1 AMR predictive performance in metagenomics

While species identification may change therapy in some settings, for example, as a result of intrinsic resistance or local epidemiology, rapid and precise antimicrobial resistance prediction linked to specific pathogens, will ultimately be required for mainstream adoption of clinical metagenomics. In a study by Anson *et al.*, categorical agreement of resistance prediction was 96%, however, very major errors was 24% and major errors was 3%. In another study by Charalampous *et al.*, categorical agreement of resistance prediction was 77%, however, very major errors was 33% and major errors was 0%. These studies performed sample pre-selection, however, there are no known species agnostic studies that have attempted resistance prediction and compared this to phenotypic results according to a recent meta-analysis in the field.[64] The FDA requires

major errors <3% and the 95% CI on the very major error rate to be ≤7.5% at the upper limit and ≤1.5% at the lower limit.[65] Several improvements are likely needed to improve performance, including increased yield of pathogen DNA, better understanding and cataloguing of the mechanisms underlying antimicrobial resistance, and finally optimised algorithms to detect these within metagenomic data. For example, large scale efforts made in TB to improve genomic prediction of antimicrobial resistance[66] may need to be replicated for other organisms and specific software, such as DeepARG is under development.[67] For plasmid-mediated resistance, this is particularly challenging which may be difficult to assign to a specific species from metagenomic data.

2.4.5 Databases

Databases are required both for species identification and for antimicrobial susceptibility prediction. Database selection and curation is vital for optimal performance. For species identification this is challenging as databases are continually updated with new strains and species and often rely on correct taxonomic identification and annotation provided by the submitter as in the case of GenBank, which for more than a quarter of a century has been the mainstream genomic repository. GenBank contains complete and draft genomes; the latter can be problematic as the chromosomes are fragmented into many pieces and have contamination sequences that may be included which may even match in metagenomic projects as they may have the similar contamination of *E. coli*, human DNA etc. RefSeq provides a curated dataset of GenBank by passing sequences though automatic filters. Currently, the number of genomes that refseq as a proportion of all sequences in genbank for bacteria is 18 788 326/59 528 335, for viral genomes 12 208/33 456 and only 474/43 387 for fungal genomes (as of 16 Oct 2019). Various reasons exist why genomes may be excluded from RefSeq including highly fragmented assemblies or

if derived from a metagenome. RefSeq records are stored by NCBI and hence can be updated independently to maintain annotations and updated information. Other curated databases do exist, but lack such vast data, however these may be high-quality such as FDA-ARGOS or the FDA Reference Viral Database. A combined approach that utilises annotated sequences from multiple databases may optimise the sensitivity and specificity of species identification.

Antimicrobial resistance prediction largely relies on curated antimicrobial resistance gene databases that link known genetic determinants of resistance to phenotypic resistance, generally from existing literature and enriched by large-scale genome-wide association studies. Public databases vary in the variety of information and the scope of resistance mechanisms covered. General databases such as the Comprehensive Antibiotic Resistance Database (CARD), Antibiotic Resistance Genes Database (ARDB) or the active Antibiotic Resistance Gene Annotation (ARG-ANNOT) cover broad-spectrum resistance genes and mechanisms; while specialised databases such as the Lahey database of β-lactamases, Mykrobe database for *S. aureus* and *M. tuberculosis* or u-CARE for *E.coli* cover in-depth information for species or specific gene groups. While general databases offer a comprehensive range of data, specific databases offer a valuable insight to understand resistance in the context of phylogenies. For instance, the CRyPTIC Consortium demonstrated resistance prediction for all first-line anti-tuberculosis drugs with over 90% sensitivity and specificity. Such predictions are highly effective especially where Horizontal Gene Transfer (HGT) events, a process where an organism transfers genetic material to another organism that is not its offspring, are rare and resistance originates mainly from chromosomal mutations.[68] As the quantity of general and species-specific databases increase, it is important to select the optimal database for AMR prediction. For agnostic diagnosis, this may involve an algorithm

which potentially selects the optimised database depending on the species initially identified. For optimal performance and confidence, databases should be continually updated and reviewed. With the growing threat of AMR, there is a need for databases to improve genotypic prediction especially for organisms such as carbapenem-resistant Enterobacteriaceae or Methicillin-resistant *S. aureus* that pose grave public health treats. However, this is challenging due to the presence of HGT, though new long-read sequencing technology may offer insight into plasmid structure and function that may ultimately improve AMR predition.[69]

2.4.6 Bioinformatic approaches

As the quantity of microbial genetic data increases, bioinformatic approaches need to be optimised to handle large datasets of complex metagenomic data while minimising computing power and memory usage. Generally, data that has been basecalled first goes through pre-processing steps which include the removal of low-complex and low-quality sequences. This step may also include the removal of human or non-target DNA. Pre-processing is required to improve the quality and confidence of variants in downstream analysis as this may suffer from base content bias, overrepresented sequences and/or adaptor contamination. Furthermore, library preparation and the sequencing process also introduce errors which may lead to inaccurate representations of the original nucleic acid sequences.

Although the vast majority of sequences thereafter undergo read-based processing, some may undergo assembly-based approaches if enough sequencing depth is reached. Although, read-based approaches are much faster and less computationally demanding, assembly-based approaches can identify novel resistance genes and capture regulatory and mobile elements. Assembling metagenomic data is much more complicated than

single-isolate assembly, as the unknown abundance of different organisms with unknown relatedness has to be accounted for.[70] Repeat sequences remain a challenge even for single genomes and their effect in metagenomics is further amplified by cross-organism and uneven levels of representation.[70] Thus conventional single-genome assembly cannot be applied directly to metagenomics and several metagenomic specific assemblers such as IDBA-UD, MEGAHIT, MetaSPAdes and MetaVelvet have been developed to overcome these challenges.[71]

In most cases species and resistance classification are then performed by read-based approaches. Numerous classification software exists with Kraken, BLAST and SURPI being the most popular.[72] These software tools can be broadly classified by their modus operandi into alignment based, k-mer based, gene-marker based or a combination of the above. Previously, BLAST was the best strategy to compare a sample to the reference database; however, with the size of the databases rapidly growing it has become computationally infeasible which has led to the development of other faster albeit less sensitive methods. Studies have attempted to benchmark the performances of the different metagenomic classification software however are limited by using simulated data.[73] This has prompted the development of The Critical Assessment of Metagenome Interpretation (CAMI) which is a new community-led initiative designed to help tackle these problems by aiming for an independent, comprehensive, and bias-free evaluation of metagenomic methods.[74] It is crucial to note that the primary objective of many databases is the cataloguing of resistance genes, rather than the direct prediction of AMR in the manner of a clinical laboratory. This focus offers an extensive record of resistance genes, providing a crucial resource for understanding the genetic basis of AMR. However, such databases may not systematically permit the elucidation of relationships between minimum inhibitory concentrations and specific genes. This is a critical aspect

of predicting clinical resistance, as the presence of a resistance gene does not always equate to phenotypic resistance. Despite this, several dedicated studies have pursued this path of inquiry, developing models to correlate genotype with MIC, and thus resistance phenotype. Notable examples of this approach include investigations into *tuberculosis* and *gonorrhoea*, among others.[66] As such, the field might benefit from a broader implementation of this strategy, facilitating a more nuanced understanding and prediction of AMR grounded not only in the presence of resistance genes but also their correlation with phenotypic resistance as represented by MIC values.

2.4.7 Nucleic Acid extraction

Nucleic acid extraction is a critical step in metagenomic workflows. However, a major bottleneck to developing a unified or optimal metagenomic approach exists as there are numerous possible methods to extract DNA that may all have distinctive outputs. Furthermore, as there is no gold standard method for DNA extraction, it is difficult to determine the true diversity of a microbial community. The ideal nucleic acid extraction method would produce the same DNA composition directly representative of the sample and with ample DNA yield. Studies have suggested using several different extraction methods to test for some of the observable loss that may occur or use a diverse mock community to serve as a control.[75] This is particularly important to prevent biasness, where an extraction method may potentially perform poorly, for instance, in Gram negative bacteria compared to Gram positive bacteria, or vice-versa which may also hold true for other unaccounted features.

The choice of extraction method is dependent on the purpose and circumstance of the study and generally involves a trade-off amongst efficacy, reduced extraction time, ease of extraction, laboratory resources available, DNA yield, cost and/or human depletion requirements. Although there is an evident need to establish optimal extraction methods, only one study with a limited number of samples has compared different methods with true clinical samples for diagnostic metagenomics. Anson *et al.* found the BiOstic Bacteraemia kit (MoBio, Qiagen, USA) to be superior to the MolYsis Plus kit (Molzym, Germany) and Amplex Hyplex Quickprep (Amplex Biosystems, Germany) both in DNA yield (being 430x higher) and having superior *Staphylococcus aureus* extraction from liquid blood culture.[3] Although studies have successfully extracted DNA for clinical metagenomics, further comparative studies of DNA extraction quality and quantity from different methods are required to determine the optimal approach.

The removal of human sequences from metagenomic sequence data can be a lengthy process and require significant computational power which may not be available at most laboratories. Moreover, a high background of human DNA may also impede sensitivity to detect low abundance pathogens in clinical specimens. In real-time sequencing, high human DNA content may limit the coverage of microbial genomes and as a function of time may impair prompt species and AMR prediction for clinical decision making. While numerous human depletion laboratory methods have been evaluated and are used pre-extraction, novel methodologies are still considerably needed that are able to extensively deplete human DNA. One such treatment that may hold promise is saponin, which is based on a mild detergent that is typically used to selectively permeabilize mammalian cells, followed by a nuclease digestion to degrade away the released human DNA. This has been reported to reduce human DNA by >95% however may perform less well on Gram-negative bacteria.[76,77] Furthermore, DASH (Depletion of Abundant

Sequences by Hybridization) may possibly remove up to 99% of unwanted high-abundance types of DNA prior to sequencing however this has only been performed on RNA.[78]Although numerous studies in literature utilised commercial kits which are readily available and quality assured, non-commercial extraction methods may be crucial for low resourced settings.

Samples post-extraction may often require an extent of amplification prior to sequencing dependent on kit requirements. PCR amplification is capable of 10^6-10^7-fold amplification of a single copy of template DNA, thus making minimal contaminants a problem.[79] Multiple displacement amplification is emerging as a new amplification technique that generates larger sized products with a lower error frequency, however challenges do exist such as GC bias that may lead to uneven genome coverage.[80]

2.4.8 Quality controls

Quality control is imperative for reliable metagenomic results and will be essential for regulatory approval for clinical applications. Negative controls are important to reduce false positives and to detect any contamination during the workflow, which is especially a problem encountered with low DNA concentrations used in metagenomics. Internal or spiked controls and positive controls are both useful to reduce false negatives. Positive controls ensure that the workflow is working and correctly configured, while internal controls may have a special role in the quality control of species representation and in demonstrating the individual samples have sequenced adequately. For example, ZymoBIOMICS have developed the first microbial community standard that consists of Gram-negative and Gram-positive bacteria and yeast with varying sizes and cell wall composition, which can potentially be used as a positive control.

The presence of controls does not ensure perfect results but rather mitigates against the risk of false positive and negative results. Some samples may still be contaminated during the extraction process even if control samples are not contaminated and hence the number of controls in proportion to the number of samples and batches processed is a significant consideration. It is recommended that all metagenomic studies consider carefully using all three forms of controls to detect and reduce errors.

2.4.9 Contamination

The ability of metagenomics to detect microbes agnostically and at very low concentrations is a major advantage however this also comes with the ability to detect contaminants much more easily which poses a major challenge, especially after DNA amplification. Contaminants may be introduced into the sample from normal flora during the sample collection process, for example flora introduced from the skin during venepuncture, or contaminants may be introduced during sample processing by contaminated reagents, inexperienced laboratory practice or by the laboratory environment. Even minute amounts of exogenous DNA can lead to an inaccurate result, especially when the initial yield of nucleic acid is low and subsequent PCR is used. This highlights the need for skilled personnel and meticulous workflow processes including the exclusive use of DNA-free reagents, a sterile and nucleic acid depleted work environment and adequate quality controls. Meticulous laboratory practice is required to minimise contamination risks. However, despite this some residual contamination is likely. Currently in literature there is no convergence on how to exclude contamination and confirm the presence of infection bioinformatically.

External contamination occurs from the outside environment and internal from cross-contamination from other samples during processing. A common approach is the removal of sequences that appear in negative controls; however cross-contamination often leads to true species sequences being found in negative controls. One refinement to this approach would be to remove these sequences below a determined abundance threshold, however relative abundance thresholds also remove rare features truly present in the sample and do not remove abundant contaminants. Other methods include excluding sequences or organisms previously identified in literature as known contaminants.

Mixture models have been used to assign the complete collection of sequencing reads while using an expectation–maximization algorithm to compute the maximum likelihood estimate of each taxon abundance.[52] Thereafter a Poisson model can be used, parameterized by the estimated background, to estimate whether levels observed in the samples exceeded those expected to originate from the environment alone determining true positivity.

Ideally a complex model that uses a combination of absolute reads, relative abundance, sequences found in negative controls, DNA concentration and known contaminants found in the literature to assign the probability of an organism being clinically relevant or a contaminant in a run may be warranted. Each organism may have a different threshold for it to be determined to be relevant which may also include the clinical relevance the individual organism may have. A database of commonly observed organisms from normal flora and the laboratory environment should be maintained;

while only organisms not on this list are reported or which have a significant higher threshold for reporting if they are clinically significant.

2.4.10 Outbreak investigations

Metagenomics may also offer enhanced capabilities during outbreak investigations, however this is still an emerging field.[81] Such outbreaks can begin in size from a few individuals, a family outbreak, or a hospital ward to epidemics across countries and continents. Studies have shown that phylogenetic tree reconstruction using direct-from-sample metagenomics can yield the same results as from sub-culture of pure isolates and whole genome sequencing.[82,83] Sequencing, including metagenomics, can deliver far greater resolutions for outbreaks than traditional methods by generating an index of genetic polymorphisms, especially single-nucleotide polymorphisms (SNPs) which can be used to characterise outbreaks. It also links genetics to epidemiology by providing insights into the evolution, structure and content of genomes including information on resistance and virulence which may be used to evaluate the nature of the outbreak.[84] Furthermore, its agnostic ability allows us to detect various unsuspecting threats and even those from an accidental release of cultured or engineered organisms or from deliberate bioterror attacks. Although numerous challenges exist from poor genome coverage in samples to devising agreeable operating processes during rapid innovation, metagenomics may prove a vital tool for outbreak surveillance and response. Accredited workflows that can be reliably used by non-expert users will be needed for metagenomics to be adopted widely. This will also involve adapting and developing laboratory information management systems to handle different workflows and classes of data.

2.4.11 Cost and time of metagenomic workflows

Various studies have shown that inadequate antimicrobial therapy is an independent risk factor for death among critically ill patients with severe infection.[85,86] In severe infection, it has been demonstrated that a time interval of >1 day before initiation of appropriate antimicrobial therapy is associated with a 2-fold increase in the risk of mortality.[87] Clinical metagenomics besides identifying fastidious organisms not identified by culture, may provide a species and antimicrobial resistance results within 4-7 hours in some cases instead of the conventional 24-72 hours with culture.[64] However, practically this may be limited by the number of times that a lab is able to perform sequencing runs within given time intervals.

In terms of cost, Charalampous *et al*[77] estimated the reagent cost per sample to be $130 (Nanopore, 6 samples/run) or $546 (Nanopore, 1 sample/run). Votintseva *et al*[88] estimated the cost per sample to be $128 (MiSeq, 12 samples/run), $685 (R9 MinION, 1 sample/run) or $134-229 (R9.4 MinION, 3-5 samples/run). Janes *et al*[89] reported cost per sample to be $132 using the Ion Torrent Proton sequencer. In summary, the average consumable cost per sample for clinical metagenomics is reported to range from $130 when samples are multiplexed up to $685 for processing a single sample at once. Other costs such as laboratory personnel, training costs, data analysis storage and analysis cost also need to be considered. In comparison, the average running cost for blood culture is estimated at <$50 per sample.[90] Although sequencing costs have dramatically reduced, the total reagent costs per sample continue to remain high. The bulk of laboratories lack robotic and automatic equipment for extraction and sequencing and hence rely on manual effort, which may incur considerable labour costs. Furthermore, the capital investment of sequencers and bioinformatic infrastructure including expertise is high

leading to an overall cost of several hundreds to thousands of dollars per sample analysed. However, it is foreseeable that a single, point-of-care tool be developed in the future that may dramatically reduce costs for infectious diseases similarly to the first FDA approved MiSeqDx NGS device that can detect multiple RAS gene mutations for colorectal cancer and variants of cystic fibrosis.[91]

Given that less than 10% of total blood cultures yield positive results[92], and considering the likely low diagnostic yield per patient, a pivotal question emerges: Who would derive the most benefit from clinical metagenomics in the healthcare journey? Indeed, metagenomic sequencing can be performed on positive blood cultures, yet to maximize its potential, sequencing should ideally be done directly from the sample. However, this approach may be feasible yet due to various technical and biological constraints, including low microbial load and overwhelming host genetic material. Exploration into more sensitive methods for amplifying and detecting early microbial growth may offer a solution to these challenges. Such advancements could enhance the feasibility of direct-from-sample sequencing, thereby maximizing the diagnostic yield of clinical metagenomics. The development and incorporation of these more refined methods into clinical practice would ultimately help identify those patients most likely to benefit from this innovative diagnostic approach. A detailed cost-benefit analysis is needed to shed light on whether this technology will be beneficial in what particular way. Implementing metagenomic sequencing in clinical settings requires careful evaluation of cost-effectiveness across varied applications. Although sequencing all blood cultures may yield the most comprehensive pathogen detection, it's likely more cost-effective to use metagenomics selectively for complex cases that defy traditional diagnostics. Hybrid models, particularly useful for critically ill patients with multiple co-morbidities, reserve metagenomic sequencing for challenging cases, allowing faster diagnosis and targeted

treatment initiation. Yet, traditional culture techniques may still be necessary to obtain vital antibiotic susceptibility data. Furthermore, if an existing molecular diagnostic is in place, the transition costs to metagenomics need to be justified by significant improvements in diagnostic yield or patient outcomes. Therefore, an understanding of the cost-benefit dynamics is crucial, including the development of effective triage protocols to identify those patients most likely to benefit from this innovative technology. Ultimately clinical trials are required to determine if metagenomic pipelines compared to conventional diagnostics may result in superior outcomes and/or reduced hospitalisation rates.

2.4.12 Regulation as a medical device

In the realm of clinical diagnostics, it's essential to distinguish between two fundamental regulatory processes: laboratory accreditation and diagnostic certification.

Laboratories are highly regulated with wide-ranging laboratory requirements that apply accreditation standards, such as that undertaken by the United Kingdom Accreditation Service (UKAS) in the UK, which pertains to the establishment and operation of the laboratory itself. In the US, the Clinical Laboratory Improvement Amendments (CLIA) law regulates laboratories.[93] Accreditation is a recognition of the lab's competency in performing tests, the integrity and aptitude of its staff, and its adherence to standardized methods and quality control procedures, as per various international standards.

On the other hand, diagnostic certification refers to the regulatory approval required for a specific in vitro diagnostic device to be utilized clinically within a given region. It

assures that the device in question has undergone stringent testing for safety, efficacy, and reliability before it is brought to market. For an in vitro diagnostic device to be used clinically it would require regulatory approvals in the region of use. Clinical laboratories are highly regulated with requirements that vary by jurisdiction, for example in the US FDA clearance is required while in the EU a CE mark is required for diagnostic platforms. In the UK, post-Brexit a UKCA marking will be required in order to place a device on the Great Britain market from 1 July 2023. The FDA intends to regulate infectious disease metagenomic high-throughput sequencing devices as systems, including all the components necessary to generate a result. These components may include the sampling device, nucleic acid extraction, library preparation, reagents, sequencing, bioinformatic process and database to the final result. FDA has developed a database entitled "FDA-ARGOS" that supplies a set of validated regulatory-grade microbial genomic sequence entries as an alternative comparator method to existing culture or composite methods. Detailed information including quality control, validation and continuous monitoring may be found in the draft guidance (noticed via 81 Federal Register 29869) released by the FDA.[94]

2.4.13 Largest metagenomic studies to date

Up to August 2021, I identified four large diagnostic metagenomic studies with more than 200 clinical samples that compared metagenomic sequencing to conventional methods. Blauwkamp *et al* demonstrated that microbial cell-free DNA (cfDNA) sequencing can reach up to 93.7% agreement with blood culture in a cohort of 350 patients in a clinically useful timeframe with a greater number of aetiological causes of sepsis than the standard-of-care.[95] However, a major limitation is that cfDNA is unable to provide adequate genome coverage for resistance prediction. Wilson *et al* showed that metagenomic sequencing of CSF obtained from patients with meningitis or encephalitis

improved diagnosis of neurologic infections in 22% (13/58) of patients that were not identified by clinical testing and provided actionable information in 7 out of 13 cases guiding treatment, however this was limited by a long turnaround time of 90 hours per sample.[49] Further to this Miller *et al* demonstrated good analytical performance in 95 patients with CSF infections with a positive percent agreement of 81% and negative percent agreement of 99% after discrepancy analysis for pan-pathogen detection.[58] Thoendel *et al* demonstrated that metagenomics is a powerful tool to identify a wide range of prosthetic joint pathogens in 94.8% (109/115) of culture-positive PJIs, with additional potential pathogens detected in 9.6% (11/115) however, the presence of human and contaminant microbial DNA from reagents and other sources was a major challenge due to low input DNA.[96]

2.5 Conclusion

Metagenomics is a revolutionary technology that could potentially disrupt traditional clinical diagnostics with the potential to offer a new reference standard in several fronts. As with most new technologies, its development has remarkably outpaced the regulatory environment with its tangible clinical value and role yet to be determined by robust clinical studies. The advantage that metagenomics promises is the agnostic detection of pathogens, wielding unparalleled and often unexpected results for the benefit of the patient. Furthermore, detecting polymicrobial infections in areas such as in prosthetic joint infections, wound infections and infective endocarditis may offer us a novel insight to treatment for these precarious conditions. Non-pathogenic strains in polymicrobial infection may influence our choice of antibiotics in the future, as resistance genes may be harboured at low-levels that act as a reservoir for resistance development. As the world witnesses the emergence of drug resistance previously unforeseen both in the

proportion and absolute number of resistant pathogens to multiple antibacterial agents[97], antimicrobial resistance is predicted to be the leading cause of death for 10 million people worldwide by 2050, higher than cancer and diabetes.[98] Metagenomics here too offers a strategic advantage. With metagenomic workflows using real-time sequencing, antimicrobial resistance prediction can be completed within 6 hours[99], offering the optimal time to hone in on treatment either at the first or subsequent antimicrobial dose. This by far surpasses traditional culture-based diagnostic times of 48-72 hours extending to 3 weeks for difficult to grow organisms. Furthermore, unlike PCR-based diagnostics, metagenomics may capture the diverse array of insertions, deletions and single-nucleotide polymorphisms (SNPs) that contribute to AMR and hence may have superior predictive value while not limited by the number or range of PCR targets. For example, PCR-based predictions of AMR in *Pseudomonas aeruginosa* and *Enterobacter cloacae* have been challenging due to the variety of resistance mechanisms.

However, in its current state, numerous limitations exist in this technology that prevent its utility at the bedside. In contrast to human sequencing diagnostics, infectious disease sequencing diagnostics generally require quick, accurate and actionable results, often within hours, as delayed or incorrect results may result in deaths. The broad range of clinical specimen types together with the large diversity of infectious agents do not allow uniform application of extraction methodologies, library preparations or bioinformatic processing. Different nucleic acid extraction methods may be required for each sample type and even for different infectious agents within the same sample type. Furthermore, the opportunity for repeat testing will be limited due to small sample quantities such as in CSF and the need to make a timely clinical decision early on. This presents a vast scientific challenge and as importantly a regulatory one. Clinical laboratories are highly regulated, with quality control being an imperative measure of accuracy. Quality control

includes measuring contamination, which is highly problematic in metagenomics due to its high sensitivity. Furthermore, analytics validation and subsequent clinical validation may prove challenging with microbial databases consistently being updated with new species and resistance genes every week. The CDC estimates that sequencing costs per bacterial sample amounts to approximately $100-250, and whether this investment will have a cost-benefit ratio to patients and in which clinical scenarios remains to be answered.[100]

In summary, beyond identifying species, clinical metagenomics may offer valuable information to predict AMR in a timely and accurate manner. Clinical metagenomics has the potential to provide quick and accurate real-time diagnostics in patients with BSI. Besides identifying fastidious organisms not identified by culture, it may provide a species and antimicrobial result within 4-6 hours instead of the conventional 24-72 hour time-to-result which then can alter antimicrobial therapy early on, reducing morbidity and mortality. In the subsequent chapter, a systematic review and meta-analysis will be conducted to assess the utility of clinical metagenomics as a diagnostic tool for infectious diseases. The focus will be on the ability of clinical metagenomics to identify microbial species and predict antimicrobial resistance (AMR) in patients, examining both its performance and limitations. The next chapter aims to address current challenges in the field, including methodological inconsistencies, interpretation difficulties, and practical application considerations. The outcomes from this review and analysis will have implications for future study design in the field of metagenomics, enhancing its utility in infectious disease diagnostics.

"Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?"

-T.S. Eliot

3 META-ANALYSIS AND SYSTEMATIC REVIEW

3.1 Introduction

Metagenomic sequencing is frequently claimed to have the potential to revolutionise microbiology through rapid species identification and antimicrobial resistance (AMR) prediction. However, culture remains the backbone of the microbiology laboratory, and is often time consuming taking several days to obtain species identification and antimicrobial susceptibility results. The complete metagenomic pipeline from sample collection to report is shown in **Figure 1**. Molecular diagnostics are increasingly being used to complement culture-based techniques, however, is limited by restricted gene panels offering limited predictive ability. In contrast, emerging real-time sequencing technologies potentially offer a pathogen and clinical syndrome agnostic platform to identify any pathogens and known resistance determinants present within hours. In this chapter, the objectives were to assess all current literature published on culture-

independent metagenomic sequencing for pathogen-agnostic infectious disease diagnostics in participants with potential infection and to describe the specific challenges and progress in the field.

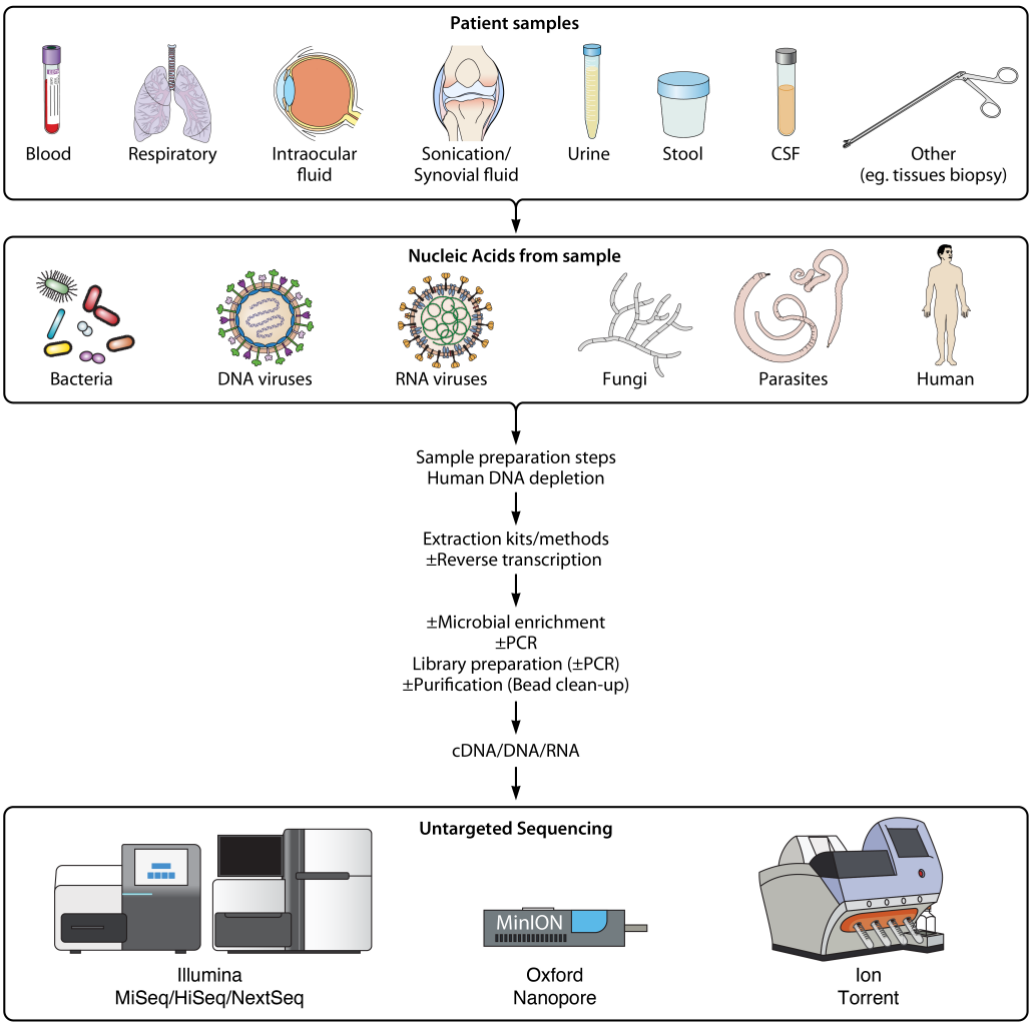

Figure 1. Metagenomic flow diagram. Patient samples undergo nucleic acid extraction. Laboratory methods may be used to reduce human DNA present, to improve pathogen sequencing efficacy. Reverse transcription may be utilised to convert RNA to cDNA for sequencing. Optional microbial enrichment is performed to further increase pathogen concentration. Random or specific PCR may be performed if a higher DNA concentration is required. Sequencer specific library preparation kits are used, which may include inherent PCR steps. Purification of the sample is often performed using bead-cleaning to remove inhibitors. Samples are then loaded onto the sequencer for sequencing.

3.2 Method

3.2.1 Search strategy and selection criteria

A systematic review was conducted on all available literature up to 13th August 2020 in accordance with the Preferred Reporting System for Systematic Reviews and Meta-Analyses for Diagnostic Test Accuracy (PRISMA-DTA) guidelines.[101] PubMed and Google Scholar were searched using a combination of relevant MeSH terms (**Table 1**). Abstracts of studies identified were screened using the inclusion and exclusion criteria specified in **Table 2**.

Table 1. Search strategy.

Database	**Query**	**Search terms**	**Number of items returned (13-08-20)**
PubMed	1	"infection"[MeSH Terms]	3434326
	2	"diagnosis"[MeSH Terms]	10192959
	3	""metagenomic"[MeSH Terms]	25024
	4	"infection"[MeSH Terms] AND "diagnosis"[MeSH Terms] AND "metagenomic"[MeSH Terms]	539

Table 2. Inclusion and exclusion criteria for this systematic review.

Inclusion criteria	**Exclusion criteria**
Original manuscripts published in peer reviewed journals	Conference abstracts/pre-print articles
Human studies	Animal and environmental studies
Studies using direct-from-sample extraction methods, that is, no sub-culturing	Case studies
Whole genome (shotgun) sequencing/metagenomics used	16S rRNA gene sequence analysis
Pathogen agnostic sequencing and analysis workflow	Microbiome community studies without clinical diagnostic purpose
Culture, molecular tests, clinical decision or a composite of the aforementioned used as a reference standard	Studies using spiked samples
Studies in English only	Studies with < 10 positive samples && <10 negative samples, and pilot studies.

In this chapter, I focused on pathogen-agnostic sequencing studies and excluded pilot studies, those with <10 samples, and those focused on diagnosis of a specific pathogen or group of pathogens. The reference lists of included manuscripts were also manually

searched for relevant studies not identified during the literature search. I assessed methodologic quality and risk of bias using the QUADAS-2 tool for diagnostic accuracy studies and excluded any study with high overall risk of bias.[102] Study inclusion, extraction of data and quality evaluation were assessed by an additional independent reviewer and any differences were resolved through discussion or by a third reviewer. Peer-reviewed articles that utilised metagenomic sequencing as a tool for clinical diagnostics in infectious diseases were appraised and relevant data was extracted where possible. Principle diagnostic accuracy measures were sensitivity and specificity per clinical sample result. Antimicrobial resistance prediction was assessed using the resistance prediction outcome over each combination of antibiotic and sample tested. This study was registered with the PROSPERO prospective register of systematic reviews (reference CRD42020163777).

3.2.2 Meta-analysis

The index test was defined as the metagenomic sequencing result, while the reference standard defining true positives and true negatives was based on a pragmatic approach of each study's best measure of the presence or absence of infection with a specific species using either culture, molecular testing, clinical decision or a combination of these factors. Additional species identified as plausible by studies including polymicrobial infections were considered additional true pathogens and contributed to the numerator and denominator in the sensitivity analysis. Specificity was determined by considering the proportion of samples deemed negative by the reference standard that were identified as negative by metagenomic sequencing. Although it is plausible to consider that any sample can contain a false positive result, including culture-positive samples, studies differed considerably in reporting this. Therefore, only reference standard negative

samples were included in specificity calculations. Studies where only sensitivity could be determined and not specificity were not included in the meta-analysis.

The R meta package for the statistical analysis was used.[103] Outcomes are reported for each study with 95% confidence intervals (CIs), and pooled estimates provided using fixed or random-effects models, depending on the absence or presence of heterogeneity using the I2 test respectively. An I2 value >50% was considered heterogenous.

3.3 Results

A total of 551 manuscripts were screened, 530 were excluded (**Figure 2**), leaving 21 eligible for inclusion. Specificity could not be determined in eight studies, leaving 13 studies to be included in the main meta-analysis (**Table 3**). An additional analysis was performed including studies that did include specificity.

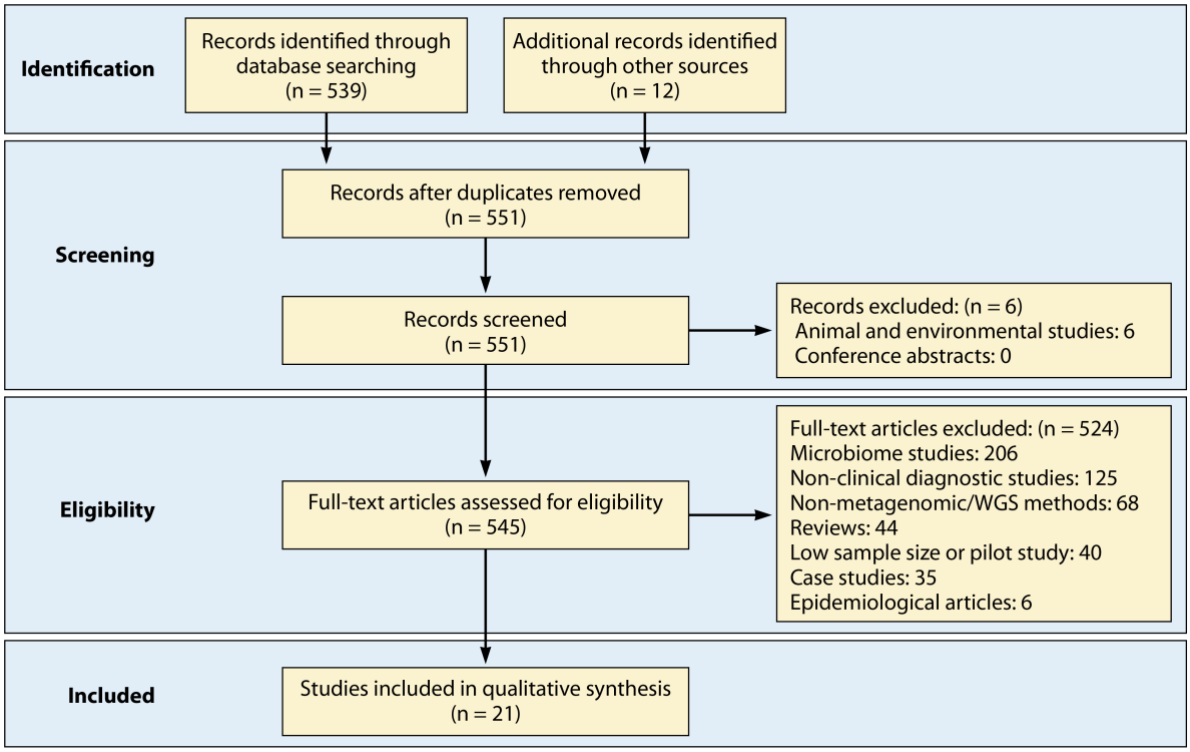

Figure 2. Flow chart showing study selection. 551 manuscripts were screened, 530 were excluded for the reasons shown, leaving 21 eligible manuscripts for inclusion. Further details available in **Supplementary Table 6**.

Table 3. Characteristics of studies included in the meta-analysis.

Study #	Study (Author and year)	Type of sample	Sequencing techniques	Range of pathogens found	Reference standard	Number of positive samples	Number of negative samples	Sensitivity (95% CI)	Specificity (95% CI)
1	Hong, NT et al. (2020)	Cerebrospinal fluid	Illumina MiSeq (CA, USA)	Viruses	PCR	19	47	0.74 (0.49-0.91)	0.66 (0.51-0.79)
2	Miller, S et al. (2019)	Cerebrospinal fluid	Illumina HiSeq	Bacteria; fungi; parasites; DNA viruses; RNA viruses	Conventional laboratory results & additional molecular testing	56	118	0.89 (0.78-0.96)	0.99 (0.95-1)
3	Wilson, M et al. (2019)	Cerebrospinal fluid	Illumina HiSeq	Bacteria; fungi; parasites; DNA viruses; RNA viruses	Conventional laboratory results & additional molecular testing	58	149	0.55 (0.42-0.68)	0.98 (0.94-1)
4	Blauwkamp, T et al. (2019)	Plasma (cfDNA)	Illumina NextSeq 500	Bacteria; DNA viruses; fungi; eukaryotic; parasites	Conventional laboratory results & additional molecular testing	182	166	0.93 (0.88-0.96)	0.63 (0.55-0.70)
5	Parize, P et al. (2017)	Plasma	Ion Proton (Life Technologies, USA)	Bacteria; viruses	Culture, serological diagnosis & PCR	19	82	0.63 (0.38-0.84)	0.71 (0.60-0.80)
6	Somasekar, S et al. (2017)	Serum	Illumina HiSeq	Viruses	PCR	26	246	0.96 (0.80-1)	1 (0.98-1)
7	Rossoff, J et al. (2019)	Plasma	Illumina NextSeq 500	Bacteria; DNA viruses; fungi; eukaryotic parasites	Clinical review	61	39	0.92 (0.82-0.97)	0.64 (0.47-0.79)
8	Schlaberg, R et al. (2017)	Respiratory samples	Illumina HiSeq 2500	Viruses	Clinical review	60	59	0.90 (0.79-0.96)	0.64 (0.51-0.76)
9	Street, T. et al. (2017)	Sonication/synovial fluid	Illumina MiSeq	Bacteria	Culture	69	97	0.88 (0.78-0.95)	0.88 (0.79-0.93)
10	Thoendel M et al. (2018)	Sonication/synovial fluid	Illumina HiSeq 2500	Bacteria	Culture	146	67	0.83 (0.76-0.89)	0.69 (0.56-0.79)
11	Ivy, M et al. (2018)	Sonication/synovial fluid	Illumina HiSeq 2500	Bacteria; fungi	Culture	82	60	0.82 (0.72-0.89)	0.35 (0.23-0.48)
12	Doan, T et al. (2017)	Intraocular fluid samples	Illumina HiSeq 4000	Bacteria; fungi; parasites; DNA viruses	PCR	31	36	0.87 (0.70-0.96)	0.78 (0.61-0.90)
13	Cheng, J et al. (2019)	Heart valves	Illumina HiSeq	Bacteria	Clinical review	41	7	0.98 (0.87-1)	0.86 (0.42-1)

3.3.1 Clinical samples and range of pathogens

The range of clinical samples included blood[51,53,55,95,104], positive blood cultures[3], urine[82], respiratory samples[77,105–107], orthopaedic device sonication fluid and other bone and joint samples[56,96,108,109], cerebrospinal fluid[49,58,110,111], intraocular fluid[112] and heart valve tissue.[113] The range of pathogens found are detailed in **Table 3**. The median number of samples sequenced per study was 67 (IQR 44-166) [Range 12-348].

3.3.2 Methodology: nucleic acid extraction and sequencing library preparation

Methods to deplete human cells prior to DNA extraction were used in 8/21 (38%) studies (**Supplementary Table 1**), including differential centrifugation[3,56,82] or filtration with a 5µm membrane[3,55,108] to remove human cells, based on their larger mass or size respectively, and the MolYsis Basic5 kit (Molzym, Bremen, Germany)[96,109] which is reported to differentially lyse human cells.

Nucleic acid extraction kits and techniques were heterogeneous across studies. Most studies utilised commercial kits, primarily based on membrane/column methods, while non-commercial methods relied mostly on ethanol precipitation or lysis followed by bead-cleaning methods. 3/21 (14%) studies performed a post-extraction enrichment step to facilitate enrichment of microbial DNA from clinical samples and 7/21 (33%) studies used a DNA amplification method prior to library preparation to achieve required input DNA concentrations. DNA clean-up with magnetic beads was used in 9/21 (43%) studies to purify extracted DNA prior to sequencing.

3.3.3 Methodology: controls

Studies were assessed for three types of controls, (i) an internal or spiked control, (ii) a negative control and (iii) a positive control. Only 4/21 (19%) studies reported internal control use; this included bacteriophages and synthetic DNA[49,51,58,95]. 13/21 (62%) studies reported use of a negative control. The median (IQR) [range] of negative control samples per every 10 samples processed was 1.5 (1.0-1.5) [0.5-1.5]. Only 6/21 (29%) studies reported use of a positive control, including using a mixture of 7 representative organisms and a single *Corynebacterium glutamicum* ATCC 13032 positive control.[49,58,95,104,105,109]

3.3.4 Sequencing technologies

Illumina technology (CA, USA) was used exclusively in 16/21 (76%) studies (**Supplementary Table 2**). One study exclusively used Oxford Nanopore Technologies MinION (ONT, UK)[77] and two studies exclusively used Ion Proton (Life Technologies, USA).[53,82] One study compared the use of Illumina sequencing versus ONT MinION.[3] The BGISEQ-500/50 (BGI, Tianjin, China) platform was used in one study.[110]

3.3.5 Bioinformatic approaches

Figure 3 outlines the bioinformatic workflows implemented. The most commonly used classification software packages for species identification were alignment based (n=11) followed by k-mer based (n=8) and marker-gene based (n=1) (see **Supplementary Table 2** for details). The most commonly used species databases originate from the NCBI and included the NCBI RefSeq database (n=9), unspecified NCBI databases (n=7), MetaPhlAn2 database (n=3), NCBI Genbank (n=1) and Karius proprietary database (n=1).

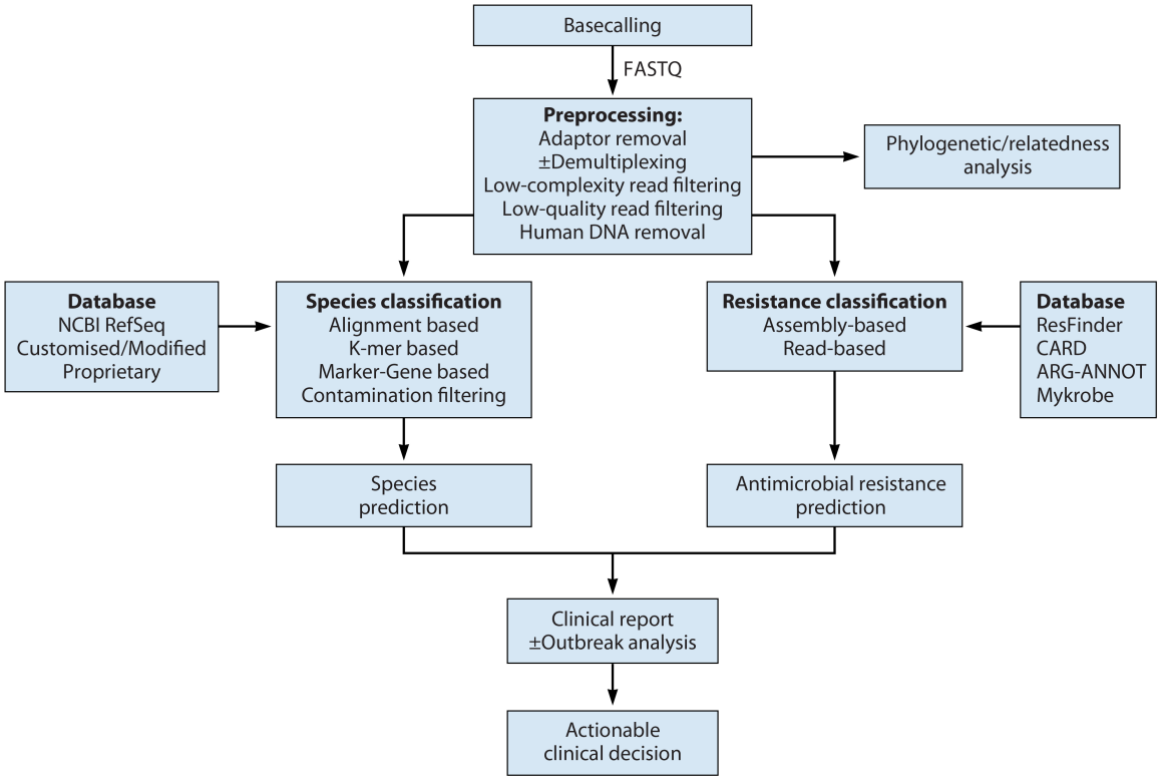

Figure 3. Metagenomic bioinformatic flow diagram. Raw data is basecalled to FASTQ files with corresponding quality scores. Preprocessing steps involve adaptor removal, demultiplexing, removing low-complexity and low-quality reads, and removing human reads. Species can then be classified using alignment, k-mer or marker-gene based software with appropriate databases. Contamination filtering may be performed thereafter. Resistance classification can be assembly- or read-based using the appropriate database. A relatedness/phylogenetic analysis may be performed for transmission analyses. A final clinical report ± outbreak analysis can be produced to provide potentially actionable clinical information.

3.3.6 Assessing true species presence versus contamination

Various strategies were employed to identify true presence of an organism verses contamination. The most commonly used methods include an empirical approach setting absolute thresholds based on sequencing known negative and positive samples (n=4) [3,49,58,109] while other studies (n=3) used a parametric approach based on the observed distribution of contaminating reads in control and other samples.[95,96,105] Other strategies are described in **Figure 4**. By use of negative controls and other approaches, 16/21 (76%) studies report contamination of some degree while 5/21 (24%) had no details of contamination available. No study reported no evidence of contamination.

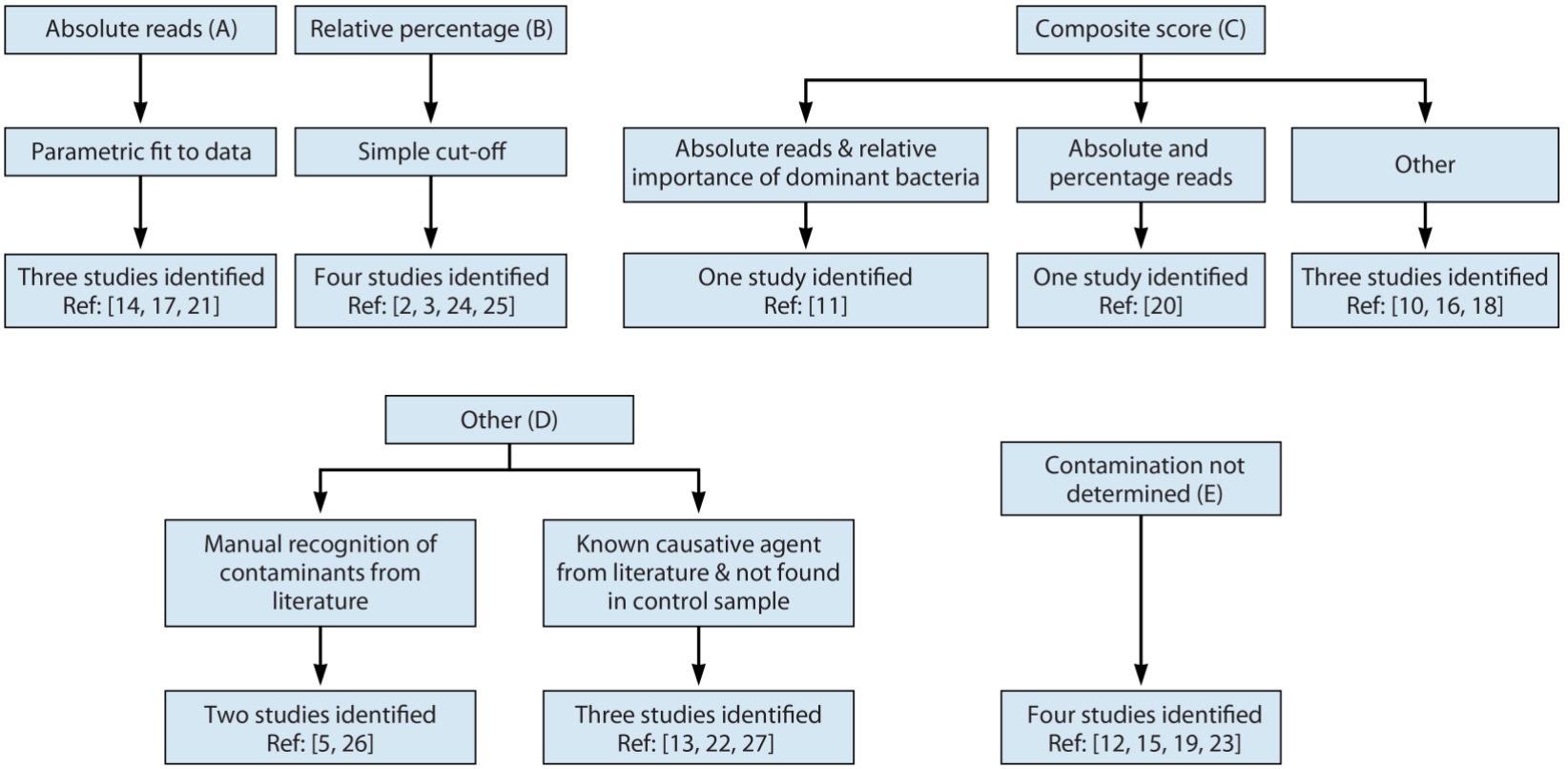

Figure 4. Common approaches to contamination control. Various strategies were utilised to identify true presence of an organism verses contamination. A: Parametric approach based on an assumed distribution for contaminating reads informed by numbers of reads in the control and other samples; B: Empirical approach to setting these absolute thresholds based on sequencing known negative and positive samples; C: Composite measures to determine contamination using factors such as absolute and percentage reads, and relative importance of bacteria among others; D: Manually from known literature or if reported to be a contaminant previously in the local laboratory; E: No description of how the presence of an organism was confirmed.

3.3.7 Performance: species identification

Methodological quality was assessed using the QUADAS-2 tool (**Figure 5**).[102]

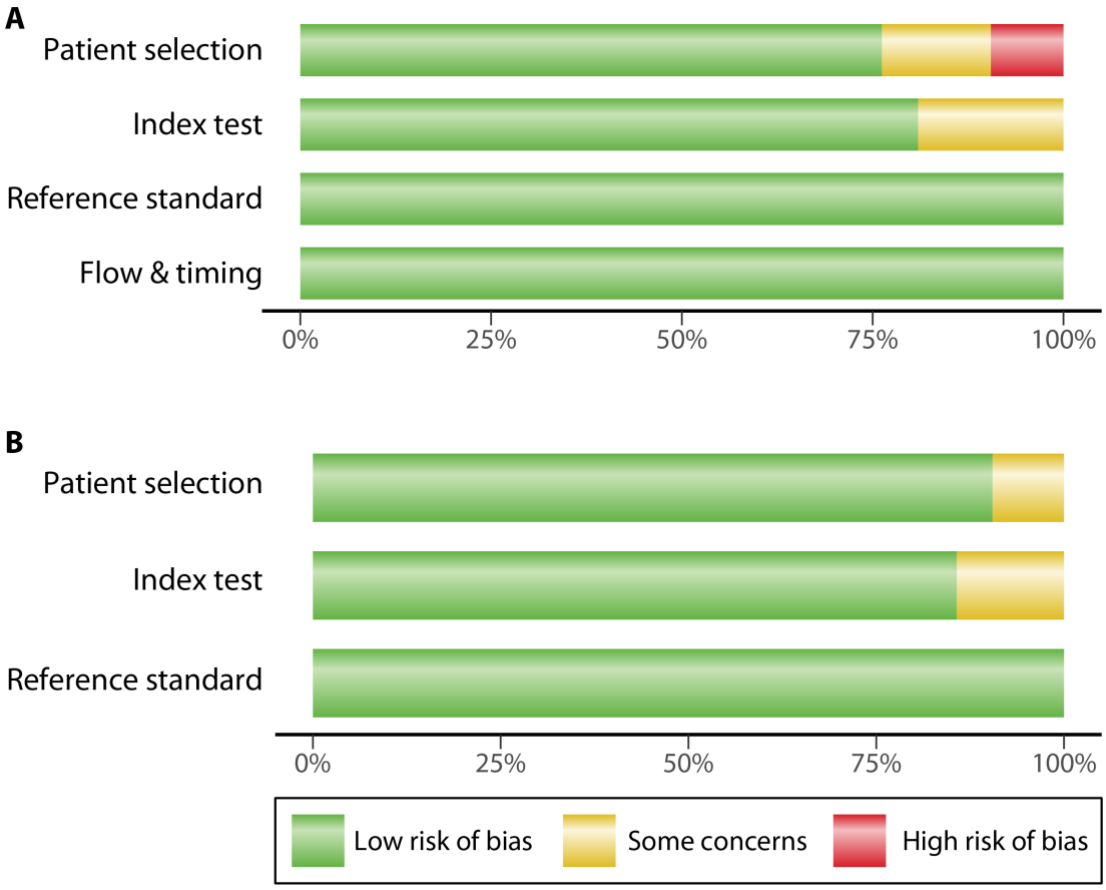

Figure 5. QUADAS-2 TOOL for the evaluation of primary research articles. The graphs present a summary of authors' judgments with respect to each domain as percentages across all included articles. A, Risk of bias of studies included. This tool comprises 4 domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of risk of bias guided by signalling question listed in **Figure 6**. B, Applicability concerns of studies included. This tool comprises 3 domains: patient selection, index test and reference standard. The findings for each included are listed in **Figure 7**.

DOMAIN 1: PATIENT SELECTION
Risk of bias: Could the selection of patients have introduced bias?
Signalling question 1: Was a consecutive or random sample of patients enrolled?
Signalling question 2: Was a case-control design avoided?
Signalling question 3: Did the study avoid inappropriate exclusions?

Concerns regarding applicability
Is there concern that the included patients do not match the review question?

DOMAIN 2: INDEX TEST
Risk of Bias: Could the conduct or interpretation of the index test have introduced bias?
Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?
Signalling question 2: If a threshold was used, was it pre-specified?

Concerns regarding applicability
Is there concern that the index test, its conduct, or interpretation differ from the review question?

DOMAIN 3: REFERENCE STANDARD
Risk of Bias: Could the reference standard, its conduct, or its interpretation have introduced bias?
Signalling question 1: Is the reference standard likely to correctly classify the target condition?
Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?

Concerns regarding applicability
Is there concern that the target condition as defined by the reference standard does not match the review question?

DOMAIN 4: FLOW AND TIMING
Risk of Bias: Could the patient flow have introduced bias?
Signalling question 1: Was there an appropriate interval between index test and reference standard?
Signalling question 2: Did all patients receive the same reference standard?
Signalling question 3: Were all patients included in the analysis?

Figure 6. QUADAS-2 questions. Risk of bias is judged as "low", "high", or "unclear". If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" or "unclear" this flags the potential for bias. The "unclear" category was used only when insufficient data were reported to permit a judgment.

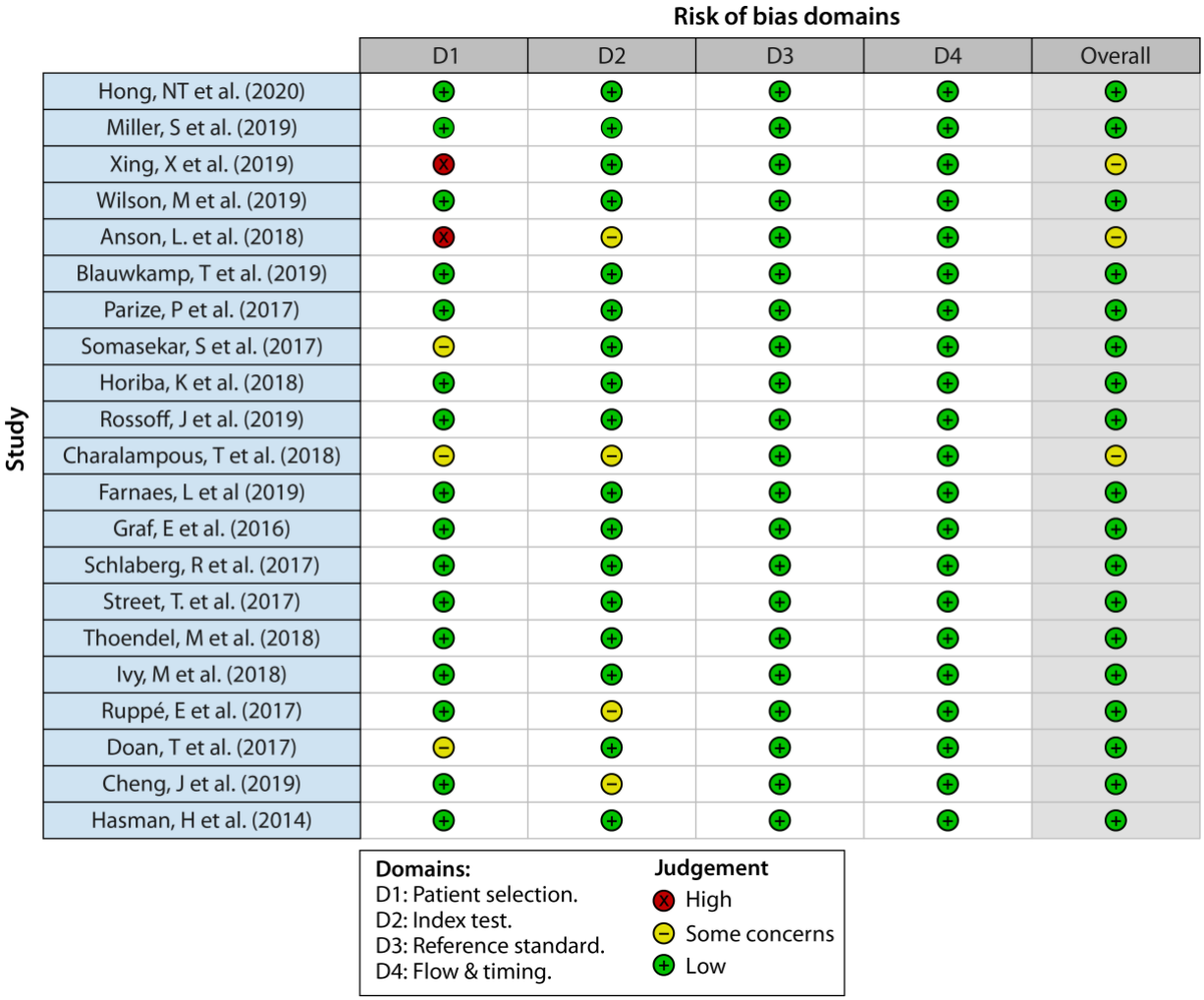

Figure 7. Traffic light plot of studies included. Individual study bias is shown below. For a detailed breakdown, refer to **Supplementary Table 4** for signalling questions.

Risk of bias was most commonly seen in the patient domain with 2/21 (10%) studies having high risk and 3/21 (14%) having some concern, all relating to selection of specific samples for study, rather than use of a random or consecutive sample. This was followed by the index test domain were 4/21 (19%) studies had some concern relating to the reporting of bioinformatic thresholds and methods to assign a result as positive or negative (**Figure 7**).

A meta-analysis of species identification performance was performed by sample type on 13/21 studies where both test sensitivity and specificity could be determined. Sensitivity in the most widely investigated sample types blood (n=288), CSF (n=133), and orthopaedic (n=297) was 90% (95%CI 78-96%), 75% (54-89%), and 84% (79-88%) respectively (**Figure 8**).

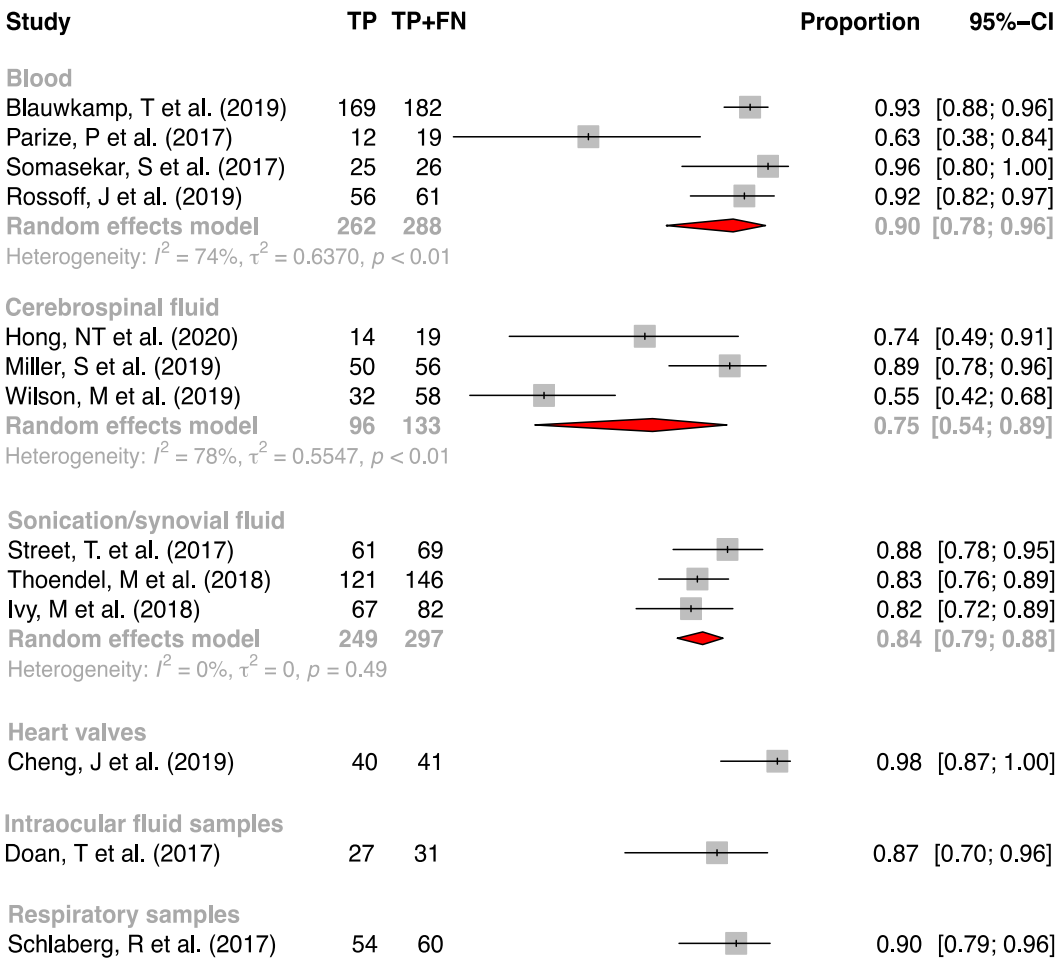

Figure 8. Pooled sensitivity of studies grouped by sample type. This analysis of sensitivity included 13 studies. A random effects model was used due to heterogeneity between studies.

Specificity in blood (n=533), CSF (n=314), and orthopaedic samples (n=224) was 86% (95%CI 45-98%), 96% (72-100%), and 67% (38-87%) respectively (**Figure 9**).

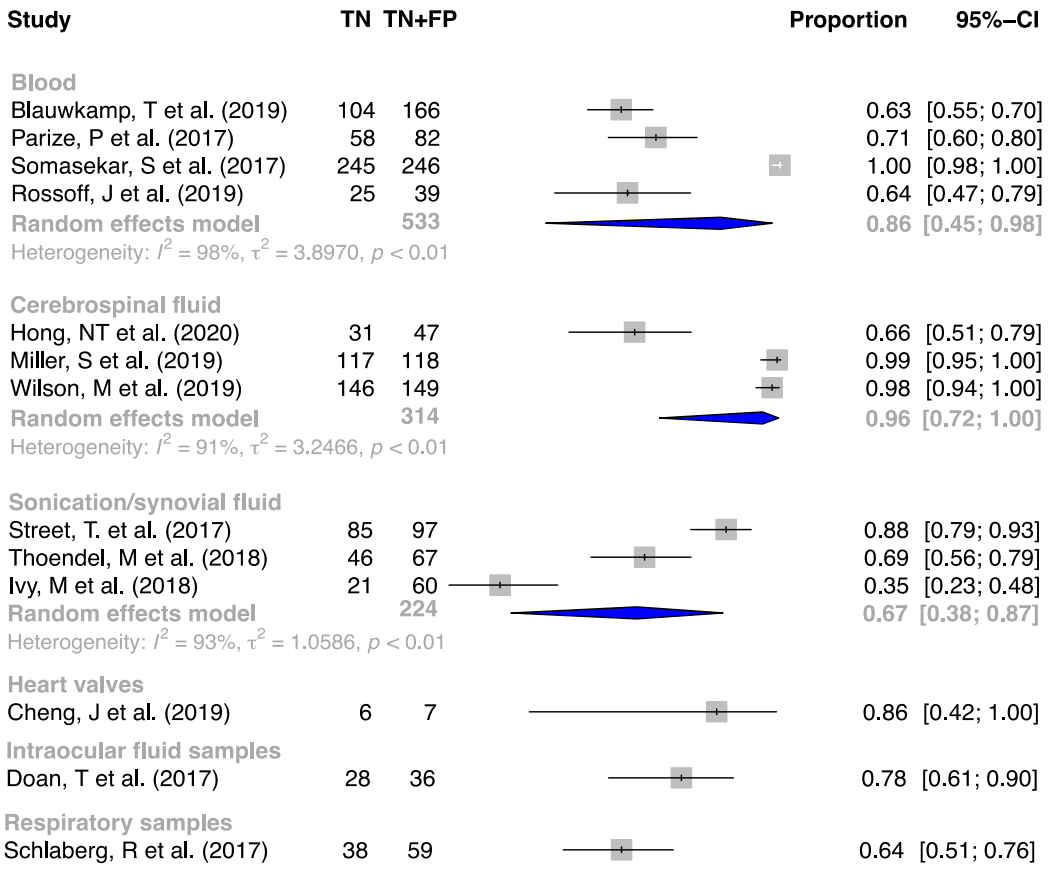

Figure 9. Pooled specificity of studies grouped by sample type. A random effects model was used due to heterogeneity between studies.

8/21 (38%) studies did not report specificity and were excluded from the meta-analysis. 20/21 (95%) studies found a median of 7 (4-15) [1-62] additional pathogens with subsequent clinical interpretation varying throughout. In this analysis, positive and negative predictive values grouped by sample type range between 72-98% and 75-90% respectively, and diagnostic accuracy between 77-95% (**Supplementary Table 5**).

Using reported per-study means, a median 91% (IQR 82-98%) [Range 76-98%] of sequence data was classified as human DNA. Blood samples had the highest median

reported of 91% (98-91) [84-98], and sonication and bone samples had a median of 87% (82-93) [76-98].

3.3.8 Performance: antimicrobial resistance prediction

Nine studies attempted genotypic drug susceptibility prediction and compared results to either drug susceptibility phenotype (n=4) or had no comparison (n=5). Four studies that compared results to drug susceptibility phenotypes were included in a meta-analysis (**Supplementary Table 3**). Performance depends on the prevalence of resistance and the specific pathogen-antimicrobial combinations tested. In the four eligible studies, the median (IQR) [range] number of clinical samples per study was 29 (22-35) [17-39]. A median of 196 (45-350) [35-369] antibiotic predictions were performed per study. Beta-lactams, quinolones and tetracyclines were some of the most commonly tested classes of antibiotics and 19% (153/795) of pathogen-antimicrobial combinations were phenotypically resistant. All studies evaluated prediction of the correct antimicrobial susceptibility as a categorical value, that is, sensitive or resistant. Categorical agreement rates were 88% (CI, 80% to 97%) across studies using a random effects model. The pooled estimate of very major error rates, defined as the number of samples with phenotypic resistance not predicted by genotype over all samples with phenotypic resistance was 24% (CI, 8% to 40%). The pooled major error rate, defined as predicted resistance by genotype but sensitive phenotype over all phenotypically sensitive samples, was 5% (CI, 0% to 12%) (**Figure 10**). There was marked heterogeneity in all three metrics (I^2 0.86-0.92; all $p<0.01$).

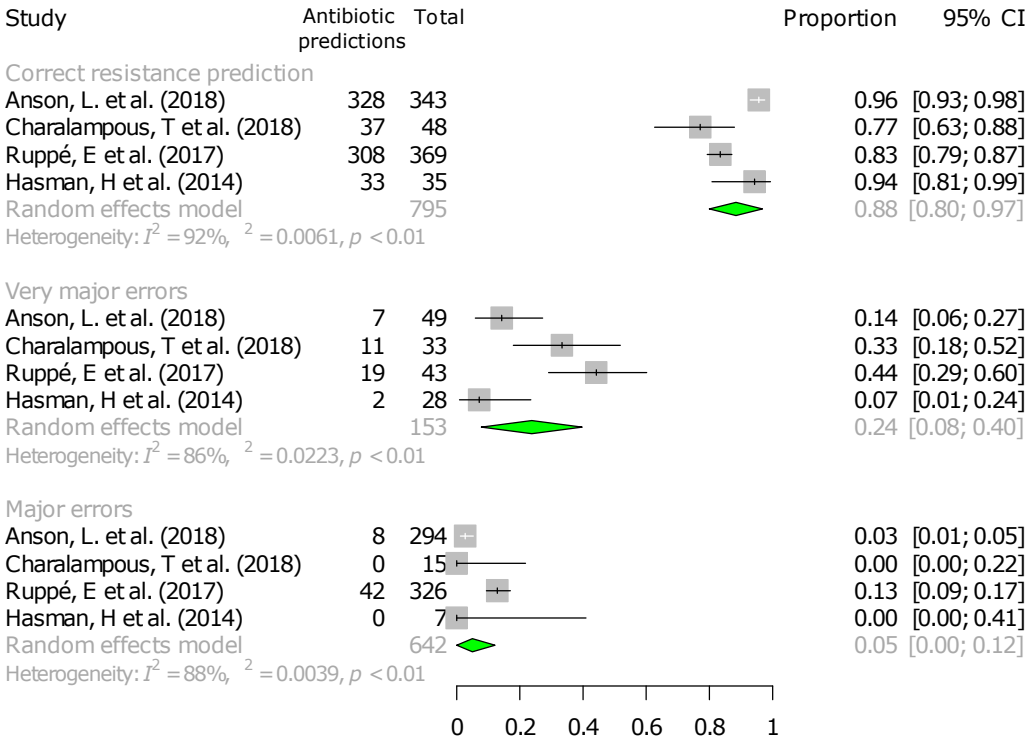

Figure 10. Pooled antimicrobial resistance prediction performance. This analysis included 4 studies that focused on species-agnostic samples and compared this to drug susceptibility phenotype. Proportions were combined using a random intercept logistic regression model. A random effects model was used due to heterogeneity between studies.

3.3.9 Time of complete process

In this study, 9/21 (43%) studies reported sample processing times (**Figure 11**). Based on the mean time provided from each study, sample extraction/preparation took a median (IQR) [range] 3 (3-3) [3-3] hours and library preparation a median 4 (4-4) [4-4] hours. Reported sequencing using Illumina took 16 (16-16) [16-16] hours. Where sequence analysis was undertaken in real-time using the Nanopore platform, first species prediction took 1 (1-1) [1-1] hour. The total time from sample extraction to speciation and AMR prediction was 29 (24-94) [4-144] hours.

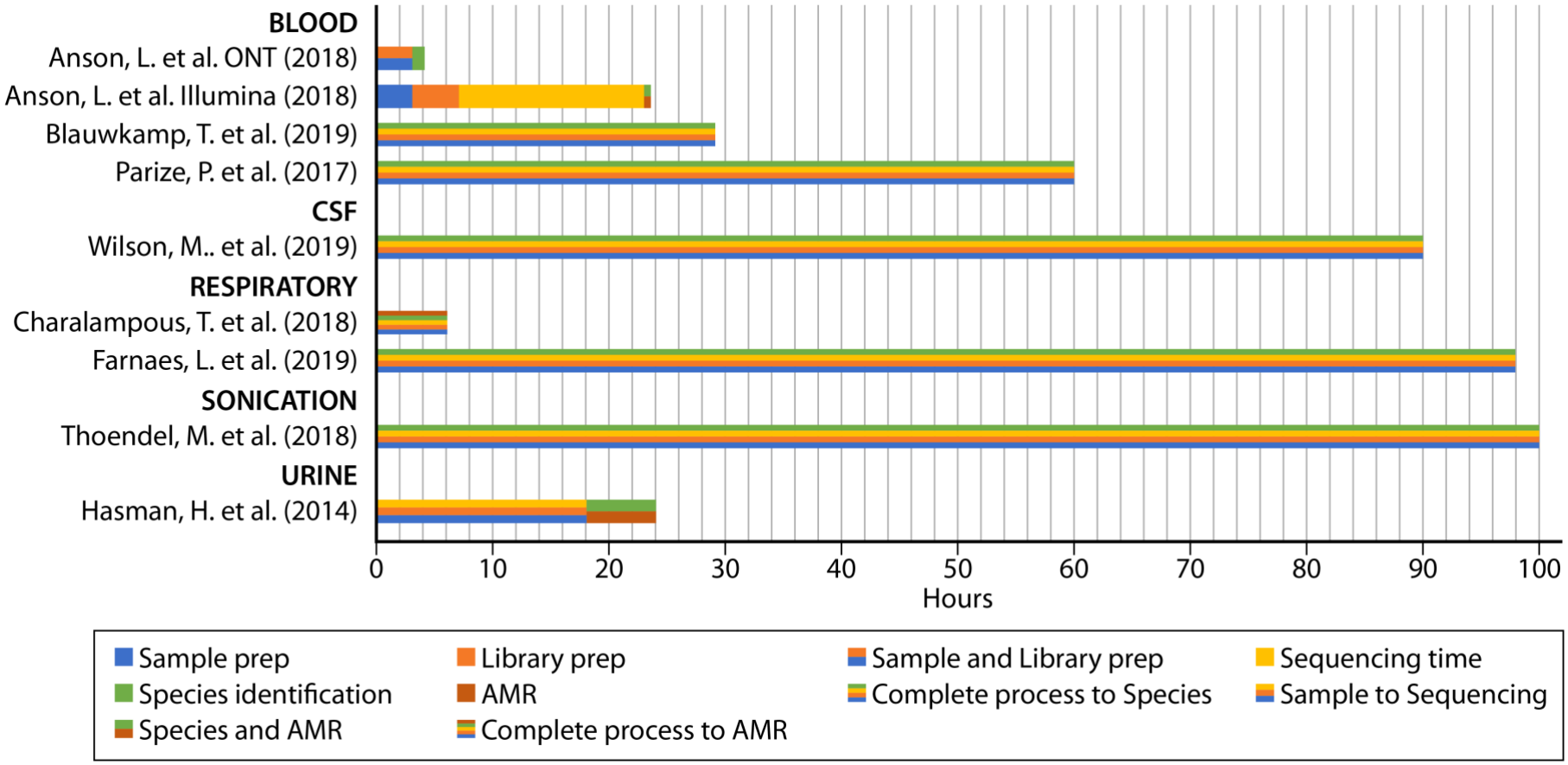

Figure 11. Time taken for sample processing grouped by sample type. The 9/21 (43%) studies shown below reported the time taken from DNA extraction to species identification or antimicrobial susceptibility prediction.

3.3.10 Cost

Only one study reported an average cost estimation per sample which varied from $130 when samples are multiplexed (Nanopore, 6 samples/run) up to $685 for processing a single sample.[77]

3.4 Discussion

In this chapter all studies found to the date of the review are described assessing the performance of metagenomic sequencing as a tool for pathogen-agnostic diagnosis of clinical infection on a variety of clinical samples including from blood, urine, respiratory samples, cerebrospinal fluid, orthopaedic sonication and synovial fluid, intraocular fluid, and cardiac valvular tissue. Sensitivity in the most widely investigated sample types blood (n=288), CSF (n=133), and orthopaedic (n=297) was 90% (78-96%), 75% (54-89%), and 84% (79-88%) respectively (**Figure 8**). Specificity in blood (n=533), CSF (n=314), and orthopaedic samples (n=224) was 86% (45-98%), 96% (72-100%), and 67% (38-87%) respectively (**Figure 9**). Analysis by sample type sub-groups demonstrate development of workflows for CSF, blood and orthopaedic fluid with reproducible performance (**Figure 8** and **Figure 9**).

While these results reflect impressive progress in the field, the level of accuracy required for clinical adoption is still being broadly assessed, having specific advantages and disadvantages reviewed elsewhere.[2,114] However, it should be remembered that the current gold standard of culture-based microbiology is imperfect, for example, two standard blood cultures will miss at least 10-18% of episodes with potentially culturable organisms.[115] Additionally, as metagenomics may detect potential pathogens not found

by culture, reported specificity may be reduced because of the use of an imperfect reference standard. For example, a median (IQR) [range] of 7 (4-15) [1-62] additional pathogens were found per study. Besides culture, metagenomics could be benchmarked against multiplex syndromic panels however, ultimately expert clinical review is required to evaluate the plausibility of additional potential pathogens detected.

Rapid and precise antimicrobial resistance prediction linked to specific pathogens would greatly improve the time to administration of targeted antimicrobials. Although the categorical agreement rate with phenotyping was 88% (CI, 80% to 97%), major and very major error rates were 5% (CI, 0% to 12%) and 24% (CI, 8% to 40%) respectively, well above the current regulatory thresholds; the FDA requires major errors <3% and the 95% CI on the very major error rate to be ≤7.5% at the upper limit and ≤1.5% at the lower limit.[65] It is clear that improvements are needed for AMR prediction to reach acceptable performance, including better understanding and cataloguing of the mechanisms underlying antimicrobial resistance and optimised algorithms to detect these within metagenomic data and link them to specific pathogens. This is particularly challenging for plasmid-mediated resistance which may be difficult to assign to a specific species from metagenomic data.

With further advances in extraction methods, sequencing technology and bioinformatic processes, the proficiency of clinical metagenomics is likely to improve while costs continue to decline. Since the mid-1800s culture-based diagnostics have remained the backbone of the microbiology lab[116]; however clinical metagenomics has the potential to be the next frontier of clinical microbiology, not only enhancing our understanding of infectious diseases but also providing us with information to act on with more precision

than ever before to impact lives. Great strides have been made both on a basic scientific level and on a regulatory level by the FDA that sets the foundation for this technology to become a routine test offered at the bedside. Though for this to happen, the value of clinical metagenomics needs to be clearly defined in the care pathway and should ideally be reduced to an easy-to-interpret, point-of-care, standalone automatic device.

Although vast progress has been made with multiple pipelines demonstrated to work on a variety of clinical samples, there still is the need for more robust studies with higher sample sizes and improved quality control metrics. Furthermore, studies should incorporate a clinically reviewed result (e.g. by an infectious disease review board) as an outcome measure as well as standard microbiology results. To date no clinical trials have been conducted to show superior patient outcome or reduced hospital stay from clinical metagenomics, however multiple anecdotal case reports have been published that show promising results especially for highly intricate and unexplained clinical cases.[117] Contamination control is extremely important and methods both in the laboratory and bioinformatically need to be optimised to reduce false positives and is especially true of samples with low pathogen titres where PCR is used. A major bottleneck has been the sequencing run time, however emerging real-time devices such as the Oxford Nanopore drastically reduce time-to-result and may provide additional information using long-reads such as plasmid sequences for improved AMR prediction.

As databases continue to expand in volume and number, novel bioinformatic challenges are posed concerning big data and optimal processing methods. This has led to the Critical Assessment of Metagenome Interpretation (CAMI) which has been launched as a community-led initiative to establish standard methods of interpretation. Furthermore,

database selection and curation are vital for optimal performance however requires further benchmarking. Metagenomics may also have an irreplaceable role in outbreak investigation with the agnostic capability to detect not only natural outbreaks with high genetic resolution but also engineered pathogens from a bioterrorist attack. As cultures are an imperfect gold standard, clinical metagenomics may have the potential to set a new standard however for this to occur studies will need to compare performance not only to standard microbiology but also to clinically reviewed results and outcomes. In this chapter I have found that the robustness of studies is lacking with many using poor quality control measures and low sample sizes. Furthermore, many studies focused on speciation while AMR prediction proved more challenging, likely due to lower genome coverages in metagenomic samples. The sequencing time bottleneck has been disrupted by newer real-time sequencers such as the Oxford Nanopore that may promise faster, clinically-actionable results including AMR prediction within 2 hours. Although sequencing costs have dramatically reduced, total costs per sample still remain high for cost-effective and routine implementation. Cost-benefit studies are required to validate this technology and where it might apply precisely in the clinical care pathway, be it as a routine test or a last resort option. Although no device based on metagenomic sequencing has been FDA approved for infectious diseases diagnostics, draft FDA regulations are already set in place that take into consideration the complexity of the technology.

The review I present in this chapter has several limitations. Firstly, the meta-analysis provided does not overcome problems inherent in the design and execution of the primary studies. Furthermore, it does not correct bias that could have arisen from selective publication. 8/21 (38%) of studies did not include negative test results limiting the ability to ascertain false-positive results or contamination within these studies and

therefore were excluded from the meta-analysis. Using the QUADAS-2 tool, the main risk of bias arose from selection of specific samples for study, rather than consecutive samples or a random subset which are recommended. Similarly, to avoid index test bias it is recommended that studies provide further clarity on bioinformatic procedures when determining a positive or negative result. Although studies were grouped by sample type, any observed difference may be confounded by other factors such as differing clinical application, laboratory and bioinformatic methods or the fact that reference standards differed among culture, molecular testing, clinical decision or a composite measure. Due to the heterogeneity between studies, it is difficult to provide specific recommendations on the ideal metagenomics workflow but anticipate by bringing studies to date together this review will provide a mechanism for improving standards and designing better studies in future. The results of this study suggest future research conform to requirements set out by the STROBE-metagenomics guideline, STARD, or other equivalent guidelines to improve the quality of study design and reporting.[118,119]

While clinical metagenomics offers us a deeper and novel insight to treating infectious diseases, it may unlikely replace established diagnostics including PCR or multiplex assays in the next several years. We are likely to see some devices approved by the FDA in the near future which may first be used in reference laboratories particularly for complex undiagnosed infectious disease cases. In the long-run as costs further decline and bioinformatic processing advance, clinical metagenomics may be established as a routine test for specific diseases and clinical scenarios from encephalitis and meningitis to sepsis and pyrexia of unknown origin. The promise of an agnostic infectious disease diagnostic tool is revolutionary for clinical microbiology and furthermore has many advantages including detecting polymicrobial infections and pathogens that are conventionally unculturable which may lead to more effective and narrow-spectrum

treatment options reducing antimicrobial resistance and leaving the microbiome less prone to dysbiosis. Clinical metagenomics may also serve as an effective tool to monitor disease outbreaks early on due to the agnostic capability of the technology and allow for a faster and more effective outbreak response. It is likely that clinical metagenomics will be a part of the clinician's armamentarium in the future to identify and guide therapy of complex infectious diseases. However, convergence around protocols and further metagenomic studies of improved quality and with higher sample sizes are required for the field to progress further.

Supplementary Table 1. Detailed Laboratory Methodology. Contains laboratory methodology of raw data extracted from eligible studies including the use of using controls and contamination reporting.

Study no.	Study (Author and year)	Type of Sample	Cell lysis methods	RNA or DNA separation	Pre-steps(1)
Included in the meta-analysis					
1	Hong, NT et al. (2020)	Cerebrospinal fluid	NA	Membrane-based column	NA
2	Miller, S et al. (2019)	Cerebrospinal fluid	Bead-beating	Membrane-based column	Bead-beating
3	Wilson, M et al. (2019)	Cerebrospinal fluid	Bead-beating	Membrane-based column	Bead-beating
4	Blauwkamp, T et al. (2019)	Plasma (cfDNA)	NA	Magnetic Beads	Centrifugation (16000g, 10min)
5	Parize, P et al. (2017)	Plasma ± others	NA	Agencourt AMPure XP beads	NA
6	Rossoff, J et al. (2019)	Plasma	NA	Magnetic Beads	Centrifugation (16000g, 10min)
7	Somasekar, S et al. (2017)	Serum	Proteinase K	Magnetic Beads	Turbo DNase
8	Schlaberg, R et al. (2017)	Respiratory samples	NA	Membrane-based column	NA
9	Street, T. et al. (2017)	Sonication/synovial fluid	bead-beating	Ethanol precipitation	Centrifugation(15000 g, 1 hour)
10	Thoendel, M et al. (2018)	Sonication/synovial fluid	Bead beating	AMPure beads	MolYsis Basic5 kit
11	Ivy, M et al. (2018)	Sonication/synovial fluid	Bead beating	AMPure beads	MolYsis Basic5 kit
12	Doan, T et al. (2017)	Intraocular fluid samples	Proteinase K	Silica membrane	NA
13	Cheng, J et al. (2019)	Heart valves	Proteinase K	Membrane-based	Proteinase K
Not included in the meta-analysis					
14	Xing, X et al. (2019)	Cerebrospinal fluid	Proteinase K	Membrane-based column	NA
15	Anson, L. et al. (2018)	Liquid blood culture (positive)	Bead-beating	Membrane-based column	Differential centrifugation (2000 g, 1 min)
16	Horiba, K et al. (2018)	Plasma; serum	Proteinase K	Membrane-based	5.0-μm filter (Merck Millipore)
17	Charalampous, T et al. (2018)	Respiratory samples	bead-beating	MagnaPure DNA Isolation Kit	Bead-beating
18	Farnaes, L et al (2019)	Respiratory samples (cfDNA)	NA	Magnetic bead-based method	NA
19	Graf, E et al. (2016)	Respiratory samples	NA	Membrane-based column	NA
20	Ruppé, E et al. (2017)	Sonication/synovial fluid	Lysis buffer	Membrane-based	Differential lysis
21	Hasman, H et al. (2014)	Urine	Proteinase K	Ethanol precipitation	Centrifugation (2,000g for 30 s)

Supplementary Table 1. Continued

Study no.	Study (Author and year)	Pre-steps(2)	Pre-steps(3)	% of human DNA reduced	Kit/method used	Post-extraction
Included in the meta-analysis						
1	Hong, NT et al. (2020)	NA	NA	NA	QIAamp viral RNA kit/DNeasy blood and tissue kit	NA
2	Miller, S et al. (2019)	NA	NA	NA	EZ1 Virus Mini Kit v2.0 (Qiagen)	DNA: NEB Microbiome Enrichment Kit/ RNA: Turbo DNAse and Baseline DNase
3	Wilson, M et al. (2019)	NA	NA	NA	EZ1 Virus Mini Kit v2.0 (Qiagen)	DNA: NEB Microbiome Enrichment Kit / RNA: Turbo DNAse and Baseline DNase
4	Blauwkamp, T et al. (2019)	NA	NA	NA	Mag-Bind DNA Kit (Omega Biotek) in	NA
5	Parize, P et al. (2017)	NA	NA	NA	MolYsis™ Complete5 & QIAamp Cador pathogen kit	NA
6	Rossoff, J et al. (2019)	NA	NA	NA	Mag-Bind cfDNA Kit (Omega Biotek)	NA
7	Somasekar, S et al. (2017)	NA	NA	NA	EZ1 Qiagen Virus Mini Kit v2.0	NA
8	Schlaberg, R et al. (2017)	NA	NA	NA	QIAamp viral RNA minikit	NA
9	Street, T. et al. (2017)	5-um-pore-size syringe filte	Beadbeating	NA	Ethanol precipitation	NEBNext microbiome DNA enrichment kit
10	Thoendel, M et al. (2018)	NA	NA	NA	BiOstic Bacteraemia	NA
11	Ivy, M et al. (2018)	NA	NA	NA	BiOstic Bacteraemia	NA
12	Doan, T et al. (2017)	NA	NA	NA	DNeasy Blood & Tissue Kit (QIAGEN, Germantown, MD)	NA
13	Cheng, J et al. (2019)	NA	NA	NA	TIANamp Micro DNA kit	NA
Not included in the meta-analysis						
14	Xing, X et al. (2019)	NA	NA	NA	TIANamp Micro DNA Kit	NA
15	Anson, L. et al. (2018)	Selective lysis with water	Filtration-5uM membrane	37–80%	BiOstic Bacteraemia; Amplex Hyplex; MolYsis Plus.	NA
16	Horiba, K et al. (2018)	NA	NA	NA	QIAamp UCP Pathogen Mini Kit (Qiagen, Hilden, Germany)	NA
17	Charalampous, T et al. (2018)	Saponin-based host DNA depletion	Proteinase K	NA	MagNA Pure Compact Nucleic Acid Isolation Kit I	NA
18	Farnaes, L et al (2019)	NA	NA	NA	Magnetic bead-based method (Omega Biotek, Norcross, GA)	NA
19	Graf, E et al. (2016)	NA	NA	NA	QIAamp viral RNA minikit	NA
20	Ruppé, E et al. (2017)	NA	NA	NA	Ultra-Deep Microbiome Prep kit (Molzym, Bremen, Germany)	NA
21	Hasman, H et al. (2014)	NA	NA	NA	Easy-DNA kit (Invitrogen)	NA

Supplementary Table 1. Continued

Study no.	Study (Author and year)	Amplification/PCR (if not part of kit/library prep)	Post-extraction (SPRI clean	Library Prep.	DNA yield	DNA purity
Included in the meta-analysis						
1	Hong, NT et al. (2020)	PCR	AMPure XP SPRI beads	Nextera XT library preparation Kit	NA	NA
2	Miller, S et al. (2019)	NA	AMPure XP beads	Nextera XT library preparation Kit	NA	NA
3	Wilson, M et al. (2019)	NA	AMPure XP beads	Nextera XT library preparation Kit	NA	NA
4	Blauwkamp, T et al. (2019)	NA	NA	Ovation Ultralow System V2 library preparation kits (NuGEN)	NA	NA
5	Parize, P et al. (2017)	PCR	Agencourt AMPure XP bead	NA	NA	NA
6	Rossoff, J et al. (2019)	NA	NA	Ovation Ultralow System V2 library preparation kits (NuGEN	NA	NA
7	Somasekar, S et al. (2017)	PCR	NA	Nextera XT library Prep Kit	NA	NA
8	Schlaberg, R et al. (2017)	NA	NA	TruSeq RNA sample prep kit	NA	NA
9	Street, T. et al. (2017)	NA	AMPure XP SPRI beads	Illumina Nextera XT protocol	NA	NA
10	Thoendel, M et al. (2018)	Multiple Displacement Amplification	AMPure XP SPRI beads	NEBNext Ultra DNA Library Prep Kit	NA	NA
11	Ivy, M et al. (2018)	Multiple Displacement Amplification	AMPure XP SPRI beads	NEBNext Ultra DNA Library Prep Kit	NA	NA
12	Doan, T et al. (2017)	NA	NA	Nextera XT DNA library preparation Kit	NA	NA
13	Cheng, J et al. (2019)	PCR	NA	NA	NA	NA
Not included in the meta-analysis						
14	Xing, X et al. (2019)	PCR	NA	NA	NA	NA
15	Anson, L. et al. (2018)	NA	AMPure XP beads	Rapid Sequencing Kit (SQK-RAD002)	94–301 ng µl-1	(260/280 ratio 0.9–2.1)
16	Horiba, K et al. (2018)	NA	AMPure XP beads	Nextera XT library Prep Kit	NA	NA
17	Charalampous, T et al. (2018	NA	NA	Rapid Low-Input by PCR Sequencing Kit (SQK-RLI001), the Rapid Low-Input Barcoding Kit (SQK-RLB001), and the Rapid PCR Barcoding Kit (SQK-RPB004).	NA	NA
18	Farnaes, L et al (2019)	NA	NA	Modified Ovation SystemV2 library kit (NuGEN, San Carlos,	NA	NA
19	Graf, E et al. (2016)	NA	NA	TruSeq RNA sample prep kit	NA	NA
20	Ruppé, E et al. (2017)	NA	NA	Nextera XT DNA Sample Preparation Kit	NA	NA
21	Hasman, H et al. (2014)	NA	NA	Ion Xpress Plus gDNA and Amplicon Library 98 preparation	NA	NA

Supplementary Table 1. Continued

Study no.	Study (Author and year)	% of human DNA	Control (Internal)	Control (Negative)	Control (Positive)	Contamination
Included in the meta-analysis						
1	Hong, NT et al. (2020)	NA	NA	Yes	NA	yes
2	Miller, S et al. (2019)	NA	Yes, spiked with bacteriophage	Yes, 1.5 controls per 10 samples processed	Yes, mixture of 7 representative organisms	Yes
3	Wilson, M et al. (2019)	NA	Yes, spiked with bacteriophage	Yes, 1.5 controls per 10 samples processed	Yes, mixture of 7 representative organisms	Yes
4	Blauwkamp, T et al. (2019)	NA	Yes, spiked with whole-assay internal normalization control (WINC) molecules	Yes	Yes, 4 microorganisms	Yes
5	Parize, P et al. (2017)	NA	NA	NA	NA	Yes
6	Rossoff, J et al. (2019)	NA	Yes, spiked with whole-assay internal normalization control (WINC) molecules	NA	NA	NA
7	Somasekar, S et al. (2017)	NA	NA	Yes, 0.5 controls per 10 samples processed	Yes, hepatitis A/B virus	Yes
8	Schlaberg, R et al. (2017)	NA	NA	NA	NA	NA
9	Street, T. et al. (2017)	NA	NA	Yes, 1 control every batch	NA	Yes
10	Thoendel, M et al. (2018)	NA	NA	Yes	NA	Yes
11	Ivy, M et al. (2018)	NA	NA	Yes, 1 control every batch	Yes, Corynebacterium glutamicum	Yes
12	Doan, T et al. (2017)	NA	NA	Yes, 1 in each library prep.	NA	NA
13	Cheng, J et al. (2019)	NA	NA	Yes, 7 controls	NA	Yes
Not included in the meta-analysis						
14	Xing, X et al. (2019)	NA	NA	Yes	NA	NA
15	Anson, L. et al. (2018)	5%	NA	NA	NA	Yes
16	Horiba, K et al. (2018)	NA	NA	NA	NA	Yes
17	Charalampous, T et al. (2018)	NA	NA	NA	NA	Yes
18	Farnaes, L et al (2019)	NA	NA	Yes	Yes	NA
19	Graf, E et al. (2016)	NA	NA	NA	NA	yes
20	Ruppé, E et al. (2017)	NA	NA	Yes, 2 controls	NA	Yes
21	Hasman, H et al. (2014)	NA	NA	NA	NA	Yes

Supplementary Table 2. Bioinformatic Details. Contains all bioinformatic related details of individual studies including classification tools and database utilised.

Study no.	Study (Author and year)	Type of sample	Sequencing techniques	Pipeline (existing; custom; commercial)	Pipeline name	Classification
Included in the meta-analysis						
1	Hong, NT et al. (2020)	Cerebrospinal fluid	Illumina MiSeq	Custom	Not found	BLAST
2	Miller, S et al. (2019)	Cerebrospinal fluid	Illumina HiSeq	Existing	SURPI+ software	SNAP and RAPSearch
3	Wilson, M et al. (2019)	Cerebrospinal fluid	Illumina HiSeq	Existing	SURPI+ software	SNAP and RAPSearch
4	Blauwkamp, T et al. (2019)	Plasma (cfDNA)	Illumina NextSeq 500	Custom	Not found	BLAST v2.2.30
5	Parize, P et al. (2017)	Plasma	Ion Proton (Life Technologies)	Commercial	PathoQuest SAS	Burrows–Wheeler alignment
6	Rossoff, J et al. (2019)	Plasma	Illumina NextSeq 500	Custom	Not found	BLAST v2.2.30
7	Somasekar, S et al. (2017)	Serum	Illumina HiSeq	Existing	SURPI+ software	SNAP and RAPSearch
8	Schlaberg, R et al. (2017)	Respiratory samples	Illumina HiSeq 2500	Existing	Not found	Taxonomer
9	Street, T. et al. (2017)	Sonication/synovial fluid	Illumina MiSeq	Custom	Not found	Kraken
10	Thoendel, M et al. (2018)	Sonication/synovial fluid	Illumina HiSeq 2500	Custom	Not found	(LMAT; version 1.2.6)/MetaPhlan2
11	Ivy, M et al. (2018)	Sonication/synovial fluid	Illumina HiSeq 2500	Custom	Not found	(LMAT; version 1.2.6)/MetaPhlan2
12	Doan, T et al. (2017)	Intraocular fluid samples	Illumina HiSeq 4000	Custom	Not found	Rapsearch2 algorithm
13	Cheng, J et al. (2019)	Heart valves	Illumina HiSeq	Custom	Not found	SNAP
Not included in the meta-analysis						
14	Xing, X et al. (2019)	Cerebrospinal fluid	BGISEQ-500/50 platform	Custom	Not found	Not found
15	Anson, L. et al. (2018)	Liquid blood culture (positive)	Illumina MiSeq; MinION	Custom	Not found	Kraken
16	Horiba, K et al. (2018)	Plasma; serum	Illumina HiSeq 2500	Existing	MePIC v2.0	MEGABLAST
17	Charalampous, T et al. (2018)	Respiratory samples	ONT MinION	Commercial	WIMP	Centrifuge
18	Farnaes, L et al (2019)	Respiratory samples (cfDNA)	Illumina NextSeq 500	Custom	Not found	Blast v2.2.30
19	Graf, E et al. (2016)	Respiratory samples	Illumina HiSeq 2500	Existing	Not found	Taxonomer
20	Ruppé, E et al. (2017)	Sonication/synovial fluid	Illumina HiSeq 2500	Custom	Not found	MetaPhlAn2
21	Hasman, H et al. (2014)	Urine	Ion Torrent PGM system	Custom	MG-RAST	Kmer/MG-RAST/BWA

Supplementary Table 2. Continued

Study no.	Study (Author and year)	Classification Heuristic	Antibiotic prediction	Mapped against single genome or multplie	Database
Included in the meta-analysis					
1	Hong, NT et al. (2020)	Alignment	Not found	Multiple (unspecified)	NCBI (RefSeq)
2	Miller, S et al. (2019)	Alignment	Not found	Multiple (bacteria, viruses, fungi/moulds and parasites)	NCBI nt (RefSeq)
3	Wilson, M et al. (2019)	Alignment	Not found	Multiple (bacteria, viruses, fungi/moulds and parasites)	NCBI nt (RefSeq)
4	Blauwkamp, T et al. (2019)	Alignment	Not found	Multiple (bacteria, viruses, fungi/moulds and eukaryotic)	NCBI (RefSeq)
5	Parize, P et al. (2017)	Alignment	Not found	Multiple (bacteria, viruses)	NCBI nt (unspecified) July 2013
6	Rossoff, J et al. (2019)	Alignment	Not found	Multiple (bacteria, viruses, fungi/moulds and eukaryotic)	NCBI (RefSeq)
7	Somasekar, S et al. (2017)	Alignment	Manual	Multiple (unspecified)	NCBI nt (unspecified)
8	Schlaberg, R et al. (2017)	Kmer	Not found	Multiple (viruses)	NCBI (unspecified)
9	Street, T. et al. (2017)	Kmer	Not found	Bacteria	NCBI (RefSeq) Jan 2017
10	Thoendel, M et al. (2018)	Kmer/Marker-gene	Not found	Multiple (bacteria, archaeal, viral, eukaryotic)	LMAT database/ MetaPhlAn2 db
11	Ivy, M et al. (2018)	Kmer/Marker-gene	Not found	Multiple (bacteria, archaeal, viral, eukaryotic)	LMAT database/ MetaPhlAn2 db
12	Doan, T et al. (2017)	Alignment	Not found	Multiple (bacteria, viruses, fungi/moulds and parasites)	NCBI nt (unspecified) July 2015
13	Cheng, J et al. (2019)	Alignment	CARD	Multiple (bacteria, archaeal, viral, eukaryotic)	NCBI (unspecified)
Not included in the meta-analysis					
14	Xing, X et al. (2019)	Not found	Not found	Multiple (bacteria, viruses, fungi/moulds and parasites)	NCBI nt (RefSeq)
15	Anson, L. et al. (2018)	Kmer	Mykrobe; resistType	Multiple (bacteria, viruses and human genomes)	NCBI (RefSeq) Jan 2015
16	Horiba, K et al. (2018)	Alignment	ResFinder	Multiple (unspecified)	NCBI nt (unspecified)
17	Charalampous, T et al. (2018)	Kmer	ARMA(CARD)	Multiple (bacteria, viruses, fungi)	NCBI (RefSeq)
18	Farnaes, L et al (2019)	Alignment	Not found	Multiple (bacteria, viruses, fungi/moulds and parasites)	Karius' reference database
19	Graf, E et al. (2016)	Kmer	Manual	Multiple (unspecified)	NCBI (unspecified)
20	Ruppé, E et al. (2017)	Marker-gene	ResFinder	Multiple (bacteria, archaeal, viral, eukaryotic)	MetaPhlAn2 database
21	Hasman, H et al. (2014)	Kmer/Alignment	ResFinder	Multiple (bacteria, fungi)	NCBI nt (GenBank)

Supplementary Table 2. Continued

Study no.	Study (Author and year)	Demultiplexing	Adaptor Removal	Quality Trimmed/ filtered	Remove low-complexity	Removing human reads	Human DNA	DeNovo Assembly	Phylogenetic analysis
Included in the meta-analysis									
1	Hong, NT et al. (2020)	Not found	Not found	Not found	Not found	Not found	Not found	Not found	Not found
2	Miller, S et al. (2019)	Not found	Not found	Not found	Not found	Not found	Not found	Not found	Not found
3	Wilson, M et al. (2019)	Not found	Not found	Not found	Not found	Not found	Not found	Not found	Not found
4	Blauwkamp, T et al. (2019)	bcl2fastq v2.17.1.14	Not found	Trimmomatic v0.32	Not found	Bowtie v2.2.4	Not found	Not found	Not found
5	Parize, P et al. (2017)	Not found	Not found	Not found	Not found	Burrows–Wheeler	hg19	Not found	Not found
6	Rossoff, J et al. (2019)	bcl2fastq v2.17.1.14	Not found	Trimmomatic v0.32	Not found	Bowtie v2.2.4	Not found	Not found	Not found
7	Somasekar, S et al. (2017)	Not found	Not found	Not found	Not found	Not found	Not found	Geneious	Not found
8	Schlaberg, R et al. (2017)	Not found	Not found	Not found	Not found	Not found	Not found	Not found	Not found
9	Street, T. et al. (2017)	Not found	BBDuk	Not found	Not found	Not found	GRCh38	Not found	Not found
10	Thoendel, M et al. (2018)	Trimmomatic (v0.36)	Not found	Not found	Not found	BioBloom tools	Not found	Not found	Not found
11	Ivy, M et al. (2018)	Trimmomatic (v0.36)	Not found	Not found	Not found	BioBloom Tools	Not found	Not found	Not found
12	Doan, T et al. (2017)	Not found	Not found	PriceSeqFilter	LZW algorithm	STAR/Bowtie2 (v2.2.4)	hg38/PanTro4	Not found	Not found
13	Cheng, J et al. (2019)	Not found	Not found	Not found	Not found	BWA	GRCh38	Not found	Not found
Not included in the meta-analysis									
14	Xing, X et al. (2019)	Not found	Not found	Not found	Not found	Not found	hg19	Not found	Not found
15	Anson, L. et al. (2018)	Not found	Not found	Not found	Not found	Not found	Not found	Not found	Not found
16	Horiba, K et al. (2018)	MePIC v2.0	Not found	Not found	Not found	MePIC v2.0	Not found	Not found	Not found
17	Charalampous, T et al. (2018)	Not found	Not found	Not found	Not found	Not found	Not found	Canu	Not found
18	Farnaes, L et al (2019)	bcl2fastq	Not found	Not found	Not found	Bowtie2	Not found	Not found	Not found
19	Graf, E et al. (2016)	Not found	Not found	Not found	Not found	Not found	Not found	Not found	Not found
20	Ruppé, E et al. (2017)	Not found	Not found	FastQC	Not found	Kraken	Not found	metaSPAdes	Not found
21	Hasman, H et al. (2014)	Not found	Not found	Not found	Not found	BWA	Not found	Not found	snpTree

Supplementary Table 3. Performance of each study. Contains sensitivity and specificity performance data of studies, including resistance prediction performance.

Study no.	Study (Author and year)	Type of Sample	True positives	False negatives	True negatives	False positives	Standard compared to:	Additional diagnosis
Included in the meta-analysis								
1	Hong, NT et al. (2020)	Cerebrospinal fluid	14	5	31	16	PCR	7 pathogens
2	Miller, S et al. (2019)	Cerebrospinal fluid	50	6	117	1	Routine & additional PCR	8 samples
3	Wilson, M et al. (2019)	Cerebrospinal fluid	32	26	146	3	Routine & additional PCR	13 pathogens
4	Blauwkamp, T et al. (2019)	Blood	169	13	104	62	Routine & additional PCR	62 samples
5	Parize, P et al. (2017)	Blood	12	7	58	24	Culture, serological & PCR	25 samples
6	Somasekar, S et al. (2017)	Blood	25	1	245	1	PCR	9 pathogens
7	Rossoff, J et al. (2019)	Blood	56	5	25	14	Clinical review	14 pathogens
8	Schlaberg, R et al. (2017)	Respiratory samples	54	6	38	21	Clinical review	17 pathogens
9	Street, T. et al. (2017)	Sonication/synovial fluid	61	8	85	12	Culture	3 samples
10	Thoendel, M et al. (2018)	Sonication/synovial fluid	121	25	46	21	Culture	21 samples
11	Ivy, M et al. (2018)	Sonication/synovial fluid	67	15	21	39	Culture	4 samples
12	Doan, T et al. (2017)	Intraocular fluid samples	27	4	28	8	PCR	6 pathogens
13	Cheng, J et al. (2019)	Heart valves	40	1	6	1	Clinical review	1 sample
Not included in the meta-analysis								
14	Xing, X et al. (2019)	Cerebrospinal fluid	13	3	Not found	Not found	Culture	4 samples
15	Anson, L. et al. (2018)	Blood	41	3	Not found	Not found	Culture	4 cases
16	Horiba, K et al. (2018)	Blood	8	4	Not found	Not found	Culture	3 samples
17	Charalampous, T et al. (2018)	Respiratory samples	32	1	Not found	Not found	Culture	27 pathogens
18	Farnaes, L et al. (2019)	Respiratory samples	13	2	Not found	Not found	Culture & multiplex PCR	7 samples
19	Graf, E et al. (2016)	Respiratory samples	48	19	Not found	Not found	Multiplex respiratory panel	7 pathogens
20	Ruppé, E et al. (2017)	Sonication/synovial fluid	40	23	Not found	Not found	Culture	Not found
21	Hasman, H et al. (2014)	Urine	17	2	Not found	Not found	Culture	2 samples

Supplementary Table 3. Continued

Study no.	Study (Author and year)	% Human reads	Drug susceptibility	Samples tested for AMR	Phenotypic comparison	Susceptible genotype	Susceptible phenotype	Resistant genotype	Resistant phenotype	Common drugs tested
Included in the meta-analysis										
1	Hong, NT et al. (2020)	Not found	Not found	Not found	Not found					Not found
2	Miller, S et al. (2019)	Not found	Not found	Not found	Not found					Not found
3	Wilson, M et al. (2019)	Not found	Not found	Not found	Not found					Not found
4	Blauwkamp, T et al. (2019)	Not found	Not found	Not found	Not found					Not found
5	Parize, P et al. (2017)	Not found	Not found	Not found	Not found					Not found
6	Somasekar, S et al. (2017)	84%	Yes	10	Not compared					Not found
7	Rossoff, J et al. (2019)	Not found	Not found	Not found	Not found					Not found
8	Schlaberg, R et al. (2017)	Not found	Not found	Not found	Not found					Not found
9	Street, T. et al. (2017)	98.40%	Not found	Not found	Not found					Not found
10	Thoendel, M et al. (2018)	Not found	Not found	Not found	Not found					Not found
11	Ivy, M et al. (2018)	76%	Not found	Not found	Not found					Not found
12	Doan, T et al. (2017)	Not found	Yes	3	Not compared					Not found
13	Cheng, J et al. (2019)	Not found	Yes	1	Not compared					Not found
Not included in the meta-analysis										
14	Xing, X et al. (2019)	Not found	Not found	Not found	Not found					Not found
15	Anson, L. et al. (2018)	Not found	Yes	39	BD Pheonix	286	294	42	49	Penicillin, amoxicillin, co-amoxiclav, cefttriacin, cirpofloxacin.
16	Horiba, K et al. (2018)	98%	Yes, 62 resistance genes	10	Not compared					Not found
17	Charalampous, T et al. (2018)	Not found	Yes	33	Routine microbiology	15	15	22	33	Amoxicillin, ciprofloxacin, gentamicin
18	Farnaes, L et al. (2019)	Not found	Not found	Not found	Not found					Not found
19	Graf, E et al. (2016)	Not found	Yes	Not found	Not compared					Not found
20	Ruppé, E et al. (2017)	Not found	Yes	24	Laboratory DST	284	326	24	43	Beta-lactamases, tetracycline, erythromycin, trimethoprim
21	Hasman, H et al. (2014)	Not found	Yes	17	Laboratory DST	7	7	26	28	Penicillin, tetracycline, ampicillin

Supplementary Table 4. QUADAS-2 results. Details the QUADAS-2 appraisal for each study in the meta-analysis.

Study no.	Study (Author and year)	Sig. Q1	Sig. Q2	Sig. Q3	D1 (ROB)	Sig. Q1	Sig. Q2	D2 (ROB)	Sig. Q1	Sig. Q2	D3 (ROB)
Included in the meta-analysis											
1	Hong, NT et al. (2020)	yes	yes	yes	low	yes	yes	low	yes	yes	low
2	Miller, S et al. (2019)	yes	yes	yes	low	yes	yes	low	yes	yes	low
3	Wilson, M et al. (2019)	yes	yes	yes	low	yes	yes	low	yes	yes	low
4	Blauwkamp, T et al. (2019)	yes	yes	yes	low	yes	yes	low	yes	yes	low
5	Parize, P et al. (2017)	yes	yes	yes	low	yes	yes	low	yes	yes	low
6	Somasekar, S et al. (2017)	yes	no	yes	Some concerns	yes	yes	low	yes	yes	low
7	Rossoff, J et al. (2019)	yes	yes	yes	low	yes	yes	low	yes	yes	low
8	Schlaberg, R et al. (2017)	yes	yes	yes	low	yes	yes	low	yes	yes	low
9	Street, T. et al. (2017)	yes	yes	yes	low	yes	yes	low	yes	yes	low
10	Thoendel, M et al. (2018)	yes	yes	yes	low	yes	yes	low	yes	yes	low
11	Ivy, M et al. (2018)	yes	yes	yes	low	yes	yes	low	yes	yes	low
12	Doan, T et al. (2017)	yes	no	yes	Some concerns	yes	yes	low	yes	yes	low
13	Cheng, J et al. (2019)	yes	yes	yes	low	unclear	unclear	Some concerns	yes	yes	low
Not included in the meta-analysis											
14	Xing, X et al. (2019)	no	no	no	high	yes	yes	low	yes	yes	low
15	Anson, L. et al. (2018)	no	no	no	high	yes	no	Some concerns	yes	yes	low
16	Horiba, K et al. (2018)	yes	yes	yes	low	yes	yes	low	yes	yes	low
17	Charalampous, T et al. (2018)	no	yes	yes	Some concerns	unclear	unclear	Some concerns	yes	yes	low
18	Farnaes, L et al (2019)	yes	yes	yes	low	yes	yes	low	yes	yes	low
19	Graf, E et al. (2016)	yes	yes	yes	low	yes	yes	low	yes	yes	low
20	Ruppé, E et al. (2017)	yes	yes	yes	low	yes	unclear	Some concerns	yes	yes	low
21	Hasman, H et al. (2014)	yes	yes	yes	low	yes	yes	low	yes	yes	low

Supplementary Table 4. Continued

Study no.	Study (Author and year)	Sig. Q1	Sig. Q2	Sig. Q3	D4 (ROB)	Overall	Random samples	D1 (App.)	D2 (App.)	D3 (App.)	Overall
Included in the meta-analysis											
1	Hong, NT et al. (2020)	yes	yes	yes	low	low	yes	low	low	low	low
2	Miller, S et al. (2019)	yes	yes	yes	low	low	yes	low	low	low	low
3	Wilson, M et al. (2019)	yes	yes	yes	low	low	yes	low	low	low	low
4	Blauwkamp, T et al. (2019)	yes	yes	yes	low	low	yes	low	low	low	low
5	Parize, P et al. (2017)	yes	yes	yes	low	low	yes	low	low	low	low
6	Somasekar, S et al. (2017)	yes	yes	yes	low	low	yes	low	low	low	low
7	Rossoff, J et al. (2019)	yes	yes	yes	low	low	yes	low	low	low	low
8	Schlaberg, R et al. (2017)	yes	yes	yes	low	low	yes	low	low	low	low
9	Street, T. et al. (2017)	yes	yes	yes	low	low	yes	low	low	low	low
10	Thoendel, M et al. (2018)	yes	yes	yes	low	low	yes	low	low	low	low
11	Ivy, M et al. (2018)	yes	yes	yes	low	low	yes	low	low	low	low
12	Doan, T et al. (2017)	yes	yes	yes	low	low	yes	low	low	low	low
13	Cheng, J et al. (2019)	yes	yes	yes	low	low	yes	low	Some concerns	low	Low
Not included in the meta-analysis											
14	Xing, X et al. (2019)	yes	yes	yes	low	Some concerns	no	low	low	low	low
15	Anson, L. et al. (2018)	yes	yes	yes	low	Some concerns	no	Some concerns	low	low	low
16	Horiba, K et al. (2018)	yes	yes	yes	low	low	yes	low	low	low	low
17	Charalampous, T et al. (2018)	yes	yes	yes	low	Some concerns	yes	Some concerns	Some concerns	low	Some concerns
18	Farnaes, L et al (2019)	yes	yes	yes	low	low	yes	low	low	low	low
19	Graf, E et al. (2016)	yes	yes	yes	low	low	yes	low	low	low	low
20	Ruppé, E et al. (2017)	yes	yes	yes	low	low	yes	low	Some concerns	low	low
21	Hasman, H et al. (2014)	yes	yes	yes	low	low	yes	low	low	low	low

Supplementary Table 5. Other meta-analysis measures. Details other important meta-analysis measures such as PPV, NPV and diagnostic accuracy.

Type of Sample	Sensitivity [95% CI]*	Specificty [95% CI]*	PPV [95% CI]*	NPV [95% CI]*	Diagnostic accuracy [95% CI]*
Cerebrospinal fluid	0.75 [0.54-0.89]	0.96 [0.72-1]	0.88 [0.50-0.98]	0.90 [0.82-0.94]	0.87 [0.69-0.95]
Blood	0.90 [0.78-0.96]	0.86 [0.45-0.98]	0.75 [0.46-0.92]	0.94 [0.79-0.98]	0.88 [0.62-0.97]
Respiratory samples	0.90 [0.79-0.95]	0.64 [0.52-0.76]	0.72 [0.61-0.81]	0.86 [0.72-0.93]	0.77 [0.68-0.83]
Sonication/synovial fluid	0.84[0.79-0.88]	0.67 [0.38-0.87]	0.79 [0.66-0.87]	0.75 [0.52-0.89]	0.77 [0.63-0.87]
Intraocular fluid samples	0.87 [0.70-0.96]	0.78 [0.61-0.90]	0.77 [0.61-0.88]	0.88 [0.71-0.95]	0.82 [0.71-0.89]
Heart valves	0.98 [0.87-1]	0.86 [0.42-1.00]	0.98 [0.85-0.99]	0.86 [0.41-0.98]	0.95 [0.84-0.98]

Note: Studies that reported no specificity were excluded from this analysis. Total clinical samples in this analysis (n=2023).

* a random effects model was used.

Supplementary Table 6. Detailed threshold abundance recommended to call additional true species and genus.

No	PMID	Citation	First Author	Reason for exclusion
1	31639732	J Appl Lab Med. 2019 Jan;3(4):643-653. doi: 10.1373/jalm.2018.026120. Epub 2018 Nov 30.	Greninger AL	7-Reviews/Commentary
2	30918369	Nat Rev Genet. 2019 Jun;20(6):341-355. doi: 10.1038/s41576-019-0113-7.	Chiu CY	7-Reviews/Commentary
3	31189036	N Engl J Med. 2019 Jun 13;380(24):2327-2340. doi: 10.1056/NEJMoa1803396.	Wilson MR	1-Included
4	30355154	Annu Rev Pathol. 2019 Jan 24;14:319-338. doi: 10.1146/annurev-pathmechdis-012418-012751. Epub 2018 Oct 24.	Gu W	7-Reviews/Commentary
5	32015508	Nature. 2020 Mar;579(7798):265-269. doi: 10.1038/s41586-020-2008-3. Epub 2020 Feb 3.	Wu F	6-Case studies/series
6	28042011	J Biotechnol. 2017 Feb 10;243:16-24. doi: 10.1016/j.jbiotec.2016.12.022. Epub 2016 Dec 29.	Deurenberg RH	7-Reviews/Commentary
7	32525600	FASEB J. 2020 Jul;34(7):8787-8795. doi: 10.1096/fj.202001115R. Epub 2020 Jun 11.	Debnath M	10-Non-clinical diagnostic studies
8	31184572	J Med Microbiol. 2019 Aug;68(8):1204-1210. doi: 10.1099/jmm.0.000994. Epub 2019 Jun 11.	Xing XW	1-Included
9	31681628	Front Cell Infect Microbiol. 2019 Oct 18;9:351. doi: 10.3389/fcimb.2019.00351. eCollection 2019.	Zhou X	4-Non-metagenomic, targeted or non-WGS methods
10	29787889	Clin Microbiol Infect. 2018 Sep;24(9):944-955. doi: 10.1016/j.cmi.2018.05.007. Epub 2018 May 19.	Peker N	7-Reviews/Commentary
11	24848255	Sci Transl Med. 2014 May 21;6(237):237ra65. doi: 10.1126/scitranslmed.3008599.	Aagaard K	5-Microbiome studies
12	30482864	Proc Natl Acad Sci U S A. 2018 Dec 26;115(52):E12353-E12362. doi: 10.1073/pnas.1809700115. Epub 2018 Nov 27.	Langelier C	5-Microbiome studies
13	29040428	Clin Infect Dis. 2018 Feb 10;66(5):778-788. doi: 10.1093/cid/cix881.	Simner PJ	7-Reviews/Commentary
14	32345639	mBio. 2020 Apr 28;11(2):e00218-20. doi: 10.1128/mBio.00218-20.	Neugent ML	5-Microbiome studies
15	29055712	Virus Res. 2018 Jan 15;244:218-229. doi: 10.1016/j.virusres.2017.10.014. Epub 2017 Oct 18.	Greninger AL	7-Reviews/Commentary
16	29305150	J Infect. 2018 Mar;76(3):225-240. doi: 10.1016/j.jinf.2017.12.014. Epub 2018 Jan 2.	Brown JR	7-Reviews/Commentary
17	30397356	Nat Med. 2018 Dec;24(12):1919-1929. doi: 10.1038/s41591-018-0222-4. Epub 2018 Nov 5.	Sun L	5-Microbiome studies
18	27628595	Microb Ecol. 2017 Feb;73(2):492-503. doi: 10.1007/s00248-016-0854-1. Epub 2016 Sep 14.	Gomez A	5-Microbiome studies
19	32379863	Clin Chem. 2020 Jul 1;66(7):966-972. doi: 10.1093/clinchem/hvaa106.	Peddu V	9-Low sample size or pilot study
20	31661299	Am J Respir Cell Mol Biol. 2020 Mar;62(3):283-299. doi: 10.1165/rcmb.2019-0273TR.	Carney SM	5-Microbiome studies
21	29648630	Clin Infect Dis. 2018 Oct 15;67(9):1333-1338. doi: 10.1093/cid/ciy303.	Thoendel MJ	1-Included
22	30992304	Genome Res. 2019 May;29(5):831-842. doi: 10.1101/gr.238170.118. Epub 2019 Apr 16.	Miller S	1-Included
23	29074157	Clin Microbiol Infect. 2018 Apr;24(4):335-341. doi: 10.1016/j.cmi.2017.10.013. Epub 2017 Oct 23.	Besser J	7-Reviews/Commentary
24	29710329	JAMA Neurol. 2018 Aug 1;75(8):947-955. doi: 10.1001/jamaneurol.2018.0463.	Wilson MR	6-Case studies/series
25	31385886	Curr Opin Organ Transplant. 2019 Oct;24(5):527-534. doi: 10.1097/MOT.0000000000000682.	Fishman JA	7-Reviews/Commentary
26	29848568	J Clin Microbiol. 2018 Aug 27;56(9):e00402-18. doi: 10.1128/JCM.00402-18. Print 2018 Sep.	Ivy MI	1-Included
27	28538734	Nature. 2017 Jun 15;546(7658):411-415. doi: 10.1038/nature22402. Epub 2017 May 24.	Metsky HC	10-Non-clinical diagnostic studies
28	31349542	Int J Mol Sci. 2019 Jul 25;20(15):3631. doi: 10.3390/ijms20153631.	Kuo MT	4-Non-metagenomic, targeted or non-WGS methods
29	31619531	J Clin Microbiol. 2019 Dec 23;58(1):e01315-19. doi: 10.1128/JCM.01315-19. Print 2019 Dec 23.	Petersen LM	7-Reviews/Commentary
30	29129921	Nat Rev Genet. 2018 Jan;19(1):9-20. doi: 10.1038/nrg.2017.88. Epub 2017 Nov 13.	Gardy JL	10-Non-clinical diagnostic studies
31	30834996	Eur J Clin Microbiol Infect Dis. 2019 Jun;38(6):1059-1070. doi: 10.1007/s10096-019-03520-3. Epub 2019 Mar 5.	Boers SA	7-Reviews/Commentary
32	30666248	Front Immunol. 2019 Jan 7;9:2868. doi: 10.3389/fimmu.2018.02868. eCollection 2018.	Malla MA	5-Microbiome studies
33	26811607	World J Gastroenterol. 2016 Jan 14;22(2):557-66. doi: 10.3748/wjg.v22.i2.557.	Nosho K	10-Non-clinical diagnostic studies
34	32174267	Emerg Microbes Infect. 2020 Mar 16;9(1):597-600. doi: 10.1080/22221751.2020.1738905. eCollection 2020.	Ai JW	6-Case studies/series
35	28628622	PLoS One. 2017 Jun 19;12(6):e0177633. doi: 10.1371/journal.pone.0177633. eCollection 2017.	Bossa L	5-Microbiome studies
36	29186529	J Infect Dis. 2018 Jan 30;217(4):628-636. doi: 10.1093/infdis/jix604.	de Gunzburg J	5-Microbiome studies
37	28169558	Arch Pathol Lab Med. 2017 Jun;141(6):776-786. doi: 10.5858/arpa.2016-0539-RA. Epub 2017 Feb 7.	Schlaberg R	4-Non-metagenomic, targeted or non-WGS methods
38	32159296	Orthop Surg. 2020 Apr;12(2):463-470. doi: 10.1111/os.12641. Epub 2020 Mar 11.	Zhang CF	4-Non-metagenomic, targeted or non-WGS methods
39	32403302	Genes (Basel). 2020 May 11;11(5):536. doi: 10.3390/genes11050536.	Françoise A	5-Microbiome studies
40	31451660	Proc Natl Acad Sci U S A. 2019 Sep 10;116(37):18738-18744. doi: 10.1073/pnas.1906320116. Epub 2019 Aug 26.	Cheng AP	10-Non-clinical diagnostic studies
41	24982315	Clin Microbiol Rev. 2014 Jul;27(3):419-40. doi: 10.1128/CMR.00092-13.	Achermann Y	10-Non-clinical diagnostic studies
42	31666364	J Clin Microbiol. 2019 Dec 23;58(1):e00963-19. doi: 10.1128/JCM.00963-19. Print 2019 Dec 23.	Lewandowski K	9-Low sample size or pilot study
43	26778104	Transfus Clin Biol. 2016 Feb;23(1):28-38. doi: 10.1016/j.tracli.2015.12.002. Epub 2016 Jan 9.	Sauvage V	7-Reviews/Commentary
44	31932713	Nat Microbiol. 2020 Mar;5(3):443-454. doi: 10.1038/s41564-019-0637-9. Epub 2020 Jan 13.	Deng X	10-Non-clinical diagnostic studies
45	31881145	N Engl J Med. 2019 Dec 26;381(26):2569-2580. doi: 10.1056/NEJMsr1813907.	Armstrong GL	10-Non-clinical diagnostic studies
46	29458686	J Med Microbiol. 2018 Mar;67(3):347-357. doi: 10.1099/jmm.0.000664. Epub 2018 Jan 10.	Anson LW	1-Included
47	27766372	J Mol Med (Berl). 2017 Jan;95(1):41-51. doi: 10.1007/s00109-016-1478-0. Epub 2016 Oct 20.	Willmann M	10-Non-clinical diagnostic studies
48	23337890	J Invest Dermatol. 2013 Sep;133(9):2152-60. doi: 10.1038/jid.2013.21. Epub 2013 Jan 21.	Fitz-Gibbon S	5-Microbiome studies
49	31367035	Nature. 2019 Aug;572(7769):329-334. doi: 10.1038/s41586-019-1451-5. Epub 2019 Jul 31.	de Goffau MC	5-Microbiome studies
50	29988504	Front Cell Infect Microbiol. 2018 Jun 25;8:205. doi: 10.3389/fcimb.2018.00205. eCollection 2018.	Li H	9-Low sample size or pilot study
51	29968805	Sci Rep. 2018 Jul 3;8(1):10056. doi: 10.1038/s41598-018-28477-9.	Nagy-Szakal D	4-Non-metagenomic, targeted or non-WGS methods

Supplementary Table 6. Continued

52	31078141	Microbiome. 2019 May 11;7(1):73. doi: 10.1186/s40168-019-0687-5.	Rocafort M	5-Microbiome studies
53	26927306	PLoS Negl Trop Dis. 2016 Feb 29;10(2):e0004470. doi: 10.1371/journal.pntd.0004470. eCollection 2016 Feb.	Decuypere S	4-Non-metagenomic, targeted or non-WGS methods
54	30388112	PLoS One. 2018 Nov 2;13(11):e0205064. doi: 10.1371/journal.pone.0205064. eCollection 2018.	Jiang ZD	5-Microbiome studies
55	32138779	Microbiome. 2020 Mar 5;8(1):28. doi: 10.1186/s40168-020-00811-2.	Heshiki Y	5-Microbiome studies
56	23033984	Clin Microbiol Infect. 2012 Dec;18(12):1185-93. doi: 10.1111/1469-0691.12023. Epub 2012 Oct 3.	Lagier JC	5-Microbiome studies
57	32690600	Gut. 2021 Feb;70(2):276-284. doi: 10.1136/gutjnl-2020-322294. Epub 2020 Jul 20.	Zuo T	5-Microbiome studies
58	30234027	Front Cell Infect Microbiol. 2018 Aug 28;8:301. doi: 10.3389/fcimb.2018.00301. eCollection 2018.	Wang Z	5-Microbiome studies
59	31436107	J Int Med Res. 2019 Oct;47(10):4878-4885. doi: 10.1177/0300060519866953. Epub 2019 Aug 22.	Liu N	9-Low sample size or pilot study
60	31001490	Front Cell Infect Microbiol. 2019 Apr 3;9:90. doi: 10.3389/fcimb.2019.00090. eCollection 2019.	Hu Y	5-Microbiome studies
61	32408110	EBioMedicine. 2020 May;55:102766. doi: 10.1016/j.ebiom.2020.102766. Epub 2020 May 11.	Wei M	5-Microbiome studies
62	31554674	J Clin Microbiol. 2019 Nov 22;57(12):e01113-19. doi: 10.1128/JCM.01113-19. Print 2019 Dec.	Carpenter ML	10-Non-clinical diagnostic studies
63	31398374	J Infect. 2019 Oct;79(4):383-388. doi: 10.1016/j.jinf.2019.08.003. Epub 2019 Aug 6.	Jerome H	9-Low sample size or pilot study
64	31332169	Nat Commun. 2019 Jul 22;10(1):3273. doi: 10.1038/s41467-019-11152-6.	Antunes KH	5-Microbiome studies
65	30239621	Clin Infect Dis. 2019 May 17;68(11):1847-1855. doi: 10.1093/cid/ciy802.	Zinter MS	5-Microbiome studies
66	28490492	J Clin Microbiol. 2017 Aug;55(8):2334-2347. doi: 10.1128/JCM.00462-17. Epub 2017 May 10.	Street TL	1-Included
67	31681624	Front Cell Infect Microbiol. 2019 Oct 18;9:346. doi: 10.3389/fcimb.2019.00346. eCollection 2019.	Zhu S	5-Microbiome studies
68	29681906	Front Immunol. 2018 Apr 6;9:712. doi: 10.3389/fimmu.2018.00712. eCollection 2018.	Iino C	5-Microbiome studies
69	31488869	Sci Rep. 2019 Sep 5;9(1):12807. doi: 10.1038/s41598-019-49189-8.	Amrane S	5-Microbiome studies
70	30478001	EBioMedicine. 2018 Dec;38:265-272. doi: 10.1016/j.ebiom.2018.11.035. Epub 2018 Nov 23.	Wei S	5-Microbiome studies
71	27468850	Genome Biol. 2016 Jul 28;17(1):163. doi: 10.1186/s13059-016-1021-1.	Yu G	5-Microbiome studies
72	23489584	J Am Acad Dermatol. 2013 Jul;69(1):143-55. doi: 10.1016/j.jaad.2013.01.016. Epub 2013 Mar 13.	Chen YE	5-Microbiome studies
73	31123265	Nat Commun. 2019 May 23;10(1):2288. doi: 10.1038/s41467-019-10294-x.	Li Y	10-Non-clinical diagnostic studies
74	28798333	Sci Rep. 2017 Aug 10;7(1):7718. doi: 10.1038/s41598-017-07546-5.	Ruppé E	1-Included
75	32448200	BMC Infect Dis. 2020 May 24;20(1):367. doi: 10.1186/s12879-020-05085-5.	Ding L	4-Non-metagenomic, targeted or non-WGS methods
76	27209148	Lancet. 2016 Jul 30;388(10043):498-503. doi: 10.1016/S0140-6736(16)30386-5. Epub 2016 May 18.	Jacobs M	6-Case studies/series
77	32251799	Int J Infect Dis. 2020 Jun;95:8-12. doi: 10.1016/j.ijid.2020.03.055. Epub 2020 Apr 3.	Wang CX	9-Low sample size or pilot study
78	31371723	Nat Commun. 2019 Aug 1;10(1):3460. doi: 10.1038/s41467-019-11494-1.	Coyne MJ	5-Microbiome studies
79	32505872	Int J Infect Dis. 2020 Jul;96:573-578. doi: 10.1016/j.ijid.2020.05.125. Epub 2020 Jun 4.	Cai Y	9-Low sample size or pilot study
80	31114866	Nucleic Acids Res. 2019 Aug 22;47(14):e83. doi: 10.1093/nar/gkz418.	Quan J	4-Non-metagenomic, targeted or non-WGS methods
81	24843156	Proc Natl Acad Sci U S A. 2014 Jun 3;111(22):E2329-38. doi: 10.1073/pnas.1319284111. Epub 2014 May 19.	Franzosa EA	5-Microbiome studies
82	28539351	Gut. 2018 Apr;67(4):634-643. doi: 10.1136/gutjnl-2017-313952. Epub 2017 May 24.	Zuo T	5-Microbiome studies
83	30110400	PLoS One. 2018 Aug 15;13(8):e0202049. doi: 10.1371/journal.pone.0202049. eCollection 2018.	Watanabe N	4-Non-metagenomic, targeted or non-WGS methods
84	32497137	PLoS One. 2020 Jun 4;15(6):e0232610. doi: 10.1371/journal.pone.0232610. eCollection 2020.	Wang H	9-Low sample size or pilot study
85	31775777	Respir Res. 2019 Nov 27;20(1):265. doi: 10.1186/s12931-019-1218-4.	Yang L	9-Low sample size or pilot study
86	31255652	Gastroenterology. 2019 Oct;157(4):1109-1122. doi: 10.1053/j.gastro.2019.06.028. Epub 2019 Jun 27.	Schwimmer JB	5-Microbiome studies
87	31164085	BMC Infect Dis. 2019 Jun 4;19(1):495. doi: 10.1186/s12879-019-4132-y.	Zhang XX	4-Non-metagenomic, targeted or non-WGS methods
88	30784601	Cell Rep. 2019 Feb 19;26(8):2227-2240.e5. doi: 10.1016/j.celrep.2019.01.091.	Nelson MT	5-Microbiome studies
89	24576467	Parasitology. 2014 Dec;141(14):1856-62. doi: 10.1017/S0031182014000134. Epub 2014 Feb 27.	Pallen MJ	7-Reviews/Commentary
90	30355340	BMC Microbiol. 2018 Oct 24;18(1):157. doi: 10.1186/s12866-018-1304-7.	Mailhe M	5-Microbiome studies
91	32314307	Infection. 2020 Aug;48(4):535-542. doi: 10.1007/s15010-020-01429-0. Epub 2020 Apr 20.	Chen X	4-Non-metagenomic, targeted or non-WGS methods
92	30406048	Front Cell Infect Microbiol. 2018 Oct 23;8:375. doi: 10.3389/fcimb.2018.00375. eCollection 2018.	Oechslin CP	10-Non-clinical diagnostic studies
93	32083024	Front Cell Infect Microbiol. 2020 Jan 29;9:470. doi: 10.3389/fcimb.2019.00470. eCollection 2019.	Chen J	5-Microbiome studies
94	30459201	mBio. 2018 Nov 20;9(6):e02248-18. doi: 10.1128/mBio.02248-18.	Hannigan GD	10-Non-clinical diagnostic studies
95	30128994	Methods Mol Biol. 2018;1838:125-140. doi: 10.1007/978-1-4939-8682-8_10.	Yang Y	4-Non-metagenomic, targeted or non-WGS methods
96	32439542	Int J Infect Dis. 2020 Aug;97:54-59. doi: 10.1016/j.ijid.2020.05.044. Epub 2020 May 19.	Wang CX	9-Low sample size or pilot study
97	28679063	Am J Respir Crit Care Med. 2017 Dec 15;196(12):1618-1619. doi: 10.1164/rccm.201706-1144LE.	Pendleton KM	7-Reviews/Commentary
98	22674328	Science. 2012 Jun 8;336(6086):1314-7. doi: 10.1126/science.1221789. Epub 2012 Jun 6.	Iliev ID	5-Microbiome studies
99	28506279	Microbiome. 2017 May 15;5(1):54. doi: 10.1186/s40168-017-0271-9.	Rath S	5-Microbiome studies
100	22310478	Genome Res. 2012 May;22(5):850-9. doi: 10.1101/gr.131029.111. Epub 2012 Feb 6.	Kong HH	5-Microbiome studies
101	30046045	Sci Rep. 2018 Jul 25;8(1):11218. doi: 10.1038/s41598-018-29571-8.	Kose SH	10-Non-clinical diagnostic studies
102	30652494	Am J Physiol Lung Cell Mol Physiol. 2019 Mar 1;316(3):L578-L584. doi: 10.1152/ajplung.00476.2018. Epub 2019 Jan 17.	Kalantar KL	5-Microbiome studies
103	32054085	Int J Mol Sci. 2020 Feb 11;21(4):1205. doi: 10.3390/ijms21041205.	Langan EA	5-Microbiome studies

Supplementary Table 6. Continued

104	31378866	Semin Neurol. 2019 Jun;39(3):297-311. doi: 10.1055/s-0039-1688441. Epub 2019 Aug 2.	Kanjilal S	4-Non-metagenomic, targeted or non-WGS methods
105	30563591	Euro Surveill. 2018 Dec;23(50):1800228. doi: 10.2807/1560-7917.ES.2018.23.50.1800228.	Kafetzopoulou LE	4-Non-metagenomic, targeted or non-WGS methods
106	31988585	World J Gastroenterol. 2020 Jan 14;26(2):199-218. doi: 10.3748/wjg.v26.i2.199.	Lv XY	4-Non-metagenomic, targeted or non-WGS methods
107	31421897	Ophthalmology. 2019 Dec;126(12):1724-1726. doi: 10.1016/j.ophtha.2019.06.013. Epub 2019 Jun 18.	Seitzman GD	6-Case studies/series
108	31286879	BMC Infect Dis. 2019 Jul 8;19(1):591. doi: 10.1186/s12879-019-4231-9.	Tschumi F	6-Case studies/series
109	30619779	Front Cell Infect Microbiol. 2018 Dec 12;8:433. doi: 10.3389/fcimb.2018.00433. eCollection 2018.	Hu YL	5-Microbiome studies
110	31738795	PLoS One. 2019 Nov 18;14(11):e0224658. doi: 10.1371/journal.pone.0224658. eCollection 2019.	Cheong HC	5-Microbiome studies
111	31127652	APMIS. 2019 Aug;127(8):577-587. doi: 10.1111/apm.12969. Epub 2019 Jun 19.	Zhao-Fleming HH	10-Non-clinical diagnostic studies
112	28085120	Viruses. 2017 Jan 12;9(1):10. doi: 10.3390/v9010010.	Pérot P	7-Reviews/Commentary
113	31806002	BMC Oral Health. 2019 Dec 5;19(1):275. doi: 10.1186/s12903-019-0956-5.	Yeoh YK	5-Microbiome studies
114	26490635	Sci Rep. 2015 Oct 22;5:14771. doi: 10.1038/srep14771.	Qin N	5-Microbiome studies
115	31287781	Emerg Microbes Infect. 2019;8(1):1043-1053. doi: 10.1080/22221751.2019.1637702.	Matsumoto Y	4-Non-metagenomic, targeted or non-WGS methods
116	28838146	J Infect Dis. 2017 Jul 15;216(2):203-209. doi: 10.1093/infdis/jix264.	Ogimi C	10-Non-clinical diagnostic studies
117	28686513	Am J Respir Crit Care Med. 2018 Feb 15;197(4):524-528. doi: 10.1164/rccm.201706-1097LE.	Langelier C	6-Case studies/series
118	30589932	JAMA Ophthalmol. 2019 Apr 1;137(4):435-438. doi: 10.1001/jamaophthalmol.2018.6185.	Gonzales JA	4-Non-metagenomic, targeted or non-WGS methods
119	28337072	J Biomol Tech. 2017 Apr;28(1):40-45. doi: 10.7171/jbt.17-2801-007. Epub 2017 Mar 21.	Afshinnekoo E	10-Non-clinical diagnostic studies
120	32339718	Int J Infect Dis. 2020 Jul;96:270-275. doi: 10.1016/j.ijid.2020.04.048. Epub 2020 Apr 24.	Yan L	4-Non-metagenomic, targeted or non-WGS methods
121	31052348	Genes (Basel). 2019 Apr 30;10(5):332. doi: 10.3390/genes10050332.	Edridge AWD	9-Low sample size or pilot study
122	31572693	Front Cell Infect Microbiol. 2019 Sep 12;9:327. doi: 10.3389/fcimb.2019.00327. eCollection 2019.	Reid G	10-Non-clinical diagnostic studies
123	28172519	Clin Infect Dis. 2017 Feb 15;64(4):476-481. doi: 10.1093/cid/ciw767.	Bouquet J	10-Non-clinical diagnostic studies
124	32570978	Int J Mol Sci. 2020 Jun 18;21(12):4351. doi: 10.3390/ijms21124351.	Abe Y	10-Non-clinical diagnostic studies
125	29076868	Curr Opin Gastroenterol. 2018 Jan;34(1):19-24. doi: 10.1097/MOG.0000000000000405.	Macfarlane-Smith LR	4-Non-metagenomic, targeted or non-WGS methods
126	27045926	Am J Gastroenterol. 2016 Jun;111(6):879-90. doi: 10.1038/ajg.2016.95. Epub 2016 Apr 5.	D'Argenio V	5-Microbiome studies
127	31894501	Mycopathologia. 2020 Apr;185(2):207-231. doi: 10.1007/s11046-019-00413-z. Epub 2020 Jan 1.	Tiew PY	5-Microbiome studies
128	28927397	BMC Infect Dis. 2017 Sep 19;17(1):631. doi: 10.1186/s12879-017-2727-8.	Abayasekara LM	4-Non-metagenomic, targeted or non-WGS methods
129	29111455	Virus Res. 2018 Jan 2;243:83-90. doi: 10.1016/j.virusres.2017.10.016. Epub 2017 Oct 27.	Shi M	4-Non-metagenomic, targeted or non-WGS methods
130	30346341	Cornea. 2019 Feb;38(2):246-248. doi: 10.1097/ICO.0000000000001790.	Seitzman GD	4-Non-metagenomic, targeted or non-WGS methods
131	24296350	Cold Spring Harb Perspect Med. 2013 Dec 1;3(12):a015362. doi: 10.1101/cshperspect.a015362.	Hannigan GD	5-Microbiome studies
132	22719852	PLoS One. 2012;7(6):e37818. doi: 10.1371/journal.pone.0037818. Epub 2012 Jun 18.	Srinivasan S	5-Microbiome studies
133	30213965	Sci Rep. 2018 Sep 13;8(1):13767. doi: 10.1038/s41598-018-31873-w.	Couto N	5-Microbiome studies
134	30373528	BMC Infect Dis. 2018 Oct 29;18(1):537. doi: 10.1186/s12879-018-3446-5.	Bal A	9-Low sample size or pilot study
135	31800638	PLoS One. 2019 Dec 4;14(12):e0225961. doi: 10.1371/journal.pone.0225961. eCollection 2019.	Choi S	5-Microbiome studies
136	31572686	Front Cell Infect Microbiol. 2019 Sep 12;9:305. doi: 10.3389/fcimb.2019.00305. eCollection 2019.	Genton L	5-Microbiome studies
137	30366549	Neurol Clin. 2018 Nov;36(4):687-703. doi: 10.1016/j.ncl.2018.07.004.	Ramachandran PS	7-Reviews/Commentary
138	31466373	Genes (Basel). 2019 Aug 28;10(9):655. doi: 10.3390/genes10090655.	Junier T	7-Reviews/Commentary
139	31536612	PLoS Pathog. 2019 Sep 19;15(9):e1008009. doi: 10.1371/journal.ppat.1008009. eCollection 2019 Sep.	Van Dycke J	10-Non-clinical diagnostic studies
140	31745012	Indian J Med Microbiol. 2019 Apr-Jun;37(2):133-140. doi: 10.4103/ijmm.IJMM_19_401.	Jacob JJ	7-Reviews/Commentary
141	27136432	Blood Transfus. 2016 Sep;14(5):400-7. doi: 10.2450/2016.0160-15. Epub 2016 Mar 16.	Sauvage V	10-Non-clinical diagnostic studies
142	32448130	BMC Infect Dis. 2020 May 24;20(1):370. doi: 10.1186/s12879-020-05092-6.	Gao LY	6-Case studies/series
143	26763966	J Clin Microbiol. 2016 Apr;54(4):919-27. doi: 10.1128/JCM.03050-15. Epub 2016 Jan 13.	Hasan MR	10-Non-clinical diagnostic studies
144	32178660	BMC Pulm Med. 2020 Mar 17;20(1):65. doi: 10.1186/s12890-020-1098-x.	Gu L	4-Non-metagenomic, targeted or non-WGS methods
145	29939994	PLoS Comput Biol. 2018 Jun 25;14(6):e1006277. doi: 10.1371/journal.pcbi.1006277. eCollection 2018 Jun.	Lu J	10-Non-clinical diagnostic studies
146	28368491	J Infect Dis. 2017 May 1;215(9):1407-1415. doi: 10.1093/infdis/jix148.	Schlaberg R	4-Non-metagenomic, targeted or non-WGS methods
147	28811246	Clin Microbiol Infect. 2018 Mar;24(3):308.e1-308.e4. doi: 10.1016/j.cmi.2017.08.006. Epub 2017 Aug 12.	Motro Y	10-Non-clinical diagnostic studies
148	29468143	Front Cell Infect Microbiol. 2018 Feb 6;8:25. doi: 10.3389/fcimb.2018.00025. eCollection 2018.	Gigliucci F	5-Microbiome studies
149	30742071	Nat Microbiol. 2019 Apr;4(4):663-674. doi: 10.1038/s41564-018-0349-6. Epub 2019 Feb 11.	Blauwkamp TA	1-Included
150	28356418	J Clin Microbiol. 2017 Jun;55(6):1789-1801. doi: 10.1128/JCM.02402-16. Epub 2017 Mar 29.	Thoendel M	10-Non-clinical diagnostic studies
151	32228597	BMC Infect Dis. 2020 Mar 30;20(1):253. doi: 10.1186/s12879-020-04955-2.	Wang C	9-Low sample size or pilot study
152	22925473	Clin Microbiol Infect. 2013 Feb;19(2):107-12. doi: 10.1111/j.1469-0691.2012.04001.x. Epub 2012 Aug 27.	Wolcott R	9-Low sample size or pilot study
153	31681625	Front Cell Infect Microbiol. 2019 Oct 11;9:347. doi: 10.3389/fcimb.2019.00347. eCollection 2019.	Acharya A	5-Microbiome studies
154	29447223	PLoS One. 2018 Feb 15;13(2):e0192878. doi: 10.1371/journal.pone.0192878. eCollection 2018.	Romero-Espinoza JA	5-Microbiome studies
155	32046792	Microbiome. 2020 Feb 11;8(1):18. doi: 10.1186/s40168-020-0793-4.	Burnham P	10-Non-clinical diagnostic studies

Supplementary Table 6. Continued

156	26879053	Parasit Vectors. 2016 Feb 16;9:86. doi: 10.1186/s13071-016-1363-2.	Gao D	6-Case studies/series
157	27352877	Am J Trop Med Hyg. 2016 Sep 7;95(3):663-669. doi: 10.4269/ajtmh.16-0062. Epub 2016 Jun 27.	Zhou Y	4-Non-metagenomic, targeted or non-WGS methods
158	30576430	Clin Infect Dis. 2019 Aug 30;69(6):941-948. doi: 10.1093/cid/ciy1020.	Casto AM	10-Non-clinical diagnostic studies
159	31267543	Scand J Immunol. 2020 Jan;91(1):e12805. doi: 10.1111/sji.12805. Epub 2019 Nov 10.	Zhang L	5-Microbiome studies
160	31585992	Appl Environ Microbiol. 2019 Nov 27;85(24):e01820-19. doi: 10.1128/AEM.01820-19. Print 2019 Dec 15.	Peña-Gonzalez A	5-Microbiome studies
161	26888896	J Clin Microbiol. 2016 May;54(5):1197-202. doi: 10.1128/JCM.00289-16. Epub 2016 Feb 17.	Polage CR	8-Epidemiological studies
162	22752862	Clin Infect Dis. 2012 Aug;55 Suppl 2(Suppl 2):S132-42. doi: 10.1093/cid/cis338.	Louie TJ	5-Microbiome studies
163	30889229	PLoS One. 2019 Mar 19;14(3):e0213384. doi: 10.1371/journal.pone.0213384. eCollection 2019.	Portillo A	5-Microbiome studies
164	32591865	Parasitol Res. 2020 Aug;119(8):2623-2629. doi: 10.1007/s00436-020-06754-9. Epub 2020 Jun 26.	Wylezich C	5-Microbiome studies
165	23331482	Clin Microbiol Infect. 2013 Sep;19(9):784-9. doi: 10.1111/1469-0691.12109. Epub 2013 Jan 18.	Price J	7-Reviews/Commentary
166	24525633	J Innate Immun. 2014;6(3):272-83. doi: 10.1159/000358835. Epub 2014 Feb 12.	Boyd JH	10-Non-clinical diagnostic studies
167	30192933	Gigascience. 2018 Oct 1;7(10):giy107. doi: 10.1093/gigascience/giy107.	Li F	5-Microbiome studies
168	32656571	New Microbiol. 2020 Jul;43(3):121-126. Epub 2020 Jul 13.	Scapaticci M	4-Non-metagenomic, targeted or non-WGS methods
169	22113000	Curr Opin Pulm Med. 2012 Jan;18(1):14-22. doi: 10.1097/MCP.0b013e32834dccc0.	Brar T	5-Microbiome studies
170	32076089	Sci Rep. 2020 Feb 19;10(1):2985. doi: 10.1038/s41598-020-59957-6.	Helmersen K	10-Non-clinical diagnostic studies
171	30697958	Cancer Med. 2019 Mar;8(3):1013-1023. doi: 10.1002/cam4.1985. Epub 2019 Jan 29.	Dharnidharka VR	4-Non-metagenomic, targeted or non-WGS methods
172	31882580	Sci Rep. 2019 Dec 27;9(1):19946. doi: 10.1038/s41598-019-55703-9.	Griffen AL	5-Microbiome studies
173	30953744	Ophthalmology. 2019 Aug;126(8):1090-1094. doi: 10.1016/j.ophtha.2019.03.039. Epub 2019 Apr 4.	Lalitha P	5-Microbiome studies
174	23585843	PLoS One. 2013 Apr 9;8(4):e60670. doi: 10.1371/journal.pone.0060670. Print 2013.	Shipitsyna E	5-Microbiome studies
175	30986208	PLoS One. 2019 Apr 15;14(4):e0209881. doi: 10.1371/journal.pone.0209881. eCollection 2019.	Lambert JS	4-Non-metagenomic, targeted or non-WGS methods
176	32127629	Sci Rep. 2020 Mar 3;10(1):3963. doi: 10.1038/s41598-020-60992-6.	Li CX	5-Microbiome studies
177	28595575	BMC Microbiol. 2017 Jun 8;17(1):133. doi: 10.1186/s12866-017-1041-3.	Andersen SC	4-Non-metagenomic, targeted or non-WGS methods
178	25337710	PLoS One. 2014 Oct 22;9(10):e110943. doi: 10.1371/journal.pone.0110943. eCollection 2014.	Fettweis JM	10-Non-clinical diagnostic studies
179	30110939	Viruses. 2018 Aug 14;10(8):429. doi: 10.3390/v10080429.	Osunmakinde CO	7-Reviews/Commentary
180	31689572	Virology. 2020 Jan 2;539:69-79. doi: 10.1016/j.virol.2019.10.012. Epub 2019 Oct 31.	Forberg K	10-Non-clinical diagnostic studies
181	27921234	Sci China Life Sci. 2017 Mar;60(3):279-286. doi: 10.1007/s11427-016-0244-8. Epub 2016 Dec 2.	Zou X	9-Low sample size or pilot study
182	30810995	Int Microbiol. 2019 Sep;22(3):337-342. doi: 10.1007/s10123-018-00053-1. Epub 2019 Jan 7.	Wiscovitch-Russo R	10-Non-clinical diagnostic studies
183	24998424	J Clin Virol. 2014 Sep;61(1):9-19. doi: 10.1016/j.jcv.2014.06.013. Epub 2014 Jun 24.	Quiñones-Mateu ME	7-Reviews/Commentary
184	33121615	Clin Lab Med. 2020 Dec;40(4):459-472. doi: 10.1016/j.cll.2020.08.004. Epub 2020 Aug 8.	Uhteg K	7-Reviews/Commentary
185	30261842	BMC Genomics. 2018 Sep 27;19(1):714. doi: 10.1186/s12864-018-5094-y.	Sanderson ND	9-Low sample size or pilot study
186	30816674	Cell Physiol Biochem. 2019;52(2):263-279. doi: 10.33594/000000020. Epub 2019 Feb 28.	Philley JV	5-Microbiome studies
187	30287413	Clin Microbiol Infect. 2019 Jun;25(6):688-695. doi: 10.1016/j.cmi.2018.09.017. Epub 2018 Oct 2.	Nutman A	10-Non-clinical diagnostic studies
188	24172157	J Clin Microbiol. 2014 Jan;52(1):139-46. doi: 10.1128/JCM.02452-13. Epub 2013 Oct 30.	Hasman H	1-Included
189	24803517	mBio. 2014 May 6;5(3):e01021-14. doi: 10.1128/mBio.01021-14.	Schubert AM	5-Microbiome studies
190	31900105	BMC Infect Dis. 2020 Jan 3;20(1):7. doi: 10.1186/s12879-019-4723-7.	Preiswerk B	4-Non-metagenomic, targeted or non-WGS methods
191	31855136	Emerg Infect Dis. 2020 Jan;26(1):11-19. doi: 10.3201/eid2601.190983.	Hattori N	4-Non-metagenomic, targeted or non-WGS methods
192	32631930	Thorax. 2020 Sep;75(9):780-790. doi: 10.1136/thoraxjnl-2019-214187. Epub 2020 Jul 6.	Nelson MT	5-Microbiome studies
193	21741763	Cancer Lett. 2011 Oct 28;309(2):119-27. doi: 10.1016/j.canlet.2011.06.004. Epub 2011 Jun 24.	Zhu Y	5-Microbiome studies
194	32709241	Parasit Vectors. 2020 Jul 24;13(1):371. doi: 10.1186/s13071-020-04228-6.	Tokarz R	10-Non-clinical diagnostic studies
195	23279287	Clin Microbiol Infect. 2013 Jan;19(1):15-22. doi: 10.1111/1469-0691.12056.	Capobianchi MR	7-Reviews/Commentary
196	31536536	PLoS One. 2019 Sep 19;14(9):e0222323. doi: 10.1371/journal.pone.0222323. eCollection 2019.	Kirst ME	5-Microbiome studies
197	28945480	Ann Am Thorac Soc. 2017 Sep;14(Supplement_3):S237-S241. doi: 10.1513/AnnalsATS.201608-634MG.	Burnham P	5-Microbiome studies
198	31566105	J Am Heart Assoc. 2019 Oct;8(19):e011797. doi: 10.1161/JAHA.118.011797. Epub 2019 Sep 28.	Amar J	5-Microbiome studies
199	31235920	Nat Biotechnol. 2019 Jul;37(7):783-792. doi: 10.1038/s41587-019-0156-5. Epub 2019 Jun 24.	Charalampous T	1-Included
200	30670612	mBio. 2019 Jan 22;10(1):e02262-18. doi: 10.1128/mBio.02262-18.	Bowers JR	4-Non-metagenomic, targeted or non-WGS methods
201	30567558	Ann Clin Microbiol Antimicrob. 2018 Dec 20;17(1):43. doi: 10.1186/s12941-018-0294-5.	Cheng J	9-Low sample size or pilot study
202	31039443	Clin Microbiol Infect. 2019 Sep;25(9):1086-1095. doi: 10.1016/j.cmi.2019.04.019. Epub 2019 Apr 27.	Mintzer V	7-Reviews/Commentary
203	31473248	Microb Pathog. 2019 Dec;137:103675. doi: 10.1016/j.micpath.2019.103675. Epub 2019 Aug 29.	Selvin J	5-Microbiome studies
204	31529842	Biomedica. 2019 Aug 1;39(Supl. 2):157-171. doi: 10.7705/biomedica.v39i4.4520.	Caguazango JC	5-Microbiome studies
205	23601179	Clin Microbiol Infect. 2013 Sep;19(9):803-13. doi: 10.1111/1469-0691.12217. Epub 2013 Apr 18.	Bertelli C	7-Reviews/Commentary
206	28192237	Clin Microbiol Infect. 2017 Aug;23(8):574.e1-574.e6. doi: 10.1016/j.cmi.2017.02.006. Epub 2017 Feb 10.	Parize P	1-Included
207	31660864	BMC Infect Dis. 2019 Oct 28;19(1):905. doi: 10.1186/s12879-019-4462-9.	Ko ER	6-Case studies/series

Supplementary Table 6. Continued

208	30429253	J Clin Microbiol. 2019 Jan 30;57(2):e01182-18. doi: 10.1128/JCM.01182-18. Print 2019 Feb.	Yan Q	9-Low sample size or pilot study
209	31577814	PLoS One. 2019 Oct 2;14(10):e0222915. doi: 10.1371/journal.pone.0222915. eCollection 2019.	Vijayvargiya P	9-Low sample size or pilot study
210	29340642	Invest Ophthalmol Vis Sci. 2018 Jan 1;59(1):280-288. doi: 10.1167/iovs.17-21617.	Li Z	9-Low sample size or pilot study
211	29491015	J Clin Microbiol. 2018 Apr 25;56(5):e01663-17. doi: 10.1128/JCM.01663-17. Print 2018 May.	Amrane S	5-Microbiome studies
212	30926543	Int J Infect Dis. 2019 Jun;83:148-153. doi: 10.1016/j.ijid.2019.03.007. Epub 2019 Mar 27.	Cheng J	1-Included
213	30152468	Rev Sci Tech. 2017 Aug;36(2):499-512. doi: 10.20506/rst.36.2.2669.	Epstein JH	7-Reviews/Commentary
214	23063405	Virology. 2012 Dec 20;434(2):151-61. doi: 10.1016/j.virol.2012.09.017. Epub 2012 Oct 12.	Henry M	7-Reviews/Commentary
215	29515160	Sci Rep. 2018 Mar 7;8(1):4126. doi: 10.1038/s41598-018-22416-4.	Kirstahler P	10-Non-clinical diagnostic studies
216	25374023	Biomed Environ Sci. 2014 Nov;27(11):898-901. doi: 10.3967/bes2014.126.	Yang DH	9-Low sample size or pilot study
217	22174638	Int J Mol Sci. 2011;12(11):7861-84. doi: 10.3390/ijms12117861. Epub 2011 Nov 14.	Barzon L	7-Reviews/Commentary
218	25759963	Curr Opin Ophthalmol. 2015 May;26(3):221-5. doi: 10.1097/ICU.0000000000000145.	Hong BK	7-Reviews/Commentary
219	25356041	World J Gastroenterol. 2014 Oct 28;20(40):14805-20. doi: 10.3748/wjg.v20.i40.14805.	Wang ZK	7-Reviews/Commentary
220	29936877	J Int Med Res. 2018 Aug;46(8):3468-3473. doi: 10.1177/0300060518780124. Epub 2018 Jun 25.	Chen ZQ	4-Non-metagenomic, targeted or non-WGS methods
221	30991963	BMC Infect Dis. 2019 Apr 16;19(1):326. doi: 10.1186/s12879-019-3953-z.	Foulex A	6-Case studies/series
222	32092386	J Infect. 2020 May;80(5):571-577. doi: 10.1016/j.jinf.2020.02.003. Epub 2020 Feb 22.	Xu Y	10-Non-clinical diagnostic studies
223	29404836	Eur J Clin Microbiol Infect Dis. 2018 May;37(5):851-857. doi: 10.1007/s10096-017-3178-2. Epub 2018 Feb 5.	Boers SA	5-Microbiome studies
224	29547097	Microb Genom. 2018 Mar;4(3):e000167. doi: 10.1099/mgen.0.000167. Epub 2018 Mar 16.	Williams D	10-Non-clinical diagnostic studies
225	30586509	N Engl J Med. 2018 Dec 27;379(26):2529-2539. doi: 10.1056/NEJMoa1803238.	Johnson RC	10-Non-clinical diagnostic studies
226	30598080	BMC Genomics. 2018 Dec 31;19(Suppl 10):876. doi: 10.1186/s12864-018-5284-7.	Chen HM	5-Microbiome studies
227	22378797	Nutr Clin Pract. 2012 Apr;27(2):215-25. doi: 10.1177/0884533611434934. Epub 2012 Feb 29.	Kelly CJ	5-Microbiome studies
228	22614624	Reprod Sci. 2012 Nov;19(11):1154-62. doi: 10.1177/1933719112446075. Epub 2012 May 21.	Macklaim JM	5-Microbiome studies
229	29772404	Virology. 2018 Jul;520:21-29. doi: 10.1016/j.virol.2018.04.006. Epub 2018 May 14.	Hultin E	6-Case studies/series
230	30592753	PLoS One. 2018 Dec 28;13(12):e0209683. doi: 10.1371/journal.pone.0209683. eCollection 2018.	Thissen JB	10-Non-clinical diagnostic studies
231	31838792	J Microbiol Biotechnol. 2020 Feb 28;30(2):248-258. doi: 10.4014/jmb.1908.08016.	Chang DH	5-Microbiome studies
232	29423379	Front Cell Infect Microbiol. 2018 Jan 23;8:5. doi: 10.3389/fcimb.2018.00005. eCollection 2018.	Li Q	5-Microbiome studies
233	32543662	Invest Ophthalmol Vis Sci. 2020 Jun 3;61(6):35. doi: 10.1167/iovs.61.6.35.	Kang Y	5-Microbiome studies
234	30559311	Cold Spring Harb Mol Case Stud. 2018 Dec 17;4(6):a003244. doi: 10.1101/mcs.a003244. Print 2018 Dec.	Sanford E	6-Case studies/series
235	28441964	Microbiome. 2017 Apr 26;5(1):44. doi: 10.1186/s40168-017-0261-y.	Nagy-Szakal D	5-Microbiome studies
236	32425641	Int J Infect Dis. 2020 Aug;97:27-29. doi: 10.1016/j.ijid.2020.05.056. Epub 2020 May 18.	Guo F	6-Case studies/series
237	32641704	Sci Rep. 2020 Jul 8;10(1):11194. doi: 10.1038/s41598-020-68159-z.	Hasan MR	9-Low sample size or pilot study
238	32615925	BMC Infect Dis. 2020 Jul 2;20(1):467. doi: 10.1186/s12879-020-05179-0.	Li M	6-Case studies/series
239	32517705	Ann Clin Microbiol Antimicrob. 2020 Jun 9;19(1):28. doi: 10.1186/s12941-020-00370-2.	Yang T	10-Non-clinical diagnostic studies
240	28720679	Clin Chem. 2017 Nov;63(11):1685-1694. doi: 10.1373/clinchem.2016.259499. Epub 2017 Jul 18.	Koh AY	6-Case studies/series
241	31170005	Foodborne Pathog Dis. 2019 Jul;16(7):474-479. doi: 10.1089/fpd.2019.2636. Epub 2019 Jun 6.	Carleton HA	7-Reviews/Commentary
242	30866714	Gut Microbes. 2019;10(6):676-687. doi: 10.1080/19490976.2019.1586037. Epub 2019 Mar 13.	Park H	5-Microbiome studies
243	30956802	BMJ Open Respir Res. 2019 Mar 12;6(1):e000374. doi: 10.1136/bmjresp-2018-000374. eCollection 2019.	Keravec M	4-Non-metagenomic, targeted or non-WGS methods
244	27931678	Int J Mycobacteriol. 2016 Dec;5(4):384-391. doi: 10.1016/j.ijmyco.2016.11.028.	Jeanes C	7-Reviews/Commentary
245	28146577	PLoS One. 2017 Feb 1;12(2):e0169310. doi: 10.1371/journal.pone.0169310. eCollection 2017.	Traykova D	5-Microbiome studies
246	21524714	Anaerobe. 2011 Aug;17(4):186-90. doi: 10.1016/j.anaerobe.2011.03.016. Epub 2011 Apr 16.	Marrazzo JM	5-Microbiome studies
247	25572898	Clin Infect Dis. 2015 Mar 15;60(6):919-23. doi: 10.1093/cid/ciu912. Epub 2015 Jan 7.	Naccache SN	6-Case studies/series
248	30992350	mBio. 2019 Apr 16;10(2):e00431-19. doi: 10.1128/mBio.00431-19.	Cobián Güemes AG	4-Non-metagenomic, targeted or non-WGS methods
249	22882284	Exp Dermatol. 2012 Oct;21(10):735-9. doi: 10.1111/j.1600-0625.2012.01552.x. Epub 2012 Aug 7.	van der Zee HH	6-Case studies/series
250	31229651	J Mol Diagn. 2019 Sep;21(5):913-923. doi: 10.1016/j.jmoldx.2019.05.002. Epub 2019 Jun 21.	Culbreath K	4-Non-metagenomic, targeted or non-WGS methods
251	32192451	BMC Infect Dis. 2020 Mar 20;20(1):237. doi: 10.1186/s12879-020-04966-z.	Hasan MR	4-Non-metagenomic, targeted or non-WGS methods
252	22897827	Clin Microbiol Infect. 2012 Sep;18(9):825-33. doi: 10.1111/j.1469-0691.2012.03952.x.	Dridi B	7-Reviews/Commentary
253	27494185	PLoS Negl Trop Dis. 2016 Aug 5;10(8):e0004861. doi: 10.1371/journal.pntd.0004861. eCollection 2016 Aug.	Firacative C	8-Epidemiological studies
254	30122883	World J Gastroenterol. 2018 Aug 14;24(30):3448-3461. doi: 10.3748/wjg.v24.i30.3448.	Zhao Y	5-Microbiome studies
255	26551842	Clin Microbiol Infect. 2016 Mar;22(3):258.e1-8. doi: 10.1016/j.cmi.2015.10.031. Epub 2015 Nov 10.	Ignacio A	5-Microbiome studies
256	24451106	Jpn J Infect Dis. 2014;67(1):62-5. doi: 10.7883/yoken.67.62.	Takeuchi F	10-Non-clinical diagnostic studies
257	28178315	PLoS One. 2017 Feb 8;12(2):e0171911. doi: 10.1371/journal.pone.0171911. eCollection 2017.	Tayabali AF	5-Microbiome studies
258	28852043	Sci Rep. 2017 Aug 29;7(1):9802. doi: 10.1038/s41598-017-09844-4.	Mitra A	5-Microbiome studies
259	29996780	BMC Infect Dis. 2018 Jul 11;18(1):319. doi: 10.1186/s12879-018-3218-2.	Birdsell DN	9-Low sample size or pilot study

Supplementary Table 6. Continued

260	31648003	Int J Infect Dis. 2020 Jan;90:1-4. doi: 10.1016/j.ijid.2019.10.020. Epub 2019 Oct 21.	Wu J	6-Case studies/series
261	32091380	Emerg Infect Dis. 2020 Mar;26(3):615-617. doi: 10.3201/eid2603.190661.	Guan Q	4-Non-metagenomic, targeted or non-WGS methods
262	31547860	J Med Case Rep. 2019 Sep 24;13(1):302. doi: 10.1186/s13256-019-2233-2.	Ikuta Y	6-Case studies/series
263	30395567	PLoS Negl Trop Dis. 2018 Nov 5;12(11):e0006889. doi: 10.1371/journal.pntd.0006889. eCollection 2018 Nov.	Koehler JW	4-Non-metagenomic, targeted or non-WGS methods
264	31423298	Antimicrob Resist Infect Control. 2019 Aug 14;8:131. doi: 10.1186/s13756-019-0583-6. eCollection 2019.	Casaburi G	5-Microbiome studies
265	29693356	Asian Pac J Cancer Prev. 2018 Apr 25;19(4):961-967. doi: 10.22034/APJCP.2018.19.4.961.	Tsuchiya Y	5-Microbiome studies
266	25532973	Emerg Infect Dis. 2015 Jan;21(1):48-57. doi: 10.3201/eid2101.140766.	Kohl C	10-Non-clinical diagnostic studies
267	23452163	Clin Microbiol Infect. 2013 Apr;19(4):338-48. doi: 10.1111/1469-0691.12140. Epub 2013 Mar 2.	Aron-Wisnewsky J	5-Microbiome studies
268	31586049	Nat Commun. 2019 Oct 4;10(1):4538. doi: 10.1038/s41467-019-12364-6.	Spohn R	10-Non-clinical diagnostic studies
269	29373675	J Infect Dis. 2018 Aug 24;218(7):1037-1044. doi: 10.1093/infdis/jiy036.	Kuiken T	10-Non-clinical diagnostic studies
270	29196717	Sci Rep. 2017 Dec 1;7(1):16788. doi: 10.1038/s41598-017-17132-4.	Hamad I	5-Microbiome studies
271	30204780	PLoS One. 2018 Sep 11;13(9):e0203503. doi: 10.1371/journal.pone.0203503. eCollection 2018.	Wang T	5-Microbiome studies
272	28945499	J Dent Res. 2018 Feb;97(2):192-200. doi: 10.1177/0022034517732754. Epub 2017 Sep 25.	Baek K	5-Microbiome studies
273	28586923	Ann Intern Med. 2017 Jul 4;167(1):1-7. doi: 10.7326/M17-0085. Epub 2017 Jun 6.	Kandathil AJ	10-Non-clinical diagnostic studies
274	23646159	PLoS One. 2013 Apr 30;8(4):e62917. doi: 10.1371/journal.pone.0062917. Print 2013.	Zemanick ET	5-Microbiome studies
275	23452186	Clin Microbiol Infect. 2013 Apr;19(4):331-7. doi: 10.1111/1469-0691.12170. Epub 2013 Mar 2.	Knaapen M	5-Microbiome studies
276	31137527	Int J Environ Res Public Health. 2019 May 20;16(10):1779. doi: 10.3390/ijerph16101779.	Lee SH	4-Non-metagenomic, targeted or non-WGS methods
277	29976592	J Clin Microbiol. 2018 Aug 27;56(9):e00850-18. doi: 10.1128/JCM.00850-18. Print 2018 Sep.	Dekker JP	7-Reviews/Commentary
278	32040516	PLoS One. 2020 Feb 10;15(2):e0228467. doi: 10.1371/journal.pone.0228467. eCollection 2020.	Toh E	4-Non-metagenomic, targeted or non-WGS methods
279	29118177	J Clin Microbiol. 2018 Jan 24;56(2):e01374-17. doi: 10.1128/JCM.01374-17. Print 2018 Feb.	Liu J	4-Non-metagenomic, targeted or non-WGS methods
280	29540734	Sci Rep. 2018 Mar 14;8(1):4479. doi: 10.1038/s41598-018-22629-7.	San-Juan-Vergara H	5-Microbiome studies
281	30409177	Microbiome. 2018 Nov 8;6(1):201. doi: 10.1186/s40168-018-0579-0.	Zhu Q	5-Microbiome studies
282	28187782	Microbiome. 2017 Feb 10;5(1):20. doi: 10.1186/s40168-017-0234-1.	Feigelman R	5-Microbiome studies
283	32075599	BMC Infect Dis. 2020 Feb 19;20(1):159. doi: 10.1186/s12879-020-4872-8.	Fang M	6-Case studies/series
284	28062553	J Pediatric Infect Dis Soc. 2017 Nov 24;6(4):393-398. doi: 10.1093/jpids/piw066.	Mongkolrattanothai K	6-Case studies/series
285	30961537	BMC Microbiol. 2019 Apr 8;19(1):75. doi: 10.1186/s12866-019-1448-0.	Frickmann H	9-Low sample size or pilot study
286	26757840	Sci Rep. 2016 Jan 13;6:19171. doi: 10.1038/srep19171.	Wallis A	5-Microbiome studies
287	27562436	Genome Med. 2016 Aug 25;8(1):90. doi: 10.1186/s13073-016-0344-6.	Doan T	9-Low sample size or pilot study
288	28494786	Microbiome. 2017 May 12;5(1):53. doi: 10.1186/s40168-017-0266-6.	Cope EK	5-Microbiome studies
289	25497414	J Virol Methods. 2015 Mar;213:139-46. doi: 10.1016/j.jviromet.2014.12.002. Epub 2014 Dec 11.	Li L	10-Non-clinical diagnostic studies
290	28646219	Sci Rep. 2017 Jun 23;7(1):4173. doi: 10.1038/s41598-017-02239-5.	Manso CF	10-Non-clinical diagnostic studies
291	28410379	PLoS One. 2017 Apr 14;12(4):e0175569. doi: 10.1371/journal.pone.0175569. eCollection 2017.	Oberbach A	10-Non-clinical diagnostic studies
292	31776265	J Virol. 2020 Jan 31;94(4):e01519-19. doi: 10.1128/JVI.01519-19. Print 2020 Jan 31.	Katusiime MG	10-Non-clinical diagnostic studies
293	29970463	mBio. 2018 Jul 3;9(4):e00898-18. doi: 10.1128/mBio.00898-18.	Iketani S	10-Non-clinical diagnostic studies
294	25654499	Virulence. 2015;6(3):183-7. doi: 10.4161/21505594.2014.980661. Epub 2015 Feb 5.	Charalampakis G	5-Microbiome studies
295	27142181	Cancer. 2016 Jul 15;122(14):2186-96. doi: 10.1002/cncr.30039. Epub 2016 May 3.	Galloway-Peña JR	5-Microbiome studies
296	31501513	Sci Rep. 2019 Sep 9;9(1):12909. doi: 10.1038/s41598-019-49372-x.	Takeuchi S	9-Low sample size or pilot study
297	26818672	J Clin Microbiol. 2016 Apr;54(4):1000-7. doi: 10.1128/JCM.03060-15. Epub 2016 Jan 27.	Graf EH	9-Low sample size or pilot study
298	29445826	Microb Ecol. 2018 Oct;76(3):801-813. doi: 10.1007/s00248-018-1155-7. Epub 2018 Feb 14.	Cabal A	10-Non-clinical diagnostic studies
299	31162021	Am J Trop Med Hyg. 2019 Jul;101(1):253-259. doi: 10.4269/ajtmh.18-0434.	Säll O	4-Non-metagenomic, targeted or non-WGS methods
300	31253993	J Infect Dis. 2019 Sep 13;220(8):1312-1324. doi: 10.1093/infdis/jiz318.	Mollerup S	10-Non-clinical diagnostic studies
301	26882501	Elife. 2016 Feb 11;5:e10012. doi: 10.7554/eLife.10012.	Repass J	10-Non-clinical diagnostic studies
302	23548716	Public Health Genomics. 2013;16(1-2):37-43. doi: 10.1159/000345607. Epub 2013 Mar 18.	Lucchi NW	4-Non-metagenomic, targeted or non-WGS methods
303	32070282	BMC Infect Dis. 2020 Feb 18;20(1):150. doi: 10.1186/s12879-020-4859-5.	Osterman A	6-Case studies/series
304	28736993	Microbiol Immunol. 2017 Sep;61(9):359-370. doi: 10.1111/1348-0421.12503.	Protasova IN	10-Non-clinical diagnostic studies
305	30232133	J Clin Microbiol. 2018 Nov 27;56(12):e01123-18. doi: 10.1128/JCM.01123-18. Print 2018 Dec.	Wylie KM	4-Non-metagenomic, targeted or non-WGS methods
306	28429164	Eur J Clin Microbiol Infect Dis. 2017 Sep;36(9):1665-1672. doi: 10.1007/s10096-017-2981-0. Epub 2017 Apr 20.	Otto CC	4-Non-metagenomic, targeted or non-WGS methods
307	28974215	Malar J. 2017 Oct 3;16(1):398. doi: 10.1186/s12936-017-2044-0.	Lalremruata A	10-Non-clinical diagnostic studies
308	30870464	PLoS One. 2019 Mar 14;14(3):e0213280. doi: 10.1371/journal.pone.0213280. eCollection 2019.	Burcham ZM	10-Non-clinical diagnostic studies
309	24223808	PLoS One. 2013 Nov 4;8(11):e78454. doi: 10.1371/journal.pone.0078454. eCollection 2013.	Oude Munnink BB	10-Non-clinical diagnostic studies
310	25564342	BMC Med. 2015 Jan 7;13:2. doi: 10.1186/s12916-014-0232-0.	Spichler A	7-Reviews/Commentary
311	28276545	Lab Chip. 2017 Mar 29;17(7):1340-1348. doi: 10.1039/c6lc01277a.	D'Amico L	4-Non-metagenomic, targeted or non-WGS methods

Supplementary Table 6. Continued

312	28930022	Emerg Infect Dis. 2017 Oct;23(10):1964-1968. doi: 10.3201/eid2310.161986.	Chiu CY	6-Case studies/series
313	28062856	Nucleic Acids Res. 2017 May 5;45(8):e61. doi: 10.1093/nar/gkw1328.	van der Helm E	4-Non-metagenomic, targeted or non-WGS methods
314	24094376	J Cyst Fibros. 2014 Jan;13(1):1-9. doi: 10.1016/j.jcf.2013.09.004. Epub 2013 Oct 3.	Burns JL	7-Reviews/Commentary
315	26269741	Microbiome. 2015 Aug 12;3:33. doi: 10.1186/s40168-015-0092-7. eCollection 2015.	Flores R	5-Microbiome studies
316	21251190	BJOG. 2011 Apr;118(5):533-49. doi: 10.1111/j.1471-0528.2010.02840.x. Epub 2011 Jan 20.	Lamont RF	5-Microbiome studies
317	27314269	Virus Genes. 2016 Oct;52(5):620-4. doi: 10.1007/s11262-016-1361-7. Epub 2016 Jun 17.	Pei N	10-Non-clinical diagnostic studies
318	21422915	Ann Surg. 2011 Jun;253(6):1094-101. doi: 10.1097/SLA.0b013e31821175d7.	Morowitz MJ	5-Microbiome studies
319	28669650	EBioMedicine. 2017 Jul;21:142-149. doi: 10.1016/j.ebiom.2017.06.026. Epub 2017 Jun 27.	Malone M	5-Microbiome studies
320	29664913	PLoS One. 2018 Apr 17;13(4):e0194648. doi: 10.1371/journal.pone.0194648. eCollection 2018.	Chai JH	4-Non-metagenomic, targeted or non-WGS methods
321	30337259	Lancet Infect Dis. 2018 Dec;18(12):1360-1367. doi: 10.1016/S1473-3099(18)30476-6. Epub 2018 Oct 15.	Bozio CH	8-Epidemiological studies
322	23015674	J Clin Microbiol. 2012 Dec;50(12):3901-8. doi: 10.1128/JCM.01237-12. Epub 2012 Sep 26.	Xu Y	5-Microbiome studies
323	22647050	Clin Microbiol Infect. 2012 Jul;18 Suppl 4:50-3. doi: 10.1111/j.1469-0691.2012.03866.x.	Cani PD	5-Microbiome studies
324	29426368	BMC Res Notes. 2018 Feb 9;11(1):120. doi: 10.1186/s13104-018-3234-8.	Sharp C	9-Low sample size or pilot study
325	21121043	Pac Symp Biocomput. 2011:154-64. doi: 10.1142/9789814335058_0017.	Moore JH	5-Microbiome studies
326	31059795	Clin Microbiol Infect. 2019 Oct;25(10):1277-1285. doi: 10.1016/j.cmi.2019.04.028. Epub 2019 May 4.	Asplund M	5-Microbiome studies
327	27269884	Clin Microbiol Infect. 2016 Aug;22(8):737.e1-7. doi: 10.1016/j.cmi.2016.05.024. Epub 2016 Jun 3.	Romano-Bertrand S	5-Microbiome studies
328	27613679	Appl Environ Microbiol. 2016 Oct 27;82(22):6672-6681. doi: 10.1128/AEM.01802-16. Print 2016 Nov 15.	Hu Y	5-Microbiome studies
329	28170403	PLoS One. 2017 Feb 7;12(2):e0170622. doi: 10.1371/journal.pone.0170622. eCollection 2017.	Cox MJ	5-Microbiome studies
330	25762776	J Clin Microbiol. 2015 May;53(5):1685-92. doi: 10.1128/JCM.00323-15. Epub 2015 Mar 11.	Zhang S	10-Non-clinical diagnostic studies
331	25761675	Sci Rep. 2015 Mar 12;5:9053. doi: 10.1038/srep09053.	Chen H	5-Microbiome studies
332	31347944	Gut Microbes. 2020;11(2):237-244. doi: 10.1080/19490976.2019.1643673. Epub 2019 Jul 26.	Brown BP	10-Non-clinical diagnostic studies
333	23281612	BMC Genomics. 2012;13 Suppl 8(Suppl 8):S4. doi: 10.1186/1471-2164-13-S8-S4. Epub 2012 Dec 17.	Harwich MD Jr	10-Non-clinical diagnostic studies
334	23026753	Physiology (Bethesda). 2012 Oct;27(5):300-7. doi: 10.1152/physiol.00023.2012.	Burcelin R	5-Microbiome studies
335	28789678	Microbiome. 2017 Aug 8;5(1):94. doi: 10.1186/s40168-017-0317-z.	Lewandowska DW	10-Non-clinical diagnostic studies
336	22558115	PLoS One. 2012;7(4):e35053. doi: 10.1371/journal.pone.0035053. Epub 2012 Apr 25.	Ogilvie LA	10-Non-clinical diagnostic studies
337	30670143	Euro Surveill. 2019 Jan;24(3):1800699. doi: 10.2807/1560-7917.ES.2019.24.3.1800699.	Bal A	10-Non-clinical diagnostic studies
338	27337976	Clin Oral Investig. 2017 May;21(4):1221-1230. doi: 10.1007/s00784-016-1876-2. Epub 2016 Jun 23.	Fernandez Y Mostajo M	5-Microbiome studies
339	24714329	PLoS One. 2014 Apr 8;9(4):e94457. doi: 10.1371/journal.pone.0094457. eCollection 2014.	Botta C	5-Microbiome studies
340	28542207	PLoS One. 2017 May 23;12(5):e0177340. doi: 10.1371/journal.pone.0177340. eCollection 2017.	Lewandowska DW	9-Low sample size or pilot study
341	29246178	Microbiome. 2017 Dec 15;5(1):160. doi: 10.1186/s40168-017-0379-y.	Razavi M	10-Non-clinical diagnostic studies
342	26334731	Microbiome. 2015 Sep 3;3:38. doi: 10.1186/s40168-015-0100-y.	Mitra S	10-Non-clinical diagnostic studies
343	27899620	Nucleic Acids Res. 2017 Jan 4;45(D1):D475-D481. doi: 10.1093/nar/gkw1095. Epub 2016 Nov 29.	Kim H	10-Non-clinical diagnostic studies
344	28719277	Am J Trop Med Hyg. 2017 Jun;96(6):1285-1295. doi: 10.4269/ajtmh.16-0667.	Robinson ML	7-Reviews/Commentary
345	27393975	Virology. 2016 Sep;496:299-305. doi: 10.1016/j.virol.2016.06.013. Epub 2016 Jul 7.	Phan TG	10-Non-clinical diagnostic studies
346	22343914	J Pediatr Gastroenterol Nutr. 2012 Sep;55(3):299-307. doi: 10.1097/MPG.0b013e31824d2548.	Preidis GA	10-Non-clinical diagnostic studies
347	28424887	Arch Virol. 2017 Jul;162(7):2003-2012. doi: 10.1007/s00705-017-3367-4. Epub 2017 Apr 19.	Gong YN	4-Non-metagenomic, targeted or non-WGS methods
348	27385709	J Clin Microbiol. 2016 Oct;54(10):2470-84. doi: 10.1128/JCM.00330-16. Epub 2016 Jul 6.	Thomson E	10-Non-clinical diagnostic studies
349	22798321	Am J Respir Crit Care Med. 2012 Sep 15;186(6):536-45. doi: 10.1164/rccm.201204-0693OC. Epub 2012 Jul 12.	Charlson ES	5-Microbiome studies
350	30535266	Clin Infect Dis. 2019 May 17;68(11):1978-1979. doi: 10.1093/cid/ciy1024.	Ramachandran PS	4-Non-metagenomic, targeted or non-WGS methods
351	23469128	PLoS One. 2013;8(2):e58005. doi: 10.1371/journal.pone.0058005. Epub 2013 Feb 28.	Skraban J	5-Microbiome studies
352	31136295	J Med Microbiol. 2019 Jul;68(7):996-1002. doi: 10.1099/jmm.0.000968. Epub 2019 May 28.	Qi C	9-Low sample size or pilot study
353	28098541	Emerg Infect Dis. 2017 Feb;23(2):363-365. doi: 10.3201/eid2302.161564.	Mollerup S	10-Non-clinical diagnostic studies
354	21324882	Genome Res. 2011 Apr;21(4):626-33. doi: 10.1101/gr.115758.110. Epub 2011 Feb 10.	Muthappan V	5-Microbiome studies
355	24568097	BMC Res Notes. 2014 Feb 25;7:111. doi: 10.1186/1756-0500-7-111.	Campana MG	10-Non-clinical diagnostic studies
356	22417294	J Clin Periodontol. 2012 May;39(5):425-33. doi: 10.1111/j.1600-051X.2012.01856.x. Epub 2012 Mar 14.	Kumar PS	4-Non-metagenomic, targeted or non-WGS methods
357	23656607	BMC Infect Dis. 2013 May 8;13:210. doi: 10.1186/1471-2334-13-210.	Boase S	5-Microbiome studies
358	23300163	J Infect Dis. 2013 Apr;207(7):1105-14. doi: 10.1093/infdis/jit005. Epub 2013 Jan 8.	Redel H	5-Microbiome studies
359	29934497	Emerg Microbes Infect. 2018 Jun 22;7(1):113. doi: 10.1038/s41426-018-0117-y.	Ji Y	5-Microbiome studies
360	22133882	J Clin Invest. 2011 Dec;121(12):4580-3. doi: 10.1172/JCI61592. Epub 2011 Dec 1.	Rompalo A	7-Reviews/Commentary
361	28067829	Int J Mol Sci. 2017 Jan 6;18(1):109. doi: 10.3390/ijms18010109.	D'Argenio V	5-Microbiome studies
362	23118957	PLoS One. 2012;7(10):e48224. doi: 10.1371/journal.pone.0048224. Epub 2012 Oct 30.	Nelson AM	5-Microbiome studies
363	30684744	Int J Infect Dis. 2019 Apr;81:1-3. doi: 10.1016/j.ijid.2019.01.032. Epub 2019 Jan 23.	Wang G	6-Case studies/series

Supplementary Table 6. Continued

364	20819230	BMC Genomics. 2010 Sep 7;11:488. doi: 10.1186/1471-2164-11-488.	Ling Z	5-Microbiome studies
365	25583170	J Infect Dis. 2015 Jun 15;211(12):1895-904. doi: 10.1093/infdis/jiv003. Epub 2015 Jan 12.	Horton JM	5-Microbiome studies
366	26996149	Sci Rep. 2016 Mar 21;6:23532. doi: 10.1038/srep23532.	Gyarmati P	9-Low sample size or pilot study
367	31208781	J Pediatr. 2019 Sep;212:228-231. doi: 10.1016/j.jpeds.2019.04.058. Epub 2019 Jun 14.	Olson CA	7-Reviews/Commentary
368	28282386	PLoS One. 2017 Mar 10;12(3):e0173022. doi: 10.1371/journal.pone.0173022. eCollection 2017.	Pages-Monteiro L	5-Microbiome studies
369	29083040	J Med Virol. 2018 Mar;90(3):412-420. doi: 10.1002/jmv.24984. Epub 2017 Nov 11.	Madi N	5-Microbiome studies
370	23007708	Sex Transm Dis. 2012 Oct;39(10):807-12. doi: 10.1097/OLQ.0b013e3182631c79.	Brotman RM	10-Non-clinical diagnostic studies
371	24570371	mBio. 2014 Feb 25;5(2):e01064-13. doi: 10.1128/mBio.01064-13.	Willner D	10-Non-clinical diagnostic studies
372	20976137	PLoS One. 2010 Oct 18;5(10):e13381. doi: 10.1371/journal.pone.0013381.	Greninger AL	10-Non-clinical diagnostic studies
373	24068019	J Clin Microbiol. 2013 Dec;51(12):4029-35. doi: 10.1128/JCM.02164-13. Epub 2013 Sep 25.	Rogers GB	5-Microbiome studies
374	31961856	PLoS Negl Trop Dis. 2020 Jan 21;14(1):e0007897. doi: 10.1371/journal.pntd.0007897. eCollection 2020 Jan.	Wise EL	10-Non-clinical diagnostic studies
375	27180802	Sci Rep. 2016 May 16;6:25497. doi: 10.1038/srep25497.	Cernada M	5-Microbiome studies
376	22133886	J Clin Invest. 2011 Dec;121(12):4610-7. doi: 10.1172/JCI57172. Epub 2011 Dec 1.	Brotman RM	5-Microbiome studies
377	25848871	ISME J. 2015 Sep;9(9):2078-93. doi: 10.1038/ismej.2015.47. Epub 2015 Apr 7.	Liu J	10-Non-clinical diagnostic studies
378	30590603	J Infect Dis. 2019 May 5;219(11):1804-1808. doi: 10.1093/infdis/jiy741.	Mansbach JM	5-Microbiome studies
379	28628941	Ann Neurol. 2017 Jul;82(1):105-114. doi: 10.1002/ana.24982.	Wilson MR	6-Case studies/series
380	26290222	Ann Neurol. 2015 Nov;78(5):722-30. doi: 10.1002/ana.24499. Epub 2015 Aug 24.	Wilson MR	6-Case studies/series
381	20433893	Biochimie. 2011 Jan;93(1):39-45. doi: 10.1016/j.biochi.2010.04.016. Epub 2010 Apr 28.	Laugerette F	10-Non-clinical diagnostic studies
382	26378926	PLoS One. 2015 Sep 17;10(9):e0137681. doi: 10.1371/journal.pone.0137681. eCollection 2015.	Rutten NB	5-Microbiome studies
383	22170749	Inflamm Bowel Dis. 2012 Oct;18(10):1799-808. doi: 10.1002/ibd.22860. Epub 2011 Dec 14.	Michail S	5-Microbiome studies
384	22278835	J Clin Microbiol. 2012 Apr;50(4):1376-83. doi: 10.1128/JCM.05852-11. Epub 2012 Jan 25.	Wolfe AJ	10-Non-clinical diagnostic studies
385	25403800	Am J Transplant. 2015 Jan;15(1):200-9. doi: 10.1111/ajt.13031. Epub 2014 Nov 17.	Young JC	10-Non-clinical diagnostic studies
386	30578963	Int J Antimicrob Agents. 2019 Apr;53(4):435-441. doi: 10.1016/j.ijantimicag.2018.12.006. Epub 2018 Dec 20.	Piewngam P	5-Microbiome studies
387	22702530	Future Microbiol. 2012 Jun;7(6):781-6. doi: 10.2217/fmb.12.40.	Fusté E	10-Non-clinical diagnostic studies
388	24475077	PLoS One. 2014 Jan 24;9(1):e86117. doi: 10.1371/journal.pone.0086117. eCollection 2014.	Wang JH	5-Microbiome studies
389	29420585	PLoS One. 2018 Feb 8;13(2):e0192277. doi: 10.1371/journal.pone.0192277. eCollection 2018.	Khan R	10-Non-clinical diagnostic studies
390	28222161	PLoS One. 2017 Feb 21;12(2):e0172605. doi: 10.1371/journal.pone.0172605. eCollection 2017.	Ijaz UZ	5-Microbiome studies
391	24690743	PLoS One. 2014 Apr 1;9(4):e93367. doi: 10.1371/journal.pone.0093367. eCollection 2014.	Fancello L	10-Non-clinical diagnostic studies
392	27122381	J Clin Microbiol. 2016 Jul;54(7):1804-1813. doi: 10.1128/JCM.02638-15. Epub 2016 Apr 27.	Andersen H	10-Non-clinical diagnostic studies
393	28041870	J Mol Diagn. 2017 Mar;19(2):303-312. doi: 10.1016/j.jmoldx.2016.10.005. Epub 2016 Dec 30.	Lewandowski K	10-Non-clinical diagnostic studies
394	30333011	BMC Infect Dis. 2018 Oct 17;18(1):521. doi: 10.1186/s12879-018-3419-8.	Petronella N	10-Non-clinical diagnostic studies
395	24189361	Am J Trop Med Hyg. 2014 Jan;90(1):124-32. doi: 10.4269/ajtmh.13-0383. Epub 2013 Nov 4.	Youmans BP	4-Non-metagenomic, targeted or non-WGS methods
396	27277524	BMC Genomics. 2016 Jun 8;17:440. doi: 10.1186/s12864-016-2777-0.	Pop M	5-Microbiome studies
397	28526549	Ophthalmology. 2017 Aug;124(8):1247-1248. doi: 10.1016/j.ophtha.2017.03.045. Epub 2017 May 16.	Doan T	1-Included
398	30715242	Clin Infect Dis. 2019 Aug 16;69(5):861-864. doi: 10.1093/cid/ciz072.	Wang J	10-Non-clinical diagnostic studies
399	22748132	Clin Microbiol Infect. 2012 Sep;18(9):841-8. doi: 10.1111/j.1469-0691.2012.03913.x. Epub 2012 Jul 2.	Khelaifia S	7-Reviews/Commentary
400	25918347	Blood. 2015 Jun 11;125(24):3789-97. doi: 10.1182/blood-2014-12-617035. Epub 2015 Apr 27.	Seo S	10-Non-clinical diagnostic studies
401	29020199	Clin Infect Dis. 2017 Oct 16;65(9):1477-1485. doi: 10.1093/cid/cix596.	Somasekar S	10-Non-clinical diagnostic studies
402	28512093	mBio. 2017 May 16;8(3):e00489-17. doi: 10.1128/mBio.00489-17.	Long SW	8-Epidemiological studies
403	28827811	PLoS One. 2017 Aug 21;12(8):e0182585. doi: 10.1371/journal.pone.0182585. eCollection 2017.	Moss EL	5-Microbiome studies
404	28030619	PLoS One. 2016 Dec 28;11(12):e0168860. doi: 10.1371/journal.pone.0168860. eCollection 2016.	Nguyen LD	5-Microbiome studies
405	29893821	Clin Infect Dis. 2019 Jan 1;68(1):139-142. doi: 10.1093/cid/ciy486.	Edridge AWD	10-Non-clinical diagnostic studies
406	25972420	J Clin Microbiol. 2015 Jul;53(7):2238-50. doi: 10.1128/JCM.02495-14. Epub 2015 May 13.	Fischer N	9-Low sample size or pilot study
407	22171685	BMC Microbiol. 2011 Dec 15;11:264. doi: 10.1186/1471-2180-11-264.	Zihler A	5-Microbiome studies
408	22792313	PLoS One. 2012;7(7):e40425. doi: 10.1371/journal.pone.0040425. Epub 2012 Jul 9.	Smith BC	5-Microbiome studies
409	24401860	ISME J. 2014 Jun;8(6):1247-58. doi: 10.1038/ismej.2013.229. Epub 2014 Jan 9.	Whiteson KL	10-Non-clinical diagnostic studies
410	24303043	PLoS One. 2013 Nov 26;8(11):e81330. doi: 10.1371/journal.pone.0081330. eCollection 2013.	Song Y	5-Microbiome studies
411	29261736	PLoS One. 2017 Dec 20;12(12):e0189768. doi: 10.1371/journal.pone.0189768. eCollection 2017.	Camacho-Ortiz A	5-Microbiome studies
412	27732961	Oncotarget. 2017 Feb 7;8(6):10609-10619. doi: 10.18632/oncotarget.12532.	Han J	5-Microbiome studies
413	21152087	PLoS One. 2010 Nov 30;5(11):e15101. doi: 10.1371/journal.pone.0015101.	Zemanick ET	4-Non-metagenomic, targeted or non-WGS methods
414	25392361	J Clin Microbiol. 2015 Jan;53(1):237-47. doi: 10.1128/JCM.02555-14. Epub 2014 Nov 12.	Stokell JR	5-Microbiome studies
415	25360611	PLoS One. 2014 Oct 31;9(10):e111245. doi: 10.1371/journal.pone.0111245. eCollection 2014.	Zarei S	10-Non-clinical diagnostic studies

Supplementary Table 6. Continued

416	23372716	PLoS One. 2013;8(1):e54379. doi: 10.1371/journal.pone.0054379. Epub 2013 Jan 23.	Drell T	5-Microbiome studies
417	25542470	J Clin Virol. 2015 Jan;62:48-53. doi: 10.1016/j.jcv.2014.11.014. Epub 2014 Nov 20.	Fisher RG	10-Non-clinical diagnostic studies
418	28884664	J Gen Virol. 2017 Oct;98(10):2425-2437. doi: 10.1099/jgv.0.000920. Epub 2017 Sep 8.	Langevin S	5-Microbiome studies
419	26076489	PLoS One. 2015 Jun 15;10(6):e0129055. doi: 10.1371/journal.pone.0129055. eCollection 2015.	Gall A	5-Microbiome studies
420	20844578	PLoS Pathog. 2010 Sep 9;6(9):e1001097. doi: 10.1371/journal.ppat.1001097.	Endt K	5-Microbiome studies
421	21881620	Obesity (Silver Spring). 2011 Nov;19(11):2113-20. doi: 10.1038/oby.2011.68. Epub 2011 Sep 1.	Lam YY	5-Microbiome studies
422	22409163	Eur J Clin Invest. 2012 Aug;42(8):848-54. doi: 10.1111/j.1365-2362.2012.02660.x. Epub 2012 Mar 13.	Flores R	5-Microbiome studies
423	26872326	J Clin Virol. 2016 Apr;77:5-8. doi: 10.1016/j.jcv.2016.01.013. Epub 2016 Jan 29.	Fridholm H	4-Non-metagenomic, targeted or non-WGS methods
424	22152152	BMC Oral Health. 2011 Dec 12;11:33. doi: 10.1186/1472-6831-11-33.	Huang S	5-Microbiome studies
425	23148445	Clin Microbiol Infect. 2012 Dec;18(12):1157-9. doi: 10.1111/1469-0691.12032.	Greub G	5-Microbiome studies
426	28128353	Sci Rep. 2017 Jan 27;7:41601. doi: 10.1038/srep41601.	Ren L	10-Non-clinical diagnostic studies
427	27669488	Transl Res. 2017 Mar;181:59-70. doi: 10.1016/j.trsl.2016.08.008. Epub 2016 Sep 9.	Rani A	5-Microbiome studies
428	23056520	PLoS One. 2012;7(10):e46901. doi: 10.1371/journal.pone.0046901. Epub 2012 Oct 8.	Keller MJ	10-Non-clinical diagnostic studies
429	26912755	J Clin Microbiol. 2016 Jul;54(7):1686-1693. doi: 10.1128/JCM.02664-15. Epub 2016 Feb 24.	Westblade LF	7-Reviews/Commentary
430	26729566	Sci Rep. 2016 Jan 5;6:18594. doi: 10.1038/srep18594.	Yang I	5-Microbiome studies
431	29020980	Microbiome. 2017 Oct 12;5(1):134. doi: 10.1186/s40168-017-0353-8.	Berglund F	10-Non-clinical diagnostic studies
432	22347512	PLoS Negl Trop Dis. 2012;6(2):e1485. doi: 10.1371/journal.pntd.0001485. Epub 2012 Feb 7.	Yozwiak NL	9-Low sample size or pilot study
433	24192040	Elife. 2013 Nov 5;2:e01608. doi: 10.7554/eLife.01608.	Mathis D	5-Microbiome studies
434	29043660	Virus Genes. 2018 Feb;54(1):33-40. doi: 10.1007/s11262-017-1514-3. Epub 2017 Oct 17.	Phan TG	4-Non-metagenomic, targeted or non-WGS methods
435	30054809	Pathol Oncol Res. 2019 Jul;25(3):1023-1033. doi: 10.1007/s12253-018-0457-x. Epub 2018 Jul 27.	Decsi G	5-Microbiome studies
436	25939040	Inflamm Bowel Dis. 2015 Jun;21(6):1419-27. doi: 10.1097/MIB.0000000000000344.	Wang W	5-Microbiome studies
437	26154788	N Engl J Med. 2015 Jul 9;373(2):154-62. doi: 10.1056/NEJMoa1415627.	Hoffmann B	10-Non-clinical diagnostic studies
438	30587202	Ann Clin Microbiol Antimicrob. 2018 Dec 26;17(1):45. doi: 10.1186/s12941-018-0298-1.	Hu Z	4-Non-metagenomic, targeted or non-WGS methods
439	21143294	Acta Paediatr. 2011 Jan;100(1):75-8. doi: 10.1111/j.1651-2227.2010.01988.x. Epub 2010 Sep 9.	Savino F	5-Microbiome studies
440	27448188	Emerg Infect Dis. 2016 Oct;22(10):1788-92. doi: 10.3201/eid2210.160663. Epub 2016 Oct 15.	Naccache SN	8-Epidemiological studies
441	27057545	Biomed Res Int. 2016;2016:6598307. doi: 10.1155/2016/6598307. Epub 2016 Feb 14.	Wang Y	10-Non-clinical diagnostic studies
442	20629946	Lett Appl Microbiol. 2010 Sep;51(3):272-7. doi: 10.1111/j.1472-765X.2010.02891.x. Epub 2010 Jun 20.	Nelson A	10-Non-clinical diagnostic studies
443	27499042	Sci Rep. 2016 Aug 8;6:30997. doi: 10.1038/srep30997.	Shiba T	10-Non-clinical diagnostic studies
444	23315732	Appl Environ Microbiol. 2013 Mar;79(6):1835-42. doi: 10.1128/AEM.02908-12. Epub 2013 Jan 11.	MacPhee RA	5-Microbiome studies
445	23488866	Lett Appl Microbiol. 2013 Jul;57(1):19-25. doi: 10.1111/lam.12068. Epub 2013 Mar 27.	Silva TM	5-Microbiome studies
446	27806082	PLoS One. 2016 Nov 2;11(11):e0165691. doi: 10.1371/journal.pone.0165691. eCollection 2016.	Miller RR	10-Non-clinical diagnostic studies
447	23715799	Reprod Sci. 2014 Jan;21(1):32-40. doi: 10.1177/1933719113488838. Epub 2013 May 28.	Hyman RW	5-Microbiome studies
448	26620704	Genome Med. 2015 Dec 1;7:113. doi: 10.1186/s13073-015-0235-2.	Greninger AL	10-Non-clinical diagnostic studies
449	25439613	Ophthalmology. 2015 Mar;122(3):524-30. doi: 10.1016/j.ophtha.2014.09.001. Epub 2014 Nov 24.	Lee AY	6-Case studies/series
450	31163013	Microb Drug Resist. 2019 Jun;25(5):677-680. doi: 10.1089/mdr.2018.0408.	Woodworth MH	4-Non-metagenomic, targeted or non-WGS methods
451	24679105	BMC Res Notes. 2014 Mar 29;7:196. doi: 10.1186/1756-0500-7-196.	Imirzalioglu C	4-Non-metagenomic, targeted or non-WGS methods
452	21987382	Inflamm Bowel Dis. 2012 May;18(5):935-42. doi: 10.1002/ibd.21874. Epub 2011 Oct 10.	Docktor MJ	5-Microbiome studies
453	24466089	PLoS One. 2014 Jan 22;9(1):e86428. doi: 10.1371/journal.pone.0086428. eCollection 2014.	Card RM	4-Non-metagenomic, targeted or non-WGS methods
454	23469240	PLoS One. 2013;8(3):e57782. doi: 10.1371/journal.pone.0057782. Epub 2013 Mar 4.	Benítez-Páez A	4-Non-metagenomic, targeted or non-WGS methods
455	22538067	J Cyst Fibros. 2012 Sep;11(5):454-5. doi: 10.1016/j.jcf.2012.03.007. Epub 2012 Apr 24.	Scanlan PD	5-Microbiome studies
456	26982117	Emerg Infect Dis. 2016 Apr;22(4):671-8. doi: 10.3201/eid2204.151812.	Bonsall D	10-Non-clinical diagnostic studies
457	23341915	PLoS One. 2013;8(1):e52876. doi: 10.1371/journal.pone.0052876. Epub 2013 Jan 14.	Mai V	5-Microbiome studies
458	23844258	PLoS One. 2013 Jul 2;8(7):e69316. doi: 10.1371/journal.pone.0069316. Print 2013.	Pushpanathan M	2-Animal and environmental studies
459	28475350	Am J Respir Crit Care Med. 2017 Dec 15;196(12):1610-1612. doi: 10.1164/rccm.201703-0537LE.	Pendleton KM	6-Case studies/series
460	25625766	Emerg Infect Dis. 2015 Feb;21(2):242-50. doi: 10.3201/eid2101.140795.	Lindsay B	5-Microbiome studies
461	22546742	Obesity (Silver Spring). 2012 Nov;20(11):2257-61. doi: 10.1038/oby.2012.110. Epub 2012 May 1.	Karlsson CL	5-Microbiome studies
462	22084916	Clin Microbiol Infect. 2012 Nov;18(11):1126-33. doi: 10.1111/j.1469-0691.2011.03690.x. Epub 2011 Nov 15.	Li SK	10-Non-clinical diagnostic studies
463	25101249	Front Cell Infect Microbiol. 2014 Jul 23;4:95. doi: 10.3389/fcimb.2014.00095. eCollection 2014.	Fernandez y Mostajo M	5-Microbiome studies
464	22355331	PLoS One. 2012;7(2):e30875. doi: 10.1371/journal.pone.0030875. Epub 2012 Feb 15.	Lysholm F	5-Microbiome studies
465	29211109	Mem Inst Oswaldo Cruz. 2018 Jan;113(1):66-67. doi: 10.1590/0074-02760170260.	Conteville LC	9-Low sample size or pilot study
466	26997279	Cell Rep. 2016 Mar 29;14(12):2912-24. doi: 10.1016/j.celrep.2016.03.015. Epub 2016 Mar 17.	Ward DV	10-Non-clinical diagnostic studies
467	22016793	PLoS One. 2011;6(10):e25959. doi: 10.1371/journal.pone.0025959. Epub 2011 Oct 10.	Mourani PM	10-Non-clinical diagnostic studies

Supplementary Table 6. Continued

468	22403423	J Clin Microbiol. 2012 May;50(5):1721-4. doi: 10.1128/JCM.00531-12. Epub 2012 Mar 7.	Rôças IN	5-Microbiome studies
469	27633952	Sci Rep. 2016 Sep 16;6:33327. doi: 10.1038/srep33327.	Rani A	5-Microbiome studies
470	29020227	Clin Infect Dis. 2018 Feb 10;66(5):789-792. doi: 10.1093/cid/cix792.	Piantadosi A	6-Case studies/series
471	21613452	J Nutr. 2011 Jul;141(7):1335-9. doi: 10.3945/jn.110.136747. Epub 2011 May 25.	Salvini F	5-Microbiome studies
472	26259788	Antimicrob Agents Chemother. 2015 Oct;59(10):6551-60. doi: 10.1128/AAC.00933-15. Epub 2015 Aug 10.	Bengtsson-Palme J	5-Microbiome studies
473	25866327	J Clin Virol. 2015 May;66:6-11. doi: 10.1016/j.jcv.2015.02.010. Epub 2015 Feb 18.	Zoll J	4-Non-metagenomic, targeted or non-WGS methods
474	22900080	PLoS One. 2012;7(8):e43009. doi: 10.1371/journal.pone.0043009. Epub 2012 Aug 10.	Paramel Jayaprakash T	5-Microbiome studies
475	28820378	Emerg Infect Dis. 2017 Sep;23(9):1604-1608. doi: 10.3201/eid2309.170168.	Boujon CL	2-Animal and environmental studies
476	27799386	Eur Respir J. 2016 Nov;48(5):1329-1339. doi: 10.1183/13993003.00152-2016. Epub 2016 Oct 6.	Hasegawa K	5-Microbiome studies
477	22363561	PLoS One. 2012;7(2):e31113. doi: 10.1371/journal.pone.0031113. Epub 2012 Feb 17.	Riboulet-Bisson E	5-Microbiome studies
478	30441786	Viruses. 2018 Nov 14;10(11):633. doi: 10.3390/v10110633.	Bal A	10-Non-clinical diagnostic studies
479	24043818	Proc Natl Acad Sci U S A. 2013 Oct 1;110(40):16151-6. doi: 10.1073/pnas.1308049110. Epub 2013 Sep 16.	Drexler JF	2-Animal and environmental studies
480	27105043	Euro Surveill. 2016 Apr 14;21(15). doi: 10.2807/1560-7917.ES.2016.21.15.30193.	Furtak V	2-Animal and environmental studies
481	21813714	J Clin Microbiol. 2011 Oct;49(10):3463-9. doi: 10.1128/JCM.00273-11. Epub 2011 Aug 3.	Yang J	9-Low sample size or pilot study
482	27413190	J Clin Microbiol. 2016 Sep;54(9):2348-53. doi: 10.1128/JCM.00877-16. Epub 2016 Jul 13.	Sardi SI	9-Low sample size or pilot study
483	27077101	EBioMedicine. 2016 Jan 29;5:14-5. doi: 10.1016/j.ebiom.2016.01.034. eCollection 2016 Mar.	Goedert JJ	5-Microbiome studies
484	26231254	Diagn Microbiol Infect Dis. 2015 Oct;83(2):133-8. doi: 10.1016/j.diagmicrobio.2015.06.017. Epub 2015 Jun 29.	Lewandowska DW	6-Case studies/series
485	26048203	J Infect. 2015 Sep;71(3):358-367. doi: 10.1016/j.jinf.2015.05.014. Epub 2015 Jun 3.	Horwitz D	5-Microbiome studies
486	22162545	J Clin Microbiol. 2012 Mar;50(3):867-75. doi: 10.1128/JCM.05176-11. Epub 2011 Dec 7.	Rea MC	5-Microbiome studies
487	23755967	Virology. 2013 Sep 1;443(2):187-96. doi: 10.1016/j.virol.2013.05.022. Epub 2013 Jun 10.	McCallin S	10-Non-clinical diagnostic studies
488	23734239	PLoS One. 2013 May 29;8(5):e65226. doi: 10.1371/journal.pone.0065226. Print 2013.	Salipante SJ	4-Non-metagenomic, targeted or non-WGS methods
489	28379472	Clin Infect Dis. 2017 Jul 15;65(2):332-335. doi: 10.1093/cid/cix296.	Thoendel M	10-Non-clinical diagnostic studies
490	21525866	Mucosal Immunol. 2011 Sep;4(5):554-63. doi: 10.1038/mi.2011.15. Epub 2011 Apr 27.	Gori A	5-Microbiome studies
491	23543647	Eur Respir J. 2013 Apr;41(4):991-3. doi: 10.1183/09031936.00052712.	Duff RM	4-Non-metagenomic, targeted or non-WGS methods
492	24857411	Emerg Infect Dis. 2014 Jun;20(6):1072-5. doi: 10.3201/eid2006.131526.	Fischer N	6-Case studies/series
493	24365140	J Allergy Clin Immunol. 2014 Apr;133(4):1220-2. doi: 10.1016/j.jaci.2013.10.049. Epub 2013 Dec 22.	Hyde ER	10-Non-clinical diagnostic studies
494	28671295	J Med Virol. 2017 Dec;89(12):2098-2107. doi: 10.1002/jmv.24895. Epub 2017 Jul 25.	Xu L	5-Microbiome studies
495	22291996	PLoS One. 2012;7(1):e30594. doi: 10.1371/journal.pone.0030594. Epub 2012 Jan 26.	Ferraris L	5-Microbiome studies
496	21858036	PLoS One. 2011;6(8):e23227. doi: 10.1371/journal.pone.0023227. Epub 2011 Aug 9.	Sannasiddappa TH	5-Microbiome studies
497	27648613	Emerg Infect Dis. 2016 Oct;22(10):1839-41. doi: 10.3201/eid2210.160052.	Halary S	10-Non-clinical diagnostic studies
498	21603636	PLoS One. 2011;6(5):e19709. doi: 10.1371/journal.pone.0019709. Epub 2011 May 13.	Dong Q	5-Microbiome studies
499	27647685	Am J Transplant. 2017 Mar;17(3):803-808. doi: 10.1111/ajt.14058. Epub 2016 Oct 21.	Wilson MR	6-Case studies/series
500	27577999	Sci Rep. 2016 Aug 31;6:32322. doi: 10.1038/srep32322.	Khan R	10-Non-clinical diagnostic studies
501	24829242	J Clin Microbiol. 2014 Jul;52(7):2583-94. doi: 10.1128/JCM.00556-14. Epub 2014 May 14.	Be NA	5-Microbiome studies
502	21962764	Virology. 2012 Jan 5;422(1):1-5. doi: 10.1016/j.virol.2011.08.024. Epub 2011 Oct 2.	McMullan LK	4-Non-metagenomic, targeted or non-WGS methods
503	32015040	Antimicrob Agents Chemother. 2020 Mar 24;64(4):e02310-19. doi: 10.1128/AAC.02310-19. Print 2020 Mar 24.	Logan LK	2-Animal and environmental studies
504	31476479	J Glob Antimicrob Resist. 2020 Mar;20:266-271. doi: 10.1016/j.jgar.2019.08.018. Epub 2019 Aug 30.	Nabti LZ	10-Non-clinical diagnostic studies
505	25412469	PLoS One. 2014 Nov 20;9(11):e113570. doi: 10.1371/journal.pone.0113570. eCollection 2014.	Taboada B	9-Low sample size or pilot study
506	24019552	J Exp Med. 2013 Sep 23;210(10):2011-24. doi: 10.1084/jem.20130728. Epub 2013 Sep 9.	Welty NE	10-Non-clinical diagnostic studies
507	31738994	J Allergy Clin Immunol. 2020 Feb;145(2):518-527.e8. doi: 10.1016/j.jaci.2019.10.034. Epub 2019 Nov 15.	Mansbach JM	5-Microbiome studies
508	26085150	J Virol. 2015 Sep;89(17):8806-15. doi: 10.1128/JVI.00653-15. Epub 2015 Jun 17.	Yu Y	10-Non-clinical diagnostic studies
509	28288568	BMC Genomics. 2017 Mar 14;18(1):231. doi: 10.1186/s12864-017-3612-y.	Shen-Gunther J	10-Non-clinical diagnostic studies
510	28910381	PLoS One. 2017 Sep 14;12(9):e0184720. doi: 10.1371/journal.pone.0184720. eCollection 2017.	Sugawara Y	10-Non-clinical diagnostic studies
511	26424618	J Virol Methods. 2015 Dec 15;226:1-6. doi: 10.1016/j.jviromet.2015.09.010. Epub 2015 Sep 28.	Perlejewski K	6-Case studies/series
512	27852235	BMC Microbiol. 2016 Nov 16;16(1):275. doi: 10.1186/s12866-016-0894-1.	Ottesen A	5-Microbiome studies
513	24157887	Intervirology. 2013;56(6):413-23. doi: 10.1159/000354563. Epub 2013 Oct 17.	Ngounga T	4-Non-metagenomic, targeted or non-WGS methods
514	29329757	Diagn Microbiol Infect Dis. 2018 Apr;90(4):241-247. doi: 10.1016/j.diagmicrobio.2017.11.023. Epub 2017 Dec 6.	Matz LM	10-Non-clinical diagnostic studies
515	28701177	Microbiome. 2017 Jul 12;5(1):75. doi: 10.1186/s40168-017-0295-1.	Stewart CJ	5-Microbiome studies
516	31657225	Am J Physiol Gastrointest Liver Physiol. 2020 Jan 1;318(1):G84-G98. doi: 10.1152/ajpgi.00118.2019. Epub 2019 Oct 28.	Albhaisi SAM	5-Microbiome studies
517	28826586	J Cyst Fibros. 2018 Mar;17(2):204-212. doi: 10.1016/j.jcf.2017.08.007. Epub 2017 Aug 18.	Lucas SK	5-Microbiome studies
518	32105620	Braz J Infect Dis. 2020 Mar-Apr;24(2):150-159. doi: 10.1016/j.bjid.2020.01.005. Epub 2020 Feb 24.	Samer S	10-Non-clinical diagnostic studies
519	28070022	J Breath Res. 2017 Feb 6;11(1):017101. doi: 10.1088/1752-7163/aa5833.	Phan J	5-Microbiome studies

Supplementary Table 6. Continued

520	24803806	World J Gastroenterol. 2014 May 7;20(17):4948-52. doi: 10.3748/wjg.v20.i17.4948.	Wang LL	5-Microbiome studies
521	24642836	PLoS One. 2014 Mar 18;9(3):e92297. doi: 10.1371/journal.pone.0092297. eCollection 2014.	Henrich B	10-Non-clinical diagnostic studies
522	31346170	Nat Commun. 2019 Jul 25;10(1):3313. doi: 10.1038/s41467-019-11306-6.	Sichtig H	10-Non-clinical diagnostic studies
523	29851032	Cochrane Database Syst Rev. 2018 May 31;5(5):CD011680. doi: 10.1002/14651858.CD011680.pub2.	Hua C	10-Non-clinical diagnostic studies
524	28483840	Cancer Prev Res (Phila). 2017 Jul;10(7):398-409. doi: 10.1158/1940-6207.CAPR-16-0178. Epub 2017 May 8.	Ye X	7-Reviews/Commentary
525	26928707	Epidemiology. 2016 Jul;27(4):602-11. doi: 10.1097/EDE.0000000000000471.	Ogino S	8-Epidemiological studies
526	31990924	PLoS One. 2020 Jan 28;15(1):e0228085. doi: 10.1371/journal.pone.0228085. eCollection 2020.	Ericsson AC	5-Microbiome studies
527	31370775	Technol Cancer Res Treat. 2019 Jan 1;18:1533033819867354. doi: 10.1177/1533033819867354.	Chattopadhyay I	5-Microbiome studies
528	21989211	BMC Bioinformatics. 2011;12 Suppl 6(Suppl 6):S1. doi: 10.1186/1471-2105-12-S6-S1. Epub 2011 Jul 28.	Astrovskaya I	4-Non-metagenomic, targeted or non-WGS methods
529	26344758	Methods. 2015 Nov 15;90:28-38. doi: 10.1016/j.ymeth.2015.09.004. Epub 2015 Sep 3.	Perry JL	10-Non-clinical diagnostic studies
530	30530602	Antimicrob Agents Chemother. 2019 Feb 26;63(3):e01924-18. doi: 10.1128/AAC.01924-18. Print 2019 Mar.	Sugawara Y	2-Animal and environmental studies
531	26288145	PLoS One. 2015 Aug 19;10(8):e0135573. doi: 10.1371/journal.pone.0135573. eCollection 2015.	Linsuwanon P	5-Microbiome studies
532	31371186	Rev Neurol (Paris). 2019 Sep-Oct;175(7-8):431-435. doi: 10.1016/j.neurol.2019.07.006. Epub 2019 Jul 29.	Seilhean D	7-Reviews/Commentary
533	32541790	Sci Rep. 2020 Jun 15;10(1):9674. doi: 10.1038/s41598-020-66738-8.	Diakite A	5-Microbiome studies
534	28766918	Photodermatol Photoimmunol Photomed. 2018 Jan;34(1):25-34. doi: 10.1111/phpp.12334. Epub 2017 Aug 30.	Sherwani MA	5-Microbiome studies
535	30374198	Nat Med. 2018 Dec;24(12):1822-1829. doi: 10.1038/s41591-018-0216-2. Epub 2018 Oct 29.	Baumann-Dudenhoeffer AM	5-Microbiome studies
536	30463925	mSphere. 2018 Nov 21;3(6):e00588-18. doi: 10.1128/mSphereDirect.00588-18.	Bulow C	5-Microbiome studies
537	27984120	J Biotechnol. 2017 May 20;250:45-50. doi: 10.1016/j.jbiotec.2016.10.020. Epub 2016 Oct 27.	Huson DH	5-Microbiome studies
538	28126350	Gastroenterology. 2017 May;152(6):1419-1433.e5. doi: 10.1053/j.gastro.2017.01.009. Epub 2017 Jan 23.	Tsoi H	10-Non-clinical diagnostic studies
539	31004827	Cell Mol Gastroenterol Hepatol. 2019;8(2):301-318.e2. doi: 10.1016/j.jcmgh.2019.04.008. Epub 2019 Apr 18.	Sung CM	5-Microbiome studies
Additional records identified through other sources				
1	32158774	Open Forum Infect Dis. 2020 Feb 8;7(3):ofaa046. doi: 10.1093/ofid/ofaa046. eCollection 2020 Mar.	Hong NTT	1-Included
2	29491477	Sci Rep. 2018 Feb 28;8(1):3784. doi: 10.1038/s41598-018-22133-y.	Horiba K	1-Included
3	31375834	Open Forum Infect Dis. 2019 Aug 1;6(8):ofz327. doi: 10.1093/ofid/ofz327.	Rossoff J	1-Included
4	29020199	Clin Infect Dis. 2017 Oct 16;65(9):1477-1485. doi: 10.1093/cid/cix596.	Somasekar S	1-Included
5	30819624	Diagn Microbiol Infect Dis. 2019 Jun;94(2):188-191. doi: 10.1016/j.diagmicrobio.2018.12.016. Epub 2019 Feb 2.	Farnaes L	1-Included
6	26818672	J Clin Microbiol. 2016 Apr;54(4):1000-7. doi: 10.1128/JCM.03060-15. Epub 2016 Jan 27.	Graf EH	1-Included
7	28368491	J Infect Dis. 2017 May 1;215(9):1407-1415. doi: 10.1093/infdis/jix148.	Schlaberg R	1-Included
8	31940319	PLoS One. 2020 Jan 15;15(1):e0226895. doi: 10.1371/journal.pone.0226895. eCollection 2020.	Fulton BD	4-Non-metagenomic, targeted or non-WGS methods
9	31107955	J Pediatric Infect Dis Soc. 2020 Jul 13;9(3):326-333. doi: 10.1093/jpids/piz032.	Haston JC	9-Low sample size or pilot study
10	29426368	BMC Res Notes. 2018 Feb 9;11(1):120. doi: 10.1186/s13104-018-3234-8.	Sharp C	9-Low sample size or pilot study
11	31220115	PLoS One. 2019 Jun 20;14(6):e0218318. doi: 10.1371/journal.pone.0218318. eCollection 2019.	Ramesh A	9-Low sample size or pilot study
12	23613733	PLoS One. 2013 Apr 17;8(4):e60595. doi: 10.1371/journal.pone.0060595. Print 2013.	Law J	6-Case studies/series

"I have not failed. I've just found 10,000 ways that won't work."

- Thomas Edison

4 OPTIMIZATION OF DNA EXTRACTION AND SEQUENCING

In chapter 3, I performed a meta-analysis and systematic review on culture-independent metagenomic sequencing for pathogen-agnostic infectious disease diagnostics. This review has helped to shape my own study design by identifying strengths and weaknesses in other studies that exist in the field. I found that there were no studies that processed clinical blood culture samples using a species-agnostic approach to predict species and resistance, which is necessary to truly estimate performance for clinical use. Furthermore, I identified only 4 studies which had more than 200 clinical samples, none of them being on blood cultures, demonstrating the need for larger and better-quality studies. My review revealed that quality controls used in clinical metagenomic studies were generally poor, which lead me to ensure that I use negative, positive and internal controls in my study design. As contamination was a major problem in numerous studies, I ensured that meticulous laboratory procedures were followed from the very

outset. Although most studies were conducted with Illumina sequencing, it was evident in my review from newer studies that using the Oxford Nanopore was feasible and was able to provide faster results due to real-time sequencing technology. In this next chapter, I use various DNA extraction kits and human DNA depletion techniques, some of which I identified using this review, to optimise DNA extraction methods to provide DNA for downstream applications in a timely manner geared for clinical use.

4.1 Introduction

High quality and high quantity extraction of genomic DNA from clinical samples carrying diverse microbial species and often various amounts of human DNA is the key challenge for downstream applications including next generation or real-time DNA sequencing. A major bottleneck in direct-from-sample clinical metagenomics is finding the optimal nuclei acid extraction methods from the existing variety of available extraction kits and techniques. A variety of microbes exists among samples with different types of cell walls and cell membranes enclosing their cytoplasm and genomic content. Harsh sample treatment could affect the quality of DNA extracted, whereas too gentle processing may only cause partial lysis (particularly for bacteria with thick layers of peptidoglycan) limiting the extraction process. Therefore, it is important to optimise the extraction method to obtain DNA from both abundant and rare groups of organisms that might contain different thicknesses and profiles of cell membranes and cell walls. Several commercial kits are available to extract DNA from samples which mostly use a silica-based column where DNA is selectively absorbed to a stationary solid phase at high pH in a high salt environment, and then subsequently eluted. A major barrier is the problem of insufficient DNA recovery from marginal amounts of clinical or environmental samples. Additionally, DNA extraction kits all have different biases,

which may produce dramatically different results even for the same sample.[120] Multiple laboratories have reported different methods for DNA extraction based on the type of sample used for analysis.[121] Therefore, currently there is no consensus on extraction methods for clinical and environmental samples in the field.

Another significant problem is that sequencing data from human derived clinical samples, including metagenomic samples, can often be contaminated from human cells of the research subject, or even the investigator.[122]As the human genome is nearly 1000x larger than most bacterial genomes, sequencing a tissue biopsy of equal human and bacterial cells would still produce a sample containing 99.9% human DNA.[123] Therefore, human contaminant DNA needs to be removed prior to downstream processing to ensure adequate coverage of pathogen DNA and also to reduce false-positive reads associated with misclassification of human reads.[124] Various human DNA depletion methods have been described in literature, however, few studies compare performance across these methods especially from clinical blood culture samples.[64] In one such study differential centrifugation was used to remove intact human cells, and a distilled water wash to lyse remaining human cells. Other reported methods were filtration, selective lysis using other compounds and solid-phase reversible immobilization bead clean-up. Another study used saponin treatment, a human detergent previously demonstrated to differentially lyse human cells by disrupting the cell membrane, which demonstrated a significant reduction in the proportion of human DNA present, improving reference genome coverage at a 10-fold depth.[57] In this chapter, I test common DNA extraction methods and human DNA depletion methods against each other with a view to develop an optimised DNA extraction pipeline directly from blood culture.

4.2 Research question 1

What is the optimal laboratory extraction method to purify relatively small amounts of pathogen DNA from samples with high levels of human DNA?

4.2.1 Methods

4.2.1.1 Sample collection and processing

Ten random positive liquid blood culture specimens were retrieved from ten unique patients at the Oxford University Hospitals National Health Service (UK) (NHS) Foundation Trust microbiology laboratory. Samples that was due to be discarded were taken immediately after routine tests were completed. Therefore, individual patient consent was not required. This study was done with Research Ethics Committee approval (REC reference 17/LO/1420). Although aerobic and anaerobic cultures are drawn together, only one was processed as this was the bottle that flagged positive during the sample collection period. Collected blood culture bottles included five BD BACTECTM Plus Aerobic, and five Lytic Anaerobic (BD, USA). The former contains resin, and the latter lytic agents for the release of intracellular pathogens. Positive samples were retrieved from the BD BACTECTM FX blood culture system by 10am each day from the previous 6 hours. The system stored liquid blood culture bottles at 35 °C until positivity or negativity. 12ml of liquid blood culture was removed and divided into 12 aliquots of 1ml each and stored at -18 °C prior to processing.

4.2.1.2 Human depletion and DNA extraction

A no depletion method and three human DNA depletion methods were tested: 1. Differential centrifugation with selective lysis, which uses differential centrifugation to

remove intact human cells and a distilled water wash to lyse remaining human cells[3]; 2. Saponin treatment, which is a mild detergents that is used to permeabilize mammalian cells followed by DNAase treatment[77]; and 3. The MolYsis™ Basic Pathogen Isolation kit (Molzym, Germany) that uses a proprietary depletion method. These different human DNA depletion methods were tested this in combination with three DNA extraction and purification kits: MolYsis™ Complete5 (Molzym, Germany), BiOstic Bacteraemia kit (MoBio, Qiagen, USA) and QIAamp UCP Pathogen Mini (MoBio, Qiagen, USA) which all use propriety membrane-based extraction technology. Four negative controls were also tested (2 water and 2 negative culture media controls) in each combination of kit and depletion method used. All commercial kits were used according to the manufacturers' protocol and were all available in the UK. The DNA produced from each kit was purified using AMPure XP solid-phase reversible immobilization (SPRI) beads following the manufacturer's protocol (Beckman Coulter, UK) and eluted in 25μl of 1x TE buffer made from 10mM Tris-HCl and 1mM EDTA. Total DNA extracted was quantified using the Qubit 2.0 fluorometer (broad range kit; Thermo Fisher Scientific, USA).

4.2.1.3 qPCR

Quantitative PCR using specific primers and probes (described in **Table 4**) were used to quantify 16S rRNA gene DNA, that is, total bacterial DNA, and human β-actin DNA. All samples were assayed in triplicate and the mean cycle threshold value used. Standards were made from a commercial stock solution of a known reference *E. coli K12* strain and human DNA from Promega Corporation (Madison, WI, USA). 16S rRNA gene DNA is conserved across bacterial species and so identified all bacterial DNA present, regardless of whether it arose from monomicrobial or polymicrobial growth or from bacterial contaminant DNA. Primers were used at 0.8 μM and the probes

at 0.4 μM for both human β-actin and 16S rRNA genes. All reactions used 10μl of the Luna Universal Probe qPCR Master Mix (New England Biolabs, Ipswich, Massachusetts, USA) with 2 μl DNA and 6μl molecular grade water to bring the reaction volume to 20 μl. Amplification was performed using the MxPro 3005P (Agilent, USA) under the following conditions: 95 C for 1 min, 40 cycles of 95 C for 15s and 60 C for 30s.

Table 4. Quantitative PCR primers and probes.

Organism	Target gene	Probe or primer	Sequence (5'→3')	Ref.
Bacteria	16S rDNA	Forward	TCCTACGGGAGGCAGCAGT	125
		Reverse	GGACTACCAGGGTATCTAATCCTGTT	
		Probe	6FAM-CGTATTACCGCGGCTGCTGGCAC-BHQ1	
Human	β-actin	Forward	CCACACTGTGCCCATCTACG	126
		Reverse	AGGATCTTCATGAGGTAGTCAGTCAG	
		Probe	6FAM-ATGCCCTCCCCCATGCCATCCTGCGT-BHQ1	

4.2.1.4 Statistical analysis

Continuous variables were summarized as mean and standard deviation (SD) or median and interquartile range (IQR) depending on the distribution. The Kruskal-Wallis test was

used for comparing two or more independent samples for non-normal data. Pairwise comparisons using the Wilcoxon signed rank test were also performed. All analyses were performed in RStudio (version 1.1.4).

4.2.2 Results

4.2.2.1 Total DNA yield

Ten unique positive blood cultures were tested. Twelve aliquots of 1ml of liquid blood culture were extracted from each clinical sample. Four different human DNA depletion approaches were tested for each kit in combination with 3 extraction methods, that is, 40 samples per extraction kit (total n=120). Absolute DNA yields for each extraction kit, across all human DNA depletion methods, are shown below in **Figure 12**/**Table 5**.

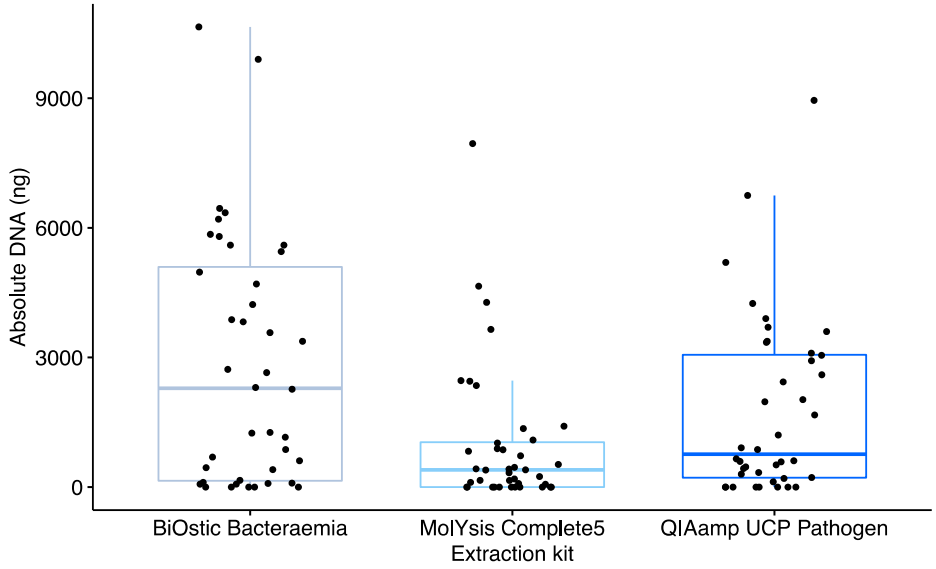

Figure 12. Total DNA by extraction kit (ng). All kits differed in absolute DNA extraction (all $p<0.001$).

Table 5. Total DNA yield by extraction kit (n=120).

Measurement	Sample numbers	Extraction kit	Median	IQR	Range
Qubit (ng)	40 positive samples per kit	BiOstic Bacteraemia kit	2285	145-5093	0-10650
		MolYsis Complete5	398	0-1038	0-7950
		QIAamp UCP Pathogen Mini	763	217-3063	0-8950

Across all samples, the BiOstic Bacteraemia kit (MoBio, Qiagen, USA) extracted 3-6x more DNA than other kits tested, yielding within 25µl a median of 2285ng (IQR 145-5093) [Range 0- 10650] compared to the QIAamp UCP Pathogen Mini kit (MoBio, Qiagen, USA) 763ng (217-3063) [0-8950] ($p<0.001$) and the MolYsis™ Complete5 (Molzym, Germany) 398ng (0-1038) [0-7950] ($p<0.001$). The QIAamp UCP Pathogen Mini (MoBio, Qiagen, USA) also produced higher yields than the MolYsis™ Complete5 (Molzym, Germany) kit ($p<0.001$).

BiOstic extracts contained detectable DNA in 35/40 (88%) samples, MolYsis™ Complete5 in 32/40 (80%) and QIAamp UCP Pathogen Mini in 29/40 (73%). 17/24 (71%) samples that extracted no DNA used the differential centrifugation with the selective lysis method, which may have resulted in excess loss of DNA during centrifugation due to only the supernatant being aliquoted out. Furthermore, other

samples along the DNA extraction process may have failed due to insufficient lysis, insufficient homogenisation of samples or loss of nucleic acids during purification. The 24 positive samples with no extracted DNA were excluded from subsequent analyses with qPCR.

4.2.2.2 Number of bacterial and human cells after depletion of human cell DNA

The relative extraction efficiency of bacterial and human DNA was assessed using qPCR by the 16S rRNA and human β-actin genes respectively. Three human DNA depletion methods and a no depletion method were tested using all three types of DNA extraction kits using qPCR, that is, 12 different methods from each of 10 positive blood cultures. Human and bacterial copy numbers per 2µl of eluted DNA grouped by depletion method and extraction kit are shown in **Figure 13**.

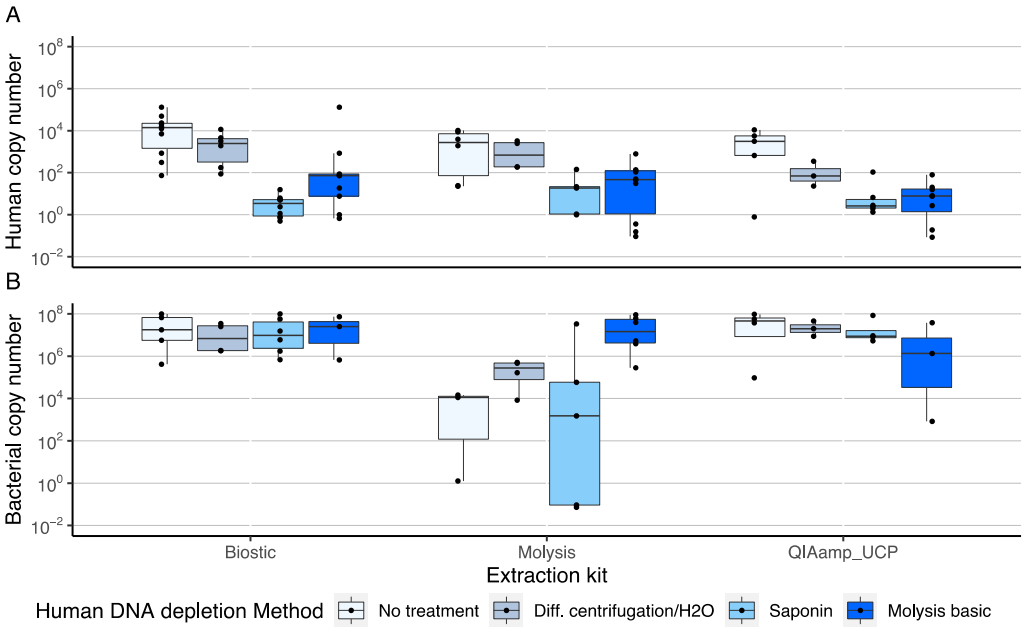

Figure 13. Human and bacterial copy number per 2µl grouped by extraction kit. 120 positive liquid blood culture samples and 48 negative controls were tested. 24 positive samples extracted no DNA and were excluded.

The BiOstic kit extracted the similar amount of bacterial DNA regardless of human depletion method (range of medians: $3x10^7$-$2x10^8$ bacterial copies per 2µl), while the MolYsis kit varied the most in bacterial DNA extraction by depletion method (range of medians: $1x10^4$-$4x10^7$ bacterial copies per 2µl) (**Figure 13.B**). The saponin depletion method combined with the BiOstic kit resulted in the least human DNA and the most bacterial DNA, with copy number per 2µl medians of 3.73 and $7.8x10^7$ respectively, followed by the saponin depletion method and the QIAamp kit of 2.64 and $4.7x10^7$ respectively (**Figure 13**).

Figure 14 displays human and bacterial copy numbers per 2µl grouped by depletion method or extraction kit. The saponin depletion method resulted in the lowest human copies per 2µl of 2.8 (1-7) [0.5- $1x10^2$] compared to the Molysis basic 20 (2-80) [9x10-2-$1x10^5$] (p=0.013), the differential centrifugation and lysis method $4x10^2$ ($2x10^2$-$3x10^3$) [$2x10^1$-$1x10^4$] (p<0.001) and the no depletion method $6x10^3$ ($6x10^2$-$1x10^4$) [$8x10^{-1}$-$1.3x10^5$] (p<0.001) (**Figure 14. A**). Despite variable performance by depletion method in reducing the amount of human DNA present (**Figure 14. A**), there was no variation seen in bacterial copy number per 2µl across the different human DNA depletion methods (p=0.38) (**Figure 14. C**).

The BiOstic kit and QIAamp kit performed comparably at bacterial DNA extraction $9.9x10^7$ ($1.7x10^7$-$2.2x10^8$) [$4.2x10^5$-$2x10^9$] and $3.8x10^7$ ($8.3x10^6$-$8.4x10^7$) [$8.3x10^2$-$2.3x10^1$] respectively (p=0.34) however, were both superior to the Molysis kit $3.7x10^5$ ($1.1x10^4$- $3.5x10^7$) [$7.2x10^{-2}$-$1.6x10^9$] (p<0.01 and p=0.03 respectively) (**Figure 14. D**).

There was no advantage seen among a specific extraction kit for reduction of human DNA (p=0.16) (**Figure 14. B**).

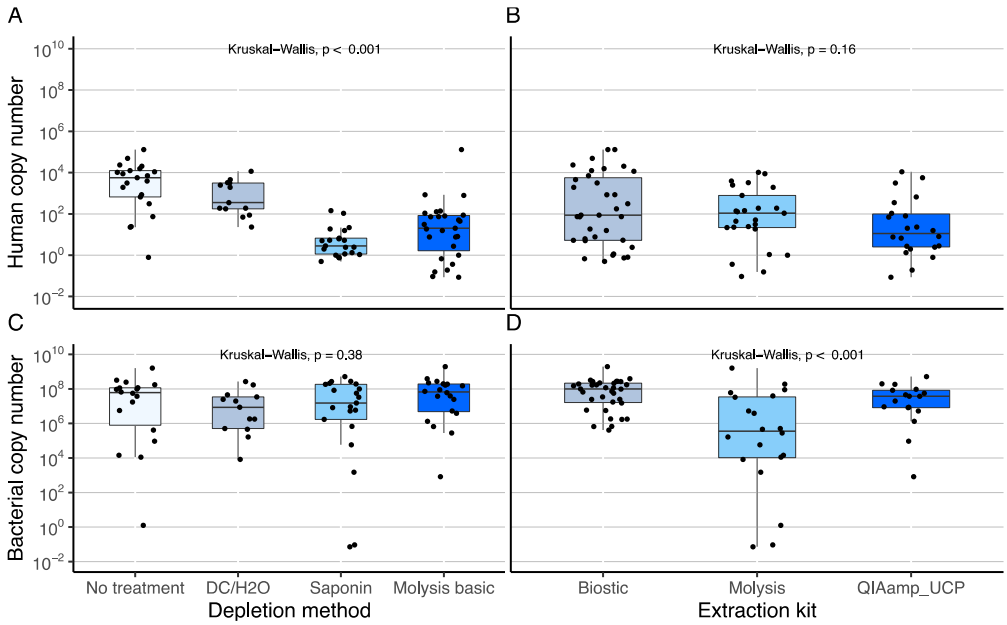

Figure 14. Human and bacterial copy number per 2µl grouped by depletion method and grouped by extraction kit.

4.2.2.3 Ratio of bacterial to human copy numbers

The ratio of bacterial copy numbers per 2µl to human copy numbers per 2µl grouped by extraction kit and grouped by depletion method is shown below in **Figure 15**. The QIAamp UCP kit $3.8x10^5$ ($9.3x10^4$-$2.7x10^6$) [$8.6x10^0$-$2.3x10^9$] extracted a higher bacterial to human ratio than the BiOstic kit $3.7x10^5$ ($1.5x10^4$-$1.1x10^7$) [$2x10^1$-$5.5x10^8$] ($p=0.03$) and the Molysis kit $2.7x10^3$ ($2.5x10^0$-$4.1x10^5$) [$3.2x10^{-4}$-$4.2x10^8$] ($p<0.01$) (**Figure 15. A**). There was no clear advantage in the bacterial/human ratio shown by using the saponin depletion method $2.9x10^6$ ($5.9x10^5$-$3.1x10^7$) [$5x10^{-4}$-$5.5x10^8$] compared to the Molysis basic $1.6x10^6$ ($8.6x10^4$-$7.5x10^6$) [$4.1x10^1$-$2.3x10^9$] ($p=0.08$), however both methods were superior to the no treatment method $7.5x10^3$ ($2x10^1$-$5.7x10^4$) [$3.2x10^{-4}$-$5.7x10^4$] ($p=0.021$ and $p=0.021$ respectively), while the saponin method was superior to the differential centrifugation and selective lysis method $1.5x10^4$ ($9.3x10^2$-$1.4x10^5$) [$3.3x10^0$-$4x10^5$] ($p=0.027$) (**Figure 15. B**).

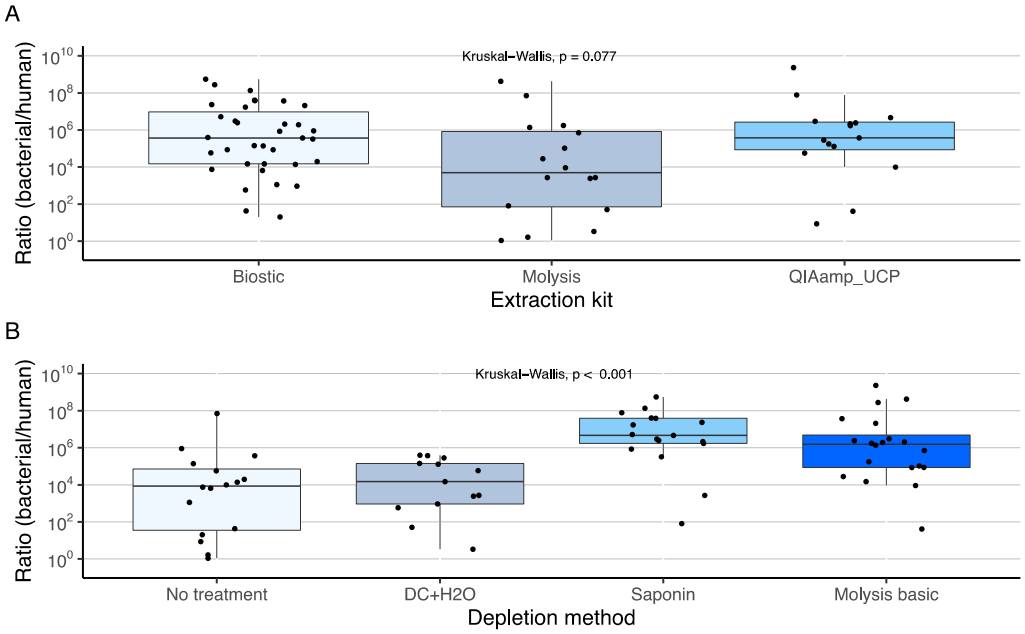

Figure 15. Ratio of bacterial copy numbers per 2µl to human copy numbers per 2µl by extraction kit and depletion method. A. Kruskal-Wallis test (overall p=0.077) B. Kruskal-Wallis test (overall p<0.001).

4.2.2.4 qPCR controls and variability

Figure 16 displays human and bacterial copy number per 2µl in controls grouped by extraction kit and further by depletion method. Controls from the BiOstic kit had the narrowest range of distribution of human copy number per 2µl of eluted DNA [Range $8.5x10^{-2}$-$5x10^{-1}$], while controls from the MolYsis kit had the highest [Range $8.5x10^{-2}$-$5.7x10^{3}$].

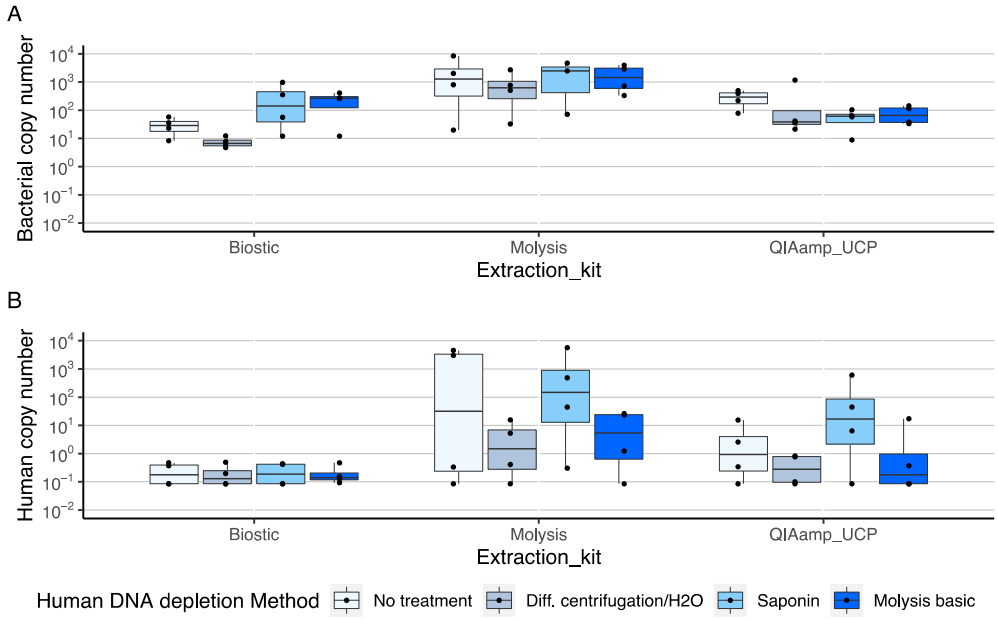

Figure 16. Displays bacterial and human copy number per 2µl in controls grouped by extraction kit and depletion method. Two water controls and two negative culture medium controls were introduced for every combination of kit and depletion method tested (n=48).

Figure 17 shows performance grouped by depletion method and grouped by extraction method. Negative controls extracted by the Molysis kit resulted in higher bacterial copy numbers per 2µl $1.4x10^3$ ($4.6x10^2$-$3.1x10^3$) [$2x10^1$-$2.6x10^4$] compared to the BiOstic kit $2.9x10^1$ ($1.1x10^1$- $2.6x10^2$) [$4.7x10^0$-$9.7x10^2$] (p=0.001) and to the QIAamp kit $7.1x10^1$ ($3.7x10^1$-$1.6x10^2$) [$8.8x10^0$-$1.2x10^3$] (p<0.001) (**Figure 17. A**). There was no clear difference seen in bacterial or human copy number per 2µl across depletion methods (p=0.371 and p=0.391 respectively) (**Figure 17. B** and **Figure 17. D**).

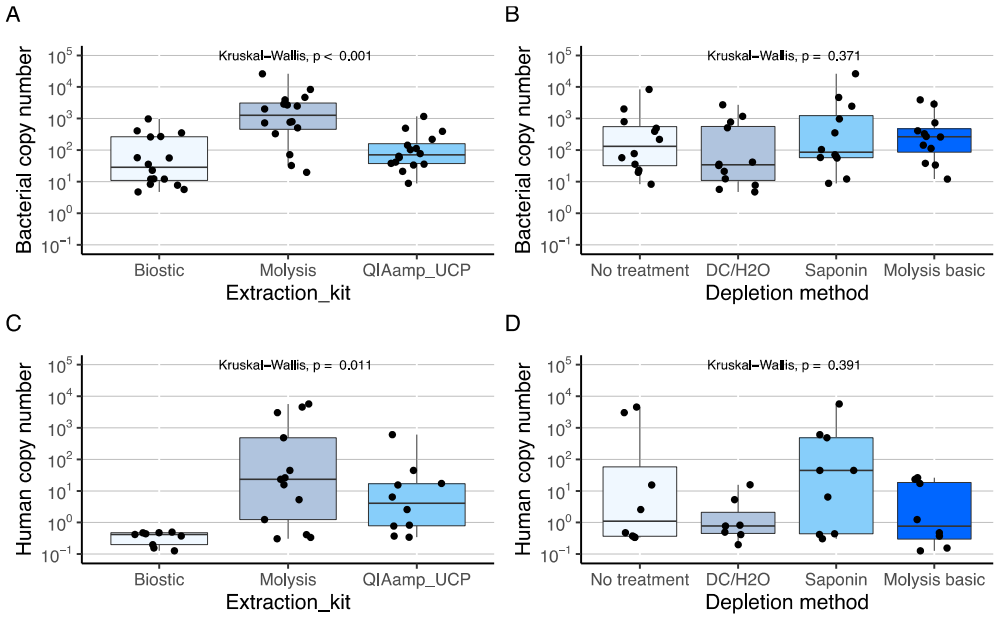

Figure 17. Displays bacterial and human copy number per 2µl in controls grouped by extraction kit and grouped by depletion method.

All controls extracted no quantifiable DNA by qubit. **Figure 18** displays human and bacterial copy number for negative controls and positive blood culture samples.

Water and negative culture-medium controls contained $6.3x10^{-1}$ ($8.5x10^{-2}$-$1.8x10^{1}$) [$8.5x10^{-2}$- $5.7x10^{3}$] human copies per 2µl and $3.2x10^{-1}$ ($8.5x10^{-2}$-$6.7x10^{-1}$) [$8.5x10^{-2}$-$4.6x10^{3}$] human copies per 2µl respectively, while combined data from all positive culture samples contained $3.1x10^{1}$ ($2.4x10^{0}$- $7.3x10^{2}$) [$8.5x10^{-2}$-$1.3x10^{5}$] ($p<0.001$) human copies per 2µl. The first to third quartile of water controls overlapped with the first quartile to the median of positive controls, while the maximum of the culture control overlapped with the minimum value to the first quartile of the positive cultures (**Figure 18. A**).

For bacterial copies per 2µl, water and negative culture-medium controls contained $1.9x10^2$ ($3.1x10^1$- $7.4x10^2$) [$4.7x10^0$-$4.7x10^3$] and $9.6x10^1$ ($3.5x10^1$-$5.5x10^2$) [$7.7x10^0$-$2.6x10^4$] respectively, while positive cultures contained $5.9x10^6$ ($1.8x10^5$-$9.7x10^7$) [$7.2x10^{-2}$-$2x10^9$] ($p<0.001$) bacterial copies per 2µl. The maximum values of both controls lied within the range of the minimum value to the first quartile of the positive cultures (**Figure 18. B**).

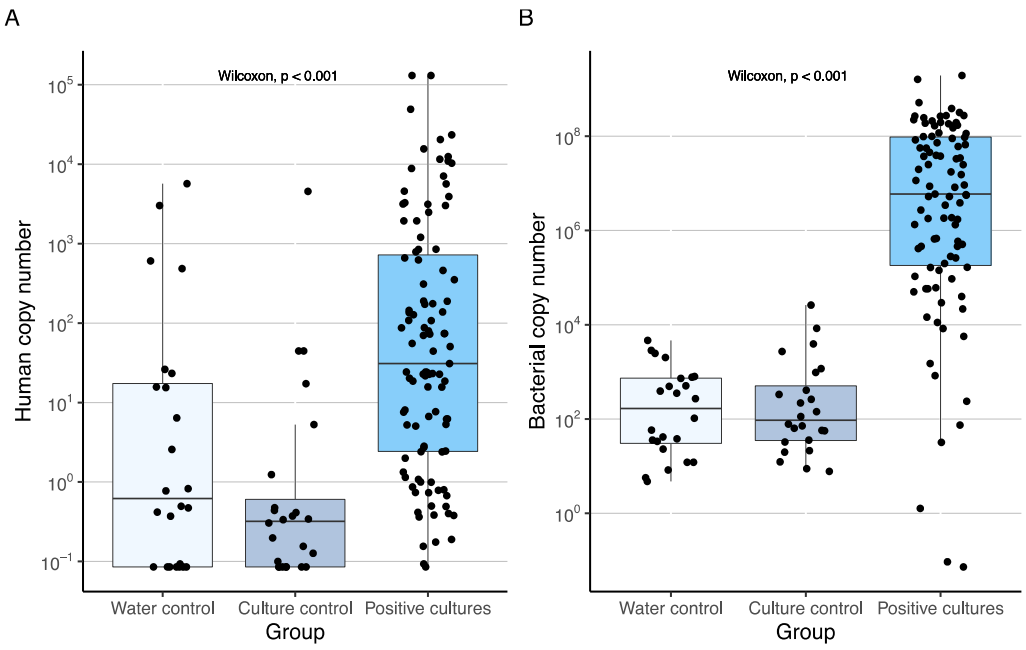

Figure 18. Displays Human and Bacterial copies per 2µl by water controls, negative culture-medium control and positive culture. For A and B, Wilcoxon test of negative controls and positive cultures ($p<0.001$).

4.2.3 Discussion

The BiOstic Bacteraemia kit extracted 3 and 6 times more total DNA compared to the QIAamp UCP Pathogen Mini kit and MolYsis™ Complete5 kits respectively. Higher DNA yields may be an important factor to maximise the sensitivity of bacterial detection, especially in samples with low amounts of bacterial DNA. However, the caveat is that higher DNA concentrations pose a risk of increasing contamination among samples and hence the optimal amount of DNA concentration to minimise contamination should be taken into consideration.

Although all human DNA depletion methods did not affect extracted bacterial copy number significantly, the saponin method resulted in a 7-142x fold depletion in human copy number compared to other methods. The QIAamp kit performed superiorly to the BiOstic kit and Molysis kit at bacterial DNA extraction. The BiOstic kit extracted almost a 137x higher bacterial/human ratio than the Molysis kit however, resulted in a similar extraction ratio compared to the QIAamp kit. Although a high ratio may not be beneficial when human DNA is at a low proportion, it may be when bacterial DNA is at a low proportion.

Negative water controls and negative culture-media controls did not differ significantly with regards to human and bacterial copy number. This indicates the absence of bacterial DNA in negative culture bottles which is promising for future metagenomic studies that may sample directly from blood culture bottles. Although meticulous attention was paid to laboratory technique and decontamination procedures including utilising a UV cabinet, frequent use of DNAse to clean pipettes and surfaces before and after use in addition to using new pipette tips for aliquoting each reagent, contamination was still a

challenge due to the very intricate nature of metagenomic handling. Furthermore, most positive extracts contained approximately $6x10^6$ bacterial copies per 2µl which means that even minor contamination during qPCR of 0.01 µl would introduce an additional $3x10^4$ bacterial copies into the contaminated sample, almost already over any maximum value found in any of the negative controls.

Although there was no evidence of human DNA depletion methods contaminating samples, it was clear that the Molysis extraction kit resulted in greater contamination across human and bacterial controls, likely due to the kit's complex and intricate protocol. The Molysis extraction kit varied the most in human and bacterial copy number per 2µl regarding negative controls [Human range: $8.5x10^{-2}$-$5.7x10^3$; bacterial range: $2x10^1$-$2.6x10^4$] which may be explained by the kit's likelihood to introduce contamination during extraction and/or a sub-optimal laboratory extraction technique. However, the range of copy numbers per 2µl found in the BiOstic negative controls were much narrower [Human range: $8.5x10^{-2}$-$5x10^{-1}$; bacterial range: $5x10^0$-$9.7x10^2$] indicating less of a contamination issue regarding this kit.

From this experiment, the BiOstic Bacteraemia kit due to its superior DNA extraction capability with relatively low contamination seen is recommended as the extraction kit of choice. The saponin method is recommended as the superior human DNA depletion method but may not be needed if working with high bacterial concentrations in positive blood cultures.

4.3 Research question 2

A pilot study: Does saponin pre-treatment improve genotypic speciation; and how can true positives verses contamination be assigned?

4.3.1 Methodology

4.3.1.1 Sample collection and processing

As above, positive samples were retrieved from the BD BACTEC™ FX blood culture system of Oxford University Hospitals Microbiology Laboratory by 10am each day from the previous 6 hours. 12 random positive liquid blood culture specimens were retrieved. 2-4ml of liquid blood culture was removed and divided into aliquots of 1ml each and stored at -18 °C prior to processing. 14 positive blood culture (from 12 unique patients) and one water control was processed using the BiOstic Bacteraemia kit with and without Saponin-based pre-treatment (n=30). Samples were eluted in 15μl. 22 samples were multiplexed with two samples in each run while 8 samples were multiplexed with 4 samples in each run. Two samples that were used when multiplexed with 2 runs were also used when multiplexed with four runs to compare the difference in sequencing depth when multiplexing. The selection of blood culture bottles was randomised and included the BD BACTEC™ Plus Aerobic (n=5), Peds Plus (n=1), and Lytic Anaerobic (BD, USA) (n=6). Although aerobic and anaerobic bottles were drawn together from the patient, only one was processed as this was the bottle that flagged during the collection period. The former two contain resin, and the latter lytic agents for the release of intracellular pathogens.

4.3.1.2 GridION sequencing

Samples were prepared for GridION sequencing using the Rapid Sequencing Kit (SQK-RBK004) following the manufacturer's protocol. Sequencing was performed using R9.4 flow cells for 48h and resulting data was live basecalled with Guppy version 3.2.6.

4.3.1.3 Sequence analysis

Ten clinical samples were sequenced with and without saponin use, each barcoded, paired and sequenced on an individual flow cell. Thereafter, four clinical samples were sequenced with and without saponin use using two individual flow cells respectively.

GridION sequencing data was processed via CRuMPIT3, an in-house pipeline that utilised prinseq-lite (v0.20.4) to remove low complexity reads, pore-chop (v0.2.3) for demultiplexing barcodes and centrifuge (v1.0.4) to remove human reads using the NCBI RefSeq human reference genome (GRCh38). Reads were classified with the metagenomic classifier, Kraken2 (update: 04/23/2019) using the minikraken2_v2_8GB_201904_UPDATE database. Species profile was compared to anonymized clinical diagnosis generated using pure culture isolates [species from Bruker microflex MALDI-TOF MS (Bruker, USA), antimicrobial resistance from BD Phoenix microbroth dilution (BD, USA)].

4.3.2 Results

Each sample was extracted from 1ml of liquid blood culture and eluted in 15μl of TE buffer. Total DNA yields for no treatment vs. saponin treatment are shown below in **Figure 19**/**Table 6**. There was no major difference seen in the total DNA yield between the no pre-treatment group 1838ng (IQR 477-3600) [Range 159-5310] compared to the saponin pre-treatment group 1620ng (417-3480) [153-5310] (p=0.09).

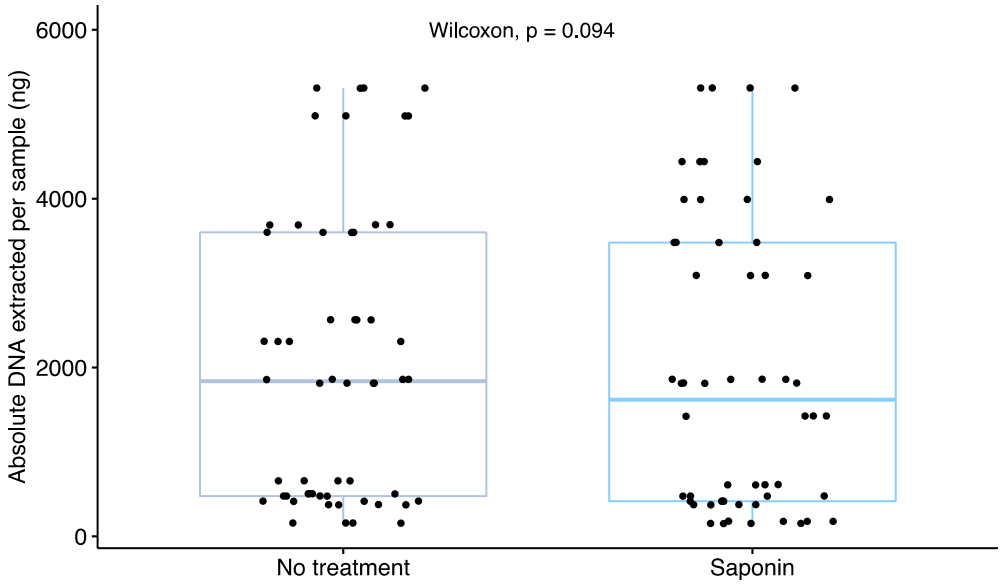

Figure 19. Total DNA extracted from no treatment vs. saponin treatment (ng).

Table 6. Total DNA yield by pre-treatment method (n=28).

Measurement	Sample numbers	Pre-treatment	Median	IQR	Range
Qubit (ng)	14 positive samples per group	No pre-treatment	1838	477-3600	159-5310
		Saponin pre-treatment	1620	417-3480	153-5310

There was no difference seen in the number of bacterial bases sequenced in the no pre-treatment group Gb 2.2 (1.1-4.2) [0.15-4.7] compared to the saponin pre-treatment group Gb 2.9 (0.92-4) [0.017-4.7] (p=0.77) (**Figure 20**). However, the number of human bases in the no pre-treatment group Mb 290 (120-1400) [0.68-4200] was much higher compared to the saponin pre-treatment group Mb 0.27 (0.2-2) [0.048-5.4] (p=0.001).

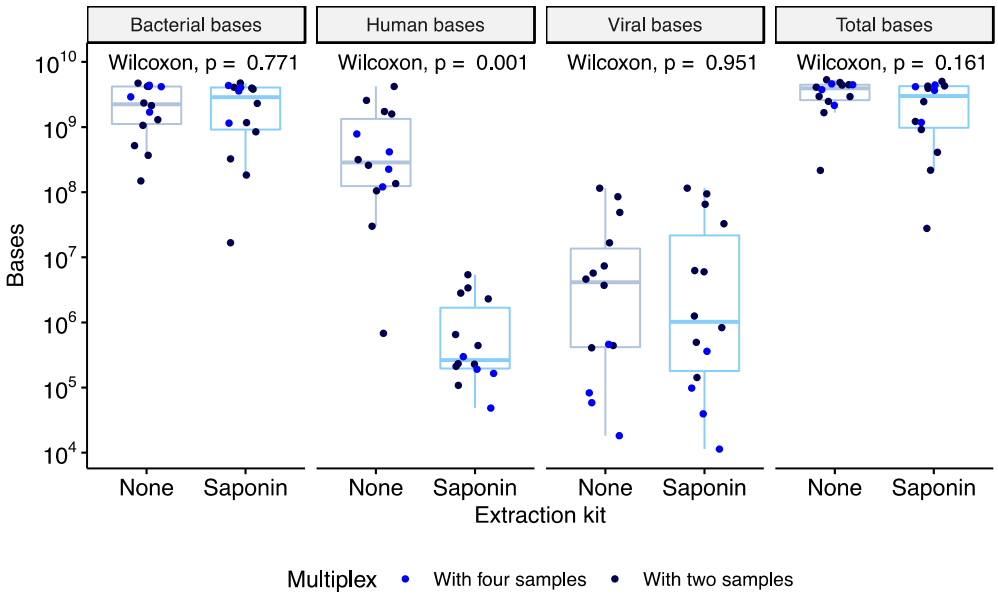

Figure 20. Bacterial, human, viral and total bases grouped by depletion method. No pre-treatment is compared with saponin pre-treatment for bacterial bases (p=0.771), human bases (p=0.001), viral bases (p=0.951) and total bases (p=0.161).

There was no difference seen in approximate bacterial sequencing depth in the no pre-treatment 448x (224-841) [30-945] compared to the saponin pre-treatment group 590x (184-808) [3-948] (p=0.77), assuming an average bacterial genome of 6 million bases. Samples when multiplexed by four did not appear to have reduced sequencing depth compared to samples when multiplexed by two (p=0.12) (**Figure 21**).

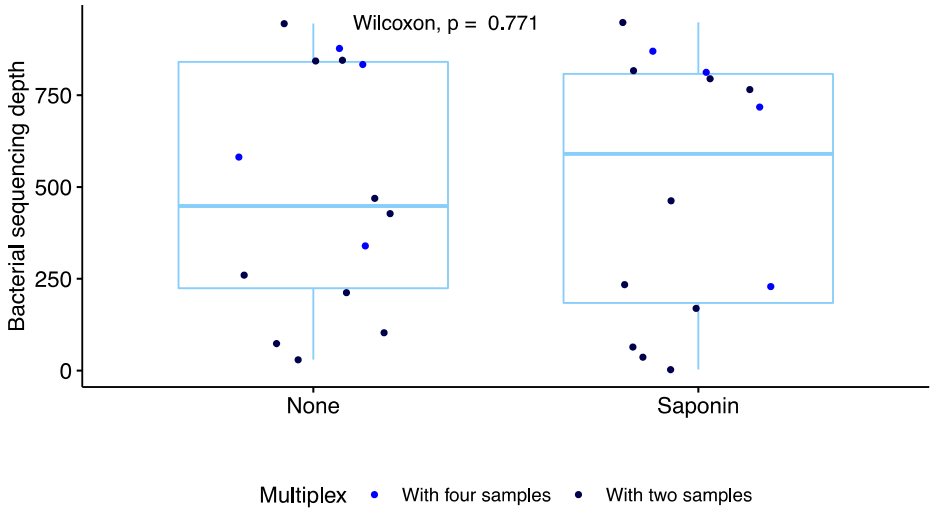

Figure 21. Bacterial sequencing depth. Calculated with assuming an average bacterial genome length of 6 million bases. No pre-treatment is compared to saponin pre-treatment for bacterial sequencing depth (p=0.771).

Next, the effect of saponin on sequencing read lengths were investigated. The bacterial read length average data per sample for no pre-treatment of samples centered at 2260 (2199-2401) [1203-2708] bases and was narrower compared to the saponin pre-treatment which centered at 2152 (1930-2442) [1243-2688] bases and was widely dispersed (**Figure 22**). However, the reverse was true regarding human read length data where the no pre-treatment group centered around 1824 (1597-2065) [1063-2620] bases and was narrowly distributed compared to the saponin pre-treatment which centered at 1598 (816-2203) [567-2797] bases where data were more dispersed.

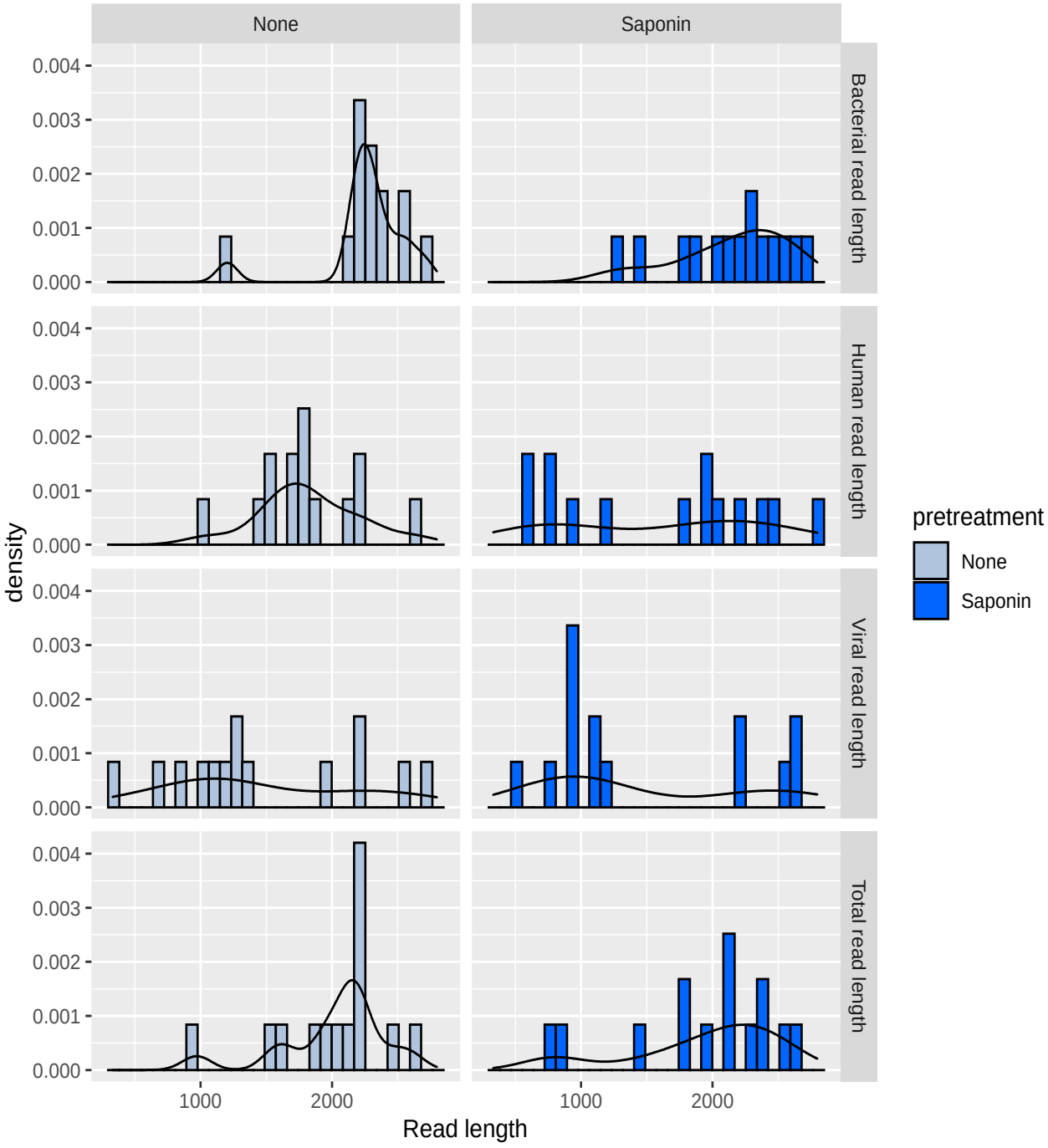

Figure 22. Histogram of average read length data per sample with density plot.

Barcode contamination was also investigated; this is where reads are assigned to the wrong barcode. Given the inherent error rate of 5-10% per base in Nanopore data[127], there is concern about the accuracy of read demultiplexing, which is based on short index sequences. The sequence yield from barcodes that were used ranged from $2.7x10^7$ to $5.3x10^9$ bases whereas the reported sequence yield from unused barcodes ranged from $7.4x10^2$ to $1.2x10^7$ (**Figure 23**). Two water control samples (no pre-treatment and saponin) that were barcoded ranged from $4.6x10^5$ to $6.5x10^5$ bases sequenced. Barcode misclassification (i.e., mis-assigned reads) when multiplexed with two samples, calculated as the total percentage of reads classified to a barcode bin not in use, was 0.075% which was similarly to other reported studies[128] and increased to 0.32% when multiplexed with four samples.

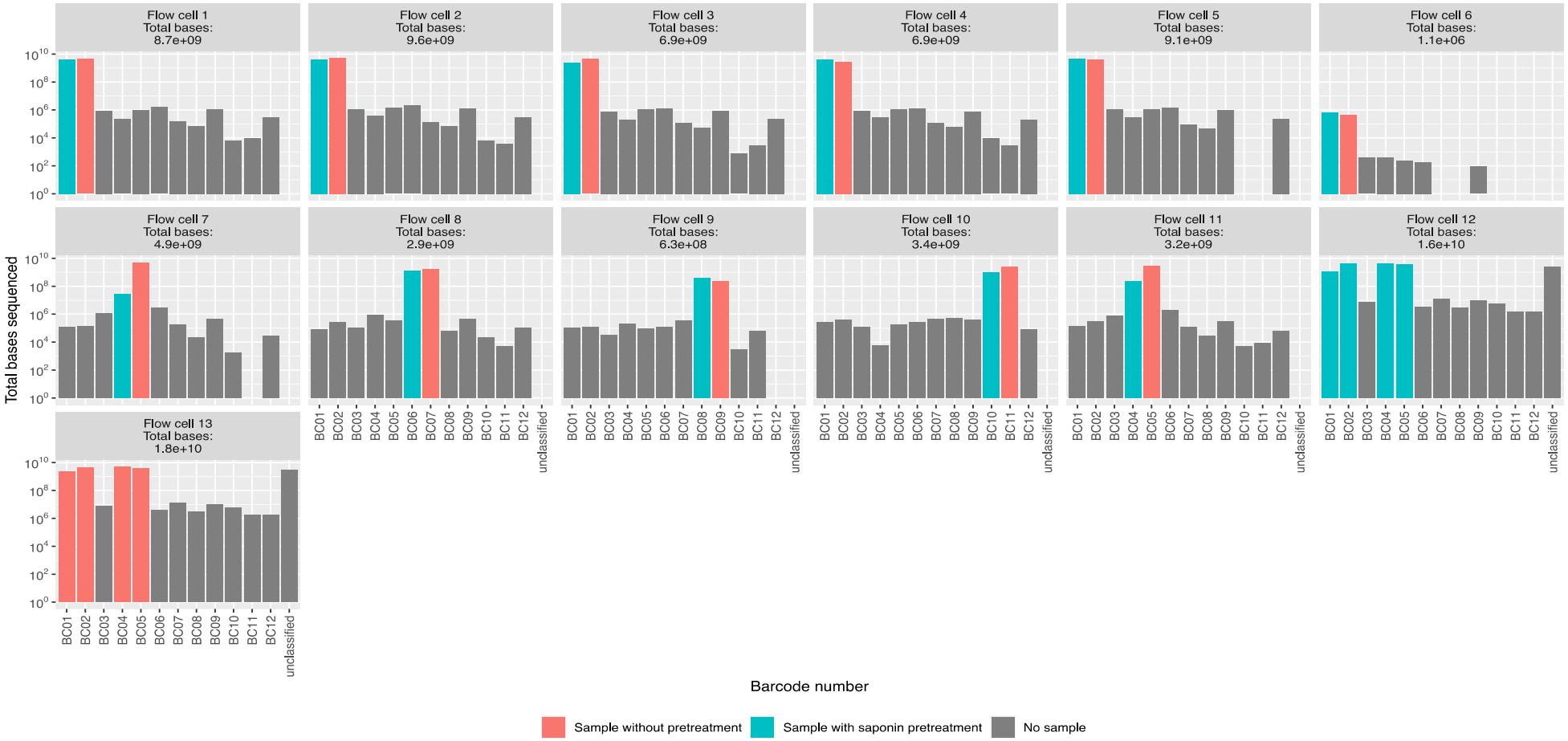

Figure 23. Total bases sequenced per barcode for each flow cell. Note the log scale on the y-axis. Flow cell 6 was a negative water control.

The Kraken2 results of samples sequenced are displayed in **Figure 24**. "Species specific reads" represent the percentage of reads classified as the correct species as determined by MALDI- TOF. Adding this to "genus only" and "right genus, wrong species" represents the correct genus assigned. "Other bacterial" refers to unclassified or the wrong bacteria. "Non-bacterial" refers to reads classified as viruses and eukaryotes.

"Species specific reads" had a median percentage of 81% (71%-84%) [22%-94%] of reads. Correct genus assigned (as previously defined) had a median percentage of 87% (84%-91%) [44%-97%] of reads. "Wrong genus, wrong species" had a median percentage of 1% (0.6%-3%) [0.1%-9%] of reads. The samples with the highest "wrong genus, wrong species" percentage of reads was *Eubacterium limosum* which had a 9% error classification and *Enterobacter cloacae* which had 4%. These errors may be explained by the following hypotheses: 1) The database used in Kraken2 was not composed of sufficient genomes of the specific species and hence was misclassified; 2) There was contamination introduced during some level of the laboratory and/or bioinformatic process; 3) There may be true polymicrobial pathogens in the sample, not cultured or detected by MALDI-TOF.

Only 6 genomes were available from the NCBI for *Eubacterium limosum* whereas 563 genomes were available for *Enterobacter cloacae.* Simulation experiments were performed with 6 genomes for *Eubacterium limosum* and 10 genomes for *Enterobacter cloacae* using the NanoSim-H tool(v3).[129] The default ecoli_R9_2D error profile was used with 10000 reads generated per sample. These experiments reveal up to 3.3% and 2.9% of the "wrong genus, wrong species" percentage of reads could be accounted for the respective organisms due to Kraken2 misclassification (**Figure 25**). Although

simulation could explain some of the misclassification in the *Enterobacter cloacae* sample, a large percentage of reads (41%) were classified as "genus only". This may be due to the fact that this species is often grouped under the *Enterobacter cloacae complex*, to which 84% of reads aligned. Further work creating a customized database and adding more specific species is required to validate this assumption.

Four samples (from two unique patients) containing *Staphylococcus hominis* had a large percentage of reads classified as "right genus, wrong species". Of note two of these samples were predominantly classified by Kraken2 as *Staphylococcus epidermidis*. Another two *Staphylococcus hominis* samples (from one unique patient) did not have the same classification problem. It is plausible that either the Kraken2 database has inadequate *Staphylococci* species genomes to accurately differentiate among specific species or that MALDI-TOF may have been incorrect in species determination as phenotypic errors have been reported previously[130] and that *Staphylococcus epidermidis* may truly exist in the sample either as a contaminant or pathogen.

An indicative threshold of >1000 bacterial reads could be used to indicate a positive blood culture. Assuming all blood positive cultures were monomicrobial as per MALDI-TOF results, the sensitivity based on the single most dominant species classified by Kraken2 was 93% (95% CI, 76-99%, 26/28) at the species level and 100% (95% CI, 88-100%, 28/28) at the genus level. Specificity was 100% (95% CI, 16-100%, 2/2) within the controls, however, this analysis was not performed using any polymicrobial samples at this stage.

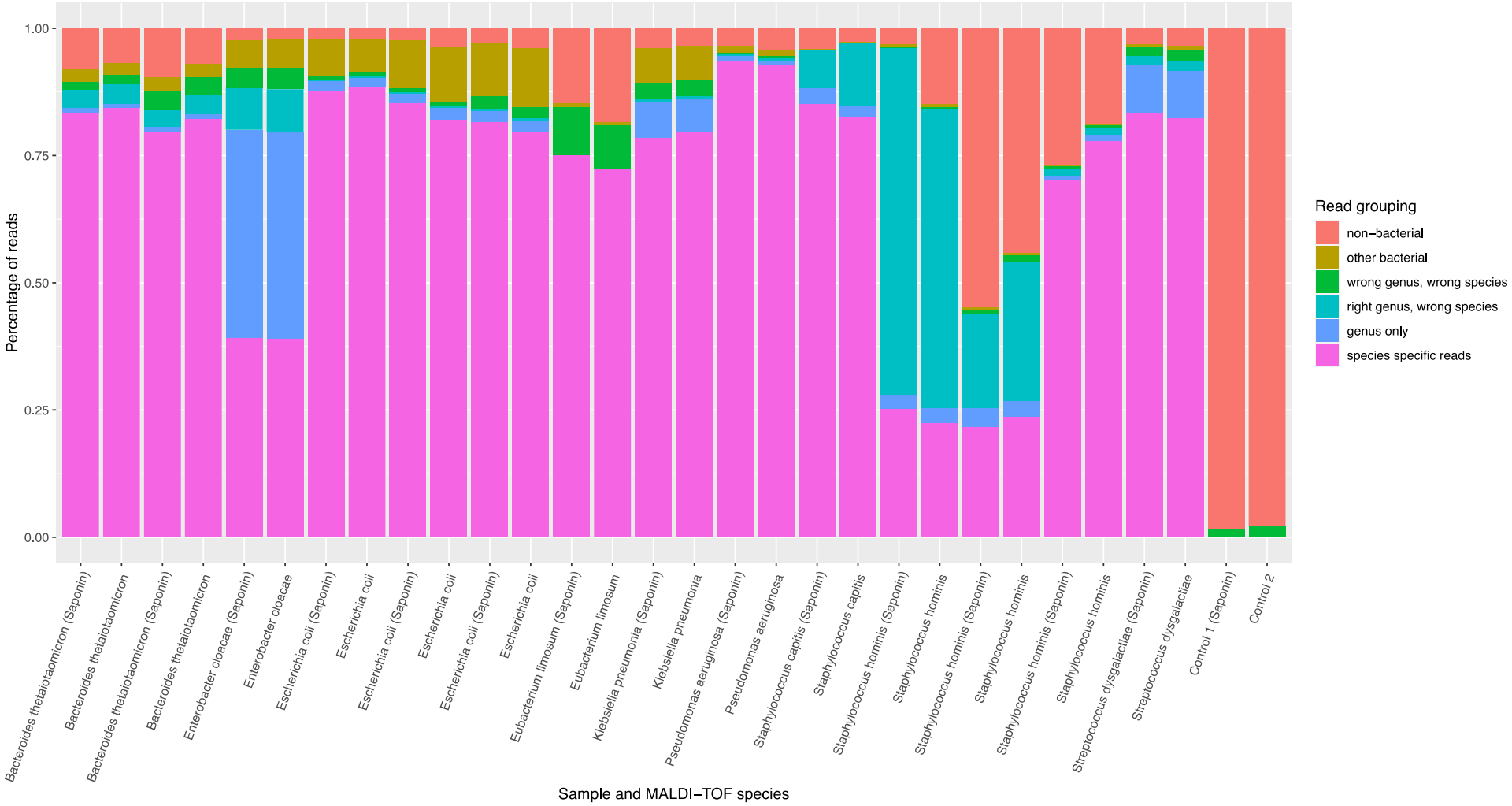

Figure 24. Kraken2 results for metagenomic samples.

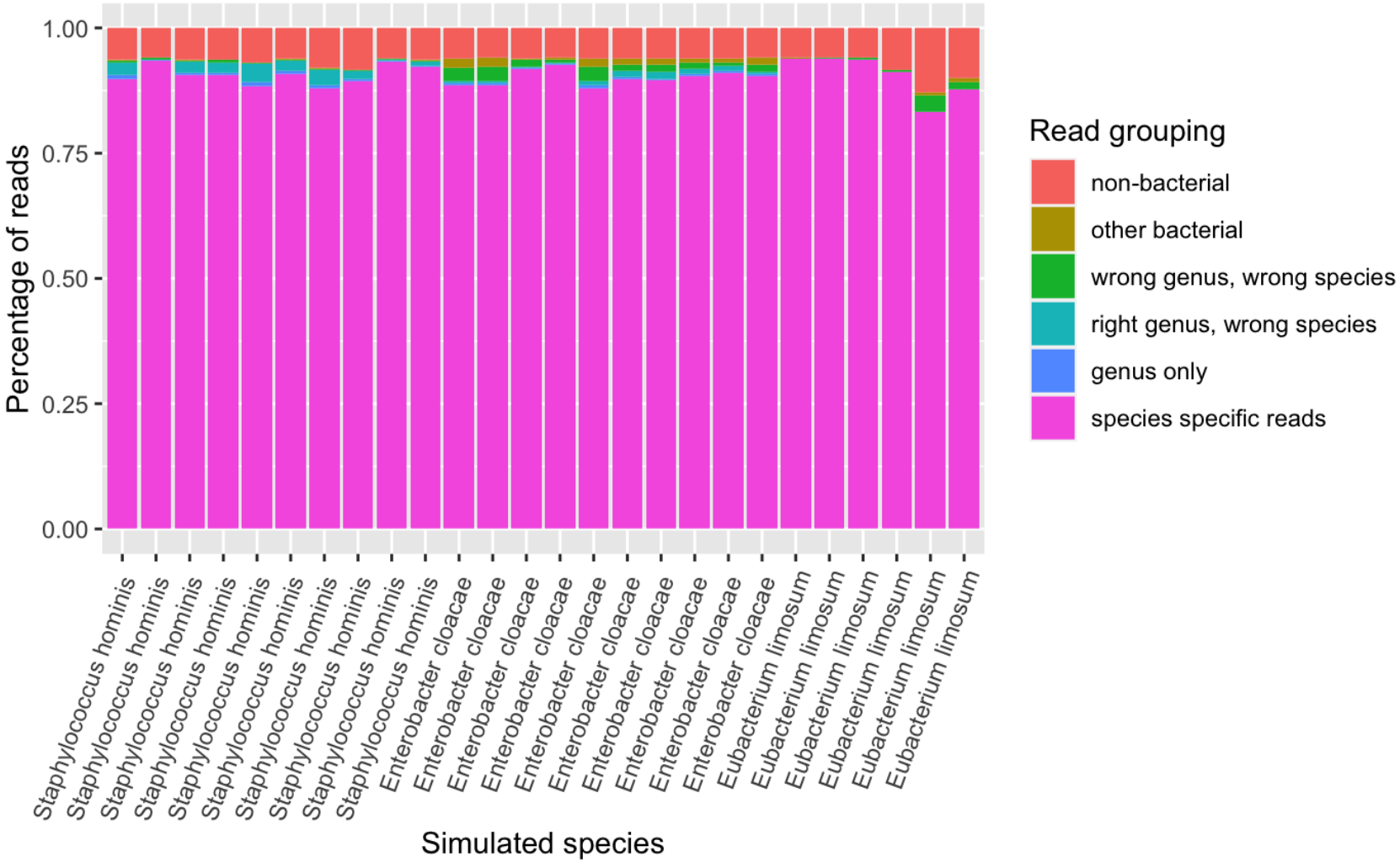

Figure 25. Kraken2 results for simulated samples from NCBI.

4.3.3 Discussion

The saponin pre-treatment method did not affect the total amount of DNA yield extracted from samples compared to no pre-treatment. Total DNA yield ranged from 153ng to 5310ng across samples. As it is recommended that an input amount of ±400ng[131] be used for the rapid barcoding kit for Nanopore sequencing, some samples (n=3) were below this threshold, however, were still sequenced successfully, although had lower sequencing output (**Figure 26**).

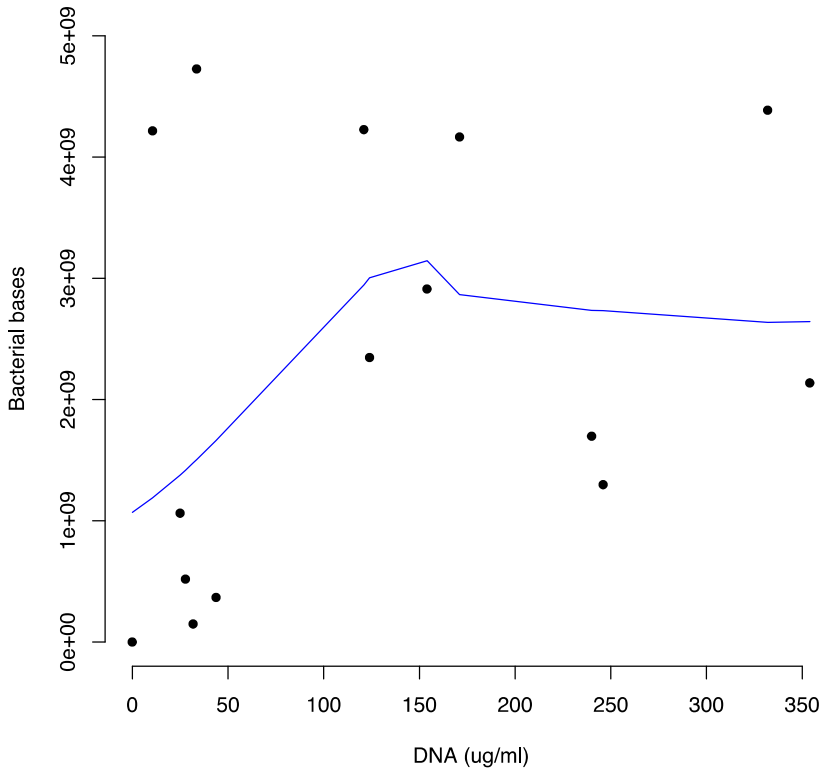

Figure 26. Extracted DNA concentration (µg/ml) plotted beside bacterial bases sequenced. A Loess regression method was used to plot a curve.

It's worth noting that, in contrast to other experimental conditions, blood culture samples initially have a higher number of bacterial reads compared to human reads. Therefore, the reduction in human bases sequenced due to saponin pre-treatment is less likely to impact the overall analysis. Although saponin pre-treatment led to a significant decrease in the number of human bases sequenced, the number of bacterial bases sequenced remained consistent with the group that did not receive pre-treatment. Additionally, there was no difference seen in the sequencing depth between the two groups. On average, the sequence depth was 519x per sample which is promising especially for (meta)assembly and increased contig accuracy. Samples when multiplexed by four did not have reduced bacterial sequencing depth compared to samples when multiplexed by two, however, barcode contamination was slightly higher with increased multiplexing.

Bacterial read lengths were broadly similar between the pre-treatment and no treatment groups. Barcode contamination was evident in non-assigned barcode bins, where no samples were added. The ratio of total bases from barcodes with samples compared to barcodes without samples is 13241 ($3.7x10^9/2.8x10^5$):1. Hence, barcode contamination was minimal. Although Kraken2 predicted the correct species in 93% (26/28) of samples and the correct genus in 100% (28/28) of samples within the first hour of sequencing, this was true assuming all positive blood cultures were monomicrobial. Furthermore, the reasons for the two incorrect predictions may be that the Kraken2 database lacks specific genomes to accurately discriminate among species or that the phenotypic MALDI-TOF result was incorrect. Further work is required to develop the optimal method to assign a species as true positive or true negative in polymicrobial samples. In conclusion, saponin pre-treatment is not necessary for achieving high bacterial sequencing depth using the Oxford Nanopore when multiplexing with 2-4 samples extracted direct-from-positive-blood culture.

4.3.4 Limitations

Although this study compares extracted metagenomic samples genotypically to the laboratory phenotypic results, I was not able to compare the sequencing results to the clinical scenario or clinical judgement in each case. Contamination in the experiment was a challenge, especially of negative controls. This has attempted to be addressed by extracting samples in a Class II biosafety cabinet and using strict laboratory procedures such as UV radiation of the cabinet and the use of DNAse spray on all equipment before and after use. Although working in a meticulous laboratory setting, there still is a risk of contamination either from laboratory handling or from the laboratory environment. A limitation to the controls is that only negative controls were used as this study did not use positive controls or spiked controls in this pipeline.

"If you cannot measure it, you cannot improve it."

-Lord Kelvin

5 BENCHMARKING TAXONOMIC CLASSIFIERS

In chapter 4, I concluded that the BiOstic Bacteraemia kit performed best at extracting direct-from-blood cultures, due to its superior DNA extraction capability with relatively low contamination observed. The saponin method was seen to be the superior human DNA depletion method, however, it may not be needed if working with high bacterial concentrations in positive blood cultures. I also ran a pilot sequencing study using clinical blood culture samples. In this pilot study, I concluded that samples when multiplexed by four did not have reduced bacterial sequencing depth compared to samples when multiplexed by two, however, barcode contamination was slightly higher with increased multiplexing, however, still minimal. In this following chapter, I work on optimising the bioinformatic pipeline by benchmarking common tools and databases used for species classification. I also perform simulations from reference genomes to control for bioinformatic contamination or misclassification.

5.1 Background

Reliable identification of bacterial species is a fundamental part of clinical microbiology. Traditionally identification has been based on growth on selective media, morphological appearances, and biochemical tests. However, many microbiology laboratories now rely largely on MALDI-TOF mass spectrometry of cultured isolates to identify most species. In clinical metagenomics, a culture-independent approach is used to identify species present in a sample based on unselective DNA sequencing.

In Chapter 3, I have shown evidence of multiple studies that compared the potential of metagenomic approaches to replicate traditional workflows, with sensitivity for species detection, regarding culture-based approaches as the reference standard, varying from 63% to 98% and specificity from 35-100%.[64] Metagenomic species identification depends on the quality and properties of the sequence data used, for example, long-read vs. short-read, the taxonomic classification algorithm and the database used to match sequences to species. Additionally approaches need to account for contamination while maximizing sensitivity and specificity. Contamination with human DNA is expected and can also arise from imperfect sampling, the laboratory environment, equipment, and reagents, and also bioinformatically. Bioinformatic contamination can occur from errors in demultiplexing where multiple samples are run at once using different indexes, and from misclassification of sequence data as the wrong genus or species.

One of the primary challenges in metagenomics is reads may be generated from multiple species rather than just one following sequencing of an isolate from culture. The creates a computational challenge, that is, how to accurately identify and differentiate all species contained within a sample, while ruling out any species detected only as artefacts? This

must also be done against an ever-increasing number of previous sequences to compare with. Metagenomic classifiers can be grouped based on their reference database type which include: DNA-to-DNA methods for example, Kraken[132], Kraken2[72], Bracken[133] and PathSeq[134]; DNA-to-protein methods for example, Kaiju[135] and Diamond[136]; and DNA-to-marker methods which include only specific gene families in their reference databases for example, MetaPhlAn[137] and mOTUs[138]. In a recent benchmarking study[139], DNA-to-DNA methods were the best-scoring methods, while DNA-to-marker methods performed more poorly.

Some studies have benchmarked classifiers and observed that the performance of different classifiers varies significantly even on the same benchmark datasets[74,139–143]. However, performance depends both on the taxonomic classifier and the reference database of sequences of known species used and few studies have compared performance using a uniform database.[139] For example, while Ye et al. previously performed classification benchmarking on short-read sequencing, previous studies have not extensively benchmarked long-read sequencing data or specifically investigated the effects of database selection or read length on classification performance. Furthermore, no catalogue of how common taxonomic misclassification is by species exists in literature to allow studies to account for bioinformatic contamination arising from the classification process.[139] Therefore, there is still a need for thorough benchmarking to shed light on the optimal selection of bioinformatic tools for metagenomics and to account for "bioinformatic contamination" which is an important and potentially underrecognized issue in the field.[144]

In this chapter, I use simulated and real-world data to define the intrinsic rates of species misclassification that occur using Illumina and Oxford Nanopore sequencing platforms with commonly used classification tools and databases. Reads from reference genomes is simulated, rather than simulating metagenomic samples, to create a controlled series of experiments that allow us to estimate the extent of taxonomic misclassification that occurs from sequencing a range of common pathogens and contaminants. These findings can be applied to metagenomic samples where there is evidence of the presence at least one bacterial species. Species specific thresholds are derived for identifying if the presence of a second, third, or further additional species arises from misclassification or true presence. Therefore, this approach allows the extent of bioinformatic contamination arising from taxonomic misclassification to be accounted for in clinical diagnostic workflows and polymicrobial infections to be identified more reliably.

5.2 Methods

5.2.1 Simulated data

Simulated sequence data was generated for ten common blood stream pathogens (*Staphylococcus aureus, Streptococcus pneumoniae, Streptococcus pyogenes, Enterococcus faecalis, Pseudomonas aeruginosa, Escherichia coli, Klebsiella pneumoniae, Proteus mirabilis, Bacteroides fragilis* and *Enterobacter cloacae*) and six common blood culture contaminants (*Staphylococcus epidermidis, Staphylococcus haemolyticus, Bacillus cereus, Cutibacterium* [*Propionibacterium*] *acnes, Streptococcus mitis* and *Micrococcus luteus*). For each species, 100 randomly-selected reference genomes were downloaded where possible (on the 10th of January 2021) from the NCBI RefSeq database[145], that is, 1568 genomes in total (only 68 genomes were

available for *Micrococcus luteus*). The scope of this evaluation was restricted to the identification of bacterial infections only. Although this study opted to not remove genomes from random selection that overlapped with the classification database (detailed in **Supplementary Table 9**), a sensitivity analysis was performed excluding these genomes to ensure that contamination was not being underestimated. However, genomes very similar to those in the reference database are likely to make up some of any genomes sequenced in any given setting. Therefore, this approach, taking a random selection (including overlapping genomes) may offer more representative results.

Simulated Illumina and Oxford Nanopore reads were generated from each genome, with separate simulations with and without sequencing errors. The relevance of simulations without sequencing errors is to allow estimates of what performance might be achieved as sequencing technologies improve and potentially per base error rates are reduced in newer releases.[146] Illumina reads were simulated using ART_Illumina v2.3.7[147] (art_illumina -i reference.fna -p -l 250 -f 20 -m 999 -s 1 -o -nf 5 -na -ss MS -na) generating reads at a 20x coverage for each reference genome (**Figure 27**), to simulate a realistic MiSeq throughput which is the around the minimum depth previously shown to be adequate for variant identification and antimicrobial resistance prediction.[148] Oxford Nanopore reads were simulated using NanoSim-H v1.1.0.4[129,149] (nanosim-h reference.fna -p ecoli_R9_1D -o -n 11000), generating reads at approximately a 20x coverage per genome with a median read length of 5921bp (IQR 2745-11709) [range 125-120218] (**Figure 28**). The default ecoli_R9_1D error profile, which mimics the R9 flow cell using 1D sequencing, and the MiSeq error profile were used for Nanopore and Illumina read simulation respectively. No pre-processing of fastq reads were performed as this was a controlled simulated experiment.

To assess potential misclassification of human reads as bacterial taxa, 100 Illumina simulations from the human genome (GRCh38 downloaded from NCBI on the 19th of April 2022) were performed using ART_Illumina and set coverage at 0.032x to approximate high human contamination in metagenomic samples (art_illumina -i GRCh38.fna -p -l 250 -f 0.032 -m 999 -s 1 -o %s -nf 5 -na -ss MS -na).

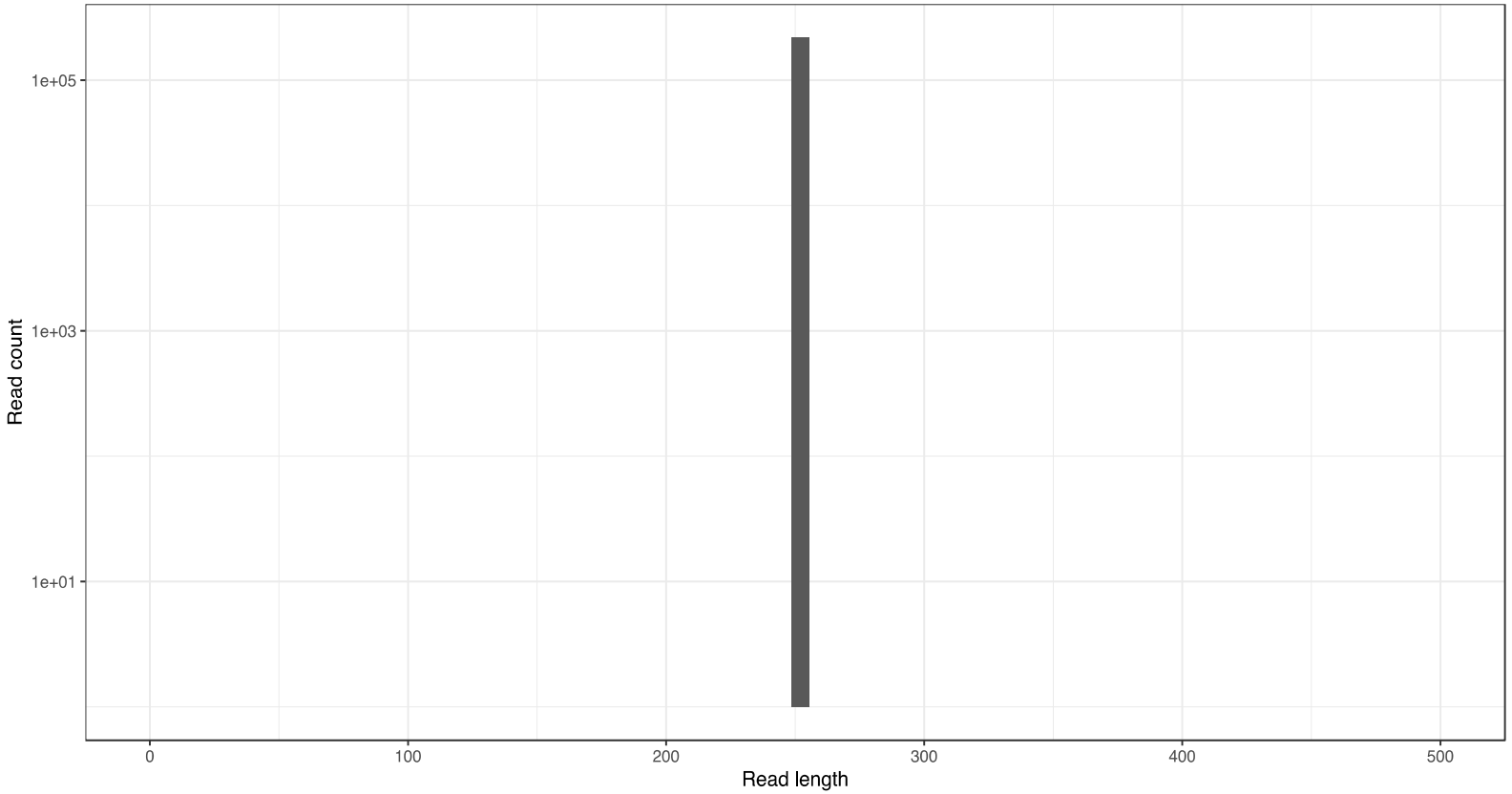

Figure 27. Illumina sequencing read distribution.

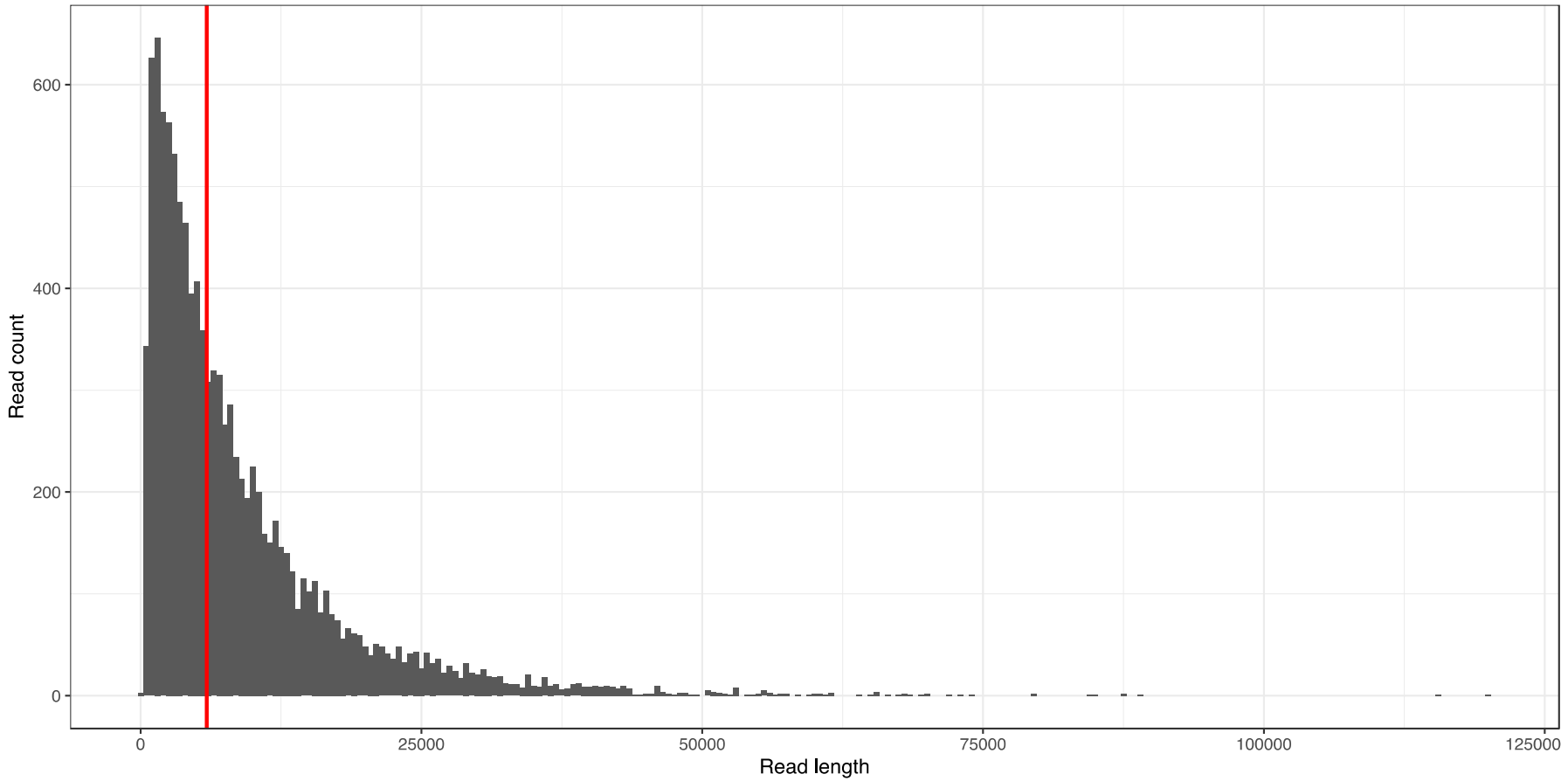

Figure 28. Nanopore sequencing read distribution. Red line shows median value.

5.2.2 Taxonomic classifiers and sequence databases

Previous studies have demonstrated that Kraken-based classification tools deliver faster classification speeds and similar accuracy compared to comparator tools.[139,150] Therefore, this study chose to evaluate Kraken2 (v2.0.7)[72] and Bracken (v2.5)[133], as well as Centrifuge[151] which was written to address the high memory requirements of Kraken. Mini (8GB) and standard (30-50GB) uniform databases were built from archaeal, bacterial, viral, and human reference genomes downloaded from NCBI RefSeq on the 10th of January 2021. Databases were built according to each tool's default setting. The mini database resembled an up-to-date version of "MiniKraken" limited to 8GB, which down-samples the standard Kraken2 database using a hash function. As Centrifuge does not allow for the restriction of size when building a database, only a standard database was built. Fastq files were classified using the default settings of Kraken2 (kraken2 --db $DBNAME sequence_data.fq --report /), Bracken (bracken -d $DBNAME -i sequence_data.fq -o /) and Centrifuge (centrifuge -k 1 -x $DBNAME -f / --report-file / -S /) however, for Centrifuge the distinct primary assignment value was set to 1 (-k 1) in order to provide a comparable standard where each read is assigned to only one taxonomic level. Kraken and Bracken produced the same output format, while the Centrifuge output was converted to a Kraken-style report using the command (centrifuge-kreport -x $DBNAME centrifuge.output.txt> krakenstyle.output.txt).

5.2.3 Real-world data

A convenience sample of real-world Illumina sequencing data was used for additional benchmarking, which consisted of reads generated from sequencing of 100 pure isolates

of *Staphylococcus aureus, Escherichia coli* and *Klebsiella pneumoniae* from previous published works [NCBI project accession numbers PRJNA690682, PRJNA604975].[152,153] Sequenced isolates originated from blood stream infections between 15-September-2008 and 01-December-2018 in Oxfordshire, UK. Available 150 bp paired-end short-read sequences were generated using the Illumina HiSeq 4000.

5.2.4 Statistical analysis

Classifier performance was measured as the proportion of reads classified as the correct species or correct genus, given the known species of the reference genome or clinical isolate from which reads were simulated or obtained. The reported read abundance for each species may be referred to as either "raw" calculated from the relative abundance of reads from each taxa or "corrected" by accounting for genome size, such that the relative number of organisms of each taxa is estimated. This study used raw abundance estimates unless the correction is automatically performed by the software as in the case of Bracken.

A generalized linear regression (with a logit link function, binomial family and robust standard errors) was used to investigate associations between the proportion of reads classified as the correct species and sequencing technology, taxonomic classifier, database and species. (R model code: m=glm (proportion_reads_correctly_classified ~ sequencer + error_profile + database + classifier + species, family=binomial(link="logit"), data=df); Robust standard errors were estimated using the sandwich library: cov.m = vcovHC(m, type = "HC1"), followed by std.err =

sqrt(diag(cov.m)).) These models were also used to present probabilities of reads being classified correctly for specific combinations of these covariates.

Similarly, to investigate the extent to which species misclassification existed because the true and misclassified genomes were similar, a linear model (R code: lm(ANI ~ pct_at_this_level_or_lower, data)) was used to investigate the relationship between the average nucleotide identity (ANI) between the true species and the misclassified species and the proportion of reads misclassified as the incorrect species. ANI values for each correct-incorrect species pair were generated by taking the median value for all pairwise comparisons between 30 genomes of each species misclassified (downloaded from NCBI RefSeq on the 24th June 2021) and 50 genomes of the correct species using FastANI (fastANI --ql [QUERY_LIST] --rl [REFERENCE_LIST] -o [OUTPUT_FILE]).[154] Analyses were performed using R (v.4.1) and R Studio (v1.4.1717).[155]

To investigate the relationship between read lengths and species classification performance, Oxford Nanopore and Illumina reads were generated of varying lengths using error and error-free profiles from 100 reference genomes of *Staphylococcus aureus* and *Escherichia coli* each, and plotted read lengths vs. the proportion of reads correctly classified by Kraken2.

5.3 Results

5.3.1 Determinants of taxonomic classification accuracy

Sequence data from 1568 genomes representing 10 common pathogens and 6 common blood culture contaminants was simulated. *Staphylococcus aureus*, classified with Kraken2, using a mini (8gb) database, with simulated Illumina reads including sequencing errors was chosen as the reference category for comparisons as this is a common pathogen and common metagenomic workflow (**Table 7** and **Table 8**). Using this combination, 89.3% (95% CI, 88.5%:90.9%) of reads were classified as the correct species, and 96.7% (95.3%:97.5%) as the correct genus. Detailed performance in is also presented in **Table 9** and **Supplementary Table 7**.

In a multivariable model, using Oxford Nanopore (including with sequencing errors) independently improved the odds of classifying reads correctly to the species, while reducing the odds of classifying reads to the genus level (adjusted odds ratio, aOR, vs. Illumina sequencing 1.26 (95% CI, 1.21:1.32) and 0.87 (0.83:0.90) respectively). As expected, simulating data without sequencing errors had a similar performance for Illumina, while significantly improving performance for Oxford Nanopore, however this level of performance cannot be currently achieved in reality (**Table 7**).

Use of the "standard", that is, larger, database independently improved species and genus level classification (aOR vs. Mini-database 1.91 (95%CI 1.83:1.98) and 3.23 (3.09:3.37) respectively). Compared to Kraken2, Centrifuge performed similarly for species and inferiorly for genus classification (aOR vs. Kraken2 1.14 (1.08:1.20) and

0.61 (0.58:0.64) respectively), while Bracken performed superiorly for both species and genus (aOR vs. Kraken2 3.26 (2.31:3.40) and 5.87 (5.57:6.18) respectively).

Classification performance varied by species. Some species, including *Cutibacterium acnes* and *Enterococcus faecalis* had a higher adjusted probability of reads being correctly classified at the species level, compared to *Staphylococcus aureus*, ranging between 92.5% to 99.7% across different sequencers, classifiers and databases (cf. 78.2-98.5% for *S. aureus*) (aOR ranging between 3.45 and 5.83 (**Table 7**)). Whereas other species were more challenging to classify correctly, for example, *E. coli* with a predicted adjusted probability of reads being classified correctly ranging between 56.1% to 96.0% (aOR vs. S. aureus 0.36 (95%CI, 0.33:0.38)). Of note, *Streptococcus mitis, Bacillus cereus* and *Enterobacter cloacae* had seemingly poor performance ranging between 28.2%-88.0%, 13.4%-74.3% and 16.6%-78.8%, however, these species belong to a group or complex of bacteria which makes genomic differentiation more complex.[156–158]

Table 7. Logistic regression model for performance of species abundance classification by sequencer, classifier, database and bacterial species.

An odds ratio >1 indicates improved odds of performance while <1 indicate poorer performance relative to the reference.

		Descriptive: Median (IQR)	Odds ratio	95% CI	p-value
Multivariable reference	Species abundance classification (Reference: *Staphylococcus aureus*, Illumina with error profile, mini database, Kraken)	89.3 (88.5:90.8)	[Reference]	-	-
Sequencing profile	Illumina with error profile [Reference]	92.6 (63.2:97.2)	[Reference]	-	-
	Illumina with error-free profile	92.6 (63.4:97.2)	1.00	0.96, 1.04	0.9
	Nanopore with error profile	92.5 (76.5:97.4)	1.26	1.21, 1.32	<0.001
	Nanopore with error-free profile	97.7 (92.9:99.2)	3.00	2.83, 3.18	<0.001
Database	Mini database	90.9 (72.2:97.7)	[Reference]	-	-
	Standard database	95.5 (81.5:98.2)	1.91	1.83, 1.98	<0.001
Classifier	Kraken	88.4 (58.1:96.3)	[Reference]	-	-
	Bracken	97.2 (90.5:98.9)	3.26	3.31, 3.40	<0.001
	Centrifuge	94.4 (72.2:97.3)	1.14	1.08, 1.20	<0.001
Species	*Cutibacterium acnes*	99.5 (98.7:99.8)	5.83	5.28, 6.44	<0.001
	Enterococcus faecalis	98.2 (96.4:99.1)	3.45	3.21, 3.70	<0.001
	Streptococcus pyogenes	97.6 (95.7:99.0)	2.00	1.86, 2.15	<0.001
	Streptococcus pneumoniae	97.3 (95.2:98.9)	1.60	1.48, 1.72	<0.001
	Proteus mirabilis	97.2 (94.9:98.6)	2.03	1.82, 2.27	<0.001
	Staphylococcus epidermidis	97.2 (94.7:98.4)	1.73	1.54, 1.95	<0.001
	Staphylococcus haemolyticus	96.5 (94.2:97.8)	1.93	1.80, 2.07	<0.001
	Bacteroides fragilis	95.4 (91.6:97.6)	1.26	1.16, 1.37	<0.001
	Staphylococcus aureus	95.2 (88.0:98.5)	[Reference]	-	-
	Micrococcus luteus	89.5 (59.5:96.3)	0.37	0.34, 0.40	<0.001
	Escherichia coli	88.7 (61.2:94.2)	0.36	0.33, 0.38	<0.001
	Pseudomonas aeruginosa	87.5 (71.4:95.0)	0.37	0.34, 0.39	<0.001
	Klebsiella pneumoniae	85.2 (55.0:91.5)	0.23	0.22, 0.25	<0.001
	Streptococcus mitis	62.3 (42.7:75.2)	0.11	0.10, 0.12	<0.001
	Bacillus cereus	33.6 (14.8:51.2)	0.04	0.04, 0.05	<0.001
	Enterobacter cloacae	11.8 (3.3:92.5)	0.06	0.05, 0.06	<0.001

Table 8. Logistic regression model for performance of genus abundance classification by sequencer, classifier, database and bacterial species.

		Descriptive: Median (IQR	Odds ratio	95% CI	p-value
Multivariable reference	Genera abundance classification	96.7 (95.3:97.5)	[Reference]	-	-
	(Reference: *Staphylococcus* , error profile, Illumina, mini database, Kraken)				
Sequencing profile	Illumina with error profile [Reference]	98.3 (95.4:99.6)	[Reference]	-	-
	Illumina with error-free profile	98.3 (95.4:99.6)	1	0.96, 1.05	>0.9
	Nanopore with error profile	98.2 (92.8:99.3)	0.87	0.83, 0.90	<0.001
	Nanopore with error-free profile	99.5 (98.5:99.9)	4.23	3.97, 4.50	<0.001
Database	Mini database	98.1 (93.6:99.5)	[Reference]	-	-
	Standard database	99.0 (97.1:99.8)	3.23	3.09, 3.37	<0.001
Classifier	Kraken	97.4 (92.6:99.3)	[Reference]	-	-
	Bracken	99.5 (98.7:99.9)	5.87	5.57, 6.18	<0.001
	Centrifuge	97.7 (95.1:99.1)	0.61	0.58, 0.64	<0.001
Genus	*Cutibacterium*	99.8 (99.0:100.0)	1.54	1.40, 1.70	<0.001
	Bacillus	99.5 (97.7:99.9)	1.11	1.04, 1.19	0.003
	Streptococcus	99.4 (98.2:99.8)	1.18	1.12, 1.24	<0.001
	Staphylococcus	99.3 (98.2:99.8)	[Reference]	-	-
	Pseudomonas	99.3 (98.0:99.8)	1.16	1.09, 1.24	<0.001
	Enterococcus	98.8 (97.5:99.6)	1.01	0.96, 1.06	0.7
	Bacteroides	97.9 (95.5:99.0)	0.62	0.59, 0.65	<0.001
	Proteus	97.6 (95.4:98.9)	0.62	0.59, 0.65	<0.001
	Micrococcus	97.4 (94.6:99.1)	0.28	0.25, 0.32	<0.001
	Klebsiella	96.8 (88.1:98.3)	0.27	0.26, 0.29	<0.001
	Enterobacter	96.3 (92.3:98.2)	0.32	0.31, 0.34	<0.001
	Escherichia	91.8 (65.2:95.9)	0.09	0.08, 0.09	<0.001

Table 9. Predicted abundance classification for *Staphylococcus aureus* and *Escherichia coli*. Rows are arranged from highest to lowest predicted performance.

Sequencer	Error profile	Database	Classifier	Predicted performance (95% CI)
Staphylococcus aureus				
Nanopore	Error-free	Standard	Bracken	98.5 (98.4:98.6)
Nanopore	Error	Standard	Bracken	96.6 (96.3:96.8)
Illumina	Error	Standard	Bracken	95.7 (95.4:96)
Nanopore	Error	Mini	Bracken	93.7 (93.2:94.1)
Nanopore	Error-free	Mini	Centrifuge	92.5 (91.8:93)
Illumina	Error-free	Mini	Bracken	92.1 (91.6:92.7)
Nanopore	Error-free	Mini	Kraken	91.5 (90.9:92.1)
Nanopore	Error	Standard	Kraken	89.6 (88.9:90.3)
Illumina	Error	Standard	Kraken	87.3 (86.4:88.1)
Nanopore	Error	Mini	Kraken	81.9 (80.8:83)
Illumina	Error	Mini	Centrifuge	80.3 (79:81.6)
Illumina	Error	Mini	Kraken	78.2 (76.9:79.5)
Illumina	Error-free	Mini	Kraken	78.2 (76.9:79.5)
Escherichia coli				
Nanopore	Error-free	Standard	Bracken	96 (95.7:96.2)
Nanopore	Error	Standard	Bracken	90.9 (90.4:91.4)
Illumina	Error	Standard	Bracken	88.8 (88.2:89.4)
Nanopore	Error	Mini	Bracken	84 (83.2:84.8)
Nanopore	Error-free	Mini	Centrifuge	81.3 (80.2:82.4)
Illumina	Error-free	Mini	Bracken	80.7 (79.7:81.6)
Nanopore	Error-free	Mini	Kraken	79.3 (78.3:80.3)
Nanopore	Error	Standard	Kraken	75.4 (74.3:76.5)
Illumina	Error	Standard	Kraken	70.9 (69.7:72.1)
Nanopore	Error	Mini	Kraken	61.7 (60.3:63.1)
Illumina	Error	Mini	Centrifuge	59.3 (57.6:60.9)
Illumina	Error-free	Mini	Kraken	56.1 (54.7:57.6)
Illumina	Error	Mini	Kraken	56.1 (54.7:57.5)

5.3.2 Best performing combinations

Overall, Bracken with the standard database was the best performing option (**Figure 29** and **Figure 30** show abundance classification details of five important common blood stream pathogens and five contaminants respectively). Across all 1,568 genomes, median performance by Bracken using a standard database for species identification was 97.8% (IQR 92.7:99.0) [range 5:100] and for genus was 99.4% (IQR 98.8:99.8) [range 95.6:100.0], and had no unclassified reads. Median performance, with a common combination, Kraken2 with a mini database, for species was 86.4% (IQR 50.5:93.7) [range 4.3:100] and for genus was 94.6% (IQR 88.2:97.5) [range 76.7:100.0], with a median of 1.4% (IQR 0.8:8.9) [range 0.0:67.6] unclassified reads.

Human read misclassification was negligible. Across 100 simulations of the human genome, all classification software and database combinations performed similarly across species, genus and unclassified reads [IQR 100.0%:100.0%; IQR 100.%:100.0% and IQR 0.0%:0.0% respectively] except for Kraken mini database [IQR 98.6%:98.6%; IQR 98.6%:98.6% and IQR 0%:0.0% respectively] (**Figure 31**).

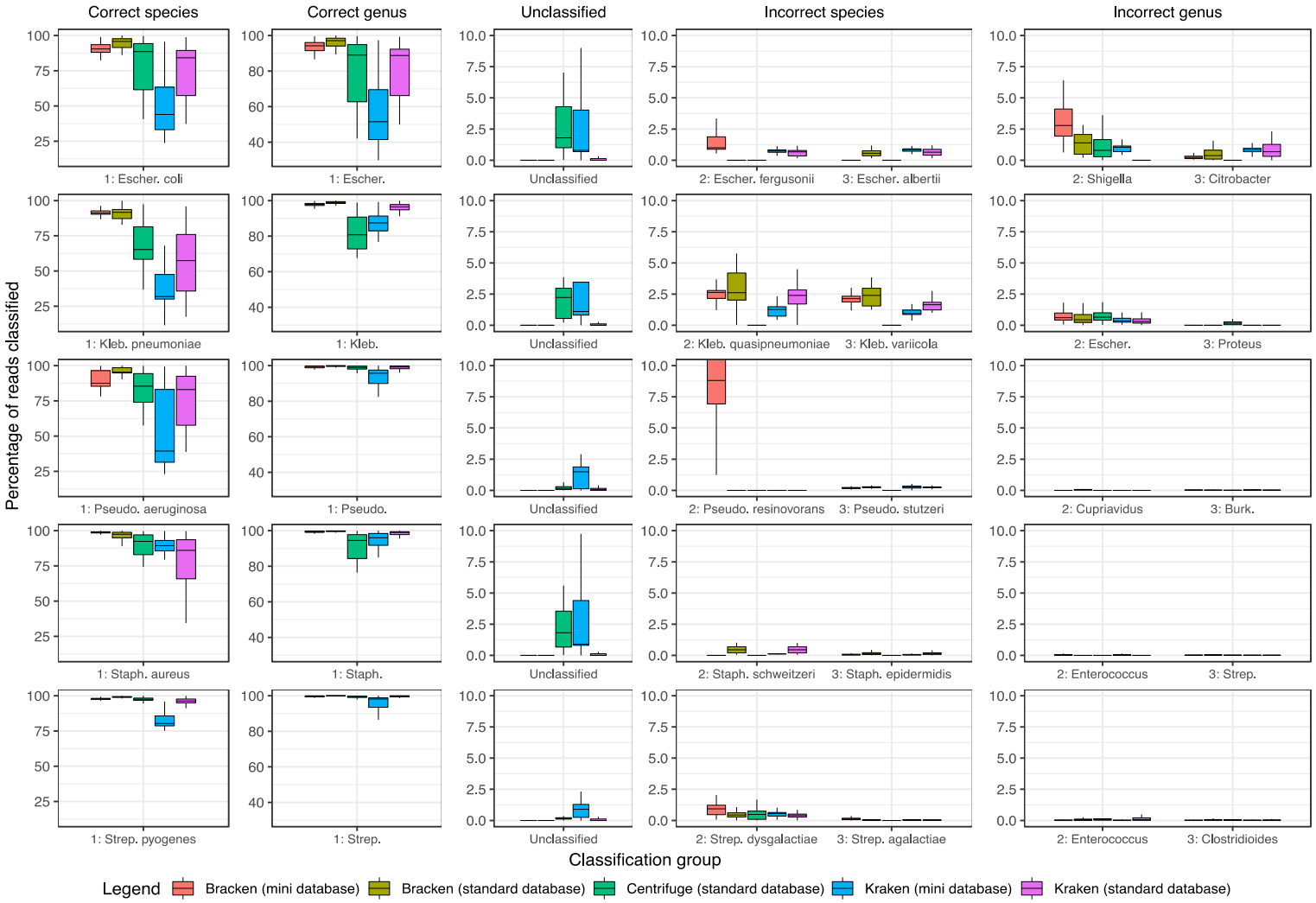

Figure 29. Percentage of reads classified correctly or incorrectly across classifiers for five important blood stream pathogens. The percentage of reads with the correct and incorrect assigned species and genus, and the percentage of reads unclassified are shown.

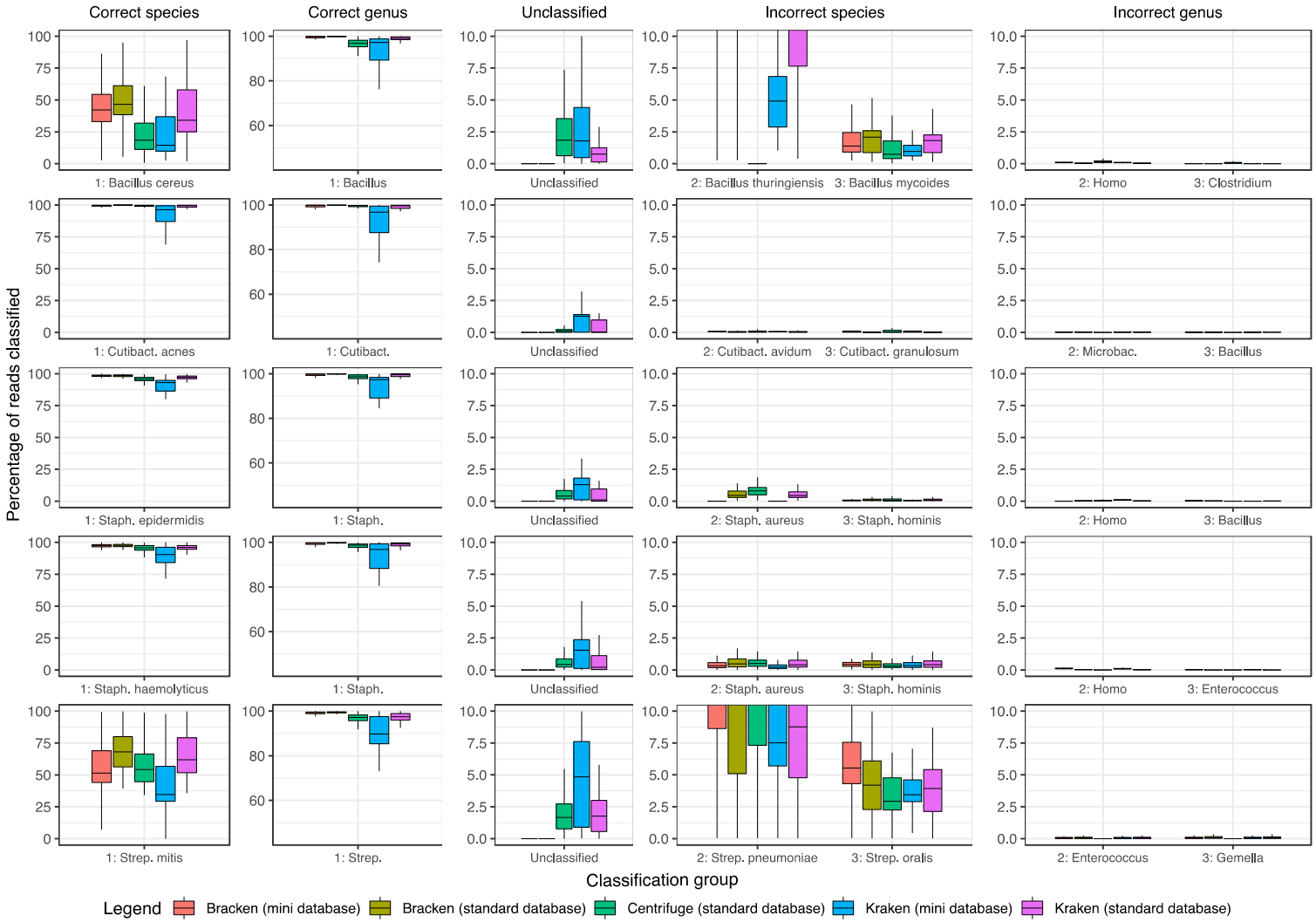

Figure 30. The percentage of reads with the correct and incorrect assigned species and genus, and the percentage of reads unclassified are shown. The percentage of reads with the correct and incorrect assigned species and genus, and the percentage of reads unclassified are shown.

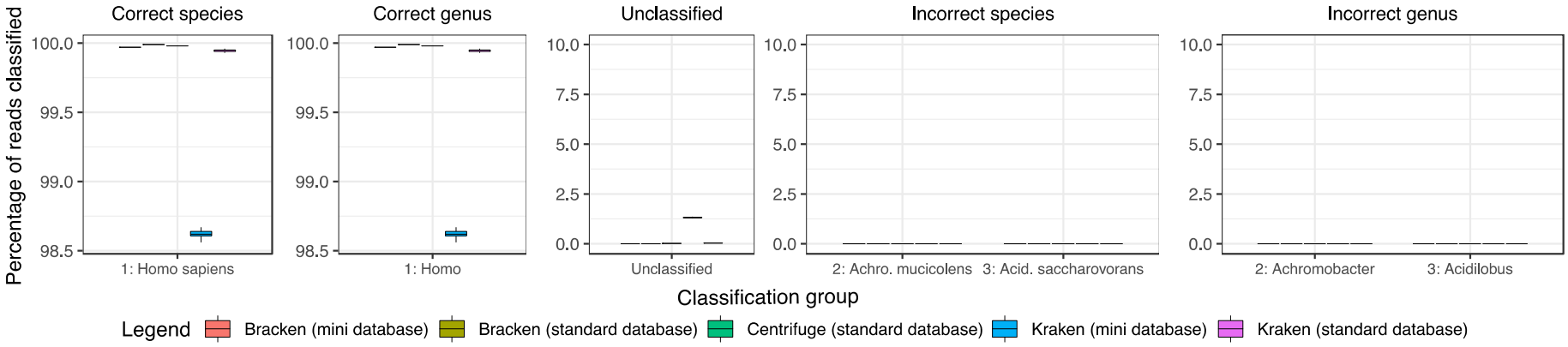

Figure 31. Percentage of reads classified from simulated human reads using Illumina sequencing. The percentage of reads with the correct and incorrect assigned species and genus, and the percentage of reads unclassified are shown.

5.3.3 Choice of sequencer

Figure 32 details correctly classified, unclassified and misclassified abundances for five common species according to sequencing technology and simulations with and without sequencing errors. Across all 1568 simulated genomes, the theoretical Nanopore error-free performed the best across species, genus, unclassified, misclassified species and misclassified genus abundance [IQR 95.0%:99.3%; IQR 98.5%:99.9%; IQR 0.0%:0.1%; IQR 0.0%:0.0% and IQR 0.0%:0.4% respectively] followed by Nanopore including simulated sequencing error [IQR 82.3%:97.6%; IQR 92.8%:99.3%; IQR 0%:1.8%; IQR 0.1%:1.9% and IQR 0%:0.4% respectively], Illumina error-free [IQR 77.0%:97.6%; IQR 95.4%:99.6%; IQR 0%:1.2%; IQR 0.1%:2.7% and IQR 0.0%:0.4% respectively] and Illumina with sequencing errors [IQR 77.0%:97.6%; IQR 95.4%:99.6%; IQR 0%:1.1%; IQR 0.1%:2.8% and IQR 0.0%:0.4% respectively].

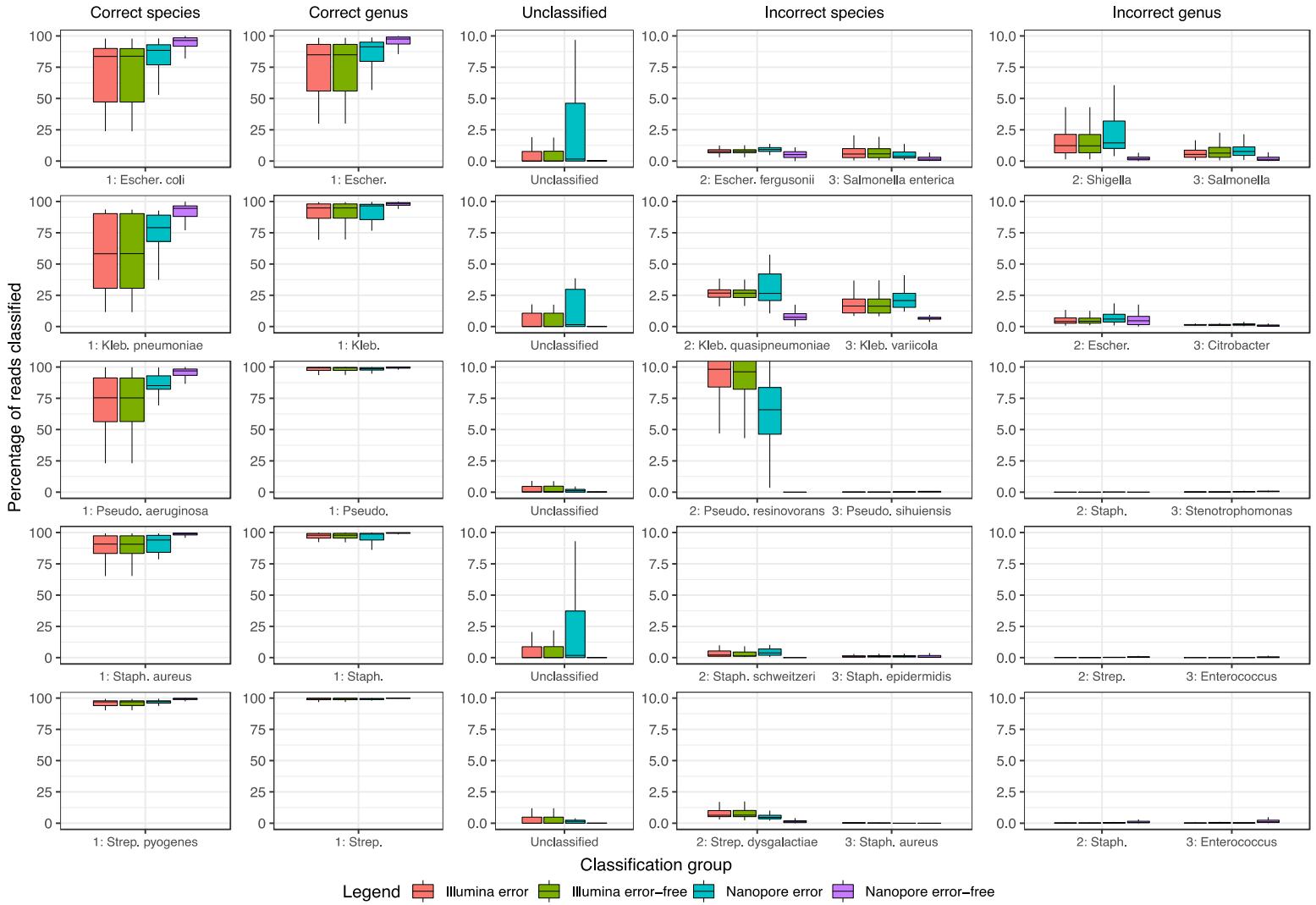

Figure 32. Percentage of reads classified correctly or incorrectly by sequencing device and error profile. The percentage of reads with correct and incorrect classification are shown as well as the percentage of unclassified reads.

5.3.4 Impact of read length.

Figure 33 shows read length vs. proportion of reads correctly classified for both Illumina and Oxford Nanopore sequence simulations, with and without simulated sequencing errors. Performance for *Staphylococcus aureus* and *Escherichia coli* was similar across Illumina error (Range 98.3%-99.1% and 86.7%-94.5% respectively) and error-free profiles (Range 98.3%-99.1% and 86.7%-94.6% respectively), however, differed for Nanopore error (Range 97.8%-99.9% and 83.5%-99.8% respectively) and error-free profiles (Range 99.5%-99.9% and 97.1%-99.5% respectively). Nanopore read lengths >8kb for *Staphylococcus aureus* and >15kb for *Escherichia coli* performed similarly regardless of error or error-free profile. For both *Escherichia coli* and *Staphylococcus aureus*, Nanopore error profile read lengths performed similarly to Illumina error-free profile by ~2000bp, and performed superiorly at higher read lengths >5000bp.

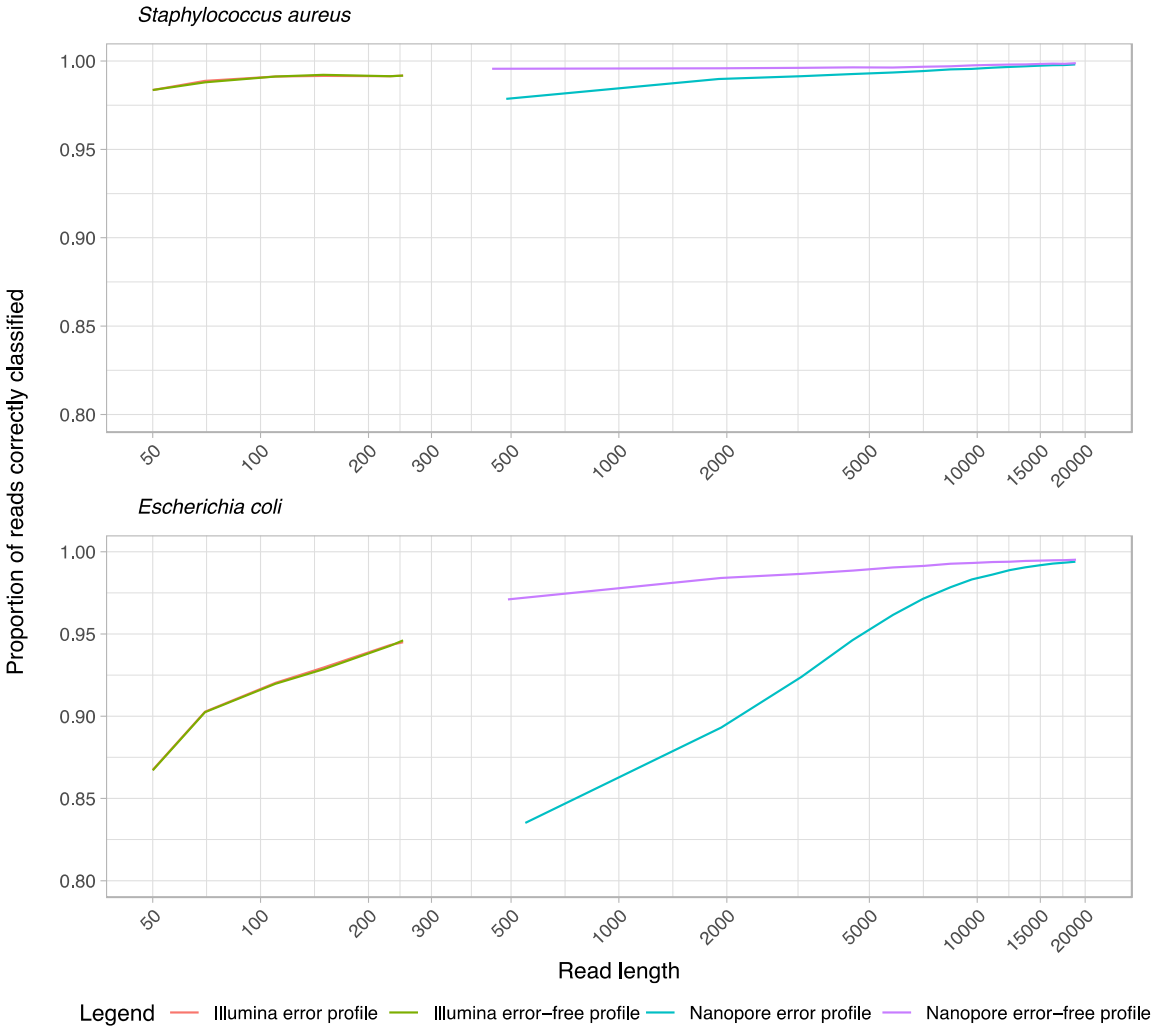

Figure 33. Read length vs. proportion of reads correctly classified. This figure shows the proportion of reads correctly classified against increasing read length in both Nanopore and Illumina sequencing with- and without simulated errors.

5.3.5 Common misclassifications

Merging data across all configurations of sequencer, classifier and database tested, the top misclassified species for *Escherichia coli* were *Escherichia fergusonii* (median across 100 reference genomes, 0.8%; IQR 0.7%:0.9%; 95th-99th percentile 2.5%:3.4%) and *Shigella flexneri* (median 0.8%; IQR 0.2%:1.2%; 95th-99th percentile 1.5%:1.6%), and for genus were *Shigella* (median 1.2%; IQR 0.6%:2.2%; 95th-99th percentile

4.8%:5.7%) and *Salmonella* (median 0.4%; IQR 0.2%:0.8%;95th-99th percentile 1.9%:3.1%) respectively. For *Staphylococcus aureus* the top misclassified species were *Staphylococcus schweitzeri* (median 0.2%; IQR 0.1%:0.5%; 95th-99th percentile 0.9%:1.0%) and *Pseudomonas aeruginosa* (median across 100 reference genomes, 0.0%; IQR 0.0%:0.0%; 95th-99th percentile 2.2%:3.6%), and for genus were *Pseudomonas* (median 0.0%; IQR 0.0%:0.0%; 95th-99th percentile 0.0%:2.7%) and *Streptococcus* (median 0.0%; IQR 0.0%:0.0%; 95th-99th percentile 0.1%:0.7%) respectively.

Of note misclassified species were more common for *Pseudomonas aeruginosa* which was classified as *Pseudomonas resinovorans* (median 8.8%; IQR 6.9%:10.7%; 95th-99th percentile 12.7%:15.1%), for *Streptococcus mitis* with misclassification as *Streptococcus pneumoniae* (median 9.8%; IQR 5.8%:15.4%; 95th-99th percentile 26.6%:31.3%), and for *Enterobacter cloacae* were *Enterobacter hormaechei* (median 2.5%; IQR 0.9%:54.8%; 95th-99th percentile 89.1%:93.8%) and *Enterobacter bugandensis* (median 0.1%; IQR 0%:0.3%; 95th-99th percentile 89.6%:95.6%) were identified instead, however, these all belonged to the same species group/complex. Further detailed information on individual species and genus misclassification can be found in **Supplementary Table 8**, while **Table 10** provides a summary with recommended 95th, 99th and 100th percentile cut-off value to determine if additional organisms are present, that is, there is polymicrobial infection.

Of the 1568 randomly selected genomes used in simulations, 7.6% overlapped with the references genomes used to build the classification database, however, when overlapping genomes were removed from this analysis this did not affect findings (data

are not shown separately as very similar). A higher sequencing depth of 40x compared to 20x did not result in any improvements in classification performance or reduction in misclassified or unclassified reads.

Table 10. Recommended threshold abundance to call additional true species. This summary table was generated by merging data across all configurations of sequencer, classifier and database tested. Detailed threshold values for specific sequencer, classifier and database combinations can be found in **Supplementary Table 8**.

Bacteria	**Correctly classified median (IQR)**	**Misclassified 1**	**95 -99th percentile**	**Misclassified 2**	**95-99th percentile**	**Misclassified 3**	**95 -99th percentile**	**Misclassified 4**	**95 -99th percentile**
Pathogens									
Bacteroides fragilis	95.4 (91.6:97.6)	*Bacteroides ovatus*	0.7:1.9	*Bacteroides xylanisolvens*	1:1.6	*Bacteroides caecimuris*	0.3:0.6	*Phocaeicola dorei*	0.7:1.4
Enterobacter cloacae	11.8 (3.3:92.5)	*Enterobacter hormaechei*	89.1:93.8	*Enterobacter bugandensis*	89.6:95.6	*Enterobacter roggenkampii*	7.2:81.6	*Enterobacter asburiae*	4.4:41.3
Enterococcus faecalis	98.2 (96.4:99.1)	*Enterococcus faecium*	0.9:1.9	*Enterococcus cecorum*	0.6:1.1	*Staphylococcus haemolyticus*	0:3	NA	NA
Escherichia coli	88.7 (61.2:94.2)	*Escherichia fergusonii*	2.5:3.4	*Shigella flexneri*	1.5:1.6	*Escherichia albertii*	1:1.2	*Escherichia marmotae*	0.6:0.8
Klebsiella pneumoniae	85.2 (55:91.5)	*Klebsiella quasipneumoniae*	4.9:5.4	*Klebsiella variicola*	3.2:3.5	*Klebsiella aerogenes*	1.9:2.7	*Escherichia coli*	1.5:2.6
Proteus mirabilis	97.2 (94.9:98.6)	*Providencia rettgeri*	1.6:1.8	*Proteus vulgaris*	0.4:6.9	*Proteus cibarius*	0.2:1.3	*Proteus terrae*	0.3:7.2
Pseudomonas aeruginosa	87.5 (71.4:95)	*Pseudomonas resinovorans*	12.7:15.1	*Pseudomonas stutzeri*	0.4:0.5	*Pseudomonas putida*	0.3:0.9	*Pseudomonas fluorescens*	0.8:1.2
Staphylococcus aureus	95.2 (88:98.5)	*Staphylococcus schweitzeri*	0.9:1	*Staphylococcus epidermidis*	0.5:0.8	*Pseudomonas aeruginosa*	2.2:3.6	*Staphylococcus pseudintermedius*	0:1.3
Streptococcus pneumoniae	97.3 (95.2:98.9)	*Streptococcus mitis*	1.8:2.8	*Streptococcus pseudopneumoniae*	1.4:2.3	*Streptococcus pyogenes*	0.1:0.4	*Homo sapiens*	0.1:1.8

Streptococcus pyogenes	97.6 (95.7:99)	*Streptococcus dysgalactiae*	1.3:1.9	*Streptococcus agalactiae*	0.6:1.1	*Filifactor alocis*	2.3:2.4	*Staphylococcus aureus*	0.2:0.7
Contaminants									
Bacillus thuringiensis	35.2 (18.5:42.3)	*Bacillus cereus*	82.8:95.3	*Bacillus mycoides*	48.2:89.4	*Bacillus albus*	4.4:9.9	*NA*	NA
Cutibacterium acnes	99.5 (98.7:99.8)	*Cutibacterium modestum*	0.4:0.7	*Cutibacterium avidum*	0.4:0.8	*Homo sapiens*	0.1:0.3	*Staphylococcus aureus*	0:0.7
Micrococcus luteus	89.5 (59.5:96.3)	*Brevibacterium luteolum*	0.5:0.5	*Kytococcus sedentarius*	0.7:0.8	*Pseudarthrobacter sulfonivorans*	0.1:12	*Rothia kristinae*	0.6:1
Staphylococcus epidermidis	97.2 (94.7:98.4)	*Staphylococcus aureus*	1.6:1.8	*Staphylococcus hominis*	0.3:62.5	*Staphylococcus warneri*	0.2:1.6	*Staphylococcus pasteuri*	1.1:6.1
Staphylococcus haemolyticus	96.5 (94.2:97.8)	*Staphylococcus aureus*	2.1:3.8	*Staphylococcus hominis*	1:2.1	*Staphylococcus epidermidis*	1:1.9	NA	NA
Streptococcus mitis	62.3 (42.7:75.2)	*Streptococcus pneumoniae*	26.6:31.3	*Streptococcus oralis*	84.1:96.3	NA	NA	NA	NA

5.3.6 Relationship between species misclassification and nucleotide identity

Misclassification at the species level was more common when the misclassified species had a higher average nucleotide identity to the true species (**Figure 34** and **Table 11**). However, there were several outliers, where misclassification was greater than expected for the given level of genome-wide nucleotide identity, for example *Bacillus thuringiensis* was misclassified at 25% when *Bacillus cereus* was the true species, and *Streptococcus pneumoniae* at 10% when *Streptococcus mitis* was the true species. This reflects the difficulty distinguishing species within known groups or complexes of bacteria, and potentially reflects that some parts of the genome may have higher nucleotide identity.

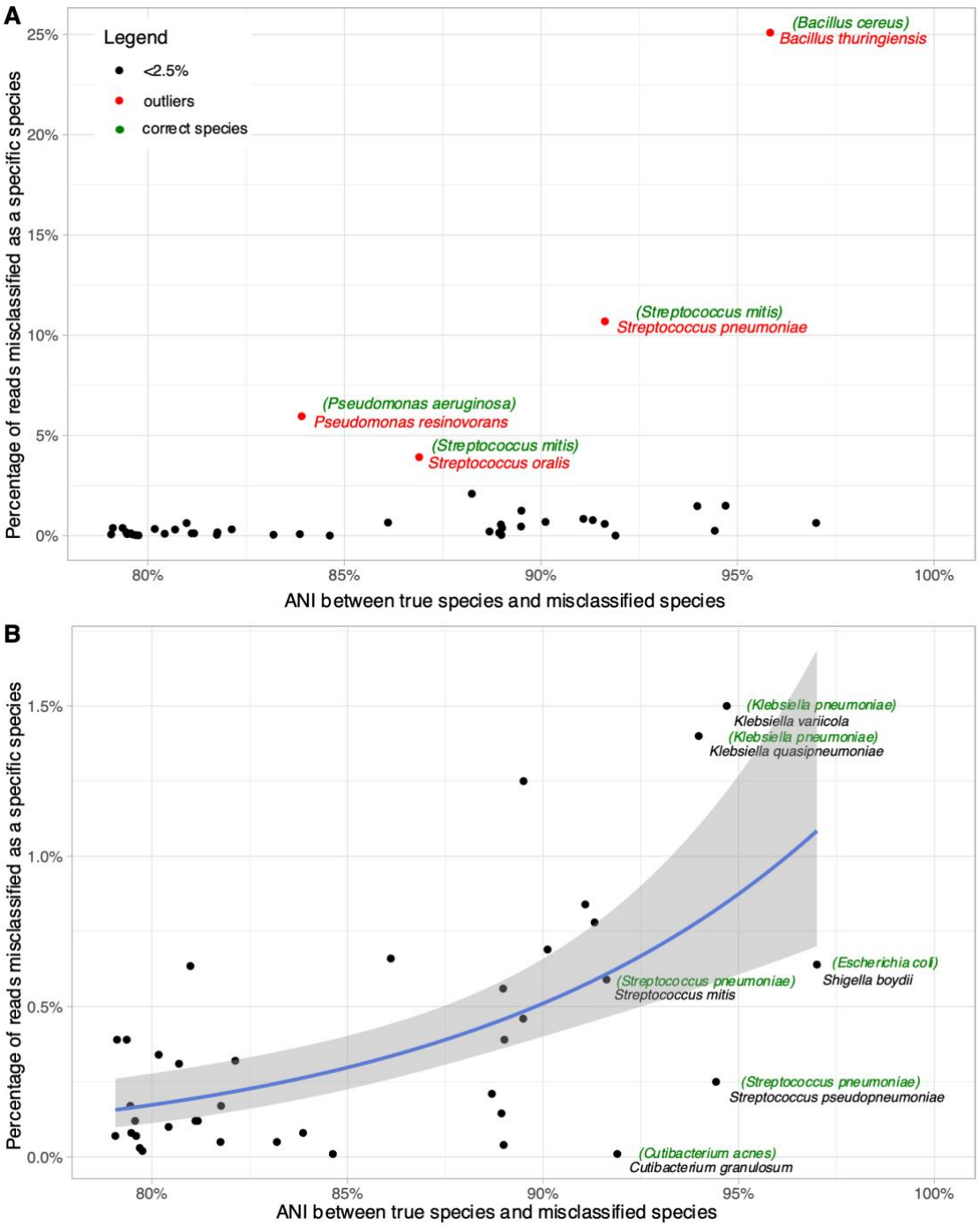

Figure 34. ANI between true species and misclassified species against percentage of reads misclassified as a specific species. Panel A shows all pairs, and panel B the same plot with the outliers from panel A (shown in red, with the correct species shown in green inside brackets) excluded. For panel B a generalized linear regression line (with a logit link function and binomial family) is plotted (the shaded area indicates the 95% confidence interval). The misclassified species corresponding to example outlying points are shown in black, with the corresponding correct species shown in green in brackets.

Table 11. ANI between true species and misclassified species against percentage of reads misclassified as a specific species.

Misclassified species	Percentage of reads misclassified as a specific species	True species	ANI between true species and misclassified species
Bacillus mycoides	0.0125	*Bacillus cereus*	89.5002
Bacteroides xylanisolvens	8.00E-04	*Bacteroides fragilis*	79.47315
Cutibacterium avidum	5.00E-04	*Cutibacterium acnes*	81.7536
Enterobacter bugandensis	0.00145	*Enterobacter cloacae*	88.9361
Enterococcus faecium	0.0017	*Enterococcus faecalis*	81.76745
Escherichia albertii	0.0069	*Escherichia coli*	90.11415
Escherichia coli	0.00635	*Klebsiella pneumoniae*	80.98335
Proteus vulgaris	8.00E-04	*Proteus mirabilis*	83.8642
Staphylococcus argenteus	0.0021	*Staphylococcus aureus*	88.6904
Staphylococcus aureus	0.0039	*Staphylococcus epidermidis*	79.10885
Staphylococcus aureus	0.0039	*Staphylococcus haemolyticus*	79.3562
Streptococcus dysgalactiae	0.0046	*Streptococcus pyogenes*	89.49165
Streptococcus mitis	0.0059	*Streptococcus pneumoniae*	91.62065
Cutibacterium granulosum	1.00E-04	*Cutibacterium acnes*	91.89555
Escherichia fergusonii	0.0084	*Escherichia coli*	91.07865
Klebsiella quasipneumoniae	0.014	*Klebsiella pneumoniae*	93.9773
Pseudomonas putida	0.0034	*Pseudomonas aeruginosa*	80.17345
Staphylococcus epidermidis	7.00E-04	*Staphylococcus aureus*	79.0649
Staphylococcus epidermidis	7.00E-04	*Staphylococcus haemolyticus*	79.5979
Staphylococcus pasteuri	2.00E-04	*Staphylococcus epidermidis*	79.75925
Streptococcus agalactiae	5.00E-04	*Streptococcus pyogenes*	83.19475
Streptococcus pseudopneumoniae	0.0025	*Streptococcus pneumoniae*	94.4205
Bacillus albus	0.0078	*Bacillus cereus*	91.31685
Bacteroides ovatus	0.0017	*Bacteroides fragilis*	79.4477
Enterobacter roggenkampii	0.0039	*Enterobacter cloacae*	89.01105
Homo sapiens	4.00E-04	*Cutibacterium acnes*	88.99195
Klebsiella variicola	0.015	*Klebsiella pneumoniae*	94.69545
Pseudomonas stutzeri	0.0031	*Pseudomonas aeruginosa*	80.6919
Salmonella enterica	0.0032	*Escherichia coli*	82.131
Staphylococcus hominis	0.0012	*Staphylococcus epidermidis*	79.57025
Staphylococcus hominis	0.0012	*Staphylococcus haemolyticus*	81.1747
Streptococcus australis	0.0012	*Streptococcus mitis*	81.11325
Bacteroides uniformis	0.001	*Bacteroides fragilis*	80.4265
Enterobacter asburiae	0.0056	*Enterobacter cloacae*	88.97905
Klebsiella aerogenes	0.0066	*Klebsiella pneumoniae*	86.10775
Providencia stuartii	7.00E-04	*Proteus mirabilis*	78.2895
Pseudomonas citronellolis	1.00E-04	*Pseudomonas aeruginosa*	84.62615
Rothia kristinae	1.00E-04	*Micrococcus luteus*	78.8284
Staphylococcus pseudintermedius	1.00E-04	*Staphylococcus aureus*	78.3567
Staphylococcus warneri	3.00E-04	*Staphylococcus epidermidis*	79.68865

5.3.7 Comparison with real-world data

Findings from Illumina simulations (using MiSeq error profiles) were very similar to real-world Illumina data (generated with the HiSeq platform, **Figure 35**). Real-world data from *Escherichia coli*, *Klebsiella pneumoniae* and *Staphylococcus aureus* performed better than simulated data both when using Kraken2 mini database (percentage of reads classified as the correct species: 46.2% vs 28.7%; 31.9% vs 19.1%; 89.3% vs 81.4% respectively) and Kraken2 standard database (84.2% vs 67.1%; 57.4% vs 53.5%; 86.1% vs 96.6% respectively). Similar bioinformatic misclassification existed in both real-world and simulated data. Misclassification in *Klebsiella pneumoniae* for example as *Klebsiella quasipneumoniae* was marginally higher in simulated data vs. real-world data in Bracken mini database (2.8% vs 2.6%), Bracken standard database (3.2% vs 2.7%) and Kraken2 standard database (2.2% vs 2.4%).

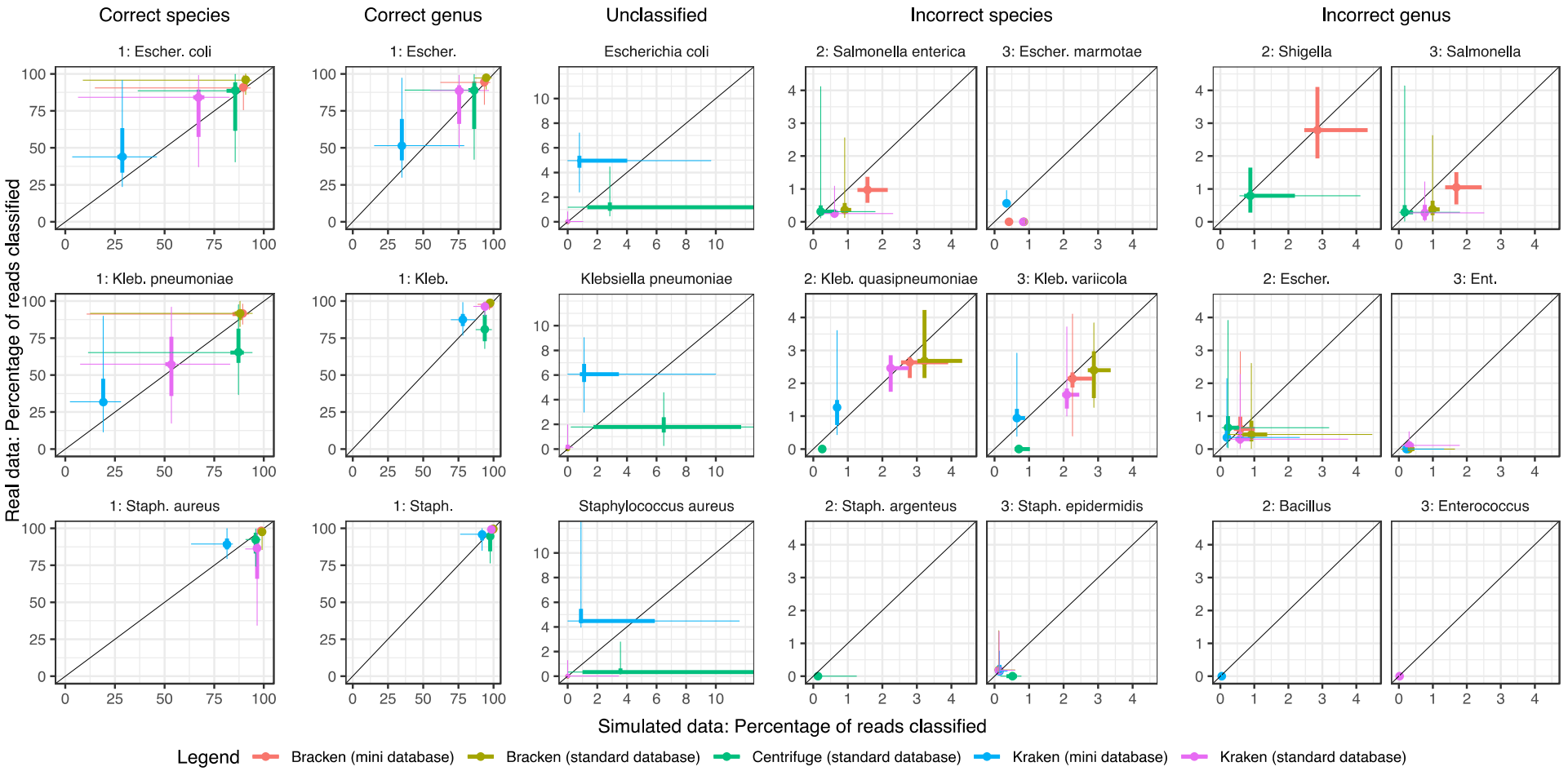

Figure 35. Percentage of reads classified correctly or incorrectly in Illumina simulated data and real-world data. The diagonal line indicates values with exact correspondence between the simulated a real-world data.

5.4 Discussion

The potential for metagenomic technology to improve diagnostics and patient outcomes has been widely recognized,[2] and there is extensive work ongoing to overcome current challenges associated with use of these approaches in clinical settings.[95] This study simulates and use real-world sequencing data to define the intrinsic rates of bioinformatic species misclassification that occur using common sequencing platforms, classification tools and databases. This enables us to quantify the extent of bioinformatic contamination in the classification process to account for this formally in diagnostic workflows, and further guide the selection of optimal bioinformatic tools to maximize classification accuracy.

Overall, across both short and long reads this study found that using the Bracken classifier with a "standard" database, that is, larger resulted in the most accurate performance across all species with almost no reads unclassified (**Figure 29** and **Figure 30**). In comparison, Kraken2 with a mini database, a commonly used combination performed worse. Previous studies with Illumina short reads have reported estimates from Kraken2 to be more accurate than Kraken1, KrakenUniq, CLARK and Centrifuge at both the genus and species levels.[72] Furthermore, studies have shown with Illumina data that Bracken which re-assigns reads in intermediate taxonomic nodes to the species level or above from Kraken1/2 outputs produce far better abundance estimates.[133,139] These results confirm that Bracken produces far more accurate abundance estimates for both long- and short- reads, especially when using a standard database. Previous metagenomic studies have rarely used Bracken or standard, that is, larger databases[64], the latter in part because of higher memory requirements, therefore these results

encourage future studies to utilise the most accurate tools available where computational resources permit.

To formally account for bioinformatic contamination in clinical diagnostic workflows, this study produced the first known catalog of misclassified species from simulating 1568 genomes representing 10 common pathogens and 6 common blood culture contaminants using different classification tools, databases, and sequencers. It should be noted that the proposed minimum cut-off thresholds for identifying the presence of additional species are based on the extent of bioinformatic contamination arising from taxonomic misclassification only. Other forms of contamination are possible during sampling and laboratory workflows which also need to be considered. These thresholds can be applied to metagenomic samples by multiplying the estimated proportion of a sample that is from a specific species by the 95%, 99% or 100% misclassification cut-off values in **Table 10**. For instance, if a sample by read proportion is found to have 50% *Streptococcus mitis* and 40% *Staphylococcus epidermidis*, then 95% threshold for the proportion of reads misclassified as *Streptococcus pneumoniae* would be 50% * 26.6% + 40% * 0% = 13.3%. Therefore, even if the remaining 10% of the sample were classified as *Streptococcus pneumoniae* this would be considered a bioinformatic contaminant, while *Streptococcus mitis* and *Staphylococcus epidermidis* would not as they both are above their respective modified cut-off values. Refinements to this heuristic approach are possible, for example probabilistically estimating the proportion of each sample made up of a given species while simultaneously accounting for misclassification.

In this chapter I used existing real-world Illumina data to show correlation with the Illumina simulator (**Figure 35**), while the Nanopore simulator used has been benchmarked extensively elsewhere.[129] Detailed 95th, 99th and 100th percentile cut-off values for species may be used to determine if there is strong evidence that additional organisms are present, that is, there is polymicrobial infection or, conversely, if only bioinformatic contamination is present. Using simulations, I show that misclassification of human reads is negligible when human DNA contamination might exist in clinical samples.

Additionally, these results show that misclassification at the species level was more common when the misclassified species had a higher average nucleotide identity to the true species. However, there were several outliers where misclassification was greater than expected for the given level of genome-wide nucleotide identity reflecting the difficulty distinguishing species within known groups or complexes of bacteria. Tools such as DeepMicrobes which use deep learning-based computational framework for taxonomic classification may offer an advantage in these situations where far fewer false positives are produced compared to other tools regardless of different degrees of nucleotide similarity.[159]

The advent of long read technologies such as the Oxford Nanopore has significant advantages for metagenomic analysis from the detection of structural variants to improved de novo assembly.[160] Furthermore, long-read sequencing of native molecules eliminate amplification bias while preserving base modifications.[161] However, these technologies are limited by a higher error-rate affecting 6-12% of all sequenced bases, with a significant fraction of insertion and deletion errors.[162] This may affect the success

of current classification methods, as there are few algorithms developed to exploit long-read data. However, even after simulating sequencing errors, the longer reads of Nanopore sequencing meant that it still could outperform Illumina sequencing. This finding is supported by previous work.[163] Following on from this observation this study directly investigated the impact of read length on the proportion of reads correctly classified. Allowing for sequencing error, for *E. coli* and *S. aureus*, Nanopore reads of ~2000bp performed similarly to Illumina reads and performed better at higher read lengths >5000bp. These findings have several implications: As technology and accuracy for long-read sequencing advances, we expect classification accuracy to significantly improve. Furthermore, filtering shorter reads (at least <1000bp) from error-prone Nanopore sequencing can provide performance either equivalent or superior to standard Illumina sequencing.

Although this study has thoroughly analyzed and discussed technical factors that correlate with taxonomic misclassification, biological factors that might be responsible for misclassification might also exist. One such biological aspect that this study has addressed here is the presence of the complexes or species which are closely related yet named differently which can result in the erroneous classification of sequence reads. Other biological factors might include the contribution of high similar genomic regions that are not specific to a particular species but shared across two or more species. These regions are well known to be prevalent in bacterial pathogens (e.g., in mobile genetic elements) and will affect read misclassification rates even if the sequencing reads are error-free. Future studies are required that ascertain how much bioinformatic contamination may be removed by building classifier databases from modified genome sequence databases, which contain exclusively the relatively "stable" genomic backbone regions after excluding regions shared among species.

In conclusion, these findings highlight the pertinent issue of the presence of bioinformatic contamination and how this varies by the combination of sequencer, classifier and database used. While this study recommends Bracken using a standard database to be utilized in metagenomic workflows, it has also produced a misclassification catalogue whereby misclassification can be accounted for using any combination of sequencers, classifiers or databases benchmarked in this study. Benchmarking metrics should inform the choice and implementation of metagenomic sequencing.

Supplementary Table 7. Predicted abundance classification for all included species.

	Species	Sequencer	Error profile	Database	Classifier	Predicted performance (95% CI)
1	*Staphylococcus aureus*	Illumina	Error	Mini	Kraken	78.2 (76.9:79.5)
2	*Staphylococcus aureus*	Illumina	Error-free	Mini	Kraken	78.2 (76.9:79.5)
3	*Staphylococcus aureus*	Illumina	Error	Mini	Centrifuge	80.3 (79:81.6)
4	*Staphylococcus aureus*	Illumina	Error-free	Mini	Centrifuge	80.3 (79:81.6)
5	*Staphylococcus aureus*	Nanopore	Error	Mini	Kraken	81.9 (80.8:83)
6	*Staphylococcus aureus*	Nanopore	Error	Mini	Centrifuge	83.7 (82.6:84.9)
7	*Staphylococcus aureus*	Illumina	Error	Standard	Kraken	87.3 (86.4:88.1)
8	*Staphylococcus aureus*	Illumina	Error-free	Standard	Kraken	87.3 (86.4:88.1)
9	*Staphylococcus aureus*	Illumina	Error	Standard	Centrifuge	88.6 (87.8:89.4)
10	*Staphylococcus aureus*	Illumina	Error-free	Standard	Centrifuge	88.6 (87.8:89.4)
11	*Staphylococcus aureus*	Nanopore	Error	Standard	Kraken	89.6 (88.9:90.3)
12	*Staphylococcus aureus*	Nanopore	Error	Standard	Centrifuge	90.8 (90.1:91.4)
13	*Staphylococcus aureus*	Nanopore	Error-free	Mini	Kraken	91.5 (90.9:92.1)
14	*Staphylococcus aureus*	Illumina	Error	Mini	Bracken	92.1 (91.6:92.7)
15	*Staphylococcus aureus*	Illumina	Error-free	Mini	Bracken	92.1 (91.6:92.7)
16	*Staphylococcus aureus*	Nanopore	Error-free	Mini	Centrifuge	92.5 (91.8:93)
17	*Staphylococcus aureus*	Nanopore	Error	Mini	Bracken	93.7 (93.2:94.1)
18	*Staphylococcus aureus*	Nanopore	Error-free	Standard	Kraken	95.4 (95:95.7)
19	*Staphylococcus aureus*	Illumina	Error	Standard	Bracken	95.7 (95.4:96)
20	*Staphylococcus aureus*	Illumina	Error-free	Standard	Bracken	95.7 (95.4:96)
21	*Staphylococcus aureus*	Nanopore	Error-free	Standard	Centrifuge	95.9 (95.6:96.2)
22	*Staphylococcus aureus*	Nanopore	Error	Standard	Bracken	96.6 (96.3:96.8)
23	*Staphylococcus aureus*	Nanopore	Error-free	Mini	Bracken	97.2 (97:97.4)
24	*Staphylococcus aureus*	Nanopore	Error-free	Standard	Bracken	98.5 (98.4:98.6)
25	*Escherichia coli*	Illumina	Error	Mini	Kraken	56.1 (54.7:57.5)
26	*Escherichia coli*	Illumina	Error-free	Mini	Kraken	56.1 (54.7:57.6)
27	*Escherichia coli*	Illumina	Error	Mini	Centrifuge	59.3 (57.6:60.9)
28	*Escherichia coli*	Illumina	Error-free	Mini	Centrifuge	59.3 (57.6:61)
29	*Escherichia coli*	Nanopore	Error	Mini	Kraken	61.7 (60.3:63.1)
30	*Escherichia coli*	Nanopore	Error	Mini	Centrifuge	64.7 (63.1:66.3)
31	*Escherichia coli*	Illumina	Error	Standard	Kraken	70.9 (69.7:72.1)
32	*Escherichia coli*	Illumina	Error-free	Standard	Kraken	70.9 (69.7:72.1)
33	*Escherichia coli*	Illumina	Error	Standard	Centrifuge	73.5 (72.3:74.7)
34	*Escherichia coli*	Illumina	Error-free	Standard	Centrifuge	73.5 (72.3:74.7)
35	*Escherichia coli*	Nanopore	Error	Standard	Kraken	75.4 (74.3:76.5)
36	*Escherichia coli*	Nanopore	Error	Standard	Centrifuge	77.7 (76.6:78.8)
37	*Escherichia coli*	Nanopore	Error-free	Mini	Kraken	79.3 (78.3:80.3)

Supplementary Table 7. Continued.

38	*Escherichia coli*	Illumina	Error	Mini	Bracken	80.7 (79.7:81.6)
39	*Escherichia coli*	Illumina	Error-free	Mini	Bracken	80.7 (79.7:81.6)
40	*Escherichia coli*	Nanopore	Error-free	Mini	Centrifuge	81.3 (80.2:82.4)
41	*Escherichia coli*	Nanopore	Error	Mini	Bracken	84 (83.2:84.8)
42	*Escherichia coli*	Nanopore	Error-free	Standard	Kraken	88 (87.3:88.6)
43	*Escherichia coli*	Illumina	Error	Standard	Bracken	88.8 (88.2:89.4)
44	*Escherichia coli*	Illumina	Error-free	Standard	Bracken	88.8 (88.2:89.4)
45	*Escherichia coli*	Nanopore	Error-free	Standard	Centrifuge	89.3 (88.6:89.9)
46	*Escherichia coli*	Nanopore	Error	Standard	Bracken	90.9 (90.4:91.4)
47	*Escherichia coli*	Nanopore	Error-free	Mini	Bracken	92.6 (92.1:93)
48	*Escherichia coli*	Nanopore	Error-free	Standard	Bracken	96 (95.7:96.2)
49	*Enterococcus faecalis*	Illumina	Error	Mini	Kraken	92.5 (91.7:93.3)
50	*Enterococcus faecalis*	Illumina	Error-free	Mini	Kraken	92.5 (91.7:93.3)
51	*Enterococcus faecalis*	Illumina	Error	Mini	Centrifuge	93.4 (92.6:94.1)
52	*Enterococcus faecalis*	Illumina	Error-free	Mini	Centrifuge	93.4 (92.6:94.1)
53	*Enterococcus faecalis*	Nanopore	Error	Mini	Kraken	94 (93.3:94.6)
54	*Enterococcus faecalis*	Nanopore	Error	Mini	Centrifuge	94.7 (94:95.3)
55	*Enterococcus faecalis*	Illumina	Error	Standard	Kraken	95.9 (95.4:96.4)
56	*Enterococcus faecalis*	Illumina	Error-free	Standard	Kraken	95.9 (95.4:96.4)
57	*Enterococcus faecalis*	Illumina	Error	Standard	Centrifuge	96.4 (96:96.8)
58	*Enterococcus faecalis*	Illumina	Error-free	Standard	Centrifuge	96.4 (96:96.8)
59	*Enterococcus faecalis*	Nanopore	Error	Standard	Kraken	96.7 (96.3:97.1)
60	*Enterococcus faecalis*	Nanopore	Error	Standard	Centrifuge	97.1 (96.8:97.4)
61	*Enterococcus faecalis*	Nanopore	Error-free	Mini	Kraken	97.4 (97:97.7)
62	*Enterococcus faecalis*	Illumina	Error	Mini	Bracken	97.6 (97.3:97.9)
63	*Enterococcus faecalis*	Illumina	Error-free	Mini	Bracken	97.6 (97.3:97.9)
64	*Enterococcus faecalis*	Nanopore	Error-free	Mini	Centrifuge	97.7 (97.4:98)
65	*Enterococcus faecalis*	Nanopore	Error	Mini	Bracken	98.1 (97.8:98.3)
66	*Enterococcus faecalis*	Nanopore	Error-free	Standard	Kraken	98.6 (98.4:98.8)
67	*Enterococcus faecalis*	Illumina	Error	Standard	Bracken	98.7 (98.6:98.9)
68	*Enterococcus faecalis*	Illumina	Error-free	Standard	Bracken	98.7 (98.6:98.9)
69	*Enterococcus faecalis*	Nanopore	Error-free	Standard	Centrifuge	98.8 (98.6:98.9)
70	*Enterococcus faecalis*	Nanopore	Error	Standard	Bracken	99 (98.8:99.1)
71	*Enterococcus faecalis*	Nanopore	Error-free	Mini	Bracken	99.2 (99.1:99.3)
72	*Enterococcus faecalis*	Nanopore	Error-free	Standard	Bracken	99.6 (99.5:99.6)
73	*Bacteroides fragilis*	Illumina	Error	Mini	Kraken	81.9 (80.7:83)
74	*Bacteroides fragilis*	Illumina	Error-free	Mini	Kraken	81.9 (80.7:83)
75	*Bacteroides fragilis*	Illumina	Error	Mini	Centrifuge	83.7 (82.5:84.9)
76	*Bacteroides fragilis*	Illumina	Error-free	Mini	Centrifuge	83.7 (82.5:84.9)
77	*Bacteroides fragilis*	Nanopore	Error	Mini	Kraken	85.1 (84:86.1)
78	*Bacteroides fragilis*	Nanopore	Error	Mini	Centrifuge	86.6 (85.5:87.6)

Supplementary Table 7. Continued.

79	*Bacteroides fragilis*	Illumina	Error	Standard	Kraken	89.6 (88.8:90.3)
80	*Bacteroides fragilis*	Illumina	Error-free	Standard	Kraken	89.6 (88.8:90.3)
81	*Bacteroides fragilis*	Illumina	Error	Standard	Centrifuge	90.7 (90:91.4)
82	*Bacteroides fragilis*	Illumina	Error-free	Standard	Centrifuge	90.7 (90:91.4)
83	*Bacteroides fragilis*	Nanopore	Error	Standard	Kraken	91.6 (90.9:92.2)
84	*Bacteroides fragilis*	Nanopore	Error	Standard	Centrifuge	92.5 (91.9:93.1)
85	*Bacteroides fragilis*	Nanopore	Error-free	Mini	Kraken	93.1 (92.6:93.6)
86	*Bacteroides fragilis*	Illumina	Error	Mini	Bracken	93.6 (93.1:94.1)
87	*Bacteroides fragilis*	Illumina	Error-free	Mini	Bracken	93.6 (93.1:94.1)
88	*Bacteroides fragilis*	Nanopore	Error-free	Mini	Centrifuge	93.9 (93.4:94.4)
89	*Bacteroides fragilis*	Nanopore	Error	Mini	Bracken	94.9 (94.5:95.3)
90	*Bacteroides fragilis*	Nanopore	Error-free	Standard	Kraken	96.3 (96:96.6)
91	*Bacteroides fragilis*	Illumina	Error	Standard	Bracken	96.6 (96.3:96.8)
92	*Bacteroides fragilis*	Illumina	Error-free	Standard	Bracken	96.6 (96.3:96.8)
93	*Bacteroides fragilis*	Nanopore	Error-free	Standard	Centrifuge	96.7 (96.4:97)
94	*Bacteroides fragilis*	Nanopore	Error	Standard	Bracken	97.3 (97:97.5)
95	*Bacteroides fragilis*	Nanopore	Error-free	Mini	Bracken	97.8 (97.6:98)
96	*Bacteroides fragilis*	Nanopore	Error-free	Standard	Bracken	98.8 (98.7:98.9)
97	*Enterobacter cloacae*	Illumina	Error	Mini	Kraken	16.6 (15.8:17.4)
98	*Enterobacter cloacae*	Illumina	Error-free	Mini	Kraken	16.6 (15.8:17.4)
99	*Enterobacter cloacae*	Illumina	Error	Mini	Centrifuge	18.5 (17.5:19.5)
100	*Enterobacter cloacae*	Illumina	Error-free	Mini	Centrifuge	18.5 (17.5:19.5)
101	*Enterobacter cloacae*	Nanopore	Error	Mini	Kraken	20 (19.2:21)
102	*Enterobacter cloacae*	Nanopore	Error	Mini	Centrifuge	22.2 (21:23.4)
103	*Enterobacter cloacae*	Illumina	Error	Standard	Kraken	27.5 (26.4:28.6)
104	*Enterobacter cloacae*	Illumina	Error-free	Standard	Kraken	27.5 (26.4:28.6)
105	*Enterobacter cloacae*	Illumina	Error	Standard	Centrifuge	30.1 (29:31.4)
106	*Enterobacter cloacae*	Illumina	Error-free	Standard	Centrifuge	30.2 (29:31.4)
107	*Enterobacter cloacae*	Nanopore	Error	Standard	Kraken	32.3 (31.1:33.5)
108	*Enterobacter cloacae*	Nanopore	Error	Standard	Centrifuge	35.2 (33.9:36.5)
109	*Enterobacter cloacae*	Nanopore	Error-free	Mini	Kraken	37.4 (36:38.7)
110	*Enterobacter cloacae*	Illumina	Error	Mini	Bracken	39.4 (38.1:40.7)
111	*Enterobacter cloacae*	Illumina	Error-free	Mini	Bracken	39.4 (38.1:40.7)
112	*Enterobacter cloacae*	Nanopore	Error-free	Mini	Centrifuge	40.4 (38.8:42.1)
113	*Enterobacter cloacae*	Nanopore	Error	Mini	Bracken	45 (43.6:46.3)
114	*Enterobacter cloacae*	Nanopore	Error-free	Standard	Kraken	53.2 (51.8:54.6)
115	*Enterobacter cloacae*	Illumina	Error	Standard	Bracken	55.3 (53.9:56.6)
116	*Enterobacter cloacae*	Illumina	Error-free	Standard	Bracken	55.3 (53.9:56.6)
117	*Enterobacter cloacae*	Nanopore	Error-free	Standard	Centrifuge	56.4 (55:57.8)
118	*Enterobacter cloacae*	Nanopore	Error	Standard	Bracken	60.9 (59.6:62.2)
119	*Enterobacter cloacae*	Nanopore	Error-free	Mini	Bracken	66 (64.8:67.3)
120	*Enterobacter cloacae*	Nanopore	Error-free	Standard	Bracken	78.8 (77.8:79.7)

Supplementary Table 7. Continued.

121	*Klebsiella pneumoniae*	Illumina	Error	Mini	Kraken	45.7 (44.4:47.1)
122	*Klebsiella pneumoniae*	Illumina	Error-free	Mini	Kraken	45.7 (44.4:47.1)
123	*Klebsiella pneumoniae*	Illumina	Error	Mini	Centrifuge	49 (47.3:50.7)
124	*Klebsiella pneumoniae*	Illumina	Error-free	Mini	Centrifuge	49 (47.3:50.7)
125	*Klebsiella pneumoniae*	Nanopore	Error	Mini	Kraken	51.5 (50.1:52.9)
126	*Klebsiella pneumoniae*	Nanopore	Error	Mini	Centrifuge	54.7 (53:56.4)
127	*Klebsiella pneumoniae*	Illumina	Error	Standard	Kraken	61.6 (60.3:63)
128	*Klebsiella pneumoniae*	Illumina	Error-free	Standard	Kraken	61.6 (60.3:63)
129	*Klebsiella pneumoniae*	Illumina	Error	Standard	Centrifuge	64.6 (63.3:66)
130	*Klebsiella pneumoniae*	Illumina	Error-free	Standard	Centrifuge	64.6 (63.3:66)
131	*Klebsiella pneumoniae*	Nanopore	Error	Standard	Kraken	66.9 (65.7:68.2)
132	*Klebsiella pneumoniae*	Nanopore	Error	Standard	Centrifuge	69.7 (68.5:71)
133	*Klebsiella pneumoniae*	Nanopore	Error-free	Mini	Kraken	71.6 (70.4:72.8)
134	*Klebsiella pneumoniae*	Illumina	Error	Mini	Bracken	73.3 (72.2:74.4)
135	*Klebsiella pneumoniae*	Illumina	Error-free	Mini	Bracken	73.3 (72.2:74.4)
136	*Klebsiella pneumoniae*	Nanopore	Error-free	Mini	Centrifuge	74.2 (72.8:75.5)
137	*Klebsiella pneumoniae*	Nanopore	Error	Mini	Bracken	77.6 (76.6:78.6)
138	*Klebsiella pneumoniae*	Nanopore	Error-free	Standard	Kraken	82.8 (81.9:83.6)
139	*Klebsiella pneumoniae*	Illumina	Error	Standard	Bracken	84 (83.2:84.7)
140	*Klebsiella pneumoniae*	Illumina	Error-free	Standard	Bracken	84 (83.2:84.7)
141	*Klebsiella pneumoniae*	Nanopore	Error-free	Standard	Centrifuge	84.6 (83.7:85.4)
142	*Klebsiella pneumoniae*	Nanopore	Error	Standard	Bracken	86.8 (86.1:87.5)
143	*Klebsiella pneumoniae*	Nanopore	Error-free	Mini	Bracken	89.2 (88.6:89.8)
144	*Klebsiella pneumoniae*	Nanopore	Error-free	Standard	Bracken	94 (93.6:94.4)
145	*Proteus mirabilis*	Illumina	Error	Mini	Kraken	87.9 (86.9:88.9)
146	*Proteus mirabilis*	Illumina	Error-free	Mini	Kraken	87.9 (86.9:88.9)
147	*Proteus mirabilis*	Illumina	Error	Mini	Centrifuge	89.2 (88.2:90.2)
148	*Proteus mirabilis*	Illumina	Error-free	Mini	Centrifuge	89.2 (88.2:90.2)
149	*Proteus mirabilis*	Nanopore	Error	Mini	Kraken	90.2 (89.3:91)
150	*Proteus mirabilis*	Nanopore	Error	Mini	Centrifuge	91.3 (90.4:92.1)
151	*Proteus mirabilis*	Illumina	Error	Standard	Kraken	93.3 (92.7:93.9)
152	*Proteus mirabilis*	Illumina	Error-free	Standard	Kraken	93.3 (92.7:93.9)
153	*Proteus mirabilis*	Illumina	Error	Standard	Centrifuge	94.1 (93.5:94.6)
154	*Proteus mirabilis*	Illumina	Error-free	Standard	Centrifuge	94.1 (93.5:94.6)
155	*Proteus mirabilis*	Nanopore	Error	Standard	Kraken	94.6 (94.1:95.1)
156	*Proteus mirabilis*	Nanopore	Error	Standard	Centrifuge	95.2 (94.8:95.7)
157	*Proteus mirabilis*	Nanopore	Error-free	Mini	Kraken	95.6 (95.2:96)
158	*Proteus mirabilis*	Illumina	Error	Mini	Bracken	96 (95.6:96.3)
159	*Proteus mirabilis*	Illumina	Error-free	Mini	Bracken	96 (95.6:96.3)
160	*Proteus mirabilis*	Nanopore	Error-free	Mini	Centrifuge	96.1 (95.7:96.5)
161	*Proteus mirabilis*	Nanopore	Error	Mini	Bracken	96.8 (96.4:97.1)
162	*Proteus mirabilis*	Nanopore	Error-free	Standard	Kraken	97.7 (97.4:97.9)

Supplementary Table 7. Continued.

163	*Proteus mirabilis*	Illumina	Error	Standard	Bracken	97.8 (97.6:98)
164	*Proteus mirabilis*	Illumina	Error-free	Standard	Bracken	97.8 (97.6:98)
165	*Proteus mirabilis*	Nanopore	Error-free	Standard	Centrifuge	97.9 (97.7:98.1)
166	*Proteus mirabilis*	Nanopore	Error	Standard	Bracken	98.3 (98.1:98.4)
167	*Proteus mirabilis*	Nanopore	Error-free	Mini	Bracken	98.6 (98.5:98.7)
168	*Proteus mirabilis*	Nanopore	Error-free	Standard	Bracken	99.3 (99.2:99.3)
169	*Pseudomonas aeruginosa*	Illumina	Error	Mini	Kraken	56.9 (55.5:58.4)
170	*Pseudomonas aeruginosa*	Illumina	Error-free	Mini	Kraken	56.9 (55.5:58.4)
171	*Pseudomonas aeruginosa*	Illumina	Error	Mini	Centrifuge	60.1 (58.4:61.7)
172	*Pseudomonas aeruginosa*	Illumina	Error-free	Mini	Centrifuge	60.1 (58.4:61.7)
173	*Pseudomonas aeruginosa*	Nanopore	Error	Mini	Kraken	62.5 (61.1:63.9)
174	*Pseudomonas aeruginosa*	Nanopore	Error	Mini	Centrifuge	65.5 (63.8:67)
175	*Pseudomonas aeruginosa*	Illumina	Error	Standard	Kraken	71.6 (70.3:72.8)
176	*Pseudomonas aeruginosa*	Illumina	Error-free	Standard	Kraken	71.6 (70.3:72.8)
177	*Pseudomonas aeruginosa*	Illumina	Error	Standard	Centrifuge	74.1 (72.9:75.3)
178	*Pseudomonas aeruginosa*	Illumina	Error-free	Standard	Centrifuge	74.1 (72.9:75.3)
179	*Pseudomonas aeruginosa*	Nanopore	Error	Standard	Kraken	76 (74.9:77.1)
180	*Pseudomonas aeruginosa*	Nanopore	Error	Standard	Centrifuge	78.3 (77.2:79.4)
181	*Pseudomonas aeruginosa*	Nanopore	Error-free	Mini	Kraken	79.8 (78.8:80.8)
182	*Pseudomonas aeruginosa*	Illumina	Error	Mini	Bracken	81.2 (80.2:82.1)
183	*Pseudomonas aeruginosa*	Illumina	Error-free	Mini	Bracken	81.2 (80.2:82.1)
184	*Pseudomonas aeruginosa*	Nanopore	Error-free	Mini	Centrifuge	81.8 (80.7:82.9)
185	*Pseudomonas aeruginosa*	Nanopore	Error	Mini	Bracken	84.4 (83.6:85.2)
186	*Pseudomonas aeruginosa*	Nanopore	Error-free	Standard	Kraken	88.3 (87.6:88.9)
187	*Pseudomonas aeruginosa*	Illumina	Error	Standard	Bracken	89.1 (88.5:89.7)
188	*Pseudomonas aeruginosa*	Illumina	Error-free	Standard	Bracken	89.1 (88.5:89.7)
189	*Pseudomonas aeruginosa*	Nanopore	Error-free	Standard	Centrifuge	89.6 (88.9:90.2)
190	*Pseudomonas aeruginosa*	Nanopore	Error	Standard	Bracken	91.2 (90.7:91.7)
191	*Pseudomonas aeruginosa*	Nanopore	Error-free	Mini	Bracken	92.8 (92.4:93.2)
192	*Pseudomonas aeruginosa*	Nanopore	Error-free	Standard	Bracken	96.1 (95.8:96.3)
193	*Streptococcus pneumoniae*	Illumina	Error	Mini	Kraken	85.1 (84:86.2)
194	*Streptococcus pneumoniae*	Illumina	Error-free	Mini	Kraken	85.1 (84:86.2)
195	*Streptococcus pneumoniae*	Illumina	Error	Mini	Centrifuge	86.7 (85.6:87.8)
196	*Streptococcus pneumoniae*	Illumina	Error-free	Mini	Centrifuge	86.7 (85.6:87.8)
197	*Streptococcus pneumoniae*	Nanopore	Error	Mini	Kraken	87.8 (86.9:88.7)
198	*Streptococcus pneumoniae*	Nanopore	Error	Mini	Centrifuge	89.1 (88.2:90)
199	*Streptococcus pneumoniae*	Illumina	Error	Standard	Kraken	91.6 (90.9:92.3)
200	*Streptococcus pneumoniae*	Illumina	Error-free	Standard	Kraken	91.6 (90.9:92.3)
201	*Streptococcus pneumoniae*	Illumina	Error	Standard	Centrifuge	92.6 (91.9:93.2)
202	*Streptococcus pneumoniae*	Illumina	Error-free	Standard	Centrifuge	92.6 (91.9:93.2)
203	*Streptococcus pneumoniae*	Nanopore	Error	Standard	Kraken	93.2 (92.6:93.8)
204	*Streptococcus pneumoniae*	Nanopore	Error	Standard	Centrifuge	94 (93.5:94.5)

Supplementary Table 7. Continued.

205	*Streptococcus pneumoniae*	Nanopore	Error-free	Mini	Kraken	94.5 (94:94.9)
206	*Streptococcus pneumoniae*	Illumina	Error	Mini	Bracken	94.9 (94.5:95.3)
207	*Streptococcus pneumoniae*	Illumina	Error-free	Mini	Bracken	94.9 (94.5:95.3)
208	*Streptococcus pneumoniae*	Nanopore	Error-free	Mini	Centrifuge	95.1 (94.7:95.6)
209	*Streptococcus pneumoniae*	Nanopore	Error	Mini	Bracken	95.9 (95.6:96.3)
210	*Streptococcus pneumoniae*	Nanopore	Error-free	Standard	Kraken	97 (96.8:97.3)
211	*Streptococcus pneumoniae*	Illumina	Error	Standard	Bracken	97.3 (97:97.5)
212	*Streptococcus pneumoniae*	Illumina	Error-free	Standard	Bracken	97.3 (97:97.5)
213	*Streptococcus pneumoniae*	Nanopore	Error-free	Standard	Centrifuge	97.4 (97.1:97.6)
214	*Streptococcus pneumoniae*	Nanopore	Error	Standard	Bracken	97.8 (97.6:98)
215	*Streptococcus pneumoniae*	Nanopore	Error-free	Mini	Bracken	98.2 (98.1:98.4)
216	*Streptococcus pneumoniae*	Nanopore	Error-free	Standard	Bracken	99.1 (99:99.2)
217	*Streptococcus pyogenes*	Illumina	Error	Mini	Kraken	87.8 (86.7:88.8)
218	*Streptococcus pyogenes*	Illumina	Error-free	Mini	Kraken	87.8 (86.7:88.8)
219	*Streptococcus pyogenes*	Illumina	Error	Mini	Centrifuge	89.1 (88.1:90.1)
220	*Streptococcus pyogenes*	Illumina	Error-free	Mini	Centrifuge	89.1 (88.1:90.1)
221	*Streptococcus pyogenes*	Nanopore	Error	Mini	Kraken	90.1 (89.2:90.9)
222	*Streptococcus pyogenes*	Nanopore	Error	Mini	Centrifuge	91.2 (90.3:91.9)
223	*Streptococcus pyogenes*	Illumina	Error	Standard	Kraken	93.2 (92.6:93.8)
224	*Streptococcus pyogenes*	Illumina	Error-free	Standard	Kraken	93.2 (92.6:93.8)
225	*Streptococcus pyogenes*	Illumina	Error	Standard	Centrifuge	94 (93.4:94.5)
226	*Streptococcus pyogenes*	Illumina	Error-free	Standard	Centrifuge	94 (93.4:94.5)
227	*Streptococcus pyogenes*	Nanopore	Error	Standard	Kraken	94.5 (94:95)
228	*Streptococcus pyogenes*	Nanopore	Error	Standard	Centrifuge	95.2 (94.7:95.6)
229	*Streptococcus pyogenes*	Nanopore	Error-free	Mini	Kraken	95.6 (95.1:96)
230	*Streptococcus pyogenes*	Illumina	Error	Mini	Bracken	95.9 (95.5:96.3)
231	*Streptococcus pyogenes*	Illumina	Error-free	Mini	Bracken	95.9 (95.5:96.3)
232	*Streptococcus pyogenes*	Nanopore	Error-free	Mini	Centrifuge	96.1 (95.7:96.5)
233	*Streptococcus pyogenes*	Nanopore	Error	Mini	Bracken	96.7 (96.4:97)
234	*Streptococcus pyogenes*	Nanopore	Error-free	Standard	Kraken	97.6 (97.4:97.8)
235	*Streptococcus pyogenes*	Illumina	Error	Standard	Bracken	97.8 (97.6:98)
236	*Streptococcus pyogenes*	Illumina	Error-free	Standard	Bracken	97.8 (97.6:98)
237	*Streptococcus pyogenes*	Nanopore	Error-free	Standard	Centrifuge	97.9 (97.7:98.1)
238	*Streptococcus pyogenes*	Nanopore	Error	Standard	Bracken	98.3 (98.1:98.4)
239	*Streptococcus pyogenes*	Nanopore	Error-free	Mini	Bracken	98.6 (98.4:98.7)
240	*Streptococcus pyogenes*	Nanopore	Error-free	Standard	Bracken	99.3 (99.2:99.3)
241	*Staphylococcus epidermidis*	Illumina	Error	Mini	Kraken	86.2 (85.1:87.2)
242	*Staphylococcus epidermidis*	Illumina	Error-free	Mini	Kraken	86.2 (85.1:87.2)
243	*Staphylococcus epidermidis*	Illumina	Error	Mini	Centrifuge	87.6 (86.5:88.7)
244	*Staphylococcus epidermidis*	Illumina	Error-free	Mini	Centrifuge	87.6 (86.5:88.7)
245	*Staphylococcus epidermidis*	Nanopore	Error	Mini	Kraken	88.7 (87.7:89.6)
246	*Staphylococcus epidermidis*	Nanopore	Error	Mini	Centrifuge	89.9 (89:90.8)

Supplementary Table 7. Continued.

247	*Staphylococcus epidermidis*	Illumina	Error	Standard	Kraken	92.2 (91.5:92.9)
248	*Staphylococcus epidermidis*	Illumina	Error-free	Standard	Kraken	92.2 (91.5:92.9)
249	*Staphylococcus epidermidis*	Illumina	Error	Standard	Centrifuge	93.1 (92.5:93.7)
250	*Staphylococcus epidermidis*	Illumina	Error-free	Standard	Centrifuge	93.1 (92.5:93.7)
251	*Staphylococcus epidermidis*	Nanopore	Error	Standard	Kraken	93.7 (93.2:94.2)
252	*Staphylococcus epidermidis*	Nanopore	Error	Standard	Centrifuge	94.4 (93.9:94.9)
253	*Staphylococcus epidermidis*	Nanopore	Error-free	Mini	Kraken	94.9 (94.4:95.3)
254	*Staphylococcus epidermidis*	Illumina	Error	Mini	Bracken	95.3 (94.9:95.7)
255	*Staphylococcus epidermidis*	Illumina	Error-free	Mini	Bracken	95.3 (94.9:95.7)
256	*Staphylococcus epidermidis*	Nanopore	Error-free	Mini	Centrifuge	95.5 (95:95.9)
257	*Staphylococcus epidermidis*	Nanopore	Error	Mini	Bracken	96.2 (95.9:96.6)
258	*Staphylococcus epidermidis*	Nanopore	Error-free	Standard	Kraken	97.3 (97:97.5)
259	*Staphylococcus epidermidis*	Illumina	Error	Standard	Bracken	97.5 (97.2:97.7)
260	*Staphylococcus epidermidis*	Illumina	Error-free	Standard	Bracken	97.5 (97.2:97.7)
261	*Staphylococcus epidermidis*	Nanopore	Error-free	Standard	Centrifuge	97.6 (97.4:97.8)
262	*Staphylococcus epidermidis*	Nanopore	Error	Standard	Bracken	98 (97.8:98.2)
263	*Staphylococcus epidermidis*	Nanopore	Error-free	Mini	Bracken	98.4 (98.2:98.5)
264	*Staphylococcus epidermidis*	Nanopore	Error-free	Standard	Bracken	99.1 (99.1:99.2)
265	*Staphylococcus haemolyticus*	Illumina	Error	Mini	Kraken	87.4 (86.3:88.4)
266	*Staphylococcus haemolyticus*	Illumina	Error-free	Mini	Kraken	87.4 (86.3:88.4)
267	*Staphylococcus haemolyticus*	Illumina	Error	Mini	Centrifuge	88.7 (87.7:89.7)
268	*Staphylococcus haemolyticus*	Illumina	Error-free	Mini	Centrifuge	88.7 (87.7:89.7)
269	*Staphylococcus haemolyticus*	Nanopore	Error	Mini	Kraken	89.7 (88.8:90.6)
270	*Staphylococcus haemolyticus*	Nanopore	Error	Mini	Centrifuge	90.9 (90:91.7)
271	*Staphylococcus haemolyticus*	Illumina	Error	Standard	Kraken	93 (92.3:93.6)
272	*Staphylococcus haemolyticus*	Illumina	Error-free	Standard	Kraken	93 (92.3:93.6)
273	*Staphylococcus haemolyticus*	Illumina	Error	Standard	Centrifuge	93.8 (93.2:94.3)
274	*Staphylococcus haemolyticus*	Illumina	Error-free	Standard	Centrifuge	93.8 (93.2:94.3)
275	*Staphylococcus haemolyticus*	Nanopore	Error	Standard	Kraken	94.3 (93.8:94.8)
276	*Staphylococcus haemolyticus*	Nanopore	Error	Standard	Centrifuge	95 (94.5:95.4)
277	*Staphylococcus haemolyticus*	Nanopore	Error-free	Mini	Kraken	95.4 (95:95.8)
278	*Staphylococcus haemolyticus*	Illumina	Error	Mini	Bracken	95.8 (95.4:96.1)
279	*Staphylococcus haemolyticus*	Illumina	Error-free	Mini	Bracken	95.8 (95.4:96.1)
280	*Staphylococcus haemolyticus*	Nanopore	Error-free	Mini	Centrifuge	95.9 (95.5:96.3)
281	*Staphylococcus haemolyticus*	Nanopore	Error	Mini	Bracken	96.6 (96.3:96.9)
282	*Staphylococcus haemolyticus*	Nanopore	Error-free	Standard	Kraken	97.5 (97.3:97.8)
283	*Staphylococcus haemolyticus*	Illumina	Error	Standard	Bracken	97.7 (97.5:97.9)
284	*Staphylococcus haemolyticus*	Illumina	Error-free	Standard	Bracken	97.7 (97.5:97.9)
285	*Staphylococcus haemolyticus*	Nanopore	Error-free	Standard	Centrifuge	97.8 (97.6:98)
286	*Staphylococcus haemolyticus*	Nanopore	Error	Standard	Bracken	98.2 (98:98.4)
287	*Staphylococcus haemolyticus*	Nanopore	Error-free	Mini	Bracken	98.5 (98.4:98.7)
288	*Staphylococcus haemolyticus*	Nanopore	Error-free	Standard	Bracken	99.2 (99.1:99.3)

Supplementary Table 7. Continued.

289	*Bacillus cereus*	Illumina	Error	Mini	Kraken	13.4 (12.8:14.1)
290	*Bacillus cereus*	Illumina	Error-free	Mini	Kraken	13.4 (12.8:14.1)
291	*Bacillus cereus*	Illumina	Error	Mini	Centrifuge	15 (14.1:15.9)
292	*Bacillus cereus*	Illumina	Error-free	Mini	Centrifuge	15 (14.1:15.9)
293	*Bacillus cereus*	Nanopore	Error	Mini	Kraken	16.3 (15.6:17.1)
294	*Bacillus cereus*	Nanopore	Error	Mini	Centrifuge	18.2 (17.2:19.2)
295	*Bacillus cereus*	Illumina	Error	Standard	Kraken	22.8 (21.8:23.8)
296	*Bacillus cereus*	Illumina	Error-free	Standard	Kraken	22.8 (21.8:23.8)
297	*Bacillus cereus*	Illumina	Error	Standard	Centrifuge	25.2 (24.1:26.3)
298	*Bacillus cereus*	Illumina	Error-free	Standard	Centrifuge	25.2 (24.1:26.3)
299	*Bacillus cereus*	Nanopore	Error	Standard	Kraken	27.1 (26:28.3)
300	*Bacillus cereus*	Nanopore	Error	Standard	Centrifuge	29.8 (28.6:31)
301	*Bacillus cereus*	Nanopore	Error-free	Mini	Kraken	31.7 (30.5:33)
302	*Bacillus cereus*	Illumina	Error	Mini	Bracken	33.6 (32.4:34.8)
303	*Bacillus cereus*	Illumina	Error-free	Mini	Bracken	33.6 (32.4:34.8)
304	*Bacillus cereus*	Nanopore	Error-free	Mini	Centrifuge	34.6 (33.1:36.2)
305	*Bacillus cereus*	Nanopore	Error	Mini	Bracken	38.9 (37.6:40.2)
306	*Bacillus cereus*	Nanopore	Error-free	Standard	Kraken	47 (45.6:48.4)
307	*Bacillus cereus*	Illumina	Error	Standard	Bracken	49.1 (47.7:50.5)
308	*Bacillus cereus*	Illumina	Error-free	Standard	Bracken	49.1 (47.7:50.5)
309	*Bacillus cereus*	Nanopore	Error-free	Standard	Centrifuge	50.2 (48.7:51.7)
310	*Bacillus cereus*	Nanopore	Error	Standard	Bracken	54.8 (53.5:56.2)
311	*Bacillus cereus*	Nanopore	Error-free	Mini	Bracken	60.3 (58.9:61.6)
312	*Bacillus cereus*	Nanopore	Error-free	Standard	Bracken	74.3 (73.2:75.4)
313	*Cutibacterium acnes*	Illumina	Error-free	Mini	Kraken	95.4 (94.7:96.1)
314	*Cutibacterium acnes*	Illumina	Error	Mini	Kraken	95.4 (94.7:96)
315	*Cutibacterium acnes*	Illumina	Error	Mini	Centrifuge	96 (95.3:96.5)
316	*Cutibacterium acnes*	Illumina	Error-free	Mini	Centrifuge	96 (95.3:96.5)
317	*Cutibacterium acnes*	Nanopore	Error	Mini	Kraken	96.3 (95.8:96.8)
318	*Cutibacterium acnes*	Nanopore	Error	Mini	Centrifuge	96.8 (96.3:97.2)
319	*Cutibacterium acnes*	Illumina	Error	Standard	Kraken	97.6 (97.2:97.9)
320	*Cutibacterium acnes*	Illumina	Error-free	Standard	Kraken	97.6 (97.2:97.9)
321	*Cutibacterium acnes*	Illumina	Error	Standard	Centrifuge	97.8 (97.5:98.1)
322	*Cutibacterium acnes*	Illumina	Error-free	Standard	Centrifuge	97.8 (97.5:98.1)
323	*Cutibacterium acnes*	Nanopore	Error	Standard	Kraken	98 (97.7:98.3)
324	*Cutibacterium acnes*	Nanopore	Error	Standard	Centrifuge	98.3 (98:98.5)
325	*Cutibacterium acnes*	Nanopore	Error-free	Mini	Kraken	98.4 (98.2:98.6)
326	*Cutibacterium acnes*	Illumina	Error	Mini	Bracken	98.6 (98.3:98.8)
327	*Cutibacterium acnes*	Illumina	Error-free	Mini	Bracken	98.6 (98.3:98.8)
328	*Cutibacterium acnes*	Nanopore	Error-free	Mini	Centrifuge	98.6 (98.4:98.8)
329	*Cutibacterium acnes*	Nanopore	Error	Mini	Bracken	98.9 (98.7:99)
330	*Cutibacterium acnes*	Nanopore	Error-free	Standard	Kraken	99.2 (99:99.3)

Supplementary Table 7. Continued.

331	*Cutibacterium acnes*	Illumina	Error	Standard	Bracken	99.2 (99.1:99.3)
332	*Cutibacterium acnes*	Illumina	Error-free	Standard	Bracken	99.2 (99.1:99.3)
333	*Cutibacterium acnes*	Nanopore	Error-free	Standard	Centrifuge	99.3 (99.2:99.4)
334	*Cutibacterium acnes*	Nanopore	Error	Standard	Bracken	99.4 (99.3:99.5)
335	*Cutibacterium acnes*	Nanopore	Error-free	Mini	Bracken	99.5 (99.4:99.6)
336	*Cutibacterium acnes*	Nanopore	Error-free	Standard	Bracken	99.7 (99.7:99.8)
337	*Streptococcus mitis*	Illumina	Error	Mini	Kraken	28.2 (27.1:29.3)
338	*Streptococcus mitis*	Illumina	Error-free	Mini	Kraken	28.2 (27.1:29.3)
339	*Streptococcus mitis*	Illumina	Error	Mini	Centrifuge	30.9 (29.5:32.3)
340	*Streptococcus mitis*	Illumina	Error-free	Mini	Centrifuge	30.9 (29.5:32.3)
341	*Streptococcus mitis*	Nanopore	Error	Mini	Kraken	33.1 (31.9:34.3)
342	*Streptococcus mitis*	Nanopore	Error	Mini	Centrifuge	36 (34.5:37.5)
343	*Streptococcus mitis*	Illumina	Error	Standard	Kraken	42.8 (41.5:44.1)
344	*Streptococcus mitis*	Illumina	Error-free	Standard	Kraken	42.8 (41.5:44.1)
345	*Streptococcus mitis*	Illumina	Error	Standard	Centrifuge	46 (44.6:47.4)
346	*Streptococcus mitis*	Illumina	Error-free	Standard	Centrifuge	46 (44.6:47.4)
347	*Streptococcus mitis*	Nanopore	Error	Standard	Kraken	48.5 (47.2:49.9)
348	*Streptococcus mitis*	Nanopore	Error	Standard	Centrifuge	51.7 (50.3:53.1)
349	*Streptococcus mitis*	Nanopore	Error-free	Mini	Kraken	54 (52.6:55.4)
350	*Streptococcus mitis*	Illumina	Error	Mini	Bracken	56.1 (54.8:57.5)
351	*Streptococcus mitis*	Illumina	Error-free	Mini	Bracken	56.1 (54.8:57.5)
352	*Streptococcus mitis*	Nanopore	Error-free	Mini	Centrifuge	57.2 (55.6:58.9)
353	*Streptococcus mitis*	Nanopore	Error	Mini	Bracken	61.7 (60.4:63)
354	*Streptococcus mitis*	Nanopore	Error-free	Standard	Kraken	69.2 (67.9:70.4)
355	*Streptococcus mitis*	Illumina	Error	Standard	Bracken	70.9 (69.8:72)
356	*Streptococcus mitis*	Illumina	Error-free	Standard	Bracken	70.9 (69.8:72)
357	*Streptococcus mitis*	Nanopore	Error-free	Standard	Centrifuge	71.8 (70.6:73)
358	*Streptococcus mitis*	Nanopore	Error	Standard	Bracken	75.4 (74.4:76.5)
359	*Streptococcus mitis*	Nanopore	Error-free	Mini	Bracken	79.3 (78.3:80.3)
360	*Streptococcus mitis*	Nanopore	Error-free	Standard	Bracken	88 (87.3:88.6)
361	*Micrococcus luteus*	Illumina	Error	Mini	Kraken	56.8 (55.1:58.5)
362	*Micrococcus luteus*	Illumina	Error-free	Mini	Kraken	56.8 (55.2:58.5)
363	*Micrococcus luteus*	Illumina	Error	Mini	Centrifuge	60 (58.1:61.8)
364	*Micrococcus luteus*	Illumina	Error-free	Mini	Centrifuge	60 (58.1:61.8)
365	*Micrococcus luteus*	Nanopore	Error	Mini	Kraken	62.4 (60.8:64)
366	*Micrococcus luteus*	Nanopore	Error	Mini	Centrifuge	65.4 (63.6:67.1)
367	*Micrococcus luteus*	Illumina	Error	Standard	Kraken	71.5 (70.1:72.9)
368	*Micrococcus luteus*	Illumina	Error-free	Standard	Kraken	71.5 (70.1:72.9)
369	*Micrococcus luteus*	Illumina	Error	Standard	Centrifuge	74.1 (72.7:75.4)
370	*Micrococcus luteus*	Illumina	Error-free	Standard	Centrifuge	74.1 (72.7:75.4)
371	*Micrococcus luteus*	Nanopore	Error	Standard	Kraken	76 (74.7:77.2)
372	*Micrococcus luteus*	Nanopore	Error	Standard	Centrifuge	78.2 (77:79.4)

Supplementary Table 7. Continued.

373	*Micrococcus luteus*	Nanopore	Error-free	Mini	Kraken	79.8 (78.6:80.9)
374	*Micrococcus luteus*	Illumina	Error	Mini	Bracken	81.1 (80:82.2)
375	*Micrococcus luteus*	Illumina	Error-free	Mini	Bracken	81.1 (80:82.2)
376	*Micrococcus luteus*	Nanopore	Error-free	Mini	Centrifuge	81.8 (80.5:83)
377	*Micrococcus luteus*	Nanopore	Error	Mini	Bracken	84.4 (83.4:85.3)
378	*Micrococcus luteus*	Nanopore	Error-free	Standard	Kraken	88.3 (87.5:89)
379	*Micrococcus luteus*	Illumina	Error	Standard	Bracken	89.1 (88.4:89.8)
380	*Micrococcus luteus*	Illumina	Error-free	Standard	Bracken	89.1 (88.4:89.8)
381	*Micrococcus luteus*	Nanopore	Error-free	Standard	Centrifuge	89.5 (88.8:90.2)
382	*Micrococcus luteus*	Nanopore	Error	Standard	Bracken	91.2 (90.6:91.7)
383	*Micrococcus luteus*	Nanopore	Error-free	Mini	Bracken	92.8 (92.3:93.3)
384	*Micrococcus luteus*	Nanopore	Error-free	Standard	Bracken	96.1 (95.8:96.4)

Supplementary Table 8. Detailed threshold abundance recommended to call additional true species and genus.

	Bacteria	Sequencing profile	Database	Tax	Correctly classified median (IQR)	Unclassified median (IQR)	Misclassified 1	Median (IQR)	95th:99th:100th percentile	Misclassified 2	Median (IQR)	95th:99th:100th percentile
1	*Bacillus*	Illumina with error profile	Bracken (mini database)	G	99.89 (99.82:99.91)	NA	*Peribacillus*	0 (0:0.01)	0.04:0.11:0.3	*Stenotrophomonas*	0 (0:0)	0.05:0.13:0.15
2	*Bacillus*	Illumina with error profile	Bracken (standard database)	G	99.9 (99.84:99.95)	NA	*Cytobacillus*	0 (0:0.01)	0.09:0.16:0.2	*Peribacillus*	0 (0:0.02)	0.13:0.22:0.23
3	*Bacillus*	Illumina with error profile	Centrifuge (standard database)	G	95.38 (94.35:96.56)	NA	*Homo*	0.15 (0.11:0.18)	0.28:0.35:0.42	*NA*	NA	NA
4	*Bacillus*	Illumina with error profile	Kraken (mini database)	G	97.34 (95.2:97.85)	NA	*Peribacillus*	0 (0:0)	0.04:0.08:0.21	*Stenotrophomonas*	0 (0:0)	0.04:0.1:0.12
5	*Bacillus*	Illumina with error profile	Kraken (standard database)	G	99.44 (98.98:99.61)	NA	*Cytobacillus*	0 (0:0)	0.08:0.16:0.2	*Peribacillus*	0 (0:0.02)	0.13:0.21:0.23
6	*Bacillus*	Illumina with error-free profile	Bracken (mini database)	G	99.89 (99.81:99.91)	NA	*Peribacillus*	0 (0:0.01)	0.05:0.11:0.26	*Stenotrophomonas*	0 (0:0)	0.1:0.14:0.15
7	*Bacillus*	Illumina with error-free profile	Bracken (standard database)	G	99.9 (99.83:99.94)	NA	*Cytobacillus*	0 (0:0)	0.09:0.14:0.18	*Peribacillus*	0 (0:0.02)	0.13:0.21:0.22
8	*Bacillus*	Illumina with error-free profile	Centrifuge (standard database)	G	95.43 (94.32:96.55)	NA	*Homo*	0.15 (0.11:0.19)	0.28:0.36:0.42	*NA*	NA	NA
9	*Bacillus*	Illumina with error-free profile	Kraken (mini database)	G	97.28 (95.26:97.87)	NA	*Peribacillus*	0 (0:0.01)	0.04:0.1:0.18	*Stenotrophomonas*	0 (0:0)	0.06:0.1:0.11
10	*Bacillus*	Illumina with error-free profile	Kraken (standard database)	G	99.44 (98.98:99.63)	NA	*Cytobacillus*	0 (0:0)	0.08:0.14:0.18	*Peribacillus*	0 (0:0.02)	0.13:0.2:0.21
11	*Bacillus*	Nanopore with error profile	Bracken (mini database)	G	99.12 (98.95:99.28)	NA	*Homo*	0.09 (0.07:0.12)	0.21:0.37:0.39	*Pseudomonas*	0.03 (0.02:0.04)	0.07:0.15:0.21
12	*Bacillus*	Nanopore with error profile	Bracken (standard database)	G	99.92 (99.85:99.96)	NA	*Cytobacillus*	0.01 (0.01:0.06)	0.1:0.13:0.14	*Homo*	0.01 (0.01:0.02)	0.04:0.1:0.11
13	*Bacillus*	Nanopore with error profile	Centrifuge (standard database)	G	97.15 (96.32:97.95)	NA	*Homo*	0.23 (0.15:0.3)	0.45:0.6:0.65	*NA*	NA	NA
14	*Bacillus*	Nanopore with error profile	Kraken (mini database)	G	88.49 (87.02:88.96)	NA	*Homo*	0.08 (0.06:0.1)	0.18:0.27:0.28	*Pseudomonas*	0.03 (0.02:0.04)	0.06:0.1:0.15
15	*Bacillus*	Nanopore with error profile	Kraken (standard database)	G	99.48 (99.09:99.62)	NA	*Cytobacillus*	0.01 (0.01:0.05)	0.1:0.13:0.14	*Homo*	0.01 (0.01:0.02)	0.04:0.08:0.1
16	*Bacillus*	Nanopore with error-free profile	Bracken (mini database)	G	99.97 (99.9:99.99)	NA	*Cytobacillus*	0.05 (0.03:0.07)	0.13:0.15:0.15	*Enterococcus*	0.01 (0.01:0.04)	0.09:0.11:0.11
17	*Bacillus*	Nanopore with error-free profile	Bracken (standard database)	G	99.95 (99.9:99.99)	NA	*Lactococcus*	0.03 (0.01:0.05)	0.15:0.17:0.18	*Priestia*	0.02 (0.02:0.06)	0.13:0.14:0.14
18	*Bacillus*	Nanopore with error-free profile	Centrifuge (standard database)	G	98.48 (97.89:99.09)	NA	*Homo*	0.11 (0.05:0.17)	0.32:0.54:0.85	*NA*	NA	NA
19	*Bacillus*	Nanopore with error-free profile	Kraken (mini database)	G	99.7 (99.35:99.87)	NA	*Cytobacillus*	0.05 (0.03:0.07)	0.13:0.15:0.15	*Enterococcus*	0.01 (0.01:0.04)	0.09:0.11:0.11
20	*Bacillus*	Nanopore with error-free profile	Kraken (standard database)	G	99.88 (99.75:99.95)	NA	*Lactococcus*	0.03 (0.01:0.05)	0.15:0.17:0.18	*Priestia*	0.02 (0.02:0.06)	0.13:0.14:0.14
21	*Bacteroides*	Illumina with error profile	Bracken (mini database)	G	99.26 (98.94:99.55)	NA	*Parabacteroides*	0.19 (0.08:0.45)	0.71:1.11:1.12	*Phocaeicola*	0.19 (0.07:0.3)	1.15:1.44:2.37
22	*Bacteroides*	Illumina with error profile	Bracken (standard database)	G	98.87 (98.63:99.33)	NA	*Parabacteroides*	0.37 (0.21:0.66)	1.33:2.05:2.29	*Phocaeicola*	0.3 (0.1:0.46)	1.12:1.5:2.13
23	*Bacteroides*	Illumina with error profile	Centrifuge (standard database)	G	95.97 (94.96:97.57)	NA	*Parabacteroides*	0.25 (0.13:1.04)	1.5:1.85:1.97	*Alistipes*	0.06 (0.02:0.11)	0.36:0.77:0.77
24	*Bacteroides*	Illumina with error profile	Kraken (mini database)	G	92.66 (91.03:94.47)	NA	*Parabacteroides*	0.17 (0.07:0.34)	0.62:0.88:0.93	*Phocaeicola*	0.16 (0.06:0.25)	0.86:1.32:1.86
25	*Bacteroides*	Illumina with error profile	Kraken (standard database)	G	95.8 (94.7:96.99)	NA	*Parabacteroides*	0.35 (0.19:0.59)	1.2:1.35:1.39	*Phocaeicola*	0.28 (0.1:0.44)	1.01:1.25:1.39
26	*Bacteroides*	Illumina with error-free profile	Bracken (mini database)	G	99.25 (98.96:99.56)	NA	*Parabacteroides*	0.19 (0.09:0.46)	0.7:1.09:1.12	*Phocaeicola*	0.18 (0.07:0.29)	1.14:1.41:2.36
27	*Bacteroides*	Illumina with error-free profile	Bracken (standard database)	G	98.85 (98.62:99.33)	NA	*Parabacteroides*	0.36 (0.22:0.67)	1.34:2:2.43	*Phocaeicola*	0.29 (0.1:0.47)	1.12:1.44:2.16
28	*Bacteroides*	Illumina with error-free profile	Centrifuge (standard database)	G	95.97 (94.96:97.59)	NA	*Parabacteroides*	0.23 (0.14:1.03)	1.49:1.81:1.95	*Alistipes*	0.06 (0.02:0.11)	0.37:0.81:0.81
29	*Bacteroides*	Illumina with error-free profile	Kraken (mini database)	G	92.66 (91.05:94.49)	NA	*Parabacteroides*	0.17 (0.08:0.36)	0.62:0.85:0.94	*Phocaeicola*	0.16 (0.06:0.26)	0.86:1.29:1.85
30	*Bacteroides*	Illumina with error-free profile	Kraken (standard database)	G	95.77 (94.69:97.01)	NA	*Parabacteroides*	0.34 (0.2:0.59)	1.22:1.3:1.34	*Phocaeicola*	0.28 (0.1:0.44)	1.03:1.27:1.39
31	*Bacteroides*	Nanopore with error profile	Bracken (mini database)	G	98.56 (98.22:98.85)	NA	*Parabacteroides*	0.2 (0.08:0.45)	0.7:0.91:1.11	*Phocaeicola*	0.13 (0.06:0.3)	1.13:1.55:2.3
32	*Bacteroides*	Nanopore with error profile	Bracken (standard database)	G	99 (98.66:99.31)	NA	*Parabacteroides*	0.31 (0.19:0.6)	1.19:1.37:1.54	*Phocaeicola*	0.27 (0.11:0.43)	1.3:1.54:1.77
33	*Bacteroides*	Nanopore with error profile	Centrifuge (standard database)	G	97 (95.88:98.32)	NA	*Parabacteroides*	0.32 (0.13:1.19)	1.71:2.19:2.39	*Alistipes*	0.07 (0.05:0.14)	0.51:0.78:0.88
34	*Bacteroides*	Nanopore with error profile	Kraken (mini database)	G	84.69 (83.45:85.99)	NA	*Parabacteroides*	0.16 (0.06:0.36)	0.59:0.74:0.85	*Phocaeicola*	0.11 (0.05:0.25)	0.87:1.31:1.78
35	*Bacteroides*	Nanopore with error profile	Kraken (standard database)	G	97.31 (96.52:97.94)	NA	*Parabacteroides*	0.3 (0.18:0.57)	1.12:1.26:1.3	*Phocaeicola*	0.26 (0.11:0.42)	1.25:1.51:1.65
36	*Bacteroides*	Nanopore with error-free profile	Bracken (mini database)	G	99.05 (98.54:99.53)	NA	*Parabacteroides*	0.23 (0.12:0.87)	1.37:1.6:1.68	*Phocaeicola*	0.18 (0.05:0.46)	1.57:2.34:2.38
37	*Bacteroides*	Nanopore with error-free profile	Bracken (standard database)	G	99.02 (98.53:99.4)	NA	*Parabacteroides*	0.32 (0.22:0.6)	1.42:1.6:1.71	*Phocaeicola*	0.29 (0.1:0.52)	1.68:2.1:2.37
38	*Bacteroides*	Nanopore with error-free profile	Centrifuge (standard database)	G	97.87 (96.82:98.95)	NA	*Parabacteroides*	0.34 (0.11:1.38)	2.1:2.48:2.57	*Alistipes*	0.11 (0.05:0.23)	0.61:1.1:1.12
39	*Bacteroides*	Nanopore with error-free profile	Kraken (mini database)	G	97.59 (96.73:98.37)	NA	*Parabacteroides*	0.23 (0.12:0.85)	1.33:1.55:1.64	*Phocaeicola*	0.17 (0.05:0.44)	1.5:2.26:2.35
40	*Bacteroides*	Nanopore with error-free profile	Kraken (standard database)	G	98.15 (97.6:98.65)	NA	*Parabacteroides*	0.32 (0.22:0.59)	1.42:1.58:1.68	*Phocaeicola*	0.28 (0.1:0.52)	1.55:2.08:2.33
41	*Cutibacterium*	Illumina with error profile	Bracken (mini database)	G	99.84 (99.68:99.86)	NA	*Propionibacterium*	0.08 (0.06:0.13)	0.4:0.54:0.62	*Bacillus*	0 (0:0)	0:4.82:6.34
42	*Cutibacterium*	Illumina with error profile	Bracken (standard database)	G	100 (99.99:100)	NA	*Paenibacillus*	0 (0:0.01)	0.08:0.15:0.17	*Corynebacterium*	0 (0:0)	0.02:0.04:0.04
43	*Cutibacterium*	Illumina with error profile	Centrifuge (standard database)	G	99.4 (98.99:99.98)	NA	*Bacillus*	0 (0:0)	5.83:10.49:11.65	*Staphylococcus*	0 (0:0.21)	0.72:0.87:0.91
44	*Cutibacterium*	Illumina with error profile	Kraken (mini database)	G	96.77 (95.67:97.31)	NA	*Propionibacterium*	0.07 (0.06:0.13)	0.38:0.52:0.59	*Bacillus*	0 (0:0)	0.01:3.37:4.67
45	*Cutibacterium*	Illumina with error profile	Kraken (standard database)	G	99.99 (99.83:100)	NA	*Paenibacillus*	0 (0:0.01)	0.08:0.14:0.16	*Corynebacterium*	0 (0:0.01)	0.02:0.04:0.04
46	*Cutibacterium*	Illumina with error-free profile	Bracken (mini database)	G	99.84 (99.66:99.86)	NA	*Propionibacterium*	0.07 (0.06:0.13)	0.37:0.54:0.66	*Bacillus*	0 (0:0)	0.32:5.14:6.34
47	*Cutibacterium*	Illumina with error-free profile	Bracken (standard database)	G	100 (99.99:100)	NA	*Corynebacterium*	0 (0:0)	0.02:0.04:0.04	*Paenibacillus*	0 (0:0.01)	0.09:0.15:0.16
48	*Cutibacterium*	Illumina with error-free profile	Centrifuge (standard database)	G	99.37 (98.96:99.98)	NA	*Bacillus*	0 (0:0)	4.64:10.19:11.58	*Staphylococcus*	0 (0:0.19)	0.66:0.87:0.92
49	*Cutibacterium*	Illumina with error-free profile	Kraken (mini database)	G	96.75 (95.72:97.3)	NA	*Propionibacterium*	0.07 (0.06:0.13)	0.36:0.52:0.63	*Bacillus*	0 (0:0)	0.01:3.46:4.61
50	*Cutibacterium*	Illumina with error-free profile	Kraken (standard database)	G	99.99 (99.83:100)	NA	*Corynebacterium*	0 (0:0.01)	0.03:0.04:0.04	*Paenibacillus*	0 (0:0.01)	0.08:0.14:0.15

Supplementary Table 8. Continued.

51	*Cutibacterium*	Nanopore with error profile	Bracken (mini database)	G	99 (98.86:99.06)	NA	*Bacillus*	0.02 (0.01:0.02)	0.04:2.59:6.41	*Staphylococcus*	0.01 (0.01:0.02)	0.17:0.68:0.84
52	*Cutibacterium*	Nanopore with error profile	Bracken (standard database)	G	99.98 (99.96:99.99)	NA	*Corynebacterium*	0.01 (0.01:0.01)	0.02:0.03:0.03	*Homo*	0.01 (0.01:0.01)	0.1:0.14:0.15
53	*Cutibacterium*	Nanopore with error profile	Centrifuge (standard database)	G	99.5 (99.12:99.65)	NA	*Propionibacterium*	0.13 (0.09:0.22)	0.58:0.66:0.96	*Homo*	0.01 (0.01:0.02)	0.23:0.4:0.44
54	*Cutibacterium*	Nanopore with error profile	Kraken (mini database)	G	86.72 (86.19:87.32)	NA	*Propionibacterium*	0.11 (0.08:0.15)	0.32:0.44:0.47	*Bacillus*	0.01 (0.01:0.02)	0.04:1.58:4.56
55	*Cutibacterium*	Nanopore with error profile	Kraken (standard database)	G	99.75 (99.7:99.79)	NA	*Corynebacterium*	0.01 (0.01:0.01)	0.02:0.03:0.03	*Homo*	0.01 (0.01:0.01)	0.09:0.13:0.14
56	*Cutibacterium*	Nanopore with error-free profile	Bracken (mini database)	G	99.99 (99.93:100)	NA	*Propionibacterium*	0.04 (0.01:0.11)	0.38:0.46:0.53	*NA*	NA	NA
57	*Cutibacterium*	Nanopore with error-free profile	Bracken (standard database)	G	100 (100:100)	NA	*Microcella*	0.02 (0.02:0.03)	0.04:0.04:0.04	*Helicobacter*	0.01 (0.01:0.01)	0.01:0.01:0.01
58	*Cutibacterium*	Nanopore with error-free profile	Centrifuge (standard database)	G	99.85 (99.64:99.98)	NA	*Propionibacterium*	0.1 (0.05:0.23)	0.32:0.33:0.33	*Acidipropionibacterium*	0.07 (0.04:0.12)	0.19:0.27:0.29
59	*Cutibacterium*	Nanopore with error-free profile	Kraken (mini database)	G	99.84 (99.73:99.94)	NA	*Propionibacterium*	0.04 (0.01:0.11)	0.38:0.46:0.53	*NA*	NA	NA
60	*Cutibacterium*	Nanopore with error-free profile	Kraken (standard database)	G	100 (99.98:100)	NA	*Microcella*	0.02 (0.01:0.03)	0.03:0.03:0.03	*Corynebacterium*	0.01 (0.01:0.01)	0.01:0.01:0.01
61	*Enterobacter*	Illumina with error profile	Bracken (mini database)	G	96.8 (95.21:97.56)	NA	*Escherichia*	0.62 (0.39:1.23)	2.87:4.63:5.34	*Lelliottia*	0.09 (0.07:0.2)	0.97:2.08:19.54
62	*Enterobacter*	Illumina with error profile	Bracken (standard database)	G	98 (96.74:98.99)	NA	*Klebsiella*	0.52 (0.33:0.79)	1.75:2.65:3.25	*Escherichia*	0.39 (0.21:1)	2.73:4.84:4.84
63	*Enterobacter*	Illumina with error profile	Centrifuge (standard database)	G	94.19 (91.75:97.01)	NA	*Klebsiella*	0.5 (0.31:0.82)	1.44:2.52:2.62	*Lelliottia*	0.06 (0.01:0.39)	2.01:18.1:41.93
64	*Enterobacter*	Illumina with error profile	Kraken (mini database)	G	85.74 (83.39:87.37)	NA	*Klebsiella*	0.46 (0.34:0.64)	1.11:1.44:2.78	*Lelliottia*	0.08 (0.06:0.16)	0.67:1.43:10
65	*Enterobacter*	Illumina with error profile	Kraken (standard database)	G	96.16 (93.28:98.12)	NA	*Klebsiella*	0.44 (0.31:0.7)	1.22:1.67:3.17	*Escherichia*	0.22 (0.13:0.39)	0.82:0.9:2.2
66	*Enterobacter*	Illumina with error-free profile	Bracken (mini database)	G	96.85 (95.06:97.53)	NA	*Salmonella*	0.44 (0.28:0.85)	2.56:3.92:5.49	*Lelliottia*	0.09 (0.07:0.2)	0.97:2.02:19.51
67	*Enterobacter*	Illumina with error-free profile	Bracken (standard database)	G	98.06 (96.8:98.99)	NA	*Klebsiella*	0.53 (0.32:0.8)	1.73:2.85:3.21	*Escherichia*	0.4 (0.19:1)	2.59:4.02:4.82
68	*Enterobacter*	Illumina with error-free profile	Centrifuge (standard database)	G	94.16 (91.92:96.98)	NA	*Klebsiella*	0.5 (0.32:0.8)	1.46:2.57:2.68	*Lelliottia*	0.06 (0.01:0.39)	2.01:18.48:42.06
69	*Enterobacter*	Illumina with error-free profile	Kraken (mini database)	G	85.75 (83.4:87.38)	NA	*Klebsiella*	0.46 (0.33:0.62)	1.13:1.45:2.78	*Lelliottia*	0.08 (0.06:0.16)	0.67:1.39:10.05
70	*Enterobacter*	Illumina with error-free profile	Kraken (standard database)	G	96.14 (93.35:98.16)	NA	*Klebsiella*	0.44 (0.31:0.7)	1.21:1.63:3.12	*Escherichia*	0.21 (0.13:0.39)	0.81:0.88:2.24
71	*Enterobacter*	Nanopore with error profile	Bracken (mini database)	G	95.73 (94.12:96.61)	NA	*Escherichia*	0.7 (0.34:1.39)	4.42:6.23:6.25	*Lelliottia*	0.11 (0.07:0.16)	0.62:1.23:12.36
72	*Enterobacter*	Nanopore with error profile	Bracken (standard database)	G	97.83 (96.31:99.12)	NA	*Klebsiella*	0.69 (0.27:1.51)	3.14:3.44:5.04	*Escherichia*	0.37 (0.15:0.9)	2.78:4.24:4.33
73	*Enterobacter*	Nanopore with error profile	Centrifuge (standard database)	G	94.88 (92.94:97.32)	NA	*Klebsiella*	0.54 (0.3:0.88)	1.87:2.63:2.98	*Lelliottia*	0.05 (0.02:0.17)	1.61:22.78:53.76
74	*Enterobacter*	Nanopore with error profile	Kraken (mini database)	G	78.53 (76.92:80.04)	NA	*Klebsiella*	0.56 (0.35:0.73)	1.38:2.55:2.69	*Lelliottia*	0.09 (0.06:0.14)	0.42:0.86:7.22
75	*Enterobacter*	Nanopore with error profile	Kraken (standard database)	G	96.68 (94.95:98.45)	NA	*Klebsiella*	0.68 (0.27:1.31)	2.74:3.22:4.76	*Escherichia*	0.26 (0.12:0.57)	1.54:2.54:2.58
76	*Enterobacter*	Nanopore with error-free profile	Bracken (mini database)	G	98.08 (96.94:99.22)	NA	*Klebsiella*	0.66 (0.27:1.27)	2.87:3.7:5.76	*Lelliottia*	0.02 (0.01:0.05)	0.33:2.39:7.55
77	*Enterobacter*	Nanopore with error-free profile	Bracken (standard database)	G	98.78 (97.83:99.52)	NA	*Klebsiella*	0.51 (0.23:1.11)	2.93:3.77:5.71	*Citrobacter*	0.11 (0.05:0.25)	0.8:2.28:2.54
78	*Enterobacter*	Nanopore with error-free profile	Centrifuge (standard database)	G	96.41 (94.67:98.3)	NA	*Escherichia*	0.32 (0.14:1)	2.14:2.8:3.9	*Lelliottia*	0.11 (0.04:0.51)	17.57:54.14:72.45
79	*Enterobacter*	Nanopore with error-free profile	Kraken (mini database)	G	96.89 (95.14:98.5)	NA	*Klebsiella*	0.72 (0.29:1.27)	2.59:3.72:5.6	*Lelliottia*	0.02 (0.01:0.05)	0.33:2.3:7.25
80	*Enterobacter*	Nanopore with error-free profile	Kraken (standard database)	G	97.97 (96.65:99.42)	NA	*Klebsiella*	0.52 (0.23:1.06)	2.34:3.74:5.44	*Citrobacter*	0.11 (0.05:0.25)	0.77:2.04:2.45
81	*Enterococcus*	Illumina with error profile	Bracken (mini database)	G	99.53 (99.19:99.65)	NA	*Streptococcus*	0.06 (0.04:0.08)	0.15:0.56:1.22	*Staphylococcus*	0.04 (0.02:0.07)	0.14:1.22:3.12
82	*Enterococcus*	Illumina with error profile	Bracken (standard database)	G	99.6 (99.27:99.77)	NA	*Streptococcus*	0.05 (0.04:0.1)	0.28:0.36:1.53	*Staphylococcus*	0.03 (0.01:0.06)	0.37:1.55:3.76
83	*Enterococcus*	Illumina with error profile	Centrifuge (standard database)	G	97.48 (96.42:98.12)	NA	*Streptococcus*	0.11 (0.05:0.24)	0.56:0.7:1.77	*Staphylococcus*	0.05 (0.01:0.09)	0.54:1.26:3.31
84	*Enterococcus*	Illumina with error profile	Kraken (mini database)	G	95.67 (94.79:96.29)	NA	*Streptococcus*	0.05 (0.03:0.07)	0.14:0.41:1.12	*Staphylococcus*	0.04 (0.03:0.07)	0.28:1.23:2.96
85	*Enterococcus*	Illumina with error profile	Kraken (standard database)	G	98.58 (97.78:98.82)	NA	*Streptococcus*	0.05 (0.04:0.09)	0.27:0.35:1.44	*Staphylococcus*	0.03 (0.01:0.06)	0.34:1.41:3.66
86	*Enterococcus*	Illumina with error-free profile	Bracken (mini database)	G	99.54 (99.19:99.63)	NA	*Streptococcus*	0.05 (0.04:0.08)	0.16:0.52:1.27	*Staphylococcus*	0.04 (0.03:0.08)	0.15:1.22:3.12
87	*Enterococcus*	Illumina with error-free profile	Bracken (standard database)	G	99.62 (99.24:99.77)	NA	*Streptococcus*	0.05 (0.03:0.1)	0.32:0.4:1.6	*Staphylococcus*	0.03 (0.01:0.05)	0.38:1.49:3.76
88	*Enterococcus*	Illumina with error-free profile	Centrifuge (standard database)	G	97.49 (96.38:98.13)	NA	*Streptococcus*	0.12 (0.04:0.23)	0.53:0.69:1.7	*Staphylococcus*	0.05 (0.01:0.08)	0.54:1.24:3.29
89	*Enterococcus*	Illumina with error-free profile	Kraken (mini database)	G	95.61 (94.79:96.29)	NA	*Staphylococcus*	0.05 (0.03:0.07)	0.3:1.21:2.97	*Streptococcus*	0.05 (0.04:0.07)	0.15:0.4:1.16
90	*Enterococcus*	Illumina with error-free profile	Kraken (standard database)	G	98.56 (97.81:98.85)	NA	*Streptococcus*	0.05 (0.03:0.09)	0.31:0.39:1.51	*Staphylococcus*	0.03 (0.01:0.05)	0.33:1.36:3.67
91	*Enterococcus*	Nanopore with error profile	Bracken (mini database)	G	98.81 (98.44:98.93)	NA	*Streptococcus*	0.06 (0.04:0.09)	0.21:0.53:1.26	*Staphylococcus*	0.05 (0.03:0.08)	0.17:1.32:3.28
92	*Enterococcus*	Nanopore with error profile	Bracken (standard database)	G	99.66 (99.41:99.75)	NA	*Streptococcus*	0.06 (0.04:0.11)	0.24:0.46:1.52	*Staphylococcus*	0.05 (0.03:0.08)	0.3:1.9:3.67
93	*Enterococcus*	Nanopore with error profile	Centrifuge (standard database)	G	98.1 (97.54:98.54)	NA	*Streptococcus*	0.17 (0.06:0.34)	0.66:1.04:1.6	*Staphylococcus*	0.07 (0.03:0.15)	0.58:1.24:3.59
94	*Enterococcus*	Nanopore with error profile	Kraken (mini database)	G	87.16 (86.35:87.55)	NA	*Staphylococcus*	0.05 (0.03:0.07)	0.28:1.14:2.86	*Streptococcus*	0.05 (0.04:0.07)	0.16:0.37:1.06
95	*Enterococcus*	Nanopore with error profile	Kraken (standard database)	G	99.02 (98.45:99.17)	NA	*Streptococcus*	0.06 (0.03:0.11)	0.22:0.45:1.51	*Staphylococcus*	0.05 (0.03:0.08)	0.26:1.87:3.65
96	*Enterococcus*	Nanopore with error-free profile	Bracken (mini database)	G	99.74 (99.43:99.83)	NA	*Streptococcus*	0.06 (0.03:0.11)	0.24:0.65:1.72	*Staphylococcus*	0.03 (0.01:0.06)	0.27:2.39:3.82
97	*Enterococcus*	Nanopore with error-free profile	Bracken (standard database)	G	99.8 (99.6:99.93)	NA	*Streptococcus*	0.08 (0.04:0.14)	0.28:1.01:1.69	*Staphylococcus*	0.05 (0.03:0.09)	0.41:2.77:4.04
98	*Enterococcus*	Nanopore with error-free profile	Centrifuge (standard database)	G	98.53 (97.89:99.03)	NA	*Streptococcus*	0.29 (0.1:0.45)	0.76:1.34:1.85	*Staphylococcus*	0.2 (0.09:0.38)	0.89:2.03:4.05
99	*Enterococcus*	Nanopore with error-free profile	Kraken (mini database)	G	99.22 (98.66:99.49)	NA	*Streptococcus*	0.06 (0.04:0.11)	0.25:0.61:1.71	*Staphylococcus*	0.03 (0.01:0.07)	0.42:2.36:3.82
100	*Enterococcus*	Nanopore with error-free profile	Kraken (standard database)	G	99.56 (99.39:99.72)	NA	*Streptococcus*	0.05 (0.04:0.13)	0.25:0.77:1.69	*Staphylococcus*	0.04 (0.03:0.08)	0.4:2.57:4.03

Supplementary Table 8. Continued

101	*Escherichia*	Illumina with error profile	Bracken (mini database)	G	92.75 (91.04:94.65)	NA	*Shigella*	3.5 (2.12:4.42)	5.38:5.55:6.41	*Salmonella*	1.38 (1.04:1.87)	2.86:4.21:12.55
102	*Escherichia*	Illumina with error profile	Bracken (standard database)	G	94.03 (92.45:97.22)	NA	*Shigella*	1.4 (0.48:2.07)	2.35:2.64:2.82	*Salmonella*	0.53 (0.34:0.7)	1.01:1.57:2.63
103	*Escherichia*	Illumina with error profile	Centrifuge (standard database)	G	62.64 (57.13:90.15)	NA	*Shigella*	1.2 (0.45:1.54)	1.89:3.41:3.61	*Salmonella*	0.26 (0.17:0.39)	0.87:2.76:3.32
104	*Escherichia*	Illumina with error profile	Kraken (mini database)	G	41.26 (36.01:51.3)	NA	*Shigella*	1.02 (0.67:1.11)	1.2:1.23:1.24	*Citrobacter*	0.84 (0.76:0.92)	1.03:1.47:1.68
105	*Escherichia*	Illumina with error profile	Kraken (standard database)	G	65.87 (58:83.78)	NA	*Citrobacter*	1.13 (0.61:1.49)	1.66:2.13:2.32	*Salmonella*	0.45 (0.3:0.56)	0.68:0.72:1.22
106	*Escherichia*	Illumina with error-free profile	Bracken (mini database)	G	92.85 (90.9:94.75)	NA	*Shigella*	3.45 (2.12:4.59)	5.31:5.87:6.15	*Salmonella*	1.35 (1.06:1.82)	2.72:3.84:13.11
107	*Escherichia*	Illumina with error-free profile	Bracken (standard database)	G	94.08 (92.39:97.27)	NA	*Shigella*	1.4 (0.48:2.05)	2.42:2.67:2.73	*Salmonella*	0.53 (0.35:0.69)	0.98:1.52:2.63
108	*Escherichia*	Illumina with error-free profile	Centrifuge (standard database)	G	62.69 (57.08:90.25)	NA	*Shigella*	1.18 (0.44:1.51)	1.88:3.44:3.46	*Salmonella*	0.26 (0.17:0.38)	0.89:2.76:3.3
109	*Escherichia*	Illumina with error-free profile	Kraken (mini database)	G	41.17 (36.05:51.35)	NA	*Shigella*	1.02 (0.67:1.11)	1.19:1.22:1.27	*Citrobacter*	0.84 (0.76:0.92)	1.04:1.49:1.61
110	*Escherichia*	Illumina with error-free profile	Kraken (standard database)	G	65.94 (58:83.72)	NA	*Citrobacter*	1.17 (0.62:1.49)	1.68:2.1:2.21	*Enterobacter*	0.84 (0.41:1.01)	1.1:1.15:1.24
111	*Escherichia*	Nanopore with error profile	Bracken (mini database)	G	93.07 (91.24:94.6)	NA	*Shigella*	2.1 (1.47:2.9)	4.67:5.63:6.06	*Salmonella*	1.06 (0.79:1.49)	2.22:3.03:5.61
112	*Escherichia*	Nanopore with error profile	Bracken (standard database)	G	96.64 (95.91:97.96)	NA	*Citrobacter*	0.7 (0.46:0.91)	1.11:1.51:1.56	*Klebsiella*	0.45 (0.33:0.6)	1.04:1.26:1.3
113	*Escherichia*	Nanopore with error profile	Centrifuge (standard database)	G	82.81 (79.66:92.69)	NA	*Shigella*	3.72 (0.81:4.51)	5.78:6.97:7.05	*Salmonella*	0.45 (0.28:0.69)	1.24:3.64:4.14
114	*Escherichia*	Nanopore with error profile	Kraken (mini database)	G	51.92 (48.18:58.13)	NA	*Shigella*	1.12 (0.74:1.28)	1.53:1.67:1.67	*Citrobacter*	1.1 (0.89:1.32)	1.57:1.89:2.03
115	*Escherichia*	Nanopore with error profile	Kraken (standard database)	G	90.81 (89.27:93.64)	NA	*Citrobacter*	0.7 (0.45:0.91)	1.11:1.51:1.55	*Klebsiella*	0.45 (0.33:0.62)	1.01:1.27:1.28
116	*Escherichia*	Nanopore with error-free profile	Bracken (mini database)	G	98.39 (97.77:99.03)	NA	*Salmonella*	0.24 (0.16:0.36)	0.51:0.9:1.42	*Citrobacter*	0.22 (0.13:0.32)	0.65:1.31:1.69
117	*Escherichia*	Nanopore with error-free profile	Bracken (standard database)	G	99.6 (99.26:99.88)	NA	*Citrobacter*	0.07 (0.03:0.12)	0.3:1.22:1.36	*Salmonella*	0.05 (0.02:0.09)	0.48:0.85:0.96
118	*Escherichia*	Nanopore with error-free profile	Centrifuge (standard database)	G	96.79 (94.46:98.3)	NA	*Salmonella*	0.22 (0.08:0.48)	1.19:3.18:4.05	*Shigella*	0.18 (0.09:0.32)	1.01:2.26:3.04
119	*Escherichia*	Nanopore with error-free profile	Kraken (mini database)	G	92.38 (90.68:95.36)	NA	*Citrobacter*	0.46 (0.27:0.62)	1:1.59:2.19	*Klebsiella*	0.23 (0.12:0.4)	0.69:1.1:1.11
120	*Escherichia*	Nanopore with error-free profile	Kraken (standard database)	G	94.42 (90.37:98.23)	NA	*Citrobacter*	0.06 (0.03:0.12)	0.3:1.2:1.35	*Salmonella*	0.05 (0.02:0.09)	0.45:0.78:0.88
121	*Klebsiella*	Illumina with error profile	Bracken (mini database)	G	97.96 (97.24:98.31)	NA	*Escherichia*	0.59 (0.42:0.92)	2:2.41:2.97	*Citrobacter*	0.14 (0.1:0.17)	1.25:1.58:1.93
122	*Klebsiella*	Illumina with error profile	Bracken (standard database)	G	98.78 (98.37:99.16)	NA	*Escherichia*	0.55 (0.32:0.88)	1.55:2.02:2.29	*Salmonella*	0.11 (0.09:0.18)	0.29:0.54:0.73
123	*Klebsiella*	Illumina with error profile	Centrifuge (standard database)	G	72.88 (70.69:75.4)	NA	*Escherichia*	0.56 (0.35:0.83)	1.51:2.33:2.58	*Citrobacter*	0.12 (0.1:0.15)	0.33:0.5:0.81
124	*Klebsiella*	Illumina with error profile	Kraken (mini database)	G	87.43 (86.61:88.63)	NA	*Escherichia*	0.29 (0.22:0.4)	0.62:0.87:1.47	*Salmonella*	0.13 (0.11:0.14)	0.2:0.53:0.53
125	*Klebsiella*	Illumina with error profile	Kraken (standard database)	G	94.82 (93.61:95.73)	NA	*Escherichia*	0.28 (0.19:0.38)	0.74:1.03:2.02	*Enterobacter*	0.11 (0.08:0.14)	0.21:0.37:0.53
126	*Klebsiella*	Illumina with error-free profile	Bracken (mini database)	G	97.83 (97.3:98.29)	NA	*Escherichia*	0.6 (0.44:0.86)	1.94:2.86:2.89	*Citrobacter*	0.13 (0.1:0.18)	1.17:1.88:1.98
127	*Klebsiella*	Illumina with error-free profile	Bracken (standard database)	G	98.76 (98.38:99.15)	NA	*Escherichia*	0.48 (0.29:0.92)	1.33:2.02:2.32	*Providencia*	0 (0:0)	0.02:0.13:0.88
128	*Klebsiella*	Illumina with error-free profile	Centrifuge (standard database)	G	72.74 (70.6:75.53)	NA	*Escherichia*	0.56 (0.35:0.81)	1.51:2.33:2.63	*Citrobacter*	0.12 (0.09:0.15)	0.32:0.53:0.83
129	*Klebsiella*	Illumina with error-free profile	Kraken (mini database)	G	87.47 (86.7:88.63)	NA	*Escherichia*	0.3 (0.22:0.38)	0.61:0.87:1.46	*Salmonella*	0.13 (0.11:0.15)	0.2:0.51:0.52
130	*Klebsiella*	Illumina with error-free profile	Kraken (standard database)	G	94.79 (93.69:95.71)	NA	*Escherichia*	0.27 (0.19:0.39)	0.75:1.02:2.04	*Enterobacter*	0.11 (0.08:0.14)	0.2:0.38:0.51
131	*Klebsiella*	Nanopore with error profile	Bracken (mini database)	G	97.19 (96.6:97.61)	NA	*Escherichia*	0.68 (0.42:1.02)	1.98:2.39:2.68	*Citrobacter*	0.16 (0.12:0.21)	0.32:1.49:1.65
132	*Klebsiella*	Nanopore with error profile	Bracken (standard database)	G	98.59 (97.98:98.99)	NA	*Escherichia*	0.59 (0.39:1.07)	1.73:2.33:2.61	*Citrobacter*	0.12 (0.08:0.17)	0.24:0.29:0.85
133	*Klebsiella*	Nanopore with error profile	Centrifuge (standard database)	G	86.8 (85.57:88.01)	NA	*Escherichia*	0.86 (0.65:1.25)	2.16:3.43:3.92	*Citrobacter*	0.25 (0.21:0.3)	0.55:0.75:1.32
134	*Klebsiella*	Nanopore with error profile	Kraken (mini database)	G	80.92 (80.11:81.44)	NA	*Escherichia*	0.38 (0.26:0.59)	0.88:1.03:1.53	*Salmonella*	0.15 (0.1:0.17)	0.25:0.48:0.53
135	*Klebsiella*	Nanopore with error profile	Kraken (standard database)	G	97.15 (96.42:97.65)	NA	*Escherichia*	0.54 (0.36:0.92)	1.45:2.21:2.25	*Citrobacter*	0.12 (0.08:0.16)	0.24:0.29:0.83
136	*Klebsiella*	Nanopore with error-free profile	Bracken (mini database)	G	98.82 (98.2:99.14)	NA	*Escherichia*	0.62 (0.41:1.03)	1.59:2.18:2.26	*Citrobacter*	0.08 (0.04:0.14)	0.34:0.94:1.25
137	*Klebsiella*	Nanopore with error-free profile	Bracken (standard database)	G	99.68 (99.43:99.82)	NA	*Escherichia*	0.16 (0.06:0.39)	0.95:1.64:2.29	*Citrobacter*	0.03 (0.01:0.07)	0.2:0.61:0.92
138	*Klebsiella*	Nanopore with error-free profile	Centrifuge (standard database)	G	96.35 (95.6:97.14)	NA	*Escherichia*	0.67 (0.38:1.03)	1.91:2.95:3.3	*Proteus*	0.16 (0.05:0.27)	0.52:0.87:1.55
139	*Klebsiella*	Nanopore with error-free profile	Kraken (mini database)	G	97.37 (96.58:97.92)	NA	*Escherichia*	0.58 (0.39:0.95)	1.48:2.12:2.15	*Citrobacter*	0.09 (0.05:0.15)	0.27:0.45:0.85
140	*Klebsiella*	Nanopore with error-free profile	Kraken (standard database)	G	98.27 (97.69:98.75)	NA	*Escherichia*	0.15 (0.06:0.36)	0.9:1.52:2.28	*Citrobacter*	0.03 (0.01:0.05)	0.2:0.6:0.9
141	*Micrococcus*	Illumina with error profile	Bracken (mini database)	G	98.59 (97.91:99.11)	NA	*Streptomyces*	0.09 (0.07:0.13)	0.2:5.71:5.84	*Pseudarthrobacter*	0.01 (0.01:0.02)	0.05:34.8:95.02
142	*Micrococcus*	Illumina with error profile	Bracken (standard database)	G	98.68 (97.28:99.53)	NA	*Streptomyces*	0.11 (0.06:0.18)	0.32:6.02:6.08	*Pseudarthrobacter*	0.02 (0.01:0.03)	0.69:55.71:90.88
143	*Micrococcus*	Illumina with error profile	Centrifuge (standard database)	G	94.24 (92.38:96.51)	NA	*Arthrobacter*	0.03 (0.02:0.05)	1.18:10.44:21.86	*Pseudarthrobacter*	0.02 (0.01:0.04)	0.39:15.88:35.4
144	*Micrococcus*	Illumina with error profile	Kraken (mini database)	G	90.08 (87.43:91.97)	NA	*Arthrobacter*	0.04 (0.03:0.04)	0.09:1.16:2.17	*Pseudarthrobacter*	0.02 (0.01:0.02)	0.05:23.9:72.03
145	*Micrococcus*	Illumina with error profile	Kraken (standard database)	G	97.2 (94.88:98.46)	NA	*Arthrobacter*	0.02 (0.01:0.05)	0.77:4.07:4.76	*Pseudarthrobacter*	0.01 (0:0.03)	0.42:48.01:82.44
146	*Micrococcus*	Illumina with error-free profile	Bracken (mini database)	G	98.56 (97.9:99.11)	NA	*Streptomyces*	0.09 (0.06:0.12)	0.2:6.09:6.2	*Pseudarthrobacter*	0.01 (0.01:0.02)	0.05:32.83:95.02
147	*Micrococcus*	Illumina with error-free profile	Bracken (standard database)	G	98.72 (97.3:99.56)	NA	*Streptomyces*	0.11 (0.05:0.19)	0.32:5.95:5.97	*Pseudarthrobacter*	0.01 (0:0.02)	0.64:53.86:90.81
148	*Micrococcus*	Illumina with error-free profile	Centrifuge (standard database)	G	94.23 (92.35:96.48)	NA	*Arthrobacter*	0.03 (0.02:0.05)	1.08:10.43:22.02	*Pseudarthrobacter*	0.02 (0.01:0.04)	0.44:15.87:35.35
149	*Micrococcus*	Illumina with error-free profile	Kraken (mini database)	G	90.13 (87.32:91.97)	NA	*Arthrobacter*	0.03 (0.02:0.04)	0.1:1.14:2.14	*Pseudarthrobacter*	0.02 (0.01:0.02)	0.05:23.84:71.91
150	*Micrococcus*	Illumina with error-free profile	Kraken (standard database)	G	97.22 (94.84:98.46)	NA	*Arthrobacter*	0.03 (0.01:0.05)	1.45:4.14:4.8	*Pseudarthrobacter*	0.01 (0:0.02)	0.34:45.49:82.33

Supplementary Table 8. Continued

151	*Micrococcus*	Nanopore with error profile	Bracken (mini database)	G	95.89 (95.15:96.56)	NA	*Streptomyces*	0.54 (0.46:0.62)	0.93:9.21:10.32	*Pseudarthrobacter*	0.01 (0.01:0.03)	0.25:42.9:90.53
152	*Micrococcus*	Nanopore with error profile	Bracken (standard database)	G	99.18 (97.86:99.68)	NA	*Arthrobacter*	0.02 (0.01:0.04)	2.61:5.31:5.94	*Pseudarthrobacter*	0.01 (0.01:0.03)	31.63:77.71:89.23
153	*Micrococcus*	Nanopore with error profile	Centrifuge (standard database)	G	96.24 (94.82:97.61)	NA	*Arthrobacter*	0.04 (0.03:0.06)	1.42:9.88:22.05	*Pseudarthrobacter*	0.02 (0.01:0.05)	0.58:21.47:43.85
154	*Micrococcus*	Nanopore with error profile	Kraken (mini database)	G	80.28 (78.6:82.08)	NA	*Streptomyces*	0.49 (0.43:0.55)	0.6:3.78:4.05	*Pseudarthrobacter*	0.02 (0.01:0.03)	0.07:27.72:65.62
155	*Micrococcus*	Nanopore with error profile	Kraken (standard database)	G	98.27 (96.55:99.13)	NA	*Arthrobacter*	0.02 (0.01:0.04)	1.67:4.28:5.33	*Pseudarthrobacter*	0.01 (0.01:0.03)	25.52:72.33:84.03
156	*Micrococcus*	Nanopore with error-free profile	Bracken (mini database)	G	99.48 (98.53:99.83)	NA	*Pseudarthrobacter*	0.66 (0.02:1.31)	69.62:87.84:92.39	*Arthrobacter*	0.03 (0.01:0.04)	3.4:4.75:5.1
157	*Micrococcus*	Nanopore with error-free profile	Bracken (standard database)	G	99.24 (97.98:99.82)	NA	*Arthrobacter*	0.04 (0.01:0.07)	1.15:5.47:7.05	*Pseudarthrobacter*	0.03 (0.02:0.08)	48.76:80.58:88.54
158	*Micrococcus*	Nanopore with error-free profile	Centrifuge (standard database)	G	97.4 (96.17:98.74)	NA	*Arthrobacter*	0.03 (0.01:0.07)	3.69:15.93:22.81	*Pseudarthrobacter*	0.03 (0.01:0.07)	0.6:41.5:64.38
159	*Micrococcus*	Nanopore with error-free profile	Kraken (mini database)	G	98.55 (97.44:99.34)	NA	*Pseudarthrobacter*	0.5 (0.02:0.95)	65:82.08:86.35	*Arthrobacter*	0.03 (0.01:0.05)	2.66:4.18:4.58
160	*Micrococcus*	Nanopore with error-free profile	Kraken (standard database)	G	99.01 (97.5:99.69)	NA	*Arthrobacter*	0.04 (0.01:0.07)	1.11:5.29:6.83	*Pseudarthrobacter*	0.02 (0.02:0.04)	34.81:76.41:86.81
161	*Proteus*	Illumina with error profile	Bracken (mini database)	G	98.9 (97.96:99.34)	NA	*Salmonella*	0.26 (0.11:0.64)	1.24:1.93:2.11	*Escherichia*	0.12 (0.07:0.4)	0.9:1.01:2.16
162	*Proteus*	Illumina with error profile	Bracken (standard database)	G	99 (98.24:99.36)	NA	*Escherichia*	0.2 (0.12:0.48)	0.9:1.13:2.3	*Providencia*	0.06 (0.04:0.23)	0.57:1.21:1.9
163	*Proteus*	Illumina with error profile	Centrifuge (standard database)	G	95.78 (94.77:96.99)	NA	*Providencia*	0.11 (0.03:0.26)	0.98:2.15:2.69	*Bacteroides*	0 (0:0)	0.01:0.95:1.3
164	*Proteus*	Illumina with error profile	Kraken (mini database)	G	94.33 (93.2:95.31)	NA	*Providencia*	0.06 (0.04:0.14)	0.35:0.99:1.17	*Klebsiella*	0.05 (0.03:0.1)	0.18:0.39:0.8
165	*Proteus*	Illumina with error profile	Kraken (standard database)	G	95.63 (94.54:96.84)	NA	*Escherichia*	0.12 (0.08:0.22)	0.43:0.51:1.37	*Providencia*	0.06 (0.04:0.23)	0.56:1.18:1.72
166	*Proteus*	Illumina with error-free profile	Bracken (mini database)	G	98.85 (98.07:99.34)	NA	*Salmonella*	0.24 (0.07:0.56)	1.37:1.98:2.08	*Escherichia*	0.13 (0.07:0.39)	0.96:1.24:1.93
167	*Proteus*	Illumina with error-free profile	Bracken (standard database)	G	99.02 (98.28:99.38)	NA	*Escherichia*	0.18 (0.11:0.48)	0.85:0.98:2.23	*Providencia*	0.07 (0.04:0.23)	0.57:1.28:1.99
168	*Proteus*	Illumina with error-free profile	Centrifuge (standard database)	G	95.78 (94.75:97.02)	NA	*Providencia*	0.12 (0.03:0.26)	1.02:2.16:2.66	*Bacteroides*	0 (0:0)	0.01:0.97:1.27
169	*Proteus*	Illumina with error-free profile	Kraken (mini database)	G	94.31 (93.22:95.31)	NA	*Providencia*	0.06 (0.04:0.14)	0.36:0.97:1.21	*Klebsiella*	0.05 (0.03:0.11)	0.19:0.38:0.8
170	*Proteus*	Illumina with error-free profile	Kraken (standard database)	G	95.66 (94.59:96.83)	NA	*Escherichia*	0.12 (0.08:0.21)	0.43:0.51:1.33	*Providencia*	0.06 (0.04:0.24)	0.57:1.2:1.79
171	*Proteus*	Nanopore with error profile	Bracken (mini database)	G	98.12 (97.06:98.59)	NA	*Salmonella*	0.39 (0.12:0.73)	1.79:2.82:2.86	*Alistipes*	0.01 (0.01:0.01)	9.33:11.82:12.44
172	*Proteus*	Nanopore with error profile	Bracken (standard database)	G	98.83 (97.62:99.09)	NA	*Escherichia*	0.35 (0.21:0.75)	1.5:1.97:3.09	*Providencia*	0.11 (0.06:0.27)	0.56:1.4:1.9
173	*Proteus*	Nanopore with error profile	Centrifuge (standard database)	G	96.9 (95.86:97.79)	NA	*Providencia*	0.13 (0.05:0.34)	1.33:2.15:2.75	*Escherichia*	0.06 (0.03:0.46)	1:1.31:1.45
174	*Proteus*	Nanopore with error profile	Kraken (mini database)	G	85.81 (84.89:86.45)	NA	*Providencia*	0.08 (0.05:0.14)	0.35:1:1.14	*Alistipes*	0.01 (0.01:0.01)	7.46:9.45:9.95
175	*Proteus*	Nanopore with error profile	Kraken (standard database)	G	97.28 (96.13:97.72)	NA	*Escherichia*	0.3 (0.16:0.55)	1.09:1.33:2.28	*Providencia*	0.11 (0.06:0.27)	0.5:1.19:1.85
176	*Proteus*	Nanopore with error-free profile	Bracken (mini database)	G	99.15 (98.15:99.5)	NA	*Escherichia*	0.26 (0.13:0.68)	1.32:1.59:2.3	*Salmonella*	0.12 (0.05:0.33)	1.23:1.62:1.84
177	*Proteus*	Nanopore with error-free profile	Bracken (standard database)	G	99.32 (98.92:99.71)	NA	*Escherichia*	0.15 (0.08:0.36)	0.54:1.49:2.34	*Providencia*	0.09 (0.02:0.3)	0.97:1.73:1.74
178	*Proteus*	Nanopore with error-free profile	Centrifuge (standard database)	G	97.57 (96.58:98.49)	NA	*Providencia*	0.17 (0.05:0.45)	1.6:2.03:2.37	*Salmonella*	0.09 (0.04:0.29)	1.25:1.81:1.87
179	*Proteus*	Nanopore with error-free profile	Kraken (mini database)	G	97.88 (96.79:98.55)	NA	*Escherichia*	0.22 (0.11:0.51)	0.97:1.13:1.8	*Providencia*	0.12 (0.06:0.26)	0.75:1.33:1.73
180	*Proteus*	Nanopore with error-free profile	Kraken (standard database)	G	98.22 (97.35:98.96)	NA	*Escherichia*	0.1 (0.06:0.27)	0.43:1.25:2.01	*Providencia*	0.09 (0.02:0.29)	0.96:1.69:1.74
181	*Pseudomonas*	Illumina with error profile	Bracken (mini database)	G	99.68 (99.49:99.77)	NA	*Burkholderia*	0.01 (0.01:0.02)	0.03:0.09:0.65	*Staphylococcus*	0 (0:0)	0.25:0.59:0.67
182	*Pseudomonas*	Illumina with error profile	Bracken (standard database)	G	99.8 (99.51:99.95)	NA	*Burkholderia*	0.01 (0.01:0.02)	0.07:0.24:0.44	*Stenotrophomonas*	0.01 (0.01:0.03)	0.04:0.22:0.42
183	*Pseudomonas*	Illumina with error profile	Centrifuge (standard database)	G	98.77 (97.75:99.62)	NA	*Burkholderia*	0.01 (0:0.01)	0.05:0.4:1.03	*Staphylococcus*	0 (0:0)	0.03:0.44:0.56
184	*Pseudomonas*	Illumina with error profile	Kraken (mini database)	G	95.75 (94.3:97.08)	NA	*Burkholderia*	0.01 (0.01:0.02)	0.03:0.08:0.59	*Staphylococcus*	0 (0:0)	0.25:0.57:0.65
185	*Pseudomonas*	Illumina with error profile	Kraken (standard database)	G	99 (97.72:99.81)	NA	*Stenotrophomonas*	0.01 (0:0.02)	0.04:0.21:0.41	*Proteus*	0 (0:0)	0:0.24:0.34
186	*Pseudomonas*	Illumina with error-free profile	Bracken (mini database)	G	99.65 (99.48:99.76)	NA	*Burkholderia*	0.01 (0.01:0.02)	0.03:0.08:0.63	*Staphylococcus*	0 (0:0)	0.03:0.52:0.65
187	*Pseudomonas*	Illumina with error-free profile	Bracken (standard database)	G	99.8 (99.51:99.95)	NA	*Burkholderia*	0.01 (0.01:0.02)	0.08:0.26:0.5	*Stenotrophomonas*	0.01 (0.01:0.02)	0.05:0.23:0.42
188	*Pseudomonas*	Illumina with error-free profile	Centrifuge (standard database)	G	98.76 (97.72:99.62)	NA	*Burkholderia*	0.01 (0:0.01)	0.05:0.37:1.01	*Staphylococcus*	0 (0:0)	0.01:0.41:0.57
189	*Pseudomonas*	Illumina with error-free profile	Kraken (mini database)	G	95.75 (94.31:97.06)	NA	*Burkholderia*	0.01 (0.01:0.02)	0.03:0.07:0.57	*Staphylococcus*	0 (0:0)	0.03:0.49:0.64
190	*Pseudomonas*	Illumina with error-free profile	Kraken (standard database)	G	99.02 (97.75:99.8)	NA	*Stenotrophomonas*	0.01 (0.01:0.02)	0.05:0.21:0.41	*Proteus*	0 (0:0)	0:0.23:0.32
191	*Pseudomonas*	Nanopore with error profile	Bracken (mini database)	G	98.38 (98.18:98.53)	NA	*Burkholderia*	0.03 (0.02:0.05)	0.11:0.17:0.43	*Staphylococcus*	0.01 (0.01:0.01)	0.02:0.54:0.69
192	*Pseudomonas*	Nanopore with error profile	Bracken (standard database)	G	99.74 (99.24:99.94)	NA	*Cupriavidus*	0.04 (0.02:0.07)	0.22:0.32:0.36	*Stenotrophomonas*	0.03 (0.02:0.05)	0.1:0.25:0.37
193	*Pseudomonas*	Nanopore with error profile	Centrifuge (standard database)	G	98.87 (98.3:99.61)	NA	*Burkholderia*	0.02 (0.01:0.04)	0.12:0.47:0.81	*NA*	NA	NA
194	*Pseudomonas*	Nanopore with error profile	Kraken (mini database)	G	86.11 (85.08:86.96)	NA	*Burkholderia*	0.04 (0.02:0.05)	0.09:0.14:0.36	*Staphylococcus*	0.01 (0.01:0.01)	0.02:0.46:0.61
195	*Pseudomonas*	Nanopore with error profile	Kraken (standard database)	G	99.32 (98.33:99.67)	NA	*Stenotrophomonas*	0.03 (0.01:0.04)	0.1:0.24:0.36	*Comamonas*	0.01 (0.01:0.02)	0.03:0.22:0.31
196	*Pseudomonas*	Nanopore with error-free profile	Bracken (mini database)	G	99.86 (99.47:99.97)	NA	*Stenotrophomonas*	0.05 (0.03:0.07)	0.12:0.37:0.5	*Burkholderia*	0.04 (0.02:0.06)	0.2:0.68:0.84
197	*Pseudomonas*	Nanopore with error-free profile	Bracken (standard database)	G	99.95 (99.77:100)	NA	*Stenotrophomonas*	0.07 (0.05:0.12)	0.32:0.49:0.53	*Burkholderia*	0.04 (0.02:0.09)	0.44:0.62:0.66
198	*Pseudomonas*	Nanopore with error-free profile	Centrifuge (standard database)	G	99.37 (98.83:99.89)	NA	*Burkholderia*	0.04 (0.01:0.05)	0.19:0.79:0.98	*Serratia*	0.02 (0.01:0.05)	0.22:0.54:0.62
199	*Pseudomonas*	Nanopore with error-free profile	Kraken (mini database)	G	99.46 (98.64:99.88)	NA	*Stenotrophomonas*	0.05 (0.03:0.07)	0.13:0.4:0.54	*Burkholderia*	0.04 (0.02:0.08)	0.21:0.67:0.84
200	*Pseudomonas*	Nanopore with error-free profile	Kraken (standard database)	G	99.53 (98.89:99.96)	NA	*Stenotrophomonas*	0.07 (0.05:0.12)	0.3:0.48:0.53	*Burkholderia*	0.03 (0.01:0.07)	0.38:0.46:0.48

Supplementary Table 8. Continued

201	*Staphylococcus*	Illumina with error profile	Bracken (mini database)	G	99.34 (99.14:99.7)	NA	*Pseudomonas*	0 (0:0)	0:0.76:2.69	*NA*	NA	NA
202	*Staphylococcus*	Illumina with error profile	Bracken (mini database)	G	99.85 (99.75:99.9)	NA	*Bradyrhizobium*	0 (0:0.02)	0.18:0.24:0.25	*NA*	NA	NA
203	*Staphylococcus*	Illumina with error profile	Bracken (mini database)	G	99.66 (99.46:99.76)	NA	*Enterobacter*	0 (0:0)	0.02:0.85:1.1	*Lentilactobacillus*	0 (0:0)	1.54:2.07:2.2
204	*Staphylococcus*	Illumina with error profile	Bracken (standard database)	G	99.52 (99.24:99.73)	NA	*Pseudomonas*	0 (0:0)	0.28:2.23:2.72	*NA*	NA	NA
205	*Staphylococcus*	Illumina with error profile	Bracken (standard database)	G	99.88 (99.47:99.93)	NA	*Mammaliicoccus*	0.01 (0:0.02)	0.09:0.12:0.17	*NA*	NA	NA
206	*Staphylococcus*	Illumina with error profile	Bracken (standard database)	G	99.76 (99.65:99.87)	NA	*Acinetobacter*	0 (0:0)	0.25:0.82:0.96	*Lentilactobacillus*	0 (0:0)	1.45:2.22:2.41
207	*Staphylococcus*	Illumina with error profile	Centrifuge (standard database)	G	84.36 (82.05:93.43)	NA	*Streptococcus*	0.01 (0.01:0.02)	0.05:0.2:0.87	*Pseudomonas*	0 (0:0)	0:1.41:2.66
208	*Staphylococcus*	Illumina with error profile	Centrifuge (standard database)	G	98.07 (97.67:99.06)	NA	*Homo*	0.04 (0.01:0.08)	0.14:0.17:0.27	*NA*	NA	NA
209	*Staphylococcus*	Illumina with error profile	Centrifuge (standard database)	G	98.4 (97.36:98.96)	NA	*Lactobacillus*	0.01 (0:0.01)	0.04:0.53:1.55	*NA*	NA	NA
210	*Staphylococcus*	Illumina with error profile	Kraken (mini database)	G	96 (95.13:96.67)	NA	*Pseudomonas*	0 (0:0)	0:0.72:2.56	*Streptococcus*	0 (0:0)	0.02:0.21:0.84
211	*Staphylococcus*	Illumina with error profile	Kraken (mini database)	G	97.37 (96.78:98.05)	NA	*Bradyrhizobium*	0 (0:0.01)	0.16:0.21:0.22	*NA*	NA	NA
212	*Staphylococcus*	Illumina with error profile	Kraken (mini database)	G	96.62 (93.48:97.17)	NA	*Enterobacter*	0 (0:0)	0.01:0.74:0.96	*Lentilactobacillus*	0 (0:0)	1.48:1.98:2.11
213	*Staphylococcus*	Illumina with error profile	Kraken (standard database)	G	97.84 (97.27:98.55)	NA	*Pseudomonas*	0 (0:0)	0.01:1.94:2.66	*Streptococcus*	0 (0:0.01)	0.04:0.41:0.86
214	*Staphylococcus*	Illumina with error profile	Kraken (standard database)	G	99.35 (98.7:99.61)	NA	*Mammaliicoccus*	0.01 (0:0.02)	0.09:0.11:0.17	*NA*	NA	NA
215	*Staphylococcus*	Illumina with error profile	Kraken (standard database)	G	99.13 (98.81:99.42)	NA	*Acinetobacter*	0 (0:0)	0.09:0.71:0.86	*Lentilactobacillus*	0 (0:0)	1.43:2.19:2.38
216	*Staphylococcus*	Illumina with error-free profile	Bracken (mini database)	G	99.35 (99.15:99.66)	NA	*Pseudomonas*	0 (0:0)	0.01:0.65:2.68	*NA*	NA	NA
217	*Staphylococcus*	Illumina with error-free profile	Bracken (mini database)	G	99.86 (99.73:99.9)	NA	*Bradyrhizobium*	0 (0:0)	0.11:0.21:0.23	*NA*	NA	NA
218	*Staphylococcus*	Illumina with error-free profile	Bracken (mini database)	G	99.67 (99.47:99.78)	NA	*Acinetobacter*	0.01 (0:0.01)	0.02:0.07:0.83	*Enterobacter*	0 (0:0)	0.01:0.89:1.12
219	*Staphylococcus*	Illumina with error-free profile	Bracken (standard database)	G	99.52 (99.28:99.73)	NA	*Pseudomonas*	0 (0:0)	0.01:2.1:2.72	*NA*	NA	NA
220	*Staphylococcus*	Illumina with error-free profile	Bracken (standard database)	G	99.88 (99.5:99.93)	NA	*Bradyrhizobium*	0 (0:0.01)	0.3:0.37:0.39	*NA*	NA	NA
221	*Staphylococcus*	Illumina with error-free profile	Bracken (standard database)	G	99.77 (99.65:99.87)	NA	*Acinetobacter*	0 (0:0)	0.44:0.85:0.95	*Lentilactobacillus*	0 (0:0)	1.09:2.14:2.4
222	*Staphylococcus*	Illumina with error-free profile	Centrifuge (standard database)	G	84.32 (82.08:93.42)	NA	*Streptococcus*	0.01 (0.01:0.02)	0.05:0.2:0.87	*Pseudomonas*	0 (0:0)	0:1.54:2.66
223	*Staphylococcus*	Illumina with error-free profile	Centrifuge (standard database)	G	98.07 (97.68:99.05)	NA	*Homo*	0.03 (0.01:0.08)	0.14:0.19:0.24	*NA*	NA	NA
224	*Staphylococcus*	Illumina with error-free profile	Centrifuge (standard database)	G	98.41 (97.31:98.97)	NA	*Lactobacillus*	0.01 (0:0.02)	0.04:0.65:1.54	*NA*	NA	NA
225	*Staphylococcus*	Illumina with error-free profile	Kraken (mini database)	G	96 (95.14:96.62)	NA	*Pseudomonas*	0 (0:0)	0.01:0.49:2.56	*Streptococcus*	0 (0:0)	0.02:0.26:0.84
226	*Staphylococcus*	Illumina with error-free profile	Kraken (mini database)	G	97.31 (96.82:98.05)	NA	*Bradyrhizobium*	0 (0:0)	0.08:0.18:0.2	*NA*	NA	NA
227	*Staphylococcus*	Illumina with error-free profile	Kraken (mini database)	G	96.62 (93.54:97.14)	NA	*Acinetobacter*	0.01 (0:0.01)	0.02:0.07:0.76	*Enterobacter*	0 (0:0)	0.01:0.76:0.96
228	*Staphylococcus*	Illumina with error-free profile	Kraken (standard database)	G	97.83 (97.27:98.49)	NA	*Pseudomonas*	0 (0:0)	0:1.73:2.66	*Streptococcus*	0 (0:0.01)	0.05:0.33:0.86
229	*Staphylococcus*	Illumina with error-free profile	Kraken (standard database)	G	99.37 (98.69:99.6)	NA	*Bradyrhizobium*	0 (0:0.01)	0.28:0.37:0.39	*NA*	NA	NA
230	*Staphylococcus*	Illumina with error-free profile	Kraken (standard database)	G	99.15 (98.81:99.43)	NA	*Enterobacter*	0 (0:0)	0.9:1.2:1.28	*Lentilactobacillus*	0 (0:0)	1.07:2.11:2.37
231	*Staphylococcus*	Nanopore with error profile	Bracken (mini database)	G	98.94 (98.79:99.07)	NA	*Pseudomonas*	0.03 (0.02:0.04)	0.08:0.33:2.51	*Streptococcus*	0.02 (0.01:0.02)	0.03:0.23:0.66
232	*Staphylococcus*	Nanopore with error profile	Bracken (mini database)	G	98.96 (98.74:99.1)	NA	*Bacillus*	0.04 (0.03:0.07)	0.17:0.25:0.33	*NA*	NA	NA
233	*Staphylococcus*	Nanopore with error profile	Bracken (mini database)	G	98.85 (98.34:98.98)	NA	*Acinetobacter*	0.02 (0.01:0.03)	0.04:0.4:1.09	*Lentilactobacillus*	0.01 (0.01:0.01)	1.5:2.68:2.98
234	*Staphylococcus*	Nanopore with error profile	Bracken (standard database)	G	99.85 (99.78:99.9)	NA	*Pseudomonas*	0.01 (0.01:0.02)	1.06:2.59:2.97	*Streptococcus*	0.01 (0.01:0.04)	0.11:0.45:0.55
235	*Staphylococcus*	Nanopore with error profile	Bracken (standard database)	G	99.91 (99.68:99.95)	NA	*Bacillus*	0.02 (0.01:0.03)	0.07:0.19:0.3	*NA*	NA	NA
236	*Staphylococcus*	Nanopore with error profile	Bracken (standard database)	G	99.87 (99.79:99.93)	NA	*Homo*	0.02 (0.01:0.02)	0.06:0.28:0.44	*Acinetobacter*	0.01 (0.01:0.02)	0.76:1.34:1.49
237	*Staphylococcus*	Nanopore with error profile	Centrifuge (standard database)	G	94.97 (94.06:96.9)	NA	*Streptococcus*	0.02 (0.01:0.03)	0.11:0.35:1.3	*Pseudomonas*	0.01 (0.01:0.01)	0.02:3.92:4.96
238	*Staphylococcus*	Nanopore with error profile	Centrifuge (standard database)	G	98.67 (97.99:99.37)	NA	*Homo*	0.04 (0.02:0.08)	0.18:0.29:0.3	*NA*	NA	NA
239	*Staphylococcus*	Nanopore with error profile	Centrifuge (standard database)	G	98.84 (98.07:99.31)	NA	*Acinetobacter*	0.01 (0.01:0.01)	0.07:0.98:1.27	*NA*	NA	NA
240	*Staphylococcus*	Nanopore with error profile	Kraken (mini database)	G	88.1 (87.48:88.49)	NA	*Pseudomonas*	0.03 (0.02:0.04)	0.05:0.2:2.15	*Streptococcus*	0.02 (0.01:0.02)	0.04:0.16:0.6
241	*Staphylococcus*	Nanopore with error profile	Kraken (mini database)	G	88.11 (87.51:88.6)	NA	*Homo*	0.1 (0.08:0.13)	0.15:0.19:0.19	*NA*	NA	NA
242	*Staphylococcus*	Nanopore with error profile	Kraken (mini database)	G	87.33 (85.24:87.97)	NA	*Acinetobacter*	0.02 (0.01:0.03)	0.04:0.31:0.94	*Lentilactobacillus*	0.01 (0.01:0.01)	1.18:2.32:2.6
243	*Staphylococcus*	Nanopore with error profile	Kraken (standard database)	G	99.22 (98.98:99.38)	NA	*Pseudomonas*	0.01 (0.01:0.01)	0.62:2.49:2.96	*Streptococcus*	0.01 (0.01:0.04)	0.1:0.43:0.55
244	*Staphylococcus*	Nanopore with error profile	Kraken (standard database)	G	99.53 (99.12:99.63)	NA	*Paenibacillus*	0.01 (0.01:0.02)	0.09:0.21:0.25	*NA*	NA	NA
245	*Staphylococcus*	Nanopore with error profile	Kraken (standard database)	G	99.39 (99.09:99.57)	NA	*Homo*	0.02 (0.01:0.02)	0.06:0.28:0.44	*Acinetobacter*	0.01 (0.01:0.02)	0.46:1.28:1.49
246	*Staphylococcus*	Nanopore with error-free profile	Bracken (mini database)	G	99.94 (99.87:99.97)	NA	*Streptococcus*	0.05 (0.02:0.07)	0.38:0.6:0.65	*Enterococcus*	0.01 (0.01:0.07)	0.14:0.18:0.19
247	*Staphylococcus*	Nanopore with error-free profile	Bracken (mini database)	G	99.97 (99.85:100)	NA	*Bacillus*	0.05 (0.02:0.08)	0.22:0.28:0.3	*NA*	NA	NA
248	*Staphylococcus*	Nanopore with error-free profile	Bracken (mini database)	G	99.93 (99.84:99.97)	NA	*Homo*	0.02 (0.01:0.04)	0.32:0.54:0.59	*Macrococcus*	0.01 (0.01:0.16)	0.23:0.24:0.24
249	*Staphylococcus*	Nanopore with error-free profile	Bracken (standard database)	G	99.97 (99.94:99.99)	NA	*Streptococcus*	0.04 (0.03:0.09)	0.42:0.65:0.71	*NA*	NA	NA
250	*Staphylococcus*	Nanopore with error-free profile	Bracken (standard database)	G	99.96 (99.82:100)	NA	*Mammaliicoccus*	0.08 (0.03:0.11)	0.15:0.2:0.21	*NA*	NA	NA

Supplementary Table 8. Continued

251	*Staphylococcus*	Nanopore with error-free profile	Bracken (standard database)	G	99.94 (99.87:99.98)	NA	*Macrococcus*	0.05 (0.01:0.15)	0.23:0.23:0.23	*Mammaliicoccus*	0.03 (0.01:0.05)	0.1:0.16:0.18
252	*Staphylococcus*	Nanopore with error-free profile	Centrifuge (standard database)	G	98.87 (98.52:99.38)	NA	*Streptococcus*	0.03 (0.01:0.06)	0.28:0.62:0.71	*NA*	NA	NA
253	*Staphylococcus*	Nanopore with error-free profile	Centrifuge (standard database)	G	99.4 (98.6:99.83)	NA	*Homo*	0.04 (0.01:0.06)	0.16:0.36:0.51	*NA*	NA	NA
254	*Staphylococcus*	Nanopore with error-free profile	Centrifuge (standard database)	G	99.45 (98.99:99.75)	NA	*Lactobacillus*	0.02 (0.01:0.07)	0.54:1.2:1.37	*Acinetobacter*	0.01 (0.01:0.03)	0.96:1.27:1.35
255	*Staphylococcus*	Nanopore with error-free profile	Kraken (mini database)	G	99.53 (99.38:99.66)	NA	*Streptococcus*	0.05 (0.02:0.07)	0.38:0.6:0.65	*Enterococcus*	0.01 (0.01:0.07)	0.13:0.17:0.18
256	*Staphylococcus*	Nanopore with error-free profile	Kraken (mini database)	G	99.85 (99.54:99.95)	NA	*Bacillus*	0.04 (0.01:0.05)	0.21:0.27:0.29	*NA*	NA	NA
257	*Staphylococcus*	Nanopore with error-free profile	Kraken (mini database)	G	99.74 (99.4:99.86)	NA	*Homo*	0.01 (0.01:0.04)	0.32:0.54:0.59	*Macrococcus*	0.01 (0.01:0.15)	0.23:0.24:0.24
258	*Staphylococcus*	Nanopore with error-free profile	Kraken (standard database)	G	99.65 (99.5:99.77)	NA	*Streptococcus*	0.04 (0.03:0.09)	0.42:0.65:0.71	*NA*	NA	NA
259	*Staphylococcus*	Nanopore with error-free profile	Kraken (standard database)	G	99.93 (99.65:99.96)	NA	*Mammaliicoccus*	0.08 (0.03:0.11)	0.15:0.2:0.21	*NA*	NA	NA
260	*Staphylococcus*	Nanopore with error-free profile	Kraken (standard database)	G	99.89 (99.78:99.94)	NA	*Mammaliicoccus*	0.03 (0.01:0.05)	0.1:0.16:0.18	*Macrococcus*	0.01 (0.01:0.06)	0.23:0.23:0.23
261	*Streptococcus*	Illumina with error profile	Bracken (mini database)	G	99.4 (99.3:99.57)	NA	*Gemella*	0.1 (0.02:0.17)	0.28:0.42:0.79	*Neisseria*	0 (0:0.01)	0.9:3.57:4.24
262	*Streptococcus*	Illumina with error profile	Bracken (mini database)	G	99.78 (99.52:99.89)	NA	*Escherichia*	0.02 (0.01:0.02)	0.89:3.09:3.64	*Homo*	0 (0:0)	0.01:0.47:1.81
263	*Streptococcus*	Illumina with error profile	Bracken (mini database)	G	99.9 (99.81:99.93)	NA	*Enterococcus*	0 (0:0.03)	0.1:0.23:0.23	*Staphylococcus*	0 (0:0.03)	0.07:0.59:0.82
264	*Streptococcus*	Illumina with error profile	Bracken (standard database)	G	99.46 (99.21:99.67)	NA	*Gemella*	0.11 (0.02:0.22)	0.32:0.55:1.13	*Neisseria*	0 (0:0.02)	1.03:4:4.74
265	*Streptococcus*	Illumina with error profile	Bracken (standard database)	G	99.82 (99.5:99.98)	NA	*Escherichia*	0.01 (0:0.02)	0.84:1.12:1.21	*Homo*	0 (0:0)	0.19:1.5:1.83
266	*Streptococcus*	Illumina with error profile	Bracken (standard database)	G	99.99 (99.95:100)	NA	*Enterococcus*	0.07 (0:0.09)	0.29:0.3:0.31	*Staphylococcus*	0.01 (0:0.03)	0.27:0.68:0.8
267	*Streptococcus*	Illumina with error profile	Centrifuge (standard database)	G	96.09 (94.67:97.32)	NA	*Filifactor*	0.01 (0:0.01)	0.04:0.52:0.64	*Neisseria*	0 (0:0.01)	0.05:0.95:4.62
268	*Streptococcus*	Illumina with error profile	Centrifuge (standard database)	G	98.88 (98.04:99.6)	NA	*Homo*	0.01 (0:0.02)	0.04:0.88:1.87	*Escherichia*	0 (0:0.1)	0.61:0.75:0.79
269	*Streptococcus*	Illumina with error profile	Centrifuge (standard database)	G	99.42 (98.69:99.7)	NA	*Filifactor*	0 (0:0)	1.81:2.29:2.41	*Homo*	0 (0:0.01)	0.37:0.79:0.89
270	*Streptococcus*	Illumina with error profile	Kraken (mini database)	G	89.93 (88.1:92.43)	NA	*Gemella*	0.09 (0.02:0.16)	0.25:0.38:0.71	*Neisseria*	0 (0:0.01)	0.05:2.73:3.62
271	*Streptococcus*	Illumina with error profile	Kraken (mini database)	G	97.63 (96.39:98.63)	NA	*Escherichia*	0.01 (0.01:0.02)	0.41:0.54:0.57	*Homo*	0 (0:0)	0.01:0.46:1.79
272	*Streptococcus*	Illumina with error profile	Kraken (mini database)	G	98.14 (97.31:98.55)	NA	*Enterococcus*	0 (0:0.03)	0.09:0.22:0.23	*Staphylococcus*	0 (0:0.03)	0.06:0.54:0.77
273	*Streptococcus*	Illumina with error profile	Kraken (standard database)	G	97.67 (96.27:98.26)	NA	*Gemella*	0.11 (0.02:0.21)	0.31:0.52:1.11	*Neisseria*	0 (0:0.01)	1.01:3.93:4.66
274	*Streptococcus*	Illumina with error profile	Kraken (standard database)	G	99.12 (98.43:99.94)	NA	*Escherichia*	0 (0:0)	0.52:0.71:0.77	*Homo*	0 (0:0)	0.19:1.5:1.83
275	*Streptococcus*	Illumina with error profile	Kraken (standard database)	G	99.78 (99.09:99.94)	NA	*Clostridioides*	0.01 (0:0.01)	0.11:0.26:0.3	*Staphylococcus*	0.01 (0:0.03)	0.24:0.67:0.79
276	*Streptococcus*	Illumina with error-free profile	Bracken (mini database)	G	99.38 (99.29:99.59)	NA	*Gemella*	0.1 (0.02:0.15)	0.28:0.4:0.76	*Neisseria*	0 (0:0.02)	1.56:3.75:4.3
277	*Streptococcus*	Illumina with error-free profile	Bracken (mini database)	G	99.76 (99.52:99.9)	NA	*Escherichia*	0.02 (0.01:0.02)	0.85:3.35:3.98	*Homo*	0 (0:0)	0.01:0.4:1.81
278	*Streptococcus*	Illumina with error-free profile	Bracken (mini database)	G	99.9 (99.83:99.93)	NA	*Enterococcus*	0 (0:0.03)	0.1:0.22:0.22	*Staphylococcus*	0 (0:0.03)	0.05:0.53:0.81
279	*Streptococcus*	Illumina with error-free profile	Bracken (standard database)	G	99.44 (99.23:99.68)	NA	*Gemella*	0.13 (0.03:0.22)	0.34:0.55:1.1	*Neisseria*	0 (0:0.02)	0.81:3.96:4.75
280	*Streptococcus*	Illumina with error-free profile	Bracken (standard database)	G	99.81 (99.52:99.98)	NA	*Escherichia*	0.01 (0.01:0.02)	0.83:1.21:1.32	*Homo*	0 (0:0)	0.19:1.51:1.84
281	*Streptococcus*	Illumina with error-free profile	Bracken (standard database)	G	99.99 (99.95:100)	NA	*Enterococcus*	0.06 (0:0.08)	0.29:0.3:0.3	*Staphylococcus*	0.01 (0:0.03)	0.25:0.69:0.81
282	*Streptococcus*	Illumina with error-free profile	Centrifuge (standard database)	G	96.09 (94.74:97.33)	NA	*Filifactor*	0.01 (0:0.01)	0.05:0.49:0.61	*Neisseria*	0 (0:0.01)	0.05:1.01:4.63
283	*Streptococcus*	Illumina with error-free profile	Centrifuge (standard database)	G	98.89 (98.02:99.6)	NA	*Homo*	0.01 (0:0.02)	0.04:0.74:1.83	*Escherichia*	0 (0:0.2)	0.62:0.74:0.77
284	*Streptococcus*	Illumina with error-free profile	Centrifuge (standard database)	G	99.42 (98.69:99.71)	NA	*Filifactor*	0 (0:0)	1.72:2.18:2.3	*Homo*	0 (0:0)	0.07:0.72:0.88
285	*Streptococcus*	Illumina with error-free profile	Kraken (mini database)	G	89.95 (88.09:92.35)	NA	*Gemella*	0.09 (0.02:0.14)	0.26:0.37:0.68	*Neisseria*	0 (0:0.01)	0.06:2.69:3.66
286	*Streptococcus*	Illumina with error-free profile	Kraken (mini database)	G	97.62 (96.45:98.6)	NA	*Escherichia*	0.02 (0.01:0.02)	0.44:0.56:0.59	*Homo*	0 (0:0)	0.01:0.39:1.79
287	*Streptococcus*	Illumina with error-free profile	Kraken (mini database)	G	98.15 (97.28:98.55)	NA	*Enterococcus*	0 (0:0.03)	0.09:0.22:0.22	*Staphylococcus*	0 (0:0.03)	0.05:0.49:0.77
288	*Streptococcus*	Illumina with error-free profile	Kraken (standard database)	G	97.67 (96.25:98.24)	NA	*Gemella*	0.13 (0.03:0.21)	0.33:0.53:1.08	*Neisseria*	0 (0:0.01)	0.33:3.79:4.66
289	*Streptococcus*	Illumina with error-free profile	Kraken (standard database)	G	99.12 (98.46:99.95)	NA	*Escherichia*	0 (0:0)	0.53:0.73:0.79	*Homo*	0 (0:0)	0.19:1.51:1.84
290	*Streptococcus*	Illumina with error-free profile	Kraken (standard database)	G	99.78 (99.11:99.94)	NA	*Enterococcus*	0.05 (0:0.08)	0.27:0.28:0.28	*Staphylococcus*	0.01 (0:0.03)	0.23:0.66:0.79
291	*Streptococcus*	Nanopore with error profile	Bracken (mini database)	G	98.27 (98.06:98.62)	NA	*Gemella*	0.05 (0.01:0.09)	0.23:0.38:0.74	*Neisseria*	0.01 (0.01:0.01)	0.03:3.36:4.53
292	*Streptococcus*	Nanopore with error profile	Bracken (mini database)	G	99.2 (99.03:99.34)	NA	*Homo*	0.06 (0.04:0.08)	0.11:0.16:2.93	*Escherichia*	0.02 (0.01:0.04)	0.15:1.48:1.97
293	*Streptococcus*	Nanopore with error profile	Bracken (mini database)	G	99.25 (99.14:99.33)	NA	*Enterococcus*	0.01 (0.01:0.03)	0.09:0.19:0.24	*Staphylococcus*	0.01 (0.01:0.03)	0.1:0.44:0.8
294	*Streptococcus*	Nanopore with error profile	Bracken (standard database)	G	99.59 (99.39:99.77)	NA	*Gemella*	0.06 (0.03:0.08)	0.22:0.48:0.87	*Neisseria*	0.01 (0.01:0.02)	3.51:4.7:5
295	*Streptococcus*	Nanopore with error profile	Bracken (standard database)	G	99.85 (99.61:99.98)	NA	*Escherichia*	0.02 (0.01:0.03)	0.37:1.18:1.38	*Homo*	0.01 (0.01:0.02)	0.03:2.17:2.84
296	*Streptococcus*	Nanopore with error profile	Bracken (standard database)	G	99.95 (99.91:99.98)	NA	*Staphylococcus*	0.07 (0.05:0.15)	0.34:0.6:0.67	*Clostridioides*	0.03 (0.02:0.04)	0.2:0.34:0.38
297	*Streptococcus*	Nanopore with error profile	Centrifuge (standard database)	G	97.44 (96.78:98.23)	NA	*Filifactor*	0.01 (0.01:0.01)	0.45:0.87:0.97	*Neisseria*	0.01 (0.01:0.02)	0.14:3.43:4.77
298	*Streptococcus*	Nanopore with error profile	Centrifuge (standard database)	G	99.31 (98.25:99.65)	NA	*Homo*	0.01 (0.01:0.03)	0.04:1.49:2.58	*Ndongobacter*	0.01 (0.01:0.01)	0.22:0.32:0.35
299	*Streptococcus*	Nanopore with error profile	Centrifuge (standard database)	G	99.57 (99.28:99.68)	NA	*Staphylococcus*	0.02 (0.01:0.04)	0.16:0.49:0.57	*Filifactor*	0.01 (0.01:0.01)	1.68:2.57:2.79
300	*Streptococcus*	Nanopore with error profile	Kraken (mini database)	G	82.79 (81.46:84.13)	NA	*Gemella*	0.04 (0.01:0.07)	0.19:0.32:0.61	*Neisseria*	0.01 (0.01:0.01)	0.03:2.49:3.64

Supplementary Table 8. Continued

301	*Streptococcus*	Nanopore with error profile	Kraken (mini database)	G	89.07 (88.52:89.67)	NA	*Homo*	0.05 (0.04:0.07)	0.09:0.15:2.65	*Escherichia*	0.01 (0.01:0.02)	0.12:0.41:0.53
302	*Streptococcus*	Nanopore with error profile	Kraken (mini database)	G	89.07 (88.6:89.46)	NA	*Staphylococcus*	0.02 (0.01:0.03)	0.07:0.41:0.73	*Enterococcus*	0.01 (0.01:0.03)	0.08:0.17:0.22
303	*Streptococcus*	Nanopore with error profile	Kraken (standard database)	G	98.34 (97.69:98.8)	NA	*Gemella*	0.05 (0.03:0.08)	0.21:0.47:0.86	*Neisseria*	0.01 (0.01:0.01)	2.98:4.56:4.95
304	*Streptococcus*	Nanopore with error profile	Kraken (standard database)	G	99.35 (99.07:99.73)	NA	*Escherichia*	0.02 (0.01:0.03)	0.35:1.05:1.22	*Homo*	0.01 (0.01:0.02)	0.03:2.17:2.84
305	*Streptococcus*	Nanopore with error profile	Kraken (standard database)	G	99.66 (99.39:99.73)	NA	*Staphylococcus*	0.06 (0.05:0.12)	0.32:0.6:0.67	*Clostridioides*	0.03 (0.02:0.04)	0.18:0.34:0.38
306	*Streptococcus*	Nanopore with error-free profile	Bracken (mini database)	G	99.68 (99.58:99.86)	NA	*Gemella*	0.06 (0.03:0.1)	0.25:0.65:1.28	*Enterococcus*	0.05 (0.02:0.11)	0.18:0.2:0.2
307	*Streptococcus*	Nanopore with error-free profile	Bracken (mini database)	G	99.87 (99.65:99.99)	NA	*Clostridioides*	0.06 (0.04:0.1)	0.14:0.16:0.17	*Escherichia*	0.01 (0.01:0.07)	0.65:1.05:1.15
308	*Streptococcus*	Nanopore with error-free profile	Bracken (mini database)	G	100 (99.94:100)	NA	*Enterococcus*	0.11 (0.05:0.19)	0.36:0.43:0.45	*Staphylococcus*	0.04 (0.02:0.05)	0.31:0.52:0.57
309	*Streptococcus*	Nanopore with error-free profile	Bracken (standard database)	G	99.74 (99.53:99.9)	NA	*Enterococcus*	0.05 (0.03:0.11)	0.2:0.25:0.26	*Gemella*	0.05 (0.02:0.09)	0.15:0.8:1.36
310	*Streptococcus*	Nanopore with error-free profile	Bracken (standard database)	G	99.94 (99.75:100)	NA	*Limosilactobacillus*	0.1 (0.08:0.13)	0.18:0.2:0.2	*Enterococcus*	0.05 (0.03:0.12)	0.25:0.25:0.25
311	*Streptococcus*	Nanopore with error-free profile	Bracken (standard database)	G	100 (99.98:100)	NA	*Enterococcus*	0.21 (0.07:0.26)	0.5:0.56:0.58	*Staphylococcus*	0.12 (0.06:0.17)	0.44:0.51:0.53
312	*Streptococcus*	Nanopore with error-free profile	Centrifuge (standard database)	G	98.94 (98.34:99.3)	NA	*Haemophilus*	0.02 (0.01:0.03)	0.44:0.66:0.71	*Neisseria*	0.01 (0.01:0.04)	2.11:3.89:4.34
313	*Streptococcus*	Nanopore with error-free profile	Centrifuge (standard database)	G	99.6 (98.66:99.92)	NA	*Bacillus*	0.03 (0.02:0.11)	0.16:0.16:0.16	*Homo*	0.01 (0.01:0.02)	0.03:1.6:2.15
314	*Streptococcus*	Nanopore with error-free profile	Centrifuge (standard database)	G	99.92 (99.72:99.96)	NA	*Staphylococcus*	0.14 (0.06:0.3)	0.43:0.5:0.52	*Enterococcus*	0.05 (0.03:0.14)	0.48:0.53:0.54
315	*Streptococcus*	Nanopore with error-free profile	Kraken (mini database)	G	98.71 (98.16:99.21)	NA	*Enterococcus*	0.05 (0.02:0.11)	0.17:0.19:0.19	*Gemella*	0.05 (0.03:0.1)	0.25:0.64:1.27
316	*Streptococcus*	Nanopore with error-free profile	Kraken (mini database)	G	99.7 (99.37:99.96)	NA	*Clostridioides*	0.06 (0.04:0.1)	0.14:0.16:0.17	*Escherichia*	0.01 (0.01:0.07)	0.56:0.89:0.97
317	*Streptococcus*	Nanopore with error-free profile	Kraken (mini database)	G	99.92 (99.66:99.97)	NA	*Enterococcus*	0.11 (0.05:0.19)	0.36:0.43:0.45	*Staphylococcus*	0.04 (0.02:0.06)	0.31:0.52:0.57
318	*Streptococcus*	Nanopore with error-free profile	Kraken (standard database)	G	99.44 (99.04:99.68)	NA	*Gemella*	0.05 (0.02:0.09)	0.15:0.79:1.35	*Lactobacillus*	0.03 (0.01:0.06)	0.2:0.24:0.25
319	*Streptococcus*	Nanopore with error-free profile	Kraken (standard database)	G	99.85 (99.56:100)	NA	*Limosilactobacillus*	0.1 (0.08:0.13)	0.18:0.2:0.2	*Enterococcus*	0.05 (0.03:0.12)	0.25:0.25:0.25
320	*Streptococcus*	Nanopore with error-free profile	Kraken (standard database)	G	99.98 (99.86:100)	NA	*Enterococcus*	0.21 (0.07:0.26)	0.5:0.56:0.58	*Staphylococcus*	0.12 (0.06:0.17)	0.44:0.51:0.53
321	*Bacillus cereus*	Illumina with error profile	Bracken (mini database)	S	39.8 (33.45:49.4)	NA	*Bacillus thuringiensis*	37.75 (22.28:42.63)	58.03:68.47:71.29	*Bacillus mycoides*	1.01 (0.81:1.41)	12.74:64.99:72.95
322	*Bacillus cereus*	Illumina with error profile	Bracken (standard database)	S	40.09 (32.46:55.3)	NA	*Bacillus mycoides*	3.19 (2.6:4.02)	35.47:88.9:90.43	*Bacillus albus*	0.6 (0.51:0.87)	5.79:9.21:77.6
323	*Bacillus cereus*	Illumina with error profile	Centrifuge (standard database)	S	13.05 (11.06:25.74)	3.58 (2.77:4.51)	*Bacillus mycoides*	0.69 (0.29:1.44)	53.11:83.07:89.34	*NA*	NA	NA
324	*Bacillus cereus*	Illumina with error profile	Kraken (mini database)	S	10.37 (8.23:13.74)	2.3 (1.8:4.2)	*Bacillus mycoides*	0.62 (0.53:0.86)	8.51:37.41:47.14	*Bacillus albus*	0.32 (0.28:0.49)	2.32:3.37:43.36
325	*Bacillus cereus*	Illumina with error profile	Kraken (standard database)	S	22.24 (16.85:32.68)	0.45 (0.31:0.83)	*Bacillus mycoides*	2.78 (2.23:3.53)	33.09:80.28:81.46	*Bacillus albus*	0.56 (0.47:0.83)	5.62:8.93:75.23
326	*Bacillus cereus*	Illumina with error-free profile	Bracken (mini database)	S	40.33 (33.56:50.61)	NA	*Bacillus thuringiensis*	37.36 (22.9:42.54)	58.04:68.83:70.75	*Bacillus mycoides*	1.02 (0.81:1.43)	12.68:64.26:72.4
327	*Bacillus cereus*	Illumina with error-free profile	Bracken (standard database)	S	40.14 (33.2:55.4)	NA	*Bacillus mycoides*	3.23 (2.59:4.16)	35.4:89.47:90.24	*Bacillus albus*	0.6 (0.5:0.87)	5.76:9.03:77.62
328	*Bacillus cereus*	Illumina with error-free profile	Centrifuge (standard database)	S	13.04 (11.04:25.75)	3.58 (2.74:4.51)	*Bacillus mycoides*	0.69 (0.3:1.46)	53.11:83.07:89.37	*NA*	NA	NA
329	*Bacillus cereus*	Illumina with error-free profile	Kraken (mini database)	S	10.34 (8.23:13.87)	2.3 (1.76:4.21)	*Bacillus mycoides*	0.64 (0.53:0.86)	8.54:37.42:46.99	*Bacillus albus*	0.31 (0.28:0.49)	2.32:3.31:43.41
330	*Bacillus cereus*	Illumina with error-free profile	Kraken (standard database)	S	22.12 (17:32.66)	0.45 (0.32:0.83)	*Bacillus mycoides*	2.8 (2.23:3.48)	33.1:80.27:81.37	*Bacillus albus*	0.56 (0.47:0.83)	5.59:8.75:75.32
331	*Bacillus cereus*	Nanopore with error profile	Bracken (mini database)	S	38.09 (32.1:47.03)	NA	*Bacillus thuringiensis*	37.22 (19.94:45.08)	55.8:63.08:74.33	*Bacillus mycoides*	1.9 (1.21:3.2)	12.22:62.85:68.98
332	*Bacillus cereus*	Nanopore with error profile	Bracken (standard database)	S	50.14 (39.89:61.92)	NA	*Bacillus mycoides*	4.7 (3.87:6.02)	43.95:93.21:93.32	*Bacillus albus*	0.88 (0.73:1.28)	4.27:8.07:79.19
333	*Bacillus cereus*	Nanopore with error profile	Centrifuge (standard database)	S	20.59 (14.83:35.98)	1.65 (1.12:2.03)	*Bacillus mycoides*	0.75 (0.28:1.59)	68.24:87.89:93.04	*NA*	NA	NA
334	*Bacillus cereus*	Nanopore with error profile	Kraken (mini database)	S	15.34 (11.34:18.77)	10.29 (9.75:11.28)	*Bacillus mycoides*	1.03 (0.88:1.19)	7.85:42.92:51.14	*Bacillus albus*	0.52 (0.45:0.78)	2.56:3.95:44.89
335	*Bacillus cereus*	Nanopore with error profile	Kraken (standard database)	S	42.97 (31.82:52.67)	0.38 (0.26:0.64)	*Bacillus mycoides*	4.46 (3.73:5.84)	40.75:86.36:86.74	*Bacillus albus*	0.86 (0.72:1.27)	4.21:7.91:77.56
336	*Bacillus cereus*	Nanopore with error-free profile	Bracken (mini database)	S	51.25 (41.24:67.94)	NA	*Bacillus thuringiensis*	27.9 (13.68:38.67)	66.39:79.44:86.99	*Bacillus mycoides*	1.39 (1.07:1.86)	16.06:90.75:93.35
337	*Bacillus cereus*	Nanopore with error-free profile	Bracken (standard database)	S	63.83 (36.78:81.77)	NA	*Bacillus mycoides*	2.01 (1.22:3.2)	54.3:97.63:98.76	*Bacillus albus*	0.43 (0.22:0.62)	4.22:9.55:87.64
338	*Bacillus cereus*	Nanopore with error-free profile	Centrifuge (standard database)	S	30.95 (15.81:49.19)	0.48 (0.34:0.66)	*Bacillus mycoides*	0.96 (0.4:1.96)	77.07:92.4:95.88	*NA*	NA	NA
339	*Bacillus cereus*	Nanopore with error-free profile	Kraken (mini database)	S	43.93 (32.33:57.75)	0.24 (0.11:0.48)	*Bacillus mycoides*	1.33 (1.02:1.76)	14.33:84.92:87.2	*Bacillus albus*	0.68 (0.5:1)	5:9.6:82.88
340	*Bacillus cereus*	Nanopore with error-free profile	Kraken (standard database)	S	60.08 (35.07:76.47)	0.06 (0.04:0.14)	*Bacillus mycoides*	1.99 (1.21:3.16)	53.63:96.72:97.83	*Bacillus albus*	0.43 (0.22:0.62)	4.21:9.54:87.43
341	*Bacteroides fragilis*	Illumina with error profile	Bracken (mini database)	S	98.17 (96.66:98.79)	NA	*Phocaeicola dorei*	0.09 (0.05:0.19)	0.58:0.97:2.23	*Bacteroides xylanisolvens*	0.04 (0.01:0.17)	1.02:1.15:3.31
342	*Bacteroides fragilis*	Illumina with error profile	Bracken (standard database)	S	97.02 (96.04:98.2)	NA	*Parabacteroides distasonis*	0.18 (0.08:0.44)	0.9:1.75:2.13	*Bacteroides xylanisolvens*	0.15 (0.06:0.25)	0.79:1.02:2.13
343	*Bacteroides fragilis*	Illumina with error profile	Centrifuge (standard database)	S	93.73 (92.14:95.89)	2.04 (1.38:2.61)	*Bacteroides ovatus*	0.17 (0.05:0.32)	0.51:0.95:1.93	*Bacteroides uniformis*	0.11 (0.02:0.24)	1.77:1.99:2
344	*Bacteroides fragilis*	Illumina with error profile	Kraken (mini database)	S	87.91 (85.48:90.3)	3.72 (2.69:5.08)	*Phocaeicola dorei*	0.07 (0.04:0.16)	0.45:0.79:1.37	*Bacteroides caecimuris*	0.01 (0:0.06)	0.24:0.32:1.12
345	*Bacteroides fragilis*	Illumina with error profile	Kraken (standard database)	S	92.65 (90.16:95.1)	0.96 (0.48:1.38)	*Bacteroides uniformis*	0.08 (0.03:0.16)	0.48:0.71:1.15	*Phocaeicola dorei*	0.06 (0.03:0.15)	0.54:0.7:1.17
346	*Bacteroides fragilis*	Illumina with error-free profile	Bracken (mini database)	S	98.19 (96.53:98.79)	NA	*Phocaeicola dorei*	0.1 (0.04:0.2)	0.62:0.97:2.22	*Bacteroides xylanisolvens*	0.03 (0.01:0.16)	1.04:1.14:3.11
347	*Bacteroides fragilis*	Illumina with error-free profile	Bracken (standard database)	S	97.06 (96.03:98.16)	NA	*Parabacteroides distasonis*	0.17 (0.09:0.44)	0.9:1.69:2.27	*Bacteroides xylanisolvens*	0.14 (0.06:0.26)	0.78:0.99:2.16
348	*Bacteroides fragilis*	Illumina with error-free profile	Centrifuge (standard database)	S	93.75 (92.17:95.89)	2.03 (1.31:2.65)	*Bacteroides ovatus*	0.16 (0.04:0.33)	0.51:0.96:1.89	*Bacteroides uniformis*	0.11 (0.02:0.23)	1.77:1.94:1.94
349	*Bacteroides fragilis*	Illumina with error-free profile	Kraken (mini database)	S	87.92 (85.43:90.26)	3.7 (2.71:5.07)	*Phocaeicola dorei*	0.07 (0.04:0.16)	0.47:0.79:1.38	*Bacteroides caecimuris*	0.01 (0:0.06)	0.24:0.3:1.12
350	*Bacteroides fragilis*	Illumina with error-free profile	Kraken (standard database)	S	92.66 (90.16:95.08)	0.93 (0.48:1.37)	*Bacteroides uniformis*	0.08 (0.03:0.16)	0.47:0.7:1.16	*Phocaeicola dorei*	0.06 (0.03:0.15)	0.54:0.7:1.16

Supplementary Table 8. Continued

351	*Bacteroides fragilis*	Nanopore with error profile	Bracken (mini database)	S	97.38 (95.98:98.14)	NA	*Bacteroides xylanisolvens*	0.08 (0.02:0.22)	1.05:1.59:2.59	*Bacteroides ovatus*	0.04 (0.02:0.14)	0.75:3.35:3.58
352	*Bacteroides fragilis*	Nanopore with error profile	Bracken (standard database)	S	97.32 (96.13:98.2)	NA	*Bacteroides xylanisolvens*	0.19 (0.08:0.44)	0.88:1.23:1.77	*Phocaeicola dorei*	0.08 (0.04:0.2)	0.81:1.08:1.57
353	*Bacteroides fragilis*	Nanopore with error profile	Centrifuge (standard database)	S	94.81 (92.96:96.55)	0.95 (0.65:1.28)	*Bacteroides ovatus*	0.27 (0.06:0.49)	0.81:1.15:2.55	*Bacteroides uniformis*	0.13 (0.03:0.49)	2.1:2.21:2.24
354	*Bacteroides fragilis*	Nanopore with error profile	Kraken (mini database)	S	80.68 (79.03:82.79)	11.79 (11.11:12.55)	*Phocaeicola dorei*	0.05 (0.03:0.14)	0.43:0.87:1.27	*Bacteroides caecimuris*	0.03 (0.01:0.15)	0.29:0.56:1.01
355	*Bacteroides fragilis*	Nanopore with error profile	Kraken (standard database)	S	94.52 (93.13:96.3)	0.52 (0.34:0.76)	*Bacteroides xylanisolvens*	0.16 (0.07:0.39)	0.81:0.93:1.12	*Phocaeicola dorei*	0.07 (0.04:0.18)	0.76:1.01:1.46
356	*Bacteroides fragilis*	Nanopore with error-free profile	Bracken (mini database)	S	97.66 (96.29:98.53)	NA	*Phocaeicola dorei*	0.11 (0.04:0.29)	0.92:1.78:2.11	*Bacteroides xylanisolvens*	0.1 (0.04:0.37)	1.51:1.73:2.17
357	*Bacteroides fragilis*	Nanopore with error-free profile	Bracken (standard database)	S	97.17 (95.93:98.21)	NA	*Phocaeicola dorei*	0.11 (0.04:0.21)	1.03:1.52:1.87	*Bacteroides uniformis*	0.11 (0.04:0.26)	1.04:1.53:1.53
358	*Bacteroides fragilis*	Nanopore with error-free profile	Centrifuge (standard database)	S	95.78 (94.34:97.45)	0.33 (0.19:0.49)	*Bacteroides ovatus*	0.29 (0.14:0.46)	0.84:1.44:2.94	*Bacteroides uniformis*	0.14 (0.05:0.57)	2.36:2.62:2.79
359	*Bacteroides fragilis*	Nanopore with error-free profile	Kraken (mini database)	S	95.28 (93.58:96.83)	0.64 (0.36:0.86)	*Bacteroides xylanisolvens*	0.1 (0.04:0.36)	1.47:1.64:1.84	*Phocaeicola dorei*	0.1 (0.03:0.24)	0.86:1.66:1.71
360	*Bacteroides fragilis*	Nanopore with error-free profile	Kraken (standard database)	S	95.71 (94.22:97.11)	0.13 (0.06:0.22)	*Phocaeicola dorei*	0.11 (0.03:0.21)	0.97:1.5:1.85	*Bacteroides uniformis*	0.11 (0.04:0.25)	1.02:1.49:1.5
361	*Cutibacterium acnes*	Illumina with error profile	Bracken (mini database)	S	99.74 (99.52:99.78)	NA	*Cutibacterium avidum*	0.07 (0.06:0.1)	0.33:0.36:0.37	*Staphylococcus aureus*	0 (0:0)	0:0.33:0.52
362	*Cutibacterium acnes*	Illumina with error profile	Bracken (standard database)	S	99.88 (99.63:99.98)	NA	*Cutibacterium modestum*	0.03 (0:0.16)	0.41:0.7:0.75	*Cutibacterium avidum*	0.02 (0.01:0.09)	0.53:0.55:0.55
363	*Cutibacterium acnes*	Illumina with error profile	Centrifuge (standard database)	S	99.15 (98.86:99.89)	0.58 (0.24:1.23)	*Cutibacterium avidum*	0.05 (0:0.11)	0.31:0.34:0.34	*Staphylococcus aureus*	0 (0:0.01)	0.68:0.87:0.91
364	*Cutibacterium acnes*	Illumina with error profile	Kraken (mini database)	S	96.17 (94.61:96.54)	1.35 (1.28:1.65)	*Cutibacterium avidum*	0.07 (0.06:0.1)	0.32:0.34:0.35	*Staphylococcus aureus*	0 (0:0)	0:0.23:0.37
365	*Cutibacterium acnes*	Illumina with error profile	Kraken (standard database)	S	99.18 (98.47:99.59)	0.17 (0.09:0.21)	*Cutibacterium modestum*	0.03 (0:0.16)	0.41:0.7:0.75	*Cutibacterium avidum*	0.02 (0.01:0.09)	0.53:0.55:0.55
366	*Cutibacterium acnes*	Illumina with error-free profile	Bracken (mini database)	S	99.72 (99.52:99.79)	NA	*Cutibacterium avidum*	0.07 (0.06:0.1)	0.35:0.37:0.4	*Staphylococcus aureus*	0 (0:0)	0.01:0.34:0.53
367	*Cutibacterium acnes*	Illumina with error-free profile	Bracken (standard database)	S	99.85 (99.62:99.98)	NA	*Cutibacterium avidum*	0.02 (0.01:0.08)	0.55:0.59:0.59	*Cutibacterium modestum*	0.01 (0:0.16)	0.44:0.72:0.73
368	*Cutibacterium acnes*	Illumina with error-free profile	Centrifuge (standard database)	S	99.12 (98.89:99.89)	0.57 (0.26:1.25)	*Cutibacterium avidum*	0.06 (0:0.11)	0.32:0.33:0.34	*Staphylococcus aureus*	0 (0:0.01)	0.69:0.87:0.92
369	*Cutibacterium acnes*	Illumina with error-free profile	Kraken (mini database)	S	96.17 (94.67:96.52)	1.34 (1.27:1.66)	*Cutibacterium avidum*	0.07 (0.06:0.1)	0.33:0.35:0.38	*Staphylococcus aureus*	0 (0:0)	0:0.22:0.35
370	*Cutibacterium acnes*	Illumina with error-free profile	Kraken (standard database)	S	99.16 (98.41:99.58)	0.18 (0.1:0.21)	*Cutibacterium avidum*	0.02 (0.01:0.08)	0.54:0.58:0.59	*Cutibacterium modestum*	0.01 (0:0.16)	0.44:0.72:0.72
371	*Cutibacterium acnes*	Nanopore with error profile	Bracken (mini database)	S	98.9 (98.73:99)	NA	*Cutibacterium avidum*	0.07 (0.05:0.1)	0.32:0.42:0.43	*Homo sapiens*	0.04 (0.03:0.05)	0.07:0.12:0.4
372	*Cutibacterium acnes*	Nanopore with error profile	Bracken (standard database)	S	99.87 (99.64:99.95)	NA	*Cutibacterium avidum*	0.05 (0.03:0.11)	0.7:0.75:0.86	*Cutibacterium modestum*	0.05 (0.02:0.11)	0.38:0.52:0.58
373	*Cutibacterium acnes*	Nanopore with error profile	Centrifuge (standard database)	S	99.38 (98.86:99.56)	0.14 (0.1:0.3)	*Cutibacterium avidum*	0.05 (0.04:0.09)	0.36:0.46:0.46	*Homo sapiens*	0.01 (0.01:0.02)	0.23:0.4:0.44
374	*Cutibacterium acnes*	Nanopore with error profile	Kraken (mini database)	S	86.22 (85.6:86.79)	10.48 (10.29:10.81)	*Cutibacterium avidum*	0.06 (0.05:0.09)	0.27:0.36:0.37	*Homo sapiens*	0.04 (0.03:0.05)	0.06:0.11:0.31
375	*Cutibacterium acnes*	Nanopore with error profile	Kraken (standard database)	S	99.32 (98.85:99.51)	0.21 (0.17:0.24)	*Cutibacterium avidum*	0.05 (0.03:0.11)	0.7:0.75:0.85	*Cutibacterium modestum*	0.05 (0.02:0.11)	0.38:0.52:0.58
376	*Cutibacterium acnes*	Nanopore with error-free profile	Bracken (mini database)	S	99.95 (99.77:99.99)	NA	*Cutibacterium granulosum*	0.1 (0.04:0.11)	0.14:0.14:0.14	*Cutibacterium avidum*	0.05 (0.02:0.06)	0.32:0.38:0.4
377	*Cutibacterium acnes*	Nanopore with error-free profile	Bracken (standard database)	S	100 (99.7:100)	NA	*Cutibacterium avidum*	0.14 (0.04:0.71)	0.95:1.03:1.05	*Cutibacterium modestum*	0.1 (0.04:0.16)	0.41:0.55:0.6
378	*Cutibacterium acnes*	Nanopore with error-free profile	Centrifuge (standard database)	S	99.84 (99.35:99.98)	0.11 (0.05:0.18)	*Cutibacterium avidum*	0.09 (0.02:0.18)	0.29:0.31:0.31	*Acidipropionibacterium acidipropionici*	0.08 (0.05:0.14)	0.21:0.27:0.29
379	*Cutibacterium acnes*	Nanopore with error-free profile	Kraken (mini database)	S	99.75 (99.45:99.9)	0.02 (0.01:0.08)	*Cutibacterium granulosum*	0.1 (0.04:0.11)	0.14:0.14:0.14	*Cutibacterium avidum*	0.05 (0.02:0.06)	0.32:0.38:0.4
380	*Cutibacterium acnes*	Nanopore with error-free profile	Kraken (standard database)	S	99.89 (99.64:99.97)	0.03 (0.02:0.04)	*Cutibacterium avidum*	0.14 (0.04:0.71)	0.95:1.03:1.05	*Cutibacterium modestum*	0.1 (0.04:0.16)	0.41:0.55:0.6
381	*Enterobacter cloacae*	Illumina with error profile	Bracken (mini database)	S	7.22 (3.23:93.21)	NA	*Enterobacter hormaechei*	3.78 (1.37:58.41)	88.8:89.45:90.15	*Enterobacter roggenkampii*	0.5 (0.25:1.64)	7.34:77.02:82.67
382	*Enterobacter cloacae*	Illumina with error profile	Bracken (standard database)	S	10.12 (5.22:95.65)	NA	*Enterobacter hormaechei*	2.21 (0.82:43.26)	75.12:86.42:88.46	*Enterobacter roggenkampii*	0.88 (0.19:2.53)	8.28:85.34:90.33
383	*Enterobacter cloacae*	Illumina with error profile	Kraken (mini database)	S	3.38 (1.75:76.28)	2.04 (1.02:2.59)	*Enterobacter asburiae*	0.28 (0.17:0.75)	3.04:37.3:37.34	*Enterobacter bugandensis*	0.16 (0.07:0.23)	50.6:52.76:52.91
384	*Enterobacter cloacae*	Illumina with error profile	Kraken (standard database)	S	5.95 (3.52:92.04)	0.25 (0.15:0.47)	*Enterobacter roggenkampii*	0.48 (0.18:0.82)	3.58:63.74:68.19	*Enterobacter bugandensis*	0.09 (0.03:0.16)	60.8:77.37:77.39
385	*Enterobacter cloacae*	Illumina with error-free profile	Bracken (mini database)	S	7.06 (3.21:93.24)	NA	*Enterobacter hormaechei*	3.74 (1.33:58.38)	88.81:89.4:90.18	*Enterobacter roggenkampii*	0.49 (0.26:1.63)	7.28:76.58:82.4
386	*Enterobacter cloacae*	Illumina with error-free profile	Bracken (standard database)	S	10.16 (5.22:95.5)	NA	*Enterobacter hormaechei*	2.24 (0.84:46.22)	77.17:86.1:88.38	*Enterobacter roggenkampii*	0.84 (0.18:2.53)	8.48:85.48:90.41
387	*Enterobacter cloacae*	Illumina with error-free profile	Kraken (mini database)	S	3.38 (1.76:76.3)	2.06 (1.03:2.61)	*Enterobacter asburiae*	0.28 (0.17:0.77)	3.05:37.3:37.4	*Enterobacter bugandensis*	0.15 (0.07:0.23)	50.73:52.7:52.78
388	*Enterobacter cloacae*	Illumina with error-free profile	Kraken (standard database)	S	5.92 (3.52:92.07)	0.26 (0.15:0.48)	*Enterobacter roggenkampii*	0.48 (0.17:0.82)	3.65:63.63:68.17	*Enterobacter bugandensis*	0.1 (0.03:0.17)	60.48:77.35:77.39
389	*Enterobacter cloacae*	Nanopore with error profile	Bracken (mini database)	S	8.59 (4.45:92.45)	NA	*Enterobacter hormaechei*	4.2 (1.33:55.92)	83.89:86.62:87.05	*Enterobacter roggenkampii*	0.84 (0.24:1.68)	8.99:77.21:81.49
390	*Enterobacter cloacae*	Nanopore with error profile	Bracken (standard database)	S	15.66 (6.79:95.33)	NA	*Enterobacter roggenkampii*	0.94 (0.17:2.12)	6.8:83.86:86.54	*Enterobacter bugandensis*	0.12 (0.04:0.22)	70.79:85.74:86.21
391	*Enterobacter cloacae*	Nanopore with error profile	Kraken (mini database)	S	4.44 (2.89:71.43)	10.07 (9.23:10.76)	*Enterobacter asburiae*	0.48 (0.15:0.95)	4.36:41.32:41.75	*Enterobacter bugandensis*	0.2 (0.06:0.32)	51.39:53.24:54.1
392	*Enterobacter cloacae*	Nanopore with error profile	Kraken (standard database)	S	14.12 (6.38:93.6)	0.22 (0.15:0.35)	*Enterobacter roggenkampii*	0.79 (0.16:1.66)	3.7:73.25:77.17	*Enterobacter bugandensis*	0.12 (0.04:0.22)	65.59:79.86:80.35
393	*Enterobacter cloacae*	Nanopore with error-free profile	Bracken (mini database)	S	8.57 (2.5:96.27)	NA	*Enterobacter hormaechei*	1.76 (0.84:56.22)	93.44:95.65:95.77	*Enterobacter bugandensis*	0.12 (0.05:0.25)	93.89:95.09:95.15
394	*Enterobacter cloacae*	Nanopore with error-free profile	Bracken (standard database)	S	8.18 (1.82:96.94)	NA	*Enterobacter hormaechei*	1.39 (0.49:45.88)	90.64:97.76:98.72	*Enterobacter bugandensis*	0.06 (0.02:21.99)	95.45:96.64:96.74
395	*Enterobacter cloacae*	Nanopore with error-free profile	Kraken (mini database)	S	7.71 (2.36:94.55)	0.1 (0.05:0.24)	*Enterobacter asburiae*	0.46 (0.13:0.75)	6.94:86.86:87.05	*Enterobacter bugandensis*	0.12 (0.05:0.24)	91.9:93.03:93.05
396	*Enterobacter cloacae*	Nanopore with error-free profile	Kraken (standard database)	S	6.62 (1.79:95.92)	0.04 (0.02:0.11)	*Enterobacter roggenkampii*	0.28 (0.11:0.86)	2.64:89.96:93.53	*Enterobacter bugandensis*	0.06 (0.02:21.8)	94.94:96.04:96.1
397	*Enterobacter cloacae complex*	Illumina with error profile	Centrifuge (standard database)	S	64.08 (32.52:91.99)	1 (0.49:1.61)	*Enterobacter cloacae*	10.95 (2.82:89.61)	95.72:97.13:99.19	*Enterobacter bugandensis*	0.09 (0.02:0.4)	93.22:93.85:94.25
398	*Enterobacter cloacae complex*	Illumina with error-free profile	Centrifuge (standard database)	S	64.31 (32.51:92.06)	1 (0.47:1.61)	*Enterobacter cloacae*	11 (2.8:89.67)	95.71:97.1:99.12	*Enterobacter bugandensis*	0.08 (0.03:0.38)	93.16:93.76:94.19
399	*Enterobacter cloacae complex*	Nanopore with error profile	Centrifuge (standard database)	S	77 (36.73:93.59)	0.57 (0.32:0.82)	*Enterobacter cloacae*	13.96 (4.53:91.39)	95.99:97.19:98.78	*Enterobacter bugandensis*	0.1 (0.03:0.35)	94.95:95.37:95.61
400	*Enterobacter cloacae complex*	Nanopore with error-free profile	Centrifuge (standard database)	S	88.89 (36.8:95.62)	0.12 (0.05:0.31)	*Enterobacter cloacae*	20.42 (3.03:93.62)	97.85:98.85:100	*Enterobacter bugandensis*	0.13 (0.03:1.13)	97.27:97.58:97.75

Supplementary Table 8. Continued

401	*Enterococcus faecalis*	Illumina with error profile	Bracken (mini database)	S	99.14 (98.56:99.38)	NA	*Enterococcus faecium*	0.16 (0.08:0.28)	0.69:1.84:2.24	*Staphylococcus haemolyticus*	0 (0:0.01)	0.68:2.22:2.6
402	*Enterococcus faecalis*	Illumina with error profile	Bracken (standard database)	S	99.09 (98.34:99.44)	NA	*Enterococcus faecium*	0.26 (0.16:0.38)	1.13:1.89:3.04	*Staphylococcus haemolyticus*	0 (0:0)	0.01:1.32:2.73
403	*Enterococcus faecalis*	Illumina with error profile	Centrifuge (standard database)	S	96.78 (95.6:97.66)	1.04 (0.62:1.92)	*Enterococcus faecium*	0.25 (0.07:0.46)	0.75:1.92:3.06	*Staphylococcus haemolyticus*	0 (0:0.01)	0.03:1.82:2.76
404	*Enterococcus faecalis*	Illumina with error profile	Kraken (mini database)	S	94.02 (92.6:94.91)	1.54 (1.08:2.45)	*Enterococcus faecium*	0.12 (0.07:0.21)	0.49:1.49:1.58	*Staphylococcus haemolyticus*	0 (0:0.01)	0.5:1.64:1.92
405	*Enterococcus faecalis*	Illumina with error profile	Kraken (standard database)	S	97.62 (96.28:98.11)	0.27 (0.06:0.89)	*Enterococcus faecium*	0.22 (0.14:0.28)	1.06:1.77:2.52	*Staphylococcus haemolyticus*	0 (0:0)	0.01:1.24:2.57
406	*Enterococcus faecalis*	Illumina with error-free profile	Bracken (mini database)	S	99.12 (98.47:99.4)	NA	*Enterococcus faecium*	0.16 (0.08:0.31)	0.72:1.84:2.21	*Staphylococcus haemolyticus*	0 (0:0)	0.04:2.08:2.59
407	*Enterococcus faecalis*	Illumina with error-free profile	Bracken (standard database)	S	99.04 (98.36:99.43)	NA	*Enterococcus faecium*	0.26 (0.15:0.37)	1.11:1.88:3.1	*Staphylococcus haemolyticus*	0 (0:0)	0.01:1.36:2.71
408	*Enterococcus faecalis*	Illumina with error-free profile	Centrifuge (standard database)	S	96.81 (95.59:97.61)	1.04 (0.65:1.88)	*Enterococcus faecium*	0.26 (0.07:0.47)	0.74:1.94:3.11	*Staphylococcus haemolyticus*	0 (0:0.01)	0.02:1.78:2.79
409	*Enterococcus faecalis*	Illumina with error-free profile	Kraken (mini database)	S	94.07 (92.62:94.89)	1.48 (1.08:2.36)	*Enterococcus faecium*	0.12 (0.06:0.21)	0.47:1.49:1.58	*Staphylococcus haemolyticus*	0 (0:0)	0.03:1.56:1.94
410	*Enterococcus faecalis*	Illumina with error-free profile	Kraken (standard database)	S	97.59 (96.24:98.15)	0.2 (0.05:0.76)	*Enterococcus faecium*	0.22 (0.13:0.28)	1:1.77:2.6	*Staphylococcus haemolyticus*	0 (0:0)	0.01:1.29:2.57
411	*Enterococcus faecalis*	Nanopore with error profile	Bracken (mini database)	S	98.48 (97.89:98.69)	NA	*Enterococcus faecium*	0.14 (0.08:0.25)	0.48:1.91:1.97	*NA*	NA	NA
412	*Enterococcus faecalis*	Nanopore with error profile	Bracken (standard database)	S	99.21 (98.6:99.48)	NA	*Enterococcus faecium*	0.23 (0.1:0.34)	0.7:1.8:2.51	*Staphylococcus haemolyticus*	0.01 (0.01:0.01)	1.45:2.83:3.18
413	*Enterococcus faecalis*	Nanopore with error profile	Centrifuge (standard database)	S	97.43 (96.67:98.07)	0.48 (0.31:0.77)	*Enterococcus faecium*	0.25 (0.06:0.58)	1.05:1.99:4.03	*Staphylococcus haemolyticus*	0.01 (0.01:0.06)	2.36:3.14:3.33
414	*Enterococcus faecalis*	Nanopore with error profile	Kraken (mini database)	S	85.92 (84.96:86.55)	10.14 (9.8:10.54)	*Enterococcus faecium*	0.11 (0.06:0.2)	0.38:1.47:1.55	*NA*	NA	NA
415	*Enterococcus faecalis*	Nanopore with error profile	Kraken (standard database)	S	98.28 (97.55:98.74)	0.29 (0.22:0.43)	*Enterococcus faecium*	0.22 (0.1:0.33)	0.68:1.77:2.38	*Staphylococcus haemolyticus*	0.01 (0.01:0.01)	1.39:2.72:3.05
416	*Enterococcus faecalis*	Nanopore with error-free profile	Bracken (mini database)	S	99.34 (98.77:99.61)	NA	*Enterococcus faecium*	0.22 (0.1:0.38)	0.67:1.62:2.22	*Staphylococcus haemolyticus*	0.03 (0.02:0.05)	2.07:3.14:3.41
417	*Enterococcus faecalis*	Nanopore with error-free profile	Bracken (standard database)	S	99.44 (99:99.71)	NA	*Enterococcus faecium*	0.17 (0.08:0.29)	0.96:1.68:2.25	*NA*	NA	NA
418	*Enterococcus faecalis*	Nanopore with error-free profile	Centrifuge (standard database)	S	97.95 (97.17:98.64)	0.14 (0.06:0.25)	*Enterococcus faecium*	0.36 (0.12:0.61)	1.25:1.83:2.73	*Enterococcus cecorum*	0.05 (0.02:0.15)	0.6:1.08:1.25
419	*Enterococcus faecalis*	Nanopore with error-free profile	Kraken (mini database)	S	98.56 (97.84:98.97)	0.1 (0.03:0.42)	*Enterococcus faecium*	0.22 (0.1:0.38)	0.66:1.61:2.19	*Staphylococcus haemolyticus*	0.03 (0.02:0.05)	1.99:3.01:3.27
420	*Enterococcus faecalis*	Nanopore with error-free profile	Kraken (standard database)	S	99.06 (98.59:99.46)	0.06 (0.02:0.17)	*Enterococcus faecium*	0.17 (0.08:0.28)	0.96:1.67:2.22	*NA*	NA	NA
421	*Escherichia coli*	Illumina with error profile	Bracken (mini database)	S	89.48 (87.66:90.74)	NA	*Salmonella enterica*	1.19 (0.95:1.66)	2.58:3.8:12.4	*Escherichia fergusonii*	0.98 (0.88:2.14)	3.41:3.99:4.31
422	*Escherichia coli*	Illumina with error profile	Bracken (standard database)	S	91.5 (89.24:96.02)	NA	*Shigella flexneri*	0.77 (0.22:1.23)	1.37:1.54:1.7	*Salmonella enterica*	0.48 (0.31:0.64)	0.94:1.08:2.56
423	*Escherichia coli*	Illumina with error profile	Centrifuge (standard database)	S	61.47 (55.4:89.61)	16.55 (2.4:22.2)	*Shigella dysenteriae*	0.38 (0.17:0.47)	1.03:1.92:2.19	*Salmonella enterica*	0.24 (0.15:0.36)	0.86:2.75:3.3
424	*Escherichia coli*	Illumina with error profile	Kraken (mini database)	S	33.2 (28.2:42.95)	0.79 (0.76:0.87)	*Escherichia fergusonii*	0.74 (0.64:0.78)	0.86:0.91:0.95	*Escherichia marmotae*	0.57 (0.52:0.59)	0.63:0.77:0.93
425	*Escherichia coli*	Illumina with error profile	Kraken (standard database)	S	57.32 (47.34:79.32)	0.03 (0.02:0.11)	*Escherichia albertii*	0.72 (0.41:0.93)	1.06:1.16:1.17	*Escherichia fergusonii*	0.66 (0.35:0.77)	0.9:0.96:1.16
426	*Escherichia coli*	Illumina with error-free profile	Bracken (mini database)	S	89.24 (87.55:90.68)	NA	*Salmonella enterica*	1.22 (0.94:1.61)	2.29:3.43:12.94	*Escherichia fergusonii*	0.98 (0.88:2.31)	3.42:3.97:4.12
427	*Escherichia coli*	Illumina with error-free profile	Bracken (standard database)	S	91.43 (89.18:96.02)	NA	*Shigella flexneri*	0.77 (0.21:1.23)	1.47:1.63:1.69	*Salmonella enterica*	0.48 (0.31:0.61)	0.9:1.1:2.56
428	*Escherichia coli*	Illumina with error-free profile	Centrifuge (standard database)	S	61.34 (55.41:89.73)	16.51 (2.43:22.31)	*Shigella dysenteriae*	0.36 (0.17:0.46)	0.99:1.93:2.14	*Salmonella enterica*	0.24 (0.16:0.36)	0.87:2.75:3.28
429	*Escherichia coli*	Illumina with error-free profile	Kraken (mini database)	S	33 (28.28:42.91)	0.79 (0.75:0.84)	*Escherichia fergusonii*	0.74 (0.66:0.8)	0.88:0.91:0.94	*Escherichia marmotae*	0.56 (0.52:0.59)	0.64:0.73:0.96
430	*Escherichia coli*	Illumina with error-free profile	Kraken (standard database)	S	57.21 (47.18:79.3)	0.03 (0.02:0.11)	*Escherichia albertii*	0.72 (0.41:0.93)	1.05:1.19:1.2	*Escherichia fergusonii*	0.68 (0.34:0.8)	0.88:0.96:1.12
431	*Escherichia coli*	Nanopore with error profile	Bracken (mini database)	S	89.38 (87.3:91.72)	NA	*Escherichia fergusonii*	1.01 (0.86:1.14)	2.37:3.07:3.13	*Salmonella enterica*	0.96 (0.71:1.22)	2.09:2.94:5.57
432	*Escherichia coli*	Nanopore with error profile	Bracken (standard database)	S	94.75 (93.48:96.79)	NA	*Escherichia albertii*	0.56 (0.36:0.75)	0.89:0.98:1.18	*Salmonella enterica*	0.29 (0.22:0.38)	0.55:0.89:1.19
433	*Escherichia coli*	Nanopore with error profile	Centrifuge (standard database)	S	80.77 (76.9:91.96)	3.89 (1.44:4.62)	*Shigella dysenteriae*	1.25 (0.34:1.54)	2.74:4.54:4.82	*Salmonella enterica*	0.42 (0.26:0.62)	1.22:3.62:4.12
434	*Escherichia coli*	Nanopore with error profile	Kraken (mini database)	S	44.83 (40.59:51.13)	8.55 (8.4:8.74)	*Escherichia fergusonii*	0.85 (0.75:0.97)	1.12:1.19:1.19	*Escherichia albertii*	0.84 (0.72:0.92)	1.07:1.14:1.15
435	*Escherichia coli*	Nanopore with error profile	Kraken (standard database)	S	86.97 (85.09:91.59)	0.16 (0.14:0.18)	*Escherichia albertii*	0.56 (0.35:0.75)	0.89:0.98:1.17	*Salmonella enterica*	0.24 (0.19:0.31)	0.48:0.82:1.09
436	*Escherichia coli*	Nanopore with error-free profile	Bracken (mini database)	S	96.56 (95.55:98.06)	NA	*Salmonella enterica*	0.23 (0.14:0.34)	0.49:0.69:1.39	*Citrobacter freundii*	0.06 (0.03:0.1)	0.29:0.67:1.38
437	*Escherichia coli*	Nanopore with error-free profile	Bracken (standard database)	S	99.41 (98.87:99.81)	NA	*Salmonella enterica*	0.05 (0.02:0.12)	0.47:0.86:0.96	*Citrobacter freundii*	0.04 (0.02:0.11)	0.47:0.78:0.87
438	*Escherichia coli*	Nanopore with error-free profile	Centrifuge (standard database)	S	96.67 (94.12:98.14)	0.63 (0.19:1.3)	*Salmonella enterica*	0.2 (0.08:0.46)	1.19:3.18:4.05	*Shigella dysenteriae*	0.05 (0.03:0.12)	0.69:1.62:2.02
439	*Escherichia coli*	Nanopore with error-free profile	Kraken (mini database)	S	88.55 (86.09:93.37)	0.02 (0.01:0.05)	*Escherichia fergusonii*	0.52 (0.3:0.75)	0.92:1.01:1.09	*Citrobacter freundii*	0.05 (0.03:0.09)	0.25:0.6:1.15
440	*Escherichia coli*	Nanopore with error-free profile	Kraken (standard database)	S	92.97 (86.35:97.66)	0.02 (0.01:0.05)	*Salmonella enterica*	0.05 (0.02:0.11)	0.44:0.78:0.86	*Citrobacter freundii*	0.02 (0.01:0.05)	0.4:0.62:0.66
441	*Klebsiella pneumoniae*	Illumina with error profile	Bracken (mini database)	S	91.12 (90.44:91.66)	NA	*Klebsiella quasipneumoniae*	2.72 (2.61:2.8)	3.09:4.07:4.41	*Klebsiella variicola*	2.2 (2.11:2.35)	2.74:3.65:3.68
442	*Klebsiella pneumoniae*	Illumina with error profile	Bracken (standard database)	S	89.56 (87.09:92.52)	NA	*Klebsiella quasipneumoniae*	3.5 (2.32:4.2)	5.01:5.12:5.16	*Klebsiella variicola*	2.26 (1.46:2.65)	3.24:3.38:3.39
443	*Klebsiella pneumoniae*	Illumina with error profile	Centrifuge (standard database)	S	58.24 (56.65:59.74)	11.7 (10.28:17.13)	*Klebsiella aerogenes*	0.54 (0.36:0.69)	1.22:1.43:1.48	*Escherichia coli*	0.54 (0.34:0.81)	1.5:2.3:2.56
444	*Klebsiella pneumoniae*	Illumina with error profile	Kraken (mini database)	S	30.05 (29.16:30.57)	1.11 (1.08:1.15)	*Klebsiella variicola*	0.93 (0.89:0.97)	1.07:1.15:1.28	*Escherichia coli*	0.25 (0.19:0.35)	0.54:0.76:1.28
445	*Klebsiella pneumoniae*	Illumina with error profile	Kraken (standard database)	S	35.61 (26.76:53.73)	0.04 (0.01:0.07)	*Klebsiella quasipneumoniae*	2.45 (1.84:2.74)	2.98:2.99:2.99	*Klebsiella variicola*	1.53 (1.14:1.71)	1.88:1.91:1.95
446	*Klebsiella pneumoniae*	Illumina with error-free profile	Bracken (mini database)	S	91.1 (90.44:91.56)	NA	*Klebsiella quasipneumoniae*	2.71 (2.63:2.8)	3.07:4.15:4.39	*Klebsiella variicola*	2.19 (2.11:2.4)	2.65:3.61:3.7
447	*Klebsiella pneumoniae*	Illumina with error-free profile	Bracken (standard database)	S	89.62 (86.96:92.52)	NA	*Klebsiella quasipneumoniae*	3.46 (2.3:4.21)	5.02:5.05:5.09	*Klebsiella variicola*	2.29 (1.47:2.67)	3.23:3.33:3.35
448	*Klebsiella pneumoniae*	Illumina with error-free profile	Centrifuge (standard database)	S	58.34 (56.75:59.68)	11.77 (10.23:17.17)	*Klebsiella aerogenes*	0.55 (0.37:0.69)	1.2:1.43:1.48	*Escherichia coli*	0.54 (0.34:0.79)	1.5:2.3:2.61
449	*Klebsiella pneumoniae*	Illumina with error-free profile	Kraken (mini database)	S	30.08 (29.19:30.47)	1.11 (1.09:1.15)	*Klebsiella variicola*	0.93 (0.89:0.98)	1.08:1.13:1.29	*Escherichia coli*	0.26 (0.19:0.34)	0.54:0.76:1.28
450	*Klebsiella pneumoniae*	Illumina with error-free profile	Kraken (standard database)	S	35.62 (26.7:53.74)	0.04 (0.01:0.09)	*Klebsiella quasipneumoniae*	2.42 (1.85:2.69)	2.96:3:3.03	*Klebsiella variicola*	1.54 (1.16:1.7)	1.85:1.91:1.93

Supplementary Table 8. Continued

451	*Klebsiella pneumoniae*	Nanopore with error profile	Bracken (mini database)	S	90.04 (89.2:90.66)	NA	*Klebsiella quasipneumoniae*	2.71 (2.42:2.96)	4.26:4.87:5.68	*Klebsiella variicola*	2.26 (2.13:2.43)	2.66:3.56:4.11
452	*Klebsiella pneumoniae*	Nanopore with error profile	Bracken (standard database)	S	88.48 (85.12:91.47)	NA	*Klebsiella quasipneumoniae*	4.06 (2.66:5.02)	5.57:5.73:5.75	*Klebsiella variicola*	2.56 (1.77:3.15)	3.49:3.77:3.84
453	*Klebsiella pneumoniae*	Nanopore with error profile	Centrifuge (standard database)	S	75.97 (74.14:78.46)	2.68 (2.35:2.98)	*Klebsiella aerogenes*	1.44 (1.21:1.72)	2.56:2.95:2.99	*Escherichia coli*	0.83 (0.63:1.21)	2.11:3.37:3.85
454	*Klebsiella pneumoniae*	Nanopore with error profile	Kraken (mini database)	S	39.49 (38.86:40.01)	9.29 (9.15:9.53)	*Klebsiella quasipneumoniae*	1.49 (1.41:1.58)	1.75:2.03:2.32	*Klebsiella variicola*	1.4 (1.32:1.45)	1.6:2.02:2.18
455	*Klebsiella pneumoniae*	Nanopore with error profile	Kraken (standard database)	S	73.25 (68.15:79.88)	0.15 (0.13:0.18)	*Klebsiella quasipneumoniae*	3.96 (2.61:4.9)	5.39:5.55:5.57	*Klebsiella variicola*	2.5 (1.73:3.06)	3.38:3.65:3.72
456	*Klebsiella pneumoniae*	Nanopore with error-free profile	Bracken (mini database)	S	96.62 (95.96:97.03)	NA	*Klebsiella quasipneumoniae*	0.74 (0.68:0.86)	1.04:3.08:3.81	*Klebsiella variicola*	0.66 (0.6:0.74)	1.09:3.01:3.08
457	*Klebsiella pneumoniae*	Nanopore with error-free profile	Bracken (standard database)	S	97.28 (95.96:99.02)	NA	*Klebsiella quasipneumoniae*	0.82 (0.14:1.28)	1.72:1.83:1.93	*Escherichia coli*	0.16 (0.06:0.37)	0.96:1.65:2.29
458	*Klebsiella pneumoniae*	Nanopore with error-free profile	Centrifuge (standard database)	S	94.73 (92.79:95.71)	0.52 (0.36:0.63)	*Escherichia coli*	0.66 (0.38:1.03)	1.91:2.95:3.28	*Klebsiella aerogenes*	0.34 (0.16:0.55)	1.75:2.6:2.7
459	*Klebsiella pneumoniae*	Nanopore with error-free profile	Kraken (mini database)	S	87.74 (86.82:88.36)	0.01 (0.01:0.03)	*Klebsiella quasipneumoniae*	0.73 (0.65:0.84)	1.01:2.93:3.61	*Klebsiella variicola*	0.65 (0.59:0.73)	1.07:2.88:2.92
460	*Klebsiella pneumoniae*	Nanopore with error-free profile	Kraken (standard database)	S	79.52 (66.4:93.49)	0.01 (0.01:0.06)	*Klebsiella quasipneumoniae*	0.8 (0.14:1.22)	1.65:1.75:1.84	*Escherichia coli*	0.15 (0.06:0.33)	0.9:1.52:2.27
461	*Micrococcus luteus*	Illumina with error profile	Bracken (mini database)	S	87.84 (86:93.71)	NA	*Microbacterium paraoxydans*	0.01 (0:0.05)	1.42:2.01:2.16	*Rothia kristinae*	0.01 (0.01:0.02)	0.09:1.04:1.08
462	*Micrococcus luteus*	Illumina with error profile	Bracken (standard database)	S	95.62 (92.52:97.3)	NA	*Microbacterium paraoxydans*	0.02 (0:0.21)	1.41:1.76:1.85	*Rothia kristinae*	0.01 (0:0.02)	0.58:0.96:0.97
463	*Micrococcus luteus*	Illumina with error profile	Centrifuge (standard database)	S	58.28 (56.56:62.98)	2.47 (1.95:3.12)	*Pseudarthrobacter sulfonivorans*	0.01 (0:0.01)	0.07:6.72:10.96	*Pseudarthrobacter chlorophenolicus*	0 (0:0.01)	0.07:1.88:2.9
464	*Micrococcus luteus*	Illumina with error profile	Kraken (mini database)	S	50.28 (48.11:57.16)	5.97 (4.08:7.43)	*Microbacterium paraoxydans*	0 (0:0.01)	1.12:1.62:1.75	*Pseudarthrobacter chlorophenolicus*	0 (0:0.01)	0.01:0.22:0.4
465	*Micrococcus luteus*	Illumina with error profile	Kraken (standard database)	S	87.87 (82.87:91.92)	1.57 (1.02:2.24)	*Microbacterium paraoxydans*	0.01 (0:0.18)	1.38:1.73:1.82	*Rothia kristinae*	0.01 (0:0.02)	0.39:0.61:0.61
466	*Micrococcus luteus*	Illumina with error-free profile	Bracken (mini database)	S	88.13 (86.22:93.58)	NA	*Microbacterium paraoxydans*	0.01 (0:0.26)	1.65:2.04:2.14	*Rothia kristinae*	0.01 (0.01:0.02)	0.09:0.96:0.98
467	*Micrococcus luteus*	Illumina with error-free profile	Bracken (standard database)	S	95.66 (92.56:97.29)	NA	*Microbacterium paraoxydans*	0.01 (0.01:0.12)	1.35:1.77:1.87	*Rothia kristinae*	0.01 (0:0.02)	0.47:0.91:0.92
468	*Micrococcus luteus*	Illumina with error-free profile	Centrifuge (standard database)	S	58.34 (56.69:63.07)	2.5 (1.92:3.2)	*Pseudarthrobacter chlorophenolicus*	0 (0:0.01)	0.07:1.95:2.87	*Pseudarthrobacter sulfonivorans*	0 (0:0.01)	0.07:6.34:10.86
469	*Micrococcus luteus*	Illumina with error-free profile	Kraken (mini database)	S	50.09 (48.01:57.05)	5.96 (4.12:7.5)	*Microbacterium paraoxydans*	0.01 (0:0.18)	1.32:1.66:1.74	*Pseudarthrobacter chlorophenolicus*	0 (0:0.01)	0.01:0.23:0.41
470	*Micrococcus luteus*	Illumina with error-free profile	Kraken (standard database)	S	87.85 (82.8:91.88)	1.57 (1.03:2.23)	*Microbacterium paraoxydans*	0.01 (0:0.1)	1.31:1.73:1.84	*Rothia kristinae*	0.01 (0:0.02)	0.34:0.57:0.57
471	*Micrococcus luteus*	Nanopore with error profile	Bracken (mini database)	S	89.18 (87.45:92.33)	NA	*Microbacterium paraoxydans*	0.03 (0.02:0.38)	1.52:1.85:1.93	*Bifidobacterium animalis*	0.02 (0.01:0.17)	3.18:6.77:7.67
472	*Micrococcus luteus*	Nanopore with error profile	Bracken (standard database)	S	97.47 (95.96:98.59)	NA	*Paenarthrobacter ureafaciens*	0.11 (0.01:0.33)	0.84:0.97:1	*Kytococcus sedentarius*	0.01 (0.01:0.03)	0.73:0.76:0.77
473	*Micrococcus luteus*	Nanopore with error profile	Centrifuge (standard database)	S	60.09 (57.22:63.04)	0.88 (0.64:1.12)	*Pseudarthrobacter phenanthrenivorans*	0.02 (0.01:0.06)	0.19:2.37:3.05	*Pseudarthrobacter sulfonivorans*	0.01 (0.01:0.03)	0.13:9.5:12.3
474	*Micrococcus luteus*	Nanopore with error profile	Kraken (mini database)	S	54.69 (52.97:58.48)	12.59 (11.57:13.58)	*Microbacterium paraoxydans*	0.01 (0.01:0.18)	1.18:1.46:1.53	*Pseudarthrobacter chlorophenolicus*	0.01 (0.01:0.01)	0.26:0.46:0.51
475	*Micrococcus luteus*	Nanopore with error profile	Kraken (standard database)	S	92.91 (90.6:94.67)	0.59 (0.32:0.9)	*Kytococcus sedentarius*	0.01 (0.01:0.03)	0.59:0.61:0.62	*Paenarthrobacter ureafaciens*	0.01 (0.01:0.02)	0.51:0.64:0.67
476	*Micrococcus luteus*	Nanopore with error-free profile	Bracken (mini database)	S	96.9 (95.52:98.45)	NA	*Sinomonas atrocyanea*	0.05 (0.02:0.57)	0.88:0.91:0.92	*Rothia kristinae*	0.01 (0.01:0.11)	1.29:1.31:1.32
477	*Micrococcus luteus*	Nanopore with error-free profile	Bracken (standard database)	S	97.91 (96.24:98.85)	NA	*Protaetiibacter intestinalis*	0.44 (0.26:0.51)	0.61:0.63:0.63	*Brevibacterium luteolum*	0.06 (0.05:0.13)	0.41:0.48:0.5
478	*Micrococcus luteus*	Nanopore with error-free profile	Centrifuge (standard database)	S	58.44 (54.81:62.55)	0.22 (0.16:0.3)	*Pseudarthrobacter phenanthrenivorans*	0.05 (0.02:0.06)	0.38:2.84:3.45	*Pseudarthrobacter sulfonivorans*	0.01 (0.01:0.07)	2.74:11.08:13.16
479	*Micrococcus luteus*	Nanopore with error-free profile	Kraken (mini database)	S	92.96 (90.75:95.12)	0.56 (0.32:0.96)	*Protaetiibacter intestinalis*	0.51 (0.33:0.54)	0.62:0.68:0.7	*Rothia kristinae*	0.01 (0.01:0.11)	0.9:0.92:0.93
480	*Micrococcus luteus*	Nanopore with error-free profile	Kraken (standard database)	S	96.91 (94.92:98.41)	0.15 (0.1:0.21)	*Protaetiibacter intestinalis*	0.44 (0.26:0.5)	0.61:0.63:0.63	*Brevibacterium luteolum*	0.06 (0.05:0.13)	0.41:0.48:0.5
481	*Proteus mirabilis*	Illumina with error profile	Bracken (mini database)	S	98.52 (97.7:99.01)	NA	*Proteus vulgaris*	0.11 (0.08:0.17)	0.38:0.62:13.92	*Proteus cibarius*	0.05 (0.03:0.1)	0.34:0.58:8.39
482	*Proteus mirabilis*	Illumina with error profile	Bracken (standard database)	S	98.72 (98.12:99.15)	NA	*Proteus vulgaris*	0.04 (0.03:0.12)	0.34:0.72:29.91	*Proteus terrae*	0.03 (0.01:0.1)	0.34:0.88:21.88
483	*Proteus mirabilis*	Illumina with error profile	Centrifuge (standard database)	S	95.68 (94.38:96.95)	0.46 (0.23:0.85)	*Proteus vulgaris*	0.11 (0.04:0.19)	0.38:1.4:89.34	*Providencia rettgeri*	0.04 (0.01:0.08)	1.24:1.78:1.8
484	*Proteus mirabilis*	Illumina with error profile	Kraken (mini database)	S	92.56 (91.53:93.47)	1.21 (1.08:1.55)	*Proteus vulgaris*	0.09 (0.07:0.14)	0.31:0.48:9.16	*Proteus cibarius*	0.01 (0.01:0.03)	0.09:0.15:1.35
485	*Proteus mirabilis*	Illumina with error profile	Kraken (standard database)	S	95.25 (93.9:96.28)	0.15 (0.08:0.25)	*Proteus vulgaris*	0.04 (0.02:0.11)	0.31:0.63:22.71	*Proteus terrae*	0.02 (0.01:0.07)	0.23:0.44:7.18
486	*Proteus mirabilis*	Illumina with error-free profile	Bracken (mini database)	S	98.48 (97.69:99.09)	NA	*Proteus vulgaris*	0.11 (0.09:0.16)	0.4:0.66:14.03	*Proteus cibarius*	0.05 (0.03:0.11)	0.32:0.56:8.34
487	*Proteus mirabilis*	Illumina with error-free profile	Bracken (standard database)	S	98.73 (98.11:99.14)	NA	*Proteus vulgaris*	0.04 (0.03:0.12)	0.35:0.74:29.93	*Proteus terrae*	0.02 (0.01:0.1)	0.34:0.65:22.36
488	*Proteus mirabilis*	Illumina with error-free profile	Centrifuge (standard database)	S	95.68 (94.37:96.91)	0.47 (0.23:0.85)	*Proteus vulgaris*	0.11 (0.04:0.18)	0.38:1.39:89.58	*Providencia rettgeri*	0.04 (0.01:0.07)	1.1:1.79:1.81
489	*Proteus mirabilis*	Illumina with error-free profile	Kraken (mini database)	S	92.55 (91.48:93.49)	1.19 (1.07:1.53)	*Proteus vulgaris*	0.09 (0.07:0.13)	0.31:0.51:9.25	*Proteus cibarius*	0.01 (0.01:0.04)	0.08:0.15:1.35
490	*Proteus mirabilis*	Illumina with error-free profile	Kraken (standard database)	S	95.25 (93.92:96.3)	0.13 (0.08:0.25)	*Proteus vulgaris*	0.04 (0.02:0.1)	0.32:0.64:22.69	*Proteus terrae*	0.01 (0.01:0.07)	0.22:0.37:7.3
491	*Proteus mirabilis*	Nanopore with error profile	Bracken (mini database)	S	97.75 (96.71:98.3)	NA	*Proteus vulgaris*	0.12 (0.09:0.16)	0.31:0.68:13.24	*Proteus cibarius*	0.06 (0.03:0.1)	0.29:1.92:6.77
492	*Proteus mirabilis*	Nanopore with error profile	Bracken (standard database)	S	98.57 (97.4:98.94)	NA	*Proteus vulgaris*	0.06 (0.03:0.08)	0.27:1.88:34.74	*Proteus terrae*	0.04 (0.02:0.09)	0.23:2.12:16.5
493	*Proteus mirabilis*	Nanopore with error profile	Centrifuge (standard database)	S	96.56 (95.54:97.41)	0.27 (0.18:0.43)	*Proteus vulgaris*	0.21 (0.12:0.3)	0.55:1.72:92.05	*Providencia rettgeri*	0.04 (0.02:0.07)	1.48:1.89:2.02
494	*Proteus mirabilis*	Nanopore with error profile	Kraken (mini database)	S	84.15 (83.16:84.72)	9.95 (9.62:10.12)	*Proteus vulgaris*	0.09 (0.07:0.13)	0.24:0.5:8.75	*Proteus cibarius*	0.02 (0.01:0.04)	0.08:0.41:1.32
495	*Proteus mirabilis*	Nanopore with error profile	Kraken (standard database)	S	96.76 (95.62:97.38)	0.23 (0.2:0.27)	*Proteus vulgaris*	0.05 (0.03:0.07)	0.26:1.63:28.98	*Proteus terrae*	0.03 (0.01:0.05)	0.17:1.04:7.26
496	*Proteus mirabilis*	Nanopore with error-free profile	Bracken (mini database)	S	99.02 (98.07:99.46)	NA	*Proteus vulgaris*	0.04 (0.01:0.06)	0.41:3.53:6.99	*Proteus cibarius*	0.02 (0.01:0.04)	0.2:2.51:3.9
497	*Proteus mirabilis*	Nanopore with error-free profile	Bracken (standard database)	S	99.14 (98.79:99.67)	NA	*Proteus vulgaris*	0.06 (0.04:0.09)	0.26:7.53:32.06	*Proteus terrae*	0.06 (0.02:0.19)	0.32:6.1:8.93
498	*Proteus mirabilis*	Nanopore with error-free profile	Centrifuge (standard database)	S	97.47 (96.34:98.42)	0.06 (0.03:0.12)	*Proteus vulgaris*	0.11 (0.04:0.23)	0.5:26.33:95.95	*Providencia rettgeri*	0.05 (0.02:0.08)	1.55:1.84:1.93
499	*Proteus mirabilis*	Nanopore with error-free profile	Kraken (mini database)	S	97.74 (96.66:98.39)	0.04 (0.01:0.11)	*Proteus vulgaris*	0.04 (0.01:0.06)	0.41:3.44:6.82	*Proteus cibarius*	0.02 (0.01:0.04)	0.19:2.03:3.13
500	*Proteus mirabilis*	Nanopore with error-free profile	Kraken (standard database)	S	98.06 (97.1:98.86)	0.03 (0.02:0.05)	*Proteus terrae*	0.05 (0.02:0.17)	0.3:5.31:7.77	*Proteus vulgaris*	0.05 (0.04:0.08)	0.26:7.36:31.33

Supplementary Table 8. Continued

501	*Pseudomonas aeruginosa*	Illumina with error profile	Bracken (mini database)	S	86.42 (84.9:87.85)	NA	*Pseudomonas resinovorans*	9.82 (8.4:11.08)	12.86:15.13:15.69	*Pseudomonas sihuiensis*	0 (0:0)	0.02:0.27:0.84
502	*Pseudomonas aeruginosa*	Illumina with error profile	Bracken (standard database)	S	95.04 (94.59:95.68)	NA	*Pseudomonas stutzeri*	0.27 (0.23:0.29)	0.34:0.51:0.71	*Pseudomonas sihuiensis*	0.01 (0.01:0.01)	0.02:0.26:0.68
503	*Pseudomonas aeruginosa*	Illumina with error profile	Centrifuge (standard database)	S	74.15 (72.16:89.04)	0.23 (0.11:0.45)	*Burkholderia cepacia complex*	0 (0:0.01)	0.03:0.47:1.03	*NA*	NA	NA
504	*Pseudomonas aeruginosa*	Illumina with error profile	Kraken (mini database)	S	31.44 (30.53:34.33)	1.55 (1.44:1.81)	*Burkholderia multivorans*	0 (0:0)	0.01:0.24:0.57	*Pseudomonas sihuiensis*	0 (0:0)	0.01:0.21:0.61
505	*Pseudomonas aeruginosa*	Illumina with error profile	Kraken (standard database)	S	57.62 (55.91:68.97)	0.08 (0.04:0.16)	*Pseudomonas stutzeri*	0.26 (0.22:0.28)	0.32:0.5:0.66	*Pseudomonas sihuiensis*	0.01 (0.01:0.01)	0.02:0.26:0.63
506	*Pseudomonas aeruginosa*	Illumina with error-free profile	Bracken (mini database)	S	86.68 (84.83:87.76)	NA	*Pseudomonas resinovorans*	9.61 (8.22:11.14)	12.97:15.84:16.41	*Pseudomonas sihuiensis*	0 (0:0)	0.02:0.28:0.85
507	*Pseudomonas aeruginosa*	Illumina with error-free profile	Bracken (standard database)	S	95.06 (94.58:95.71)	NA	*Pseudomonas stutzeri*	0.27 (0.22:0.29)	0.34:0.52:0.75	*Pseudomonas sihuiensis*	0.01 (0.01:0.01)	0.01:0.27:0.68
508	*Pseudomonas aeruginosa*	Illumina with error-free profile	Centrifuge (standard database)	S	74.1 (72.16:88.86)	0.24 (0.11:0.44)	*Burkholderia cepacia complex*	0 (0:0.01)	0.04:0.47:1.01	*NA*	NA	NA
509	*Pseudomonas aeruginosa*	Illumina with error-free profile	Kraken (mini database)	S	31.37 (30.48:34.34)	1.54 (1.43:1.79)	*Burkholderia multivorans*	0 (0:0)	0.01:0.23:0.56	*Pseudomonas sihuiensis*	0 (0:0)	0.02:0.2:0.61
510	*Pseudomonas aeruginosa*	Illumina with error-free profile	Kraken (standard database)	S	57.62 (55.95:69)	0.08 (0.04:0.16)	*Pseudomonas stutzeri*	0.26 (0.21:0.28)	0.33:0.5:0.7	*Pseudomonas sihuiensis*	0.01 (0.01:0.01)	0.01:0.26:0.64
511	*Pseudomonas aeruginosa*	Nanopore with error profile	Bracken (mini database)	S	86.2 (84.05:88.2)	NA	*Pseudomonas resinovorans*	6.58 (4.62:8.36)	11.64:12.77:14	*Pseudomonas sihuiensis*	0.02 (0.01:0.02)	0.07:0.6:1.15
512	*Pseudomonas aeruginosa*	Nanopore with error profile	Bracken (standard database)	S	95.04 (94.63:95.61)	NA	*Pseudomonas stutzeri*	0.26 (0.21:0.31)	0.38:0.55:0.84	*Pseudomonas sihuiensis*	0.02 (0.01:0.02)	0.05:0.63:1.09
513	*Pseudomonas aeruginosa*	Nanopore with error profile	Centrifuge (standard database)	S	85.3 (83.84:91.52)	0.16 (0.11:0.25)	*Pseudomonas citronellolis*	0.03 (0.02:0.05)	0.17:0.48:0.91	*NA*	NA	NA
514	*Pseudomonas aeruginosa*	Nanopore with error profile	Kraken (mini database)	S	41.1 (40.24:43.41)	10 (9.8:10.22)	*Pseudomonas stutzeri*	0.33 (0.29:0.37)	0.45:0.52:0.84	*Pseudomonas sihuiensis*	0.02 (0.01:0.02)	0.05:0.48:0.91
515	*Pseudomonas aeruginosa*	Nanopore with error profile	Kraken (standard database)	S	84.44 (83.69:86.22)	0.18 (0.15:0.22)	*Pseudomonas stutzeri*	0.25 (0.2:0.3)	0.38:0.53:0.8	*Pseudomonas sihuiensis*	0.02 (0.01:0.02)	0.05:0.62:1.08
516	*Pseudomonas aeruginosa*	Nanopore with error-free profile	Bracken (mini database)	S	97.76 (97.47:98.13)	NA	*Pseudomonas stutzeri*	0.18 (0.15:0.23)	0.32:0.41:0.7	*Pseudomonas sihuiensis*	0.04 (0.02:0.16)	0.78:0.93:0.97
517	*Pseudomonas aeruginosa*	Nanopore with error-free profile	Bracken (standard database)	S	99.66 (99.43:99.9)	NA	*Pseudomonas fluorescens*	0.04 (0.01:0.05)	0.11:0.53:0.9	*Pseudomonas putida*	0.04 (0.01:0.09)	0.28:0.68:0.87
518	*Pseudomonas aeruginosa*	Nanopore with error-free profile	Centrifuge (standard database)	S	96.16 (94.86:97.94)	0.04 (0.02:0.09)	*Pseudomonas fluorescens*	0.18 (0.06:0.48)	0.86:1.95:2.32	*NA*	NA	NA
519	*Pseudomonas aeruginosa*	Nanopore with error-free profile	Kraken (mini database)	S	89.5 (88.83:90.21)	0.04 (0.02:0.09)	*Pseudomonas stutzeri*	0.18 (0.15:0.23)	0.32:0.41:0.69	*Pseudomonas sihuiensis*	0.04 (0.02:0.16)	0.78:0.92:0.96
520	*Pseudomonas aeruginosa*	Nanopore with error-free profile	Kraken (standard database)	S	94.73 (93.89:96.08)	0.02 (0.01:0.03)	*Pseudomonas putida*	0.04 (0.01:0.09)	0.28:0.67:0.85	*Pseudomonas fluorescens*	0.04 (0.01:0.05)	0.11:0.53:0.9
521	*Staphylococcus aureus*	Illumina with error profile	Bracken (mini database)	S	98.71 (98.47:99.06)	NA	*Staphylococcus epidermidis*	0.05 (0.04:0.08)	0.16:0.45:0.53	*NA*	NA	NA
522	*Staphylococcus aureus*	Illumina with error profile	Bracken (standard database)	S	95.24 (90.62:97.54)	NA	*Staphylococcus schweitzeri*	0.46 (0.21:0.81)	0.91:0.93:0.98	*NA*	NA	NA
523	*Staphylococcus aureus*	Illumina with error profile	Centrifuge (standard database)	S	82.84 (79.44:92.07)	13.75 (4.86:15.12)	*Pseudomonas aeruginosa*	0 (0:0)	1.46:2.09:2.25	*NA*	NA	NA
524	*Staphylococcus aureus*	Illumina with error profile	Kraken (mini database)	S	89.33 (88.45:90.78)	0.9 (0.87:0.94)	*Staphylococcus schweitzeri*	0.12 (0.11:0.13)	0.16:0.25:0.26	*Staphylococcus epidermidis*	0.05 (0.03:0.07)	0.14:0.32:0.44
525	*Staphylococcus aureus*	Illumina with error profile	Kraken (standard database)	S	65.8 (39.43:83.32)	0.05 (0.04:0.34)	*Staphylococcus schweitzeri*	0.45 (0.2:0.78)	0.88:0.9:0.95	*Staphylococcus epidermidis*	0.15 (0.11:0.2)	0.32:0.54:0.91
526	*Staphylococcus aureus*	Illumina with error-free profile	Bracken (mini database)	S	98.72 (98.48:99.02)	NA	*Staphylococcus epidermidis*	0.05 (0.04:0.08)	0.16:0.43:0.95	*NA*	NA	NA
527	*Staphylococcus aureus*	Illumina with error-free profile	Bracken (standard database)	S	95.16 (90.47:97.46)	NA	*Staphylococcus epidermidis*	0.15 (0.12:0.19)	0.33:0.92:1.09	*NA*	NA	NA
528	*Staphylococcus aureus*	Illumina with error-free profile	Centrifuge (standard database)	S	82.9 (79.6:92.12)	13.62 (4.89:14.91)	*Pseudomonas aeruginosa*	0 (0:0)	1.46:2.09:2.25	*NA*	NA	NA
529	*Staphylococcus aureus*	Illumina with error-free profile	Kraken (mini database)	S	89.39 (88.43:90.75)	0.9 (0.88:0.95)	*Staphylococcus schweitzeri*	0.12 (0.11:0.13)	0.16:0.24:0.27	*Staphylococcus epidermidis*	0.05 (0.04:0.07)	0.12:0.34:0.41
530	*Staphylococcus aureus*	Illumina with error-free profile	Kraken (standard database)	S	65.82 (39.31:83.36)	0.05 (0.04:0.33)	*Staphylococcus schweitzeri*	0.45 (0.21:0.77)	0.89:0.91:0.91	*Staphylococcus epidermidis*	0.15 (0.12:0.19)	0.33:0.53:0.91
531	*Staphylococcus aureus*	Nanopore with error profile	Bracken (mini database)	S	98.21 (98.03:98.41)	NA	*Staphylococcus epidermidis*	0.07 (0.05:0.1)	0.18:0.54:1.1	*Pseudomonas aeruginosa*	0.01 (0.01:0.02)	1.1:2.14:2.4
532	*Staphylococcus aureus*	Nanopore with error profile	Bracken (standard database)	S	97.43 (95.56:98.32)	NA	*Staphylococcus schweitzeri*	0.38 (0.19:0.69)	0.97:0.99:1.02	*Staphylococcus epidermidis*	0.15 (0.1:0.24)	0.6:0.78:1.19
533	*Staphylococcus aureus*	Nanopore with error profile	Centrifuge (standard database)	S	92.87 (91.2:95.43)	3.34 (1.95:3.74)	*Pseudomonas aeruginosa*	0.01 (0.01:0.01)	3.25:4.35:4.63	*NA*	NA	NA
534	*Staphylococcus aureus*	Nanopore with error profile	Kraken (mini database)	S	82.66 (81.7:83.36)	9.09 (8.9:9.25)	*Staphylococcus epidermidis*	0.06 (0.04:0.09)	0.15:0.39:0.48	*Pseudomonas aeruginosa*	0.01 (0.01:0.01)	0.32:0.63:0.71
535	*Staphylococcus aureus*	Nanopore with error profile	Kraken (standard database)	S	88.86 (83.21:92.92)	0.18 (0.15:0.21)	*Staphylococcus schweitzeri*	0.38 (0.19:0.68)	0.96:0.99:0.99	*Staphylococcus epidermidis*	0.15 (0.1:0.24)	0.6:0.76:1.18
536	*Staphylococcus aureus*	Nanopore with error-free profile	Bracken (mini database)	S	99.74 (99.54:99.83)	NA	*Staphylococcus epidermidis*	0.04 (0.02:0.11)	0.58:0.68:0.77	*Staphylococcus pseudintermedius*	0.01 (0.01:0.02)	0.05:0.24:0.33
537	*Staphylococcus aureus*	Nanopore with error-free profile	Bracken (standard database)	S	99.62 (98.89:99.79)	NA	*Staphylococcus epidermidis*	0.05 (0.01:0.2)	0.5:1.38:1.39	*Staphylococcus pseudintermedius*	0.01 (0.01:0.03)	0.24:1.11:1.33
538	*Staphylococcus aureus*	Nanopore with error-free profile	Centrifuge (standard database)	S	98.42 (97.77:98.98)	0.55 (0.15:0.71)	*NA*	NA	NA	*NA*	NA	NA
539	*Staphylococcus aureus*	Nanopore with error-free profile	Kraken (mini database)	S	98.85 (98.35:99.2)	0.02 (0.01:0.03)	*Staphylococcus epidermidis*	0.04 (0.02:0.11)	0.57:0.68:0.77	*Staphylococcus pseudintermedius*	0.01 (0.01:0.02)	0.05:0.24:0.32
540	*Staphylococcus aureus*	Nanopore with error-free profile	Kraken (standard database)	S	96.41 (87.69:98.53)	0.09 (0.04:0.23)	*Staphylococcus epidermidis*	0.05 (0.01:0.2)	0.5:1.1:1.37	*Staphylococcus pseudintermedius*	0.01 (0.01:0.03)	0.24:1.1:1.32
541	*Staphylococcus epidermidis*	Illumina with error profile	Bracken (mini database)	S	98.51 (98.08:98.87)	NA	*Staphylococcus hominis*	0.05 (0.04:0.09)	0.21:1.01:74.06	*Staphylococcus warneri*	0.03 (0.02:0.06)	0.17:0.4:6.12
542	*Staphylococcus epidermidis*	Illumina with error profile	Bracken (standard database)	S	97.65 (96.98:98.43)	NA	*Staphylococcus hominis*	0.1 (0.06:0.15)	0.35:1.17:73.4	*Staphylococcus warneri*	0.03 (0.01:0.06)	0.2:0.32:4.58
543	*Staphylococcus epidermidis*	Illumina with error profile	Centrifuge (standard database)	S	94.97 (94.05:96.12)	0.63 (0.32:1.32)	*Staphylococcus aureus*	0.86 (0.58:1.1)	1.59:1.77:1.9	*Staphylococcus hominis*	0.07 (0.01:0.2)	0.4:3.51:71.54
544	*Staphylococcus epidermidis*	Illumina with error profile	Kraken (mini database)	S	93.04 (91.67:93.63)	1.44 (1.26:1.82)	*Staphylococcus hominis*	0.05 (0.03:0.08)	0.18:0.94:68.72	*Staphylococcus warneri*	0.02 (0.01:0.04)	0.11:0.29:4.03
545	*Staphylococcus epidermidis*	Illumina with error profile	Kraken (standard database)	S	96.29 (95.38:97.13)	0.1 (0.04:0.25)	*Staphylococcus hominis*	0.1 (0.06:0.15)	0.35:1.16:72.46	*Staphylococcus warneri*	0.03 (0.01:0.06)	0.19:0.3:4.33
546	*Staphylococcus epidermidis*	Illumina with error-free profile	Bracken (mini database)	S	98.5 (98.04:98.84)	NA	*Staphylococcus hominis*	0.05 (0.03:0.08)	0.2:1.06:74	*Staphylococcus warneri*	0.03 (0.01:0.06)	0.17:0.41:6.15
547	*Staphylococcus epidermidis*	Illumina with error-free profile	Bracken (standard database)	S	97.66 (96.98:98.39)	NA	*Staphylococcus hominis*	0.11 (0.06:0.14)	0.38:1.19:73.41	*Staphylococcus warneri*	0.04 (0.01:0.06)	0.22:0.4:4.53
548	*Staphylococcus epidermidis*	Illumina with error-free profile	Centrifuge (standard database)	S	94.99 (94.02:96.1)	0.59 (0.32:1.34)	*Staphylococcus aureus*	0.86 (0.58:1.09)	1.61:1.82:1.89	*Staphylococcus hominis*	0.06 (0.01:0.2)	0.4:2.07:71.52
549	*Staphylococcus epidermidis*	Illumina with error-free profile	Kraken (mini database)	S	93.01 (91.68:93.67)	1.43 (1.24:1.84)	*Staphylococcus hominis*	0.05 (0.03:0.08)	0.18:0.98:68.66	*Staphylococcus warneri*	0.02 (0.01:0.04)	0.11:0.29:4.05
550	*Staphylococcus epidermidis*	Illumina with error-free profile	Kraken (standard database)	S	96.25 (95.38:97.15)	0.1 (0.04:0.23)	*Staphylococcus hominis*	0.1 (0.06:0.14)	0.37:1.17:72.45	*Staphylococcus warneri*	0.03 (0.01:0.05)	0.21:0.38:4.28

Supplementary Table 8. Continued

551	*Staphylococcus epidermidis*	Nanopore with error profile	Bracken (mini database)	S	97.59 (97.12:98.04)	NA	*Staphylococcus hominis*	0.06 (0.04:0.1)	0.21:1.69:71.19	*Staphylococcus warneri*	0.03 (0.01:0.04)	0.12:1.33:8.05
552	*Staphylococcus epidermidis*	Nanopore with error profile	Bracken (standard database)	S	98.6 (97.87:99.09)	NA	*Staphylococcus hominis*	0.06 (0.04:0.11)	0.24:1.79:71.92	*Staphylococcus pasteuri*	0.02 (0.01:0.05)	1.03:2.91:6.05
553	*Staphylococcus epidermidis*	Nanopore with error profile	Centrifuge (standard database)	S	95.8 (94.82:96.69)	0.34 (0.23:0.54)	*Staphylococcus hominis*	0.07 (0.03:0.23)	0.6:5.3:73.65	*Staphylococcus pasteuri*	0.02 (0.01:0.05)	0.23:1.76:4.12
554	*Staphylococcus epidermidis*	Nanopore with error profile	Kraken (mini database)	S	84.68 (83.59:85.39)	10.13 (9.84:10.47)	*Staphylococcus hominis*	0.05 (0.04:0.08)	0.17:1.45:60.97	*Staphylococcus warneri*	0.02 (0.01:0.03)	0.09:0.92:5.54
555	*Staphylococcus epidermidis*	Nanopore with error profile	Kraken (standard database)	S	97.75 (97:98.29)	0.24 (0.19:0.29)	*Staphylococcus hominis*	0.06 (0.04:0.11)	0.24:1.77:71.16	*Staphylococcus pasteuri*	0.01 (0.01:0.04)	0.98:2.7:5.82
556	*Staphylococcus epidermidis*	Nanopore with error-free profile	Bracken (mini database)	S	98.95 (98.32:99.34)	NA	*Staphylococcus hominis*	0.06 (0.02:0.14)	0.29:19.25:73.02	*Staphylococcus pasteuri*	0.02 (0.01:0.06)	1.05:3.92:6.37
557	*Staphylococcus epidermidis*	Nanopore with error-free profile	Bracken (standard database)	S	98.71 (98.15:99.3)	NA	*Staphylococcus hominis*	0.05 (0.02:0.14)	0.31:11.4:73.14	*Staphylococcus pasteuri*	0.04 (0.03:0.14)	1.3:5.71:7.49
558	*Staphylococcus epidermidis*	Nanopore with error-free profile	Centrifuge (standard database)	S	96.68 (95.88:97.51)	0.06 (0.03:0.15)	*Staphylococcus hominis*	0.15 (0.03:0.41)	0.74:22.11:73.39	*Staphylococcus pasteuri*	0.04 (0.03:0.13)	0.31:4.15:5.3
559	*Staphylococcus epidermidis*	Nanopore with error-free profile	Kraken (mini database)	S	98.28 (97.45:98.83)	0.04 (0.02:0.08)	*Staphylococcus hominis*	0.06 (0.02:0.14)	0.29:19.12:72.53	*Staphylococcus pasteuri*	0.02 (0.01:0.06)	0.94:3.79:6.14
560	*Staphylococcus epidermidis*	Nanopore with error-free profile	Kraken (standard database)	S	98.46 (97.8:99.05)	0.03 (0.01:0.03)	*Staphylococcus hominis*	0.05 (0.02:0.14)	0.31:11.39:73.02	*Staphylococcus pasteuri*	0.04 (0.02:0.14)	1.29:5.63:7.38
561	*Staphylococcus haemolyticus*	Illumina with error profile	Bracken (mini database)	S	97.15 (96.21:97.97)	NA	*Staphylococcus aureus*	0.42 (0.24:0.7)	2.24:3.92:4.4	*Staphylococcus epidermidis*	0.3 (0.2:0.44)	0.66:2.91:6.7
562	*Staphylococcus haemolyticus*	Illumina with error profile	Bracken (standard database)	S	96.78 (95.83:97.8)	NA	*Staphylococcus aureus*	1.13 (0.55:1.7)	2.21:2.9:3.58	*Staphylococcus epidermidis*	0.38 (0.2:0.66)	1.16:1.52:4.15
563	*Staphylococcus haemolyticus*	Illumina with error profile	Centrifuge (standard database)	S	95.16 (93.12:96.19)	0.66 (0.42:1.34)	*Staphylococcus aureus*	0.71 (0.46:1.06)	2.28:2.89:3.99	*Staphylococcus hominis*	0.3 (0.2:0.46)	0.84:2.86:2.92
564	*Staphylococcus haemolyticus*	Illumina with error profile	Kraken (mini database)	S	89.96 (87.01:91.07)	2.27 (1.84:2.91)	*Staphylococcus hominis*	0.41 (0.25:0.66)	1.07:1.42:1.98	*Staphylococcus aureus*	0.3 (0.18:0.46)	0.68:2.09:3.16
565	*Staphylococcus haemolyticus*	Illumina with error profile	Kraken (standard database)	S	95.39 (94.38:96.27)	0.21 (0.1:0.53)	*Staphylococcus epidermidis*	0.38 (0.2:0.55)	1.07:1.2:1.85	*Staphylococcus hominis*	0.36 (0.21:0.71)	1.03:1.7:2.46
566	*Staphylococcus haemolyticus*	Illumina with error-free profile	Bracken (mini database)	S	97.2 (96.17:97.92)	NA	*Staphylococcus aureus*	0.44 (0.28:0.74)	2.02:3.6:4.28	*Staphylococcus epidermidis*	0.3 (0.22:0.44)	0.74:2.2:8.05
567	*Staphylococcus haemolyticus*	Illumina with error-free profile	Bracken (standard database)	S	96.87 (95.89:97.73)	NA	*Staphylococcus aureus*	1.15 (0.54:1.74)	2.26:2.92:3.52	*Staphylococcus epidermidis*	0.38 (0.2:0.64)	1.15:1.49:3.48
568	*Staphylococcus haemolyticus*	Illumina with error-free profile	Centrifuge (standard database)	S	95.16 (93.13:96.16)	0.65 (0.4:1.31)	*Staphylococcus aureus*	0.7 (0.45:1.07)	2.25:2.98:3.98	*Staphylococcus hominis*	0.3 (0.21:0.44)	0.85:2.87:3.01
569	*Staphylococcus haemolyticus*	Illumina with error-free profile	Kraken (mini database)	S	89.98 (87.07:91.12)	2.28 (1.88:2.88)	*Staphylococcus hominis*	0.41 (0.26:0.65)	1.06:1.42:1.93	*Staphylococcus aureus*	0.28 (0.17:0.46)	0.66:2.18:3.15
570	*Staphylococcus haemolyticus*	Illumina with error-free profile	Kraken (standard database)	S	95.39 (94.37:96.3)	0.21 (0.09:0.51)	*Staphylococcus epidermidis*	0.38 (0.2:0.58)	1.06:1.18:1.78	*Staphylococcus hominis*	0.35 (0.2:0.7)	1.02:1.67:2.39
571	*Staphylococcus haemolyticus*	Nanopore with error profile	Bracken (mini database)	S	96.35 (95.32:97.12)	NA	*Staphylococcus aureus*	0.55 (0.32:1.01)	2.62:3.82:4.56	*Staphylococcus epidermidis*	0.24 (0.18:0.36)	0.63:1.11:6.42
572	*Staphylococcus haemolyticus*	Nanopore with error profile	Bracken (standard database)	S	97.94 (97.33:98.62)	NA	*Staphylococcus aureus*	0.69 (0.32:1.12)	1.76:2.25:3.04	*Staphylococcus epidermidis*	0.28 (0.12:0.45)	0.83:1.35:2.9
573	*Staphylococcus haemolyticus*	Nanopore with error profile	Centrifuge (standard database)	S	96.34 (94.29:97.17)	0.39 (0.3:0.58)	*Staphylococcus aureus*	0.56 (0.34:0.83)	1.74:3.74:3.8	*Staphylococcus hominis*	0.36 (0.25:0.54)	1.02:3.45:3.51
574	*Staphylococcus haemolyticus*	Nanopore with error profile	Kraken (mini database)	S	82.44 (79.68:83.47)	10.81 (10.34:11.32)	*Staphylococcus hominis*	0.29 (0.15:0.44)	0.82:1.45:1.47	*Staphylococcus aureus*	0.25 (0.19:0.42)	0.75:2.36:3.07
575	*Staphylococcus haemolyticus*	Nanopore with error profile	Kraken (standard database)	S	97.12 (96.48:97.78)	0.28 (0.22:0.42)	*Staphylococcus epidermidis*	0.26 (0.12:0.45)	0.82:1.33:2.01	*Staphylococcus hominis*	0.22 (0.11:0.44)	0.72:1.6:1.62
576	*Staphylococcus haemolyticus*	Nanopore with error-free profile	Bracken (mini database)	S	97.97 (97.34:98.6)	NA	*Staphylococcus aureus*	0.37 (0.19:0.75)	1.79:4.2:4.69	*Staphylococcus epidermidis*	0.28 (0.14:0.5)	0.96:1.76:4.04
577	*Staphylococcus haemolyticus*	Nanopore with error-free profile	Bracken (standard database)	S	98.43 (97.75:99.06)	NA	*Staphylococcus aureus*	0.48 (0.14:0.83)	1.47:2.32:2.82	*Staphylococcus epidermidis*	0.28 (0.09:0.53)	1.02:1.53:3.9
578	*Staphylococcus haemolyticus*	Nanopore with error-free profile	Centrifuge (standard database)	S	97.19 (95.47:98.32)	0.06 (0.04:0.13)	*Staphylococcus aureus*	0.6 (0.26:1.1)	2.09:4.71:4.86	*Staphylococcus hominis*	0.25 (0.09:0.5)	1.16:5.75:5.78
579	*Staphylococcus haemolyticus*	Nanopore with error-free profile	Kraken (mini database)	S	97.18 (96.54:97.87)	0.08 (0.05:0.28)	*Staphylococcus aureus*	0.36 (0.18:0.68)	1.35:3.99:4.64	*Staphylococcus hominis*	0.36 (0.15:0.6)	1.12:2.03:2.39
580	*Staphylococcus haemolyticus*	Nanopore with error-free profile	Kraken (standard database)	S	98.24 (97.54:98.89)	0.03 (0.01:0.08)	*Staphylococcus epidermidis*	0.26 (0.09:0.5)	1.02:1.52:3.61	*Staphylococcus hominis*	0.22 (0.08:0.41)	0.72:1.47:1.77
581	*Streptococcus mitis*	Illumina with error profile	Bracken (mini database)	S	46.03 (43.55:54.66)	NA	*Streptococcus pneumoniae*	20.55 (16.7:26.79)	31.02:41.04:41.19	*Streptococcus oralis*	6.32 (5.54:8.18)	74.49:89.77:97.23
582	*Streptococcus mitis*	Illumina with error profile	Bracken (standard database)	S	67.67 (64.03:73.19)	NA	*Streptococcus pneumoniae*	10.48 (7.54:16.56)	18.24:25.66:30.03	*Streptococcus oralis*	5.2 (4.68:7.18)	87.53:93.84:99.54
583	*Streptococcus mitis*	Illumina with error profile	Centrifuge (standard database)	S	49.27 (44.79:55.55)	2.5 (1.65:3.74)	*Streptococcus pneumoniae*	14.47 (11.34:19.51)	20.58:26.39:30.05	*Streptococcus oralis*	3.29 (2.67:4.94)	58.72:84.63:98.67
584	*Streptococcus mitis*	Illumina with error profile	Kraken (mini database)	S	29.66 (28.56:36.83)	8.64 (5.96:10.47)	*Streptococcus pneumoniae*	7.69 (6.1:10.19)	11.12:15.01:16.54	*Streptococcus oralis*	3.63 (3.23:4.78)	49.42:70.01:81.79
585	*Streptococcus mitis*	Illumina with error profile	Kraken (standard database)	S	65.33 (60.85:70.7)	1.6 (1.04:2.91)	*Streptococcus pneumoniae*	9.26 (6.69:14.41)	15.99:22.26:26.56	*Streptococcus oralis*	4.96 (4.41:6.73)	85.21:92.39:98.22
586	*Streptococcus mitis*	Illumina with error-free profile	Bracken (mini database)	S	46.58 (43.43:54.44)	NA	*Streptococcus pneumoniae*	20.6 (15.86:26.52)	31.32:41.5:41.63	*Streptococcus oralis*	6.36 (5.5:8.15)	75:89.57:96.97
587	*Streptococcus mitis*	Illumina with error-free profile	Bracken (standard database)	S	67.68 (63.95:73.44)	NA	*Streptococcus pneumoniae*	10.3 (7.51:16.42)	18.25:25.71:30.05	*Streptococcus oralis*	5.22 (4.6:7.11)	87.54:93.81:99.45
588	*Streptococcus mitis*	Illumina with error-free profile	Centrifuge (standard database)	S	49.22 (44.7:55.53)	2.51 (1.6:3.67)	*Streptococcus pneumoniae*	14.45 (11.44:19.49)	20.61:26.35:30.05	*Streptococcus oralis*	3.3 (2.68:4.9)	58.45:84.5:98.79
589	*Streptococcus mitis*	Illumina with error-free profile	Kraken (mini database)	S	29.84 (28.54:36.97)	8.53 (5.94:10.42)	*Streptococcus pneumoniae*	7.7 (6.09:10.16)	11.25:14.82:16.44	*Streptococcus oralis*	3.66 (3.24:4.8)	49.72:69.46:82.07
590	*Streptococcus mitis*	Illumina with error-free profile	Kraken (standard database)	S	65.48 (60.76:70.9)	1.61 (1.05:2.87)	*Streptococcus pneumoniae*	9.2 (6.64:14.2)	16.07:22.33:26.68	*Streptococcus oralis*	4.97 (4.4:6.65)	85.14:92.31:98.21
591	*Streptococcus mitis*	Nanopore with error profile	Bracken (mini database)	S	49.48 (45.65:59.12)	NA	*Streptococcus pneumoniae*	20.34 (14.28:27.47)	33.03:43.03:43.94	*Streptococcus oralis*	5.5 (4.59:7.02)	79.01:90.65:96.99
592	*Streptococcus mitis*	Nanopore with error profile	Bracken (standard database)	S	75.47 (71.22:81.38)	NA	*Streptococcus pneumoniae*	8.47 (5.57:15)	17.84:28.51:34.17	*Streptococcus oralis*	3.58 (2.8:4.91)	91.56:95.2:99.17
593	*Streptococcus mitis*	Nanopore with error profile	Centrifuge (standard database)	S	61.24 (52.99:67.52)	1.44 (0.96:1.79)	*Streptococcus pneumoniae*	8.68 (6.12:11.93)	13.91:17.96:21.15	*Streptococcus oralis*	2.6 (2.08:3.87)	66.44:88.93:98.55
594	*Streptococcus mitis*	Nanopore with error profile	Kraken (mini database)	S	32.42 (30.55:39.44)	14.88 (13.63:15.8)	*Streptococcus pneumoniae*	7.32 (5.56:10.57)	11.52:17.2:17.69	*Streptococcus oralis*	3.34 (2.94:4.51)	52.28:68.05:76.05
595	*Streptococcus mitis*	Nanopore with error profile	Kraken (standard database)	S	72.59 (68.44:78.54)	1.08 (0.72:1.58)	*Streptococcus pneumoniae*	7.12 (4.69:12.76)	14.39:24.26:29.17	*Streptococcus oralis*	3.42 (2.67:4.68)	88.76:93.35:97.75
596	*Streptococcus mitis*	Nanopore with error-free profile	Bracken (mini database)	S	75.37 (71.72:82.01)	NA	*Streptococcus pneumoniae*	8.61 (5.22:14.88)	16.43:29.75:35.39	*Streptococcus oralis*	2.62 (1.61:4.84)	90.5:95.12:99.57
597	*Streptococcus mitis*	Nanopore with error-free profile	Bracken (standard database)	S	85.35 (82.27:89.7)	NA	*Streptococcus pneumoniae*	5.42 (2.62:9.63)	12.04:22.64:28.02	*Streptococcus oralis*	1.74 (0.89:3.13)	94.38:97.04:99.98
598	*Streptococcus mitis*	Nanopore with error-free profile	Centrifuge (standard database)	S	68.58 (56.17:76.17)	0.34 (0.16:0.48)	*Streptococcus pseudopneumoniae*	2.48 (1.73:8.16)	10.44:22.44:27.82	*Streptococcus oralis*	1.93 (1.41:3.71)	72.09:92.81:99.97
599	*Streptococcus mitis*	Nanopore with error-free profile	Kraken (mini database)	S	72.46 (67.89:79.22)	0.79 (0.41:1.3)	*Streptococcus pneumoniae*	7.1 (4.47:12.15)	13.83:25.2:30.45	*Streptococcus oralis*	2.54 (1.54:4.56)	88.31:93.93:98.94
600	*Streptococcus mitis*	Nanopore with error-free profile	Kraken (standard database)	S	84.79 (81.8:89.44)	0.19 (0.11:0.4)	*Streptococcus pneumoniae*	5.18 (2.61:9.32)	11.75:21.61:27.24	*Streptococcus oralis*	1.73 (0.88:3.12)	94.13:96.73:99.75

Supplementary Table 8. Continued

601	*Streptococcus pneumoniae*	Illumina with error profile	Bracken (mini database)	S	95.67 (94.83:95.95)	NA	*Streptococcus mitis*	1.44 (1.38:1.55)	2.42:3.2:3.42	*Escherichia coli*	0.02 (0.01:0.02)	0.86:3.04:3.58
602	*Streptococcus pneumoniae*	Illumina with error profile	Bracken (standard database)	S	99.22 (98.11:99.52)	NA	*Streptococcus mitis*	0.14 (0.11:0.27)	1.46:3.16:3.55	*Homo sapiens*	0 (0:0)	0.19:1.5:1.83
603	*Streptococcus pneumoniae*	Illumina with error profile	Centrifuge (standard database)	S	97.45 (96.39:98.4)	0.29 (0.16:0.37)	*Streptococcus mitis*	0.15 (0.07:0.32)	1.15:2.57:2.75	*Streptococcus pseudopneumoniae*	0.15 (0.08:0.26)	1.23:1.79:2.28
604	*Streptococcus pneumoniae*	Illumina with error profile	Kraken (mini database)	S	75.48 (74.15:76.37)	0.94 (0.89:1.11)	*Streptococcus mitis*	1.25 (1.2:1.34)	2.02:2.68:2.83	*Homo sapiens*	0 (0:0)	0.01:0.46:1.79
605	*Streptococcus pneumoniae*	Illumina with error profile	Kraken (standard database)	S	97.44 (96.05:98.37)	0.18 (0.09:0.49)	*Streptococcus mitis*	0.14 (0.1:0.27)	1.44:3.12:3.5	*Homo sapiens*	0 (0:0)	0.19:1.5:1.83
606	*Streptococcus pneumoniae*	Illumina with error-free profile	Bracken (mini database)	S	95.6 (94.84:96)	NA	*Streptococcus mitis*	1.46 (1.39:1.56)	2.47:3.29:3.5	*Escherichia coli*	0.02 (0.01:0.02)	0.8:3.3:3.92
607	*Streptococcus pneumoniae*	Illumina with error-free profile	Bracken (standard database)	S	99.2 (98.09:99.53)	NA	*Streptococcus mitis*	0.13 (0.11:0.26)	1.48:3.24:3.62	*Homo sapiens*	0 (0:0)	0.19:1.51:1.84
608	*Streptococcus pneumoniae*	Illumina with error-free profile	Centrifuge (standard database)	S	97.44 (96.44:98.34)	0.29 (0.16:0.36)	*Streptococcus mitis*	0.15 (0.07:0.32)	1.19:2.41:2.87	*Streptococcus pseudopneumoniae*	0.15 (0.09:0.24)	1.26:1.78:2.21
609	*Streptococcus pneumoniae*	Illumina with error-free profile	Kraken (mini database)	S	75.55 (74.16:76.34)	0.92 (0.88:1.1)	*Streptococcus mitis*	1.27 (1.21:1.35)	2.1:2.75:2.88	*Homo sapiens*	0 (0:0)	0.01:0.39:1.79
610	*Streptococcus pneumoniae*	Illumina with error-free profile	Kraken (standard database)	S	97.43 (96.05:98.35)	0.18 (0.1:0.47)	*Streptococcus mitis*	0.13 (0.11:0.26)	1.46:3.2:3.57	*Homo sapiens*	0 (0:0)	0.19:1.51:1.84
611	*Streptococcus pneumoniae*	Nanopore with error profile	Bracken (mini database)	S	95.53 (95.1:95.8)	NA	*Streptococcus mitis*	1.26 (1.17:1.36)	2.02:2.63:2.65	*Homo sapiens*	0.06 (0.04:0.08)	0.11:0.16:2.93
612	*Streptococcus pneumoniae*	Nanopore with error profile	Bracken (standard database)	S	98.98 (98.35:99.25)	NA	*Streptococcus mitis*	0.24 (0.2:0.31)	0.88:1.88:2.24	*Homo sapiens*	0.01 (0.01:0.02)	0.03:2.17:2.84
613	*Streptococcus pneumoniae*	Nanopore with error profile	Centrifuge (standard database)	S	95.94 (94.93:96.82)	0.16 (0.14:0.23)	*Streptococcus mitis*	0.7 (0.55:1)	1.87:3:4.03	*Streptococcus pseudopneumoniae*	0.48 (0.33:0.63)	1.87:2.94:3.25
614	*Streptococcus pneumoniae*	Nanopore with error profile	Kraken (mini database)	S	72.05 (71:72.48)	9.16 (8.93:9.36)	*Streptococcus mitis*	1.08 (1:1.15)	1.67:2.19:2.22	*Homo sapiens*	0.05 (0.04:0.07)	0.09:0.15:2.65
615	*Streptococcus pneumoniae*	Nanopore with error profile	Kraken (standard database)	S	97.08 (96.34:97.58)	0.2 (0.15:0.24)	*Streptococcus mitis*	0.24 (0.2:0.31)	0.87:1.87:2.23	*Homo sapiens*	0.01 (0.01:0.02)	0.03:2.17:2.84
616	*Streptococcus pneumoniae*	Nanopore with error-free profile	Bracken (mini database)	S	99.34 (98.75:99.6)	NA	*Streptococcus mitis*	0.15 (0.1:0.2)	0.67:1.73:1.96	*Streptococcus pyogenes*	0.02 (0.01:0.03)	0.06:0.27:1.45
617	*Streptococcus pneumoniae*	Nanopore with error-free profile	Bracken (standard database)	S	99.78 (99.18:99.98)	NA	*Streptococcus pseudopneumoniae*	0.08 (0.02:0.19)	0.69:1.17:1.39	*Streptococcus mitis*	0.04 (0.02:0.26)	0.96:2.01:2.02
618	*Streptococcus pneumoniae*	Nanopore with error-free profile	Centrifuge (standard database)	S	98.95 (97.78:99.56)	0.03 (0.01:0.06)	*Streptococcus pseudopneumoniae*	0.05 (0.02:0.18)	1.03:2.1:2.14	*Homo sapiens*	0.01 (0.01:0.02)	0.03:1.6:2.15
619	*Streptococcus pneumoniae*	Nanopore with error-free profile	Kraken (mini database)	S	98.4 (97.58:98.85)	0.02 (0.01:0.06)	*Streptococcus mitis*	0.15 (0.1:0.2)	0.66:1.73:1.95	*Streptococcus pyogenes*	0.02 (0.01:0.03)	0.05:0.25:1.45
620	*Streptococcus pneumoniae*	Nanopore with error-free profile	Kraken (standard database)	S	99.47 (98.85:99.88)	0.04 (0.01:0.07)	*Streptococcus pseudopneumoniae*	0.08 (0.02:0.19)	0.69:1.17:1.39	*Streptococcus mitis*	0.04 (0.02:0.26)	0.96:2.01:2.02
621	*Streptococcus pyogenes*	Illumina with error profile	Bracken (mini database)	S	97.52 (97.23:97.74)	NA	*Streptococcus dysgalactiae*	1.22 (1.09:1.35)	1.67:3.04:3.11	*Streptococcus agalactiae*	0.16 (0.13:0.21)	0.69:1:2.05
622	*Streptococcus pyogenes*	Illumina with error profile	Bracken (standard database)	S	98.94 (98.74:99.08)	NA	*Streptococcus dysgalactiae*	0.58 (0.48:0.69)	0.99:1.95:2.49	*Streptococcus agalactiae*	0.03 (0.02:0.05)	0.13:0.91:1.97
623	*Streptococcus pyogenes*	Illumina with error profile	Centrifuge (standard database)	S	97.28 (96.75:97.83)	0.18 (0.15:0.47)	*Filifactor alocis*	0 (0:0)	1.81:2.29:2.41	*NA*	NA	NA
624	*Streptococcus pyogenes*	Illumina with error profile	Kraken (mini database)	S	80.32 (79.18:81.11)	0.89 (0.86:0.93)	*Streptococcus dysgalactiae*	0.63 (0.59:0.68)	0.75:1.24:1.51	*Staphylococcus aureus*	0.02 (0:0.04)	0.21:0.61:0.72
625	*Streptococcus pyogenes*	Illumina with error profile	Kraken (standard database)	S	94.94 (94.32:95.63)	0.03 (0.01:0.09)	*Streptococcus dysgalactiae*	0.48 (0.4:0.55)	0.76:1.41:2.05	*Streptococcus agalactiae*	0.02 (0.01:0.03)	0.08:0.64:1.08
626	*Streptococcus pyogenes*	Illumina with error-free profile	Bracken (mini database)	S	97.49 (97.25:97.68)	NA	*Streptococcus dysgalactiae*	1.2 (1.14:1.35)	1.72:2.77:2.97	*Streptococcus agalactiae*	0.16 (0.13:0.21)	0.68:1.23:2.08
627	*Streptococcus pyogenes*	Illumina with error-free profile	Bracken (standard database)	S	98.96 (98.71:99.1)	NA	*Streptococcus dysgalactiae*	0.58 (0.48:0.69)	0.9:1.98:2.48	*Streptococcus agalactiae*	0.03 (0.02:0.06)	0.15:0.89:1.8
628	*Streptococcus pyogenes*	Illumina with error-free profile	Centrifuge (standard database)	S	97.35 (96.72:97.81)	0.18 (0.15:0.48)	*Filifactor alocis*	0 (0:0)	1.72:2.18:2.3	*NA*	NA	NA
629	*Streptococcus pyogenes*	Illumina with error-free profile	Kraken (mini database)	S	80.24 (79.41:81.06)	0.9 (0.87:0.94)	*Streptococcus dysgalactiae*	0.63 (0.6:0.69)	0.77:1.2:1.56	*Staphylococcus aureus*	0.01 (0:0.03)	0.15:0.59:0.73
630	*Streptococcus pyogenes*	Illumina with error-free profile	Kraken (standard database)	S	95 (94.28:95.58)	0.02 (0.01:0.08)	*Streptococcus dysgalactiae*	0.48 (0.4:0.56)	0.75:1.46:2.06	*Streptococcus agalactiae*	0.02 (0.02:0.04)	0.09:0.63:1.05
631	*Streptococcus pyogenes*	Nanopore with error profile	Bracken (mini database)	S	97.72 (97.52:97.91)	NA	*Streptococcus dysgalactiae*	0.65 (0.56:0.74)	0.92:1.27:1.66	*Streptococcus agalactiae*	0.07 (0.05:0.1)	0.23:1.1:1.41
632	*Streptococcus pyogenes*	Nanopore with error profile	Bracken (standard database)	S	99.19 (98.91:99.32)	NA	*Streptococcus dysgalactiae*	0.37 (0.32:0.43)	0.75:0.97:1.6	*Streptococcus agalactiae*	0.03 (0.02:0.05)	0.14:1:1.29
633	*Streptococcus pyogenes*	Nanopore with error profile	Centrifuge (standard database)	S	97.06 (96.44:97.43)	0.17 (0.15:0.21)	*Streptococcus dysgalactiae*	0.66 (0.49:1.04)	1.82:2.45:3.05	*NA*	NA	NA
634	*Streptococcus pyogenes*	Nanopore with error profile	Kraken (mini database)	S	78.33 (77.4:78.77)	9.27 (9.05:9.54)	*Streptococcus dysgalactiae*	0.44 (0.39:0.49)	0.62:0.84:1.24	*Streptococcus agalactiae*	0.05 (0.03:0.06)	0.12:0.57:0.69
635	*Streptococcus pyogenes*	Nanopore with error profile	Kraken (standard database)	S	97.04 (96.64:97.33)	0.2 (0.16:0.22)	*Streptococcus dysgalactiae*	0.36 (0.32:0.42)	0.73:0.94:1.55	*Streptococcus agalactiae*	0.03 (0.02:0.05)	0.13:0.95:1.17
636	*Streptococcus pyogenes*	Nanopore with error-free profile	Bracken (mini database)	S	99.64 (99.39:99.74)	NA	*Streptococcus dysgalactiae*	0.16 (0.12:0.22)	0.5:0.82:1.42	*Streptococcus agalactiae*	0.03 (0.01:0.05)	0.44:0.82:0.98
637	*Streptococcus pyogenes*	Nanopore with error-free profile	Bracken (standard database)	S	99.93 (99.72:99.97)	NA	*Streptococcus dysgalactiae*	0.05 (0.02:0.1)	0.3:0.72:1.28	*Streptococcus agalactiae*	0.01 (0.01:0.16)	1.04:1.13:1.15
638	*Streptococcus pyogenes*	Nanopore with error-free profile	Centrifuge (standard database)	S	99.72 (99.25:99.85)	0.04 (0.02:0.06)	*Streptococcus dysgalactiae*	0.05 (0.02:0.13)	0.62:1.27:1.94	*NA*	NA	NA
639	*Streptococcus pyogenes*	Nanopore with error-free profile	Kraken (mini database)	S	98.45 (97.9:98.71)	0.01 (0.01:0.03)	*Streptococcus dysgalactiae*	0.16 (0.12:0.22)	0.5:0.81:1.41	*Streptococcus agalactiae*	0.03 (0.01:0.05)	0.44:0.82:0.96
640	*Streptococcus pyogenes*	Nanopore with error-free profile	Kraken (standard database)	S	99.66 (99.35:99.77)	0.02 (0.01:0.04)	*Streptococcus dysgalactiae*	0.05 (0.02:0.1)	0.3:0.71:1.27	*Streptococcus agalactiae*	0.01 (0.01:0.16)	1.04:1.11:1.13

Supplementary Table 9. Reference list of genomes randomly selected from NCBI to perform simulation.

	Accession_number	Species	In classification database		Accession_number	Species	In classification database
1	NZ_CM000716.1	*Bacillus_cereus*	TRUE	51	NZ_NTVX01000008.1	*Bacillus_cereus*	FALSE
2	NZ_JH792262.1	*Bacillus_cereus*	FALSE	52	NZ_NTUQ01000100.1	*Bacillus_cereus*	FALSE
3	NZ_JH792120.1	*Bacillus_cereus*	FALSE	53	NZ_NTWE01000100.1	*Bacillus_cereus*	FALSE
4	NZ_JH792080.1	*Bacillus_cereus*	FALSE	54	NZ_NTWK01000100.1	*Bacillus_cereus*	FALSE
5	NZ_KB976008.1	*Bacillus_cereus*	FALSE	55	NZ_NTWL01000100.1	*Bacillus_cereus*	FALSE
6	NZ_KB976125.1	*Bacillus_cereus*	FALSE	56	NZ_NTWQ01000100.1	*Bacillus_cereus*	FALSE
7	NZ_KB976152.1	*Bacillus_cereus*	FALSE	57	NZ_NTXL01000100.1	*Bacillus_cereus*	FALSE
8	NZ_KN049962.1	*Bacillus_cereus*	FALSE	58	NZ_NUVD01000076.1	*Bacillus_cereus*	FALSE
9	NZ_CP009605.1	*Bacillus_cereus*	TRUE	59	NZ_NUXJ01000010.1	*Bacillus_cereus*	FALSE
10	NZ_JYPH01000003.1	*Bacillus_cereus*	FALSE	60	NZ_NUYB01000100.1	*Bacillus_cereus*	FALSE
11	NZ_CYHM01000001.1	*Bacillus_cereus*	FALSE	61	NZ_NUVB01000100.1	*Bacillus_cereus*	FALSE
12	NZ_LOQB01000001.1	*Bacillus_cereus*	FALSE	62	NZ_NUVU01000100.1	*Bacillus_cereus*	FALSE
13	NZ_LOMP01000001.1	*Bacillus_cereus*	FALSE	63	NZ_NUWK01000100.1	*Bacillus_cereus*	FALSE
14	NZ_LJIT01000055.1	*Bacillus_cereus*	FALSE	64	NZ_NUXG01000003.1	*Bacillus_cereus*	FALSE
15	NZ_LJKJ01000056.1	*Bacillus_cereus*	FALSE	65	NZ_NUXM01000010.1	*Bacillus_cereus*	FALSE
16	NZ_LJKD01000033.1	*Bacillus_cereus*	FALSE	66	NZ_NUXQ01000058.1	*Bacillus_cereus*	FALSE
17	NZ_CP011155.1	*Bacillus_cereus*	TRUE	67	NZ_NUYL01000100.1	*Bacillus_cereus*	FALSE
18	NZ_MPON01000001.1	*Bacillus_cereus*	FALSE	68	NZ_NUVI01000100.1	*Bacillus_cereus*	FALSE
19	NZ_MIFO01000001.1	*Bacillus_cereus*	FALSE	69	NZ_NUYM01000100.1	*Bacillus_cereus*	FALSE
20	NZ_MIFK01000001.1	*Bacillus_cereus*	FALSE	70	NZ_NVCB01000067.1	*Bacillus_cereus*	FALSE
21	NZ_MIFH01000001.1	*Bacillus_cereus*	FALSE	71	NZ_NVAI01000100.1	*Bacillus_cereus*	FALSE
22	NZ_MIFP01000001.1	*Bacillus_cereus*	FALSE	72	NZ_NVAK01000100.1	*Bacillus_cereus*	FALSE
23	NZ_MUAO01000100.1	*Bacillus_cereus*	FALSE	73	NZ_NVAM01000012.1	*Bacillus_cereus*	FALSE
24	NZ_NBNG01000001.1	*Bacillus_cereus*	FALSE	74	NZ_NVAS01000010.1	*Bacillus_cereus*	FALSE
25	NZ_NHUV01000001.1	*Bacillus_cereus*	FALSE	75	NZ_NVAU01000006.1	*Bacillus_cereus*	FALSE
26	NZ_NHUD01000001.1	*Bacillus_cereus*	FALSE	76	NZ_NVBF01000100.1	*Bacillus_cereus*	FALSE
27	NZ_NHUH01000001.1	*Bacillus_cereus*	FALSE	77	NZ_NVBV01000100.1	*Bacillus_cereus*	FALSE
28	NZ_NHYG01000001.1	*Bacillus_cereus*	FALSE	78	NZ_NVCT01000079.1	*Bacillus_cereus*	FALSE
29	NZ_NHTZ01000001.1	*Bacillus_cereus*	FALSE	79	NZ_NVBO01000142.1	*Bacillus_cereus*	FALSE
30	NZ_NHUF01000001.1	*Bacillus_cereus*	FALSE	80	NZ_NVDA01000010.1	*Bacillus_cereus*	FALSE
31	NZ_NHUQ01000001.1	*Bacillus_cereus*	FALSE	81	NZ_NVDN01000100.1	*Bacillus_cereus*	FALSE
32	NZ_NIQZ01000001.1	*Bacillus_cereus*	FALSE	82	NZ_NUFR01000010.1	*Bacillus_cereus*	FALSE
33	NZ_NINJ01000001.1	*Bacillus_cereus*	FALSE	83	NZ_NUHO01000100.1	*Bacillus_cereus*	FALSE
34	NZ_CP017234.1	*Bacillus_cereus*	TRUE	84	NZ_NUKL01000010.1	*Bacillus_cereus*	FALSE
35	NZ_NVLD01000010.1	*Bacillus_cereus*	FALSE	85	NZ_NUKY01000010.1	*Bacillus_cereus*	FALSE
36	NZ_NVMN01000010.1	*Bacillus_cereus*	FALSE	86	NZ_NUJS01000068.1	*Bacillus_cereus*	FALSE
37	NZ_NVLL01000010.1	*Bacillus_cereus*	FALSE	87	NZ_NULB01000100.1	*Bacillus_cereus*	FALSE
38	NZ_NVLA01000100.1	*Bacillus_cereus*	FALSE	88	NZ_NUME01000100.1	*Bacillus_cereus*	FALSE
39	NZ_NTXQ01000100.1	*Bacillus_cereus*	FALSE	89	NZ_NUMW01000100.1	*Bacillus_cereus*	FALSE
40	NZ_NTXZ01000100.1	*Bacillus_cereus*	FALSE	90	NZ_NUMP01000100.1	*Bacillus_cereus*	FALSE
41	NZ_NTZH01000100.1	*Bacillus_cereus*	FALSE	91	NZ_NUNE01000100.1	*Bacillus_cereus*	FALSE
42	NZ_NTZL01000100.1	*Bacillus_cereus*	FALSE	92	NZ_NUPJ01000100.1	*Bacillus_cereus*	FALSE
43	NZ_NTYJ01000100.1	*Bacillus_cereus*	FALSE	93	NZ_CP020383.1	*Bacillus_cereus*	TRUE
44	NZ_NTZZ01000100.1	*Bacillus_cereus*	FALSE	94	NZ_NCKX01000001.1	*Bacillus_cereus*	FALSE
45	NZ_NTSZ01000100.1	*Bacillus_cereus*	FALSE	95	NZ_CP039269.1	*Bacillus_cereus*	TRUE
46	NZ_NTPZ01000100.1	*Bacillus_cereus*	FALSE	96	NZ_SZNY01000418.1	*Bacillus_cereus*	FALSE
47	NZ_NTTC01000100.1	*Bacillus_cereus*	FALSE	97	NZ_SZOI01000910.1	*Bacillus_cereus*	FALSE
48	NZ_NTQT01000100.1	*Bacillus_cereus*	FALSE	98	NZ_CP034551.1	*Bacillus_cereus*	TRUE
49	NZ_NTQX01000100.1	*Bacillus_cereus*	FALSE	99	NZ_VEPT01000100.1	*Bacillus_cereus*	FALSE
50	NZ_NTTH01000100.1	*Bacillus_cereus*	FALSE	100	NZ_CP029454.1	*Bacillus_cereus*	TRUE

Supplementary Table 9. Continued.

101	NZ_EQ973245.1	*Bacteroides_fragilis*	FALSE	**151**	NZ_JUOA01000001.1	*Bacteroides_fragilis*	FALSE
102	NZ_GG705226.1	*Bacteroides_fragilis*	FALSE	**152**	NZ_LN877293.1	*Bacteroides_fragilis*	TRUE
103	NZ_JH724206.1	*Bacteroides_fragilis*	FALSE	**153**	NZ_KQ971009.1	*Bacteroides_fragilis*	FALSE
104	NZ_AKBY01000001.1	*Bacteroides_fragilis*	FALSE	**154**	NZ_CP012706.1	*Bacteroides_fragilis*	TRUE
105	NZ_JH724188.1	*Bacteroides_fragilis*	FALSE	**155**	NZ_LGTH01000001.1	*Bacteroides_fragilis*	FALSE
106	NZ_JH815491.1	*Bacteroides_fragilis*	FALSE	**156**	NZ_LIDT01000001.1	*Bacteroides_fragilis*	FALSE
107	NZ_JH815524.1	*Bacteroides_fragilis*	FALSE	**157**	NZ_CM004522.1	*Bacteroides_fragilis*	FALSE
108	NZ_JGCO01000001.1	*Bacteroides_fragilis*	FALSE	**158**	NZ_LIDS01000001.1	*Bacteroides_fragilis*	FALSE
109	NZ_JGCV01000001.1	*Bacteroides_fragilis*	FALSE	**159**	NZ_MTGH01000001.1	*Bacteroides_fragilis*	FALSE
110	NZ_JGCZ01000001.1	*Bacteroides_fragilis*	FALSE	**160**	NZ_PDCU01000001.1	*Bacteroides_fragilis*	FALSE
111	NZ_JGDA01000001.1	*Bacteroides_fragilis*	FALSE	**161**	NZ_PDCW01000001.1	*Bacteroides_fragilis*	FALSE
112	NZ_JGDD01000001.1	*Bacteroides_fragilis*	FALSE	**162**	NZ_PDCV01000001.1	*Bacteroides_fragilis*	FALSE
113	NZ_JGCN01000001.1	*Bacteroides_fragilis*	FALSE	**163**	NZ_CP018937.1	*Bacteroides_fragilis*	TRUE
114	NZ_JGCR01000001.1	*Bacteroides_fragilis*	FALSE	**164**	NZ_QRBX01000001.1	*Bacteroides_fragilis*	FALSE
115	NZ_JGCS01000001.1	*Bacteroides_fragilis*	FALSE	**165**	NZ_QSOR01000001.1	*Bacteroides_fragilis*	FALSE
116	NZ_JGCT01000001.1	*Bacteroides_fragilis*	FALSE	**166**	NZ_QSRG01000001.1	*Bacteroides_fragilis*	FALSE
117	NZ_JGCQ01000001.1	*Bacteroides_fragilis*	FALSE	**167**	NZ_QSSJ01000001.1	*Bacteroides_fragilis*	FALSE
118	NZ_JGCX01000001.1	*Bacteroides_fragilis*	FALSE	**168**	NZ_QSTV01000001.1	*Bacteroides_fragilis*	FALSE
119	NZ_JGDE01000001.1	*Bacteroides_fragilis*	FALSE	**169**	NZ_QSUD01000001.1	*Bacteroides_fragilis*	FALSE
120	NZ_JGDH01000001.1	*Bacteroides_fragilis*	FALSE	**170**	NZ_QSVD01000001.1	*Bacteroides_fragilis*	FALSE
121	NZ_JGDK01000001.1	*Bacteroides_fragilis*	FALSE	**171**	NZ_QSVT01000001.1	*Bacteroides_fragilis*	FALSE
122	NZ_JGDQ01000001.1	*Bacteroides_fragilis*	FALSE	**172**	NZ_QSWB01000001.1	*Bacteroides_fragilis*	FALSE
123	NZ_JGDR01000001.1	*Bacteroides_fragilis*	FALSE	**173**	NZ_QSWE01000001.1	*Bacteroides_fragilis*	FALSE
124	NZ_JGDV01000001.1	*Bacteroides_fragilis*	FALSE	**174**	NZ_QRTQ01000001.1	*Bacteroides_fragilis*	FALSE
125	NZ_JGEN01000001.1	*Bacteroides_fragilis*	FALSE	**175**	NZ_QRYS01000001.1	*Bacteroides_fragilis*	FALSE
126	NZ_JGDG01000001.1	*Bacteroides_fragilis*	FALSE	**176**	NZ_QRZH01000001.1	*Bacteroides_fragilis*	FALSE
127	NZ_JGDT01000001.1	*Bacteroides_fragilis*	FALSE	**177**	NZ_QRZO01000001.1	*Bacteroides_fragilis*	FALSE
128	NZ_JGEC01000001.1	*Bacteroides_fragilis*	FALSE	**178**	NZ_QSJF01000001.1	*Bacteroides_fragilis*	FALSE
129	NZ_JGEI01000001.1	*Bacteroides_fragilis*	FALSE	**179**	NZ_QRJE01000001.1	*Bacteroides_fragilis*	FALSE
130	NZ_JGEJ01000001.1	*Bacteroides_fragilis*	FALSE	**180**	NZ_QRJX01000001.1	*Bacteroides_fragilis*	FALSE
131	NZ_JGEK01000001.1	*Bacteroides_fragilis*	FALSE	**181**	NZ_QRKL01000001.1	*Bacteroides_fragilis*	FALSE
132	NZ_JGEM01000001.1	*Bacteroides_fragilis*	FALSE	**182**	NZ_QRKP01000001.1	*Bacteroides_fragilis*	FALSE
133	NZ_JGEQ01000001.1	*Bacteroides_fragilis*	FALSE	**183**	NZ_QRLH01000001.1	*Bacteroides_fragilis*	FALSE
134	NZ_JGEU01000001.1	*Bacteroides_fragilis*	FALSE	**184**	NZ_CP036555.1	*Bacteroides_fragilis*	TRUE
135	NZ_JGDI01000001.1	*Bacteroides_fragilis*	FALSE	**185**	NZ_SPGY01000010.1	*Bacteroides_fragilis*	FALSE
136	NZ_JGDO01000001.1	*Bacteroides_fragilis*	FALSE	**186**	NZ_VOHV01000010.1	*Bacteroides_fragilis*	FALSE
137	NZ_JGEO01000001.1	*Bacteroides_fragilis*	FALSE	**187**	NZ_VOHY01000010.1	*Bacteroides_fragilis*	FALSE
138	NZ_JGES01000001.1	*Bacteroides_fragilis*	FALSE	**188**	NZ_VOHT01000010.1	*Bacteroides_fragilis*	FALSE
139	NZ_JHEF01000001.1	*Bacteroides_fragilis*	FALSE	**189**	NZ_VOIJ01000010.1	*Bacteroides_fragilis*	FALSE
140	NZ_JGVE01000001.1	*Bacteroides_fragilis*	FALSE	**190**	NZ_CP043610.1	*Bacteroides_fragilis*	TRUE
141	NZ_JGVF01000001.1	*Bacteroides_fragilis*	FALSE	**191**	NZ_WCHZ01000092.1	*Bacteroides_fragilis*	FALSE
142	NZ_JGVG01000001.1	*Bacteroides_fragilis*	FALSE	**192**	NZ_WCII01000010.1	*Bacteroides_fragilis*	FALSE
143	NZ_CP036546.1	*Bacteroides_fragilis*	TRUE	**193**	NZ_JABAGK010000001.1	*Bacteroides_fragilis*	FALSE
144	NZ_CP036553.1	*Bacteroides_fragilis*	TRUE	**194**	NZ_JACENG010000010.1	*Bacteroides_fragilis*	FALSE
145	NZ_CP037440.1	*Bacteroides_fragilis*	TRUE	**195**	NZ_JACFSX010000010.1	*Bacteroides_fragilis*	FALSE
146	NZ_CDQM01000001.1	*Bacteroides_fragilis*	FALSE	**196**	NZ_JACFSU010000010.1	*Bacteroides_fragilis*	FALSE
147	NZ_CP011073.1	*Bacteroides_fragilis*	TRUE	**197**	NZ_JACWLM010000005.	*Bacteroides_fragilis*	FALSE
148	NZ_JVLP01000024.1	*Bacteroides_fragilis*	FALSE	**198**	NZ_JADDIJ010000001.1	*Bacteroides_fragilis*	FALSE
149	NZ_JVLQ01000008.1	*Bacteroides_fragilis*	FALSE	**199**	NZ_JADMVI010000072.1	*Bacteroides_fragilis*	FALSE
150	NZ_JUOZ01000315.1	*Bacteroides_fragilis*	FALSE	**200**	NZ_JADMXT010000010.1	*Bacteroides_fragilis*	FALSE

Supplementary Table 9. Continued.

201	NC_006085.1	*Cutibacterium_acnes*	TRUE	**251**	NZ_MVCN01000010.1	*Cutibacterium_acnes*	FALSE
202	NZ_GL383087.1	*Cutibacterium_acnes*	FALSE	**252**	NZ_MVCP01000010.1	*Cutibacterium_acnes*	FALSE
203	NZ_GL383117.1	*Cutibacterium_acnes*	FALSE	**253**	NZ_NJFY01000003.1	*Cutibacterium_acnes*	FALSE
204	NZ_GL383195.1	*Cutibacterium_acnes*	FALSE	**254**	NZ_NJNM01000009.1	*Cutibacterium_acnes*	FALSE
205	NZ_GL383232.1	*Cutibacterium_acnes*	FALSE	**255**	NZ_LKVE01000001.1	*Cutibacterium_acnes*	FALSE
206	NZ_GL383285.1	*Cutibacterium_acnes*	FALSE	**256**	NZ_NXAQ01000001.1	*Cutibacterium_acnes*	FALSE
207	NZ_GL383310.1	*Cutibacterium_acnes*	FALSE	**257**	NZ_NXAW01000001.1	*Cutibacterium_acnes*	FALSE
208	NZ_GL383383.1	*Cutibacterium_acnes*	FALSE	**258**	NZ_NXAU01000001.1	*Cutibacterium_acnes*	FALSE
209	NZ_GL383398.1	*Cutibacterium_acnes*	FALSE	**259**	NZ_PCNH01000010.1	*Cutibacterium_acnes*	FALSE
210	NZ_GL383478.1	*Cutibacterium_acnes*	FALSE	**260**	NZ_PCNI01000001.1	*Cutibacterium_acnes*	FALSE
211	NZ_GL383624.1	*Cutibacterium_acnes*	FALSE	**261**	NZ_PCNM01000001.1	*Cutibacterium_acnes*	FALSE
212	NZ_GL383637.1	*Cutibacterium_acnes*	FALSE	**262**	NZ_PCNN01000001.1	*Cutibacterium_acnes*	FALSE
213	NZ_GL383671.1	*Cutibacterium_acnes*	FALSE	**263**	NZ_PCNO01000001.1	*Cutibacterium_acnes*	FALSE
214	NZ_GL383689.1	*Cutibacterium_acnes*	FALSE	**264**	NZ_PCNP01000001.1	*Cutibacterium_acnes*	FALSE
215	NZ_GL383745.1	*Cutibacterium_acnes*	FALSE	**265**	NZ_PCNT01000001.1	*Cutibacterium_acnes*	FALSE
216	NZ_GL383800.1	*Cutibacterium_acnes*	FALSE	**266**	NZ_PCNV01000010.1	*Cutibacterium_acnes*	FALSE
217	NZ_GL383825.1	*Cutibacterium_acnes*	FALSE	**267**	NZ_PCNW01000001.1	*Cutibacterium_acnes*	FALSE
218	NZ_GL383855.1	*Cutibacterium_acnes*	FALSE	**268**	NZ_NMRI01000001.1	*Cutibacterium_acnes*	FALSE
219	NZ_GL383913.1	*Cutibacterium_acnes*	FALSE	**269**	NZ_NMRH01000001.1	*Cutibacterium_acnes*	FALSE
220	NZ_GL383933.1	*Cutibacterium_acnes*	FALSE	**270**	NZ_PCND01000010.1	*Cutibacterium_acnes*	FALSE
221	NZ_GL384189.1	*Cutibacterium_acnes*	FALSE	**271**	NZ_RAYH01000028.1	*Cutibacterium_acnes*	FALSE
222	NZ_GL384376.1	*Cutibacterium_acnes*	FALSE	**272**	NZ_CP033718.1	*Cutibacterium_acnes*	TRUE
223	NZ_GL384526.1	*Cutibacterium_acnes*	FALSE	**273**	NZ_CP025934.1	*Cutibacterium_acnes*	TRUE
224	NZ_GL383839.1	*Cutibacterium_acnes*	FALSE	**274**	NZ_RCYO01000001.1	*Cutibacterium_acnes*	FALSE
225	NZ_ADJL01000062.1	*Cutibacterium_acnes*	FALSE	**275**	NZ_VBXS01000010.1	*Cutibacterium_acnes*	FALSE
226	NZ_ADJM01000037.1	*Cutibacterium_acnes*	FALSE	**276**	NZ_VBXT01000001.1	*Cutibacterium_acnes*	FALSE
227	NZ_GL878455.1	*Cutibacterium_acnes*	FALSE	**277**	NZ_VBXW01000001.1	*Cutibacterium_acnes*	FALSE
228	NZ_GL883031.1	*Cutibacterium_acnes*	FALSE	**278**	NZ_VBXV01000001.1	*Cutibacterium_acnes*	FALSE
229	NZ_AFUM01000014.1	*Cutibacterium_acnes*	FALSE	**279**	NZ_VBYA01000001.1	*Cutibacterium_acnes*	FALSE
230	NC_016512.1	*Cutibacterium_acnes*	TRUE	**280**	NZ_VBYB01000011.1	*Cutibacterium_acnes*	FALSE
231	NZ_AIJP01000012.1	*Cutibacterium_acnes*	FALSE	**281**	NZ_VBYC01000001.1	*Cutibacterium_acnes*	FALSE
232	NC_018707.1	*Cutibacterium_acnes*	TRUE	**282**	NZ_VBYE01000002.1	*Cutibacterium_acnes*	FALSE
233	NZ_AOGR01000001.1	*Cutibacterium_acnes*	FALSE	**283**	NZ_VBYD01000010.1	*Cutibacterium_acnes*	FALSE
234	NZ_AWZZ01000001.1	*Cutibacterium_acnes*	FALSE	**284**	NZ_VBYH01000001.1	*Cutibacterium_acnes*	FALSE
235	NZ_AXZT01000001.1	*Cutibacterium_acnes*	FALSE	**285**	NZ_VBYM01000001.1	*Cutibacterium_acnes*	FALSE
236	NZ_JNHS01000001.1	*Cutibacterium_acnes*	FALSE	**286**	NZ_VBYN01000001.1	*Cutibacterium_acnes*	FALSE
237	NZ_CP012351.1	*Cutibacterium_acnes*	TRUE	**287**	NZ_VBYO01000001.1	*Cutibacterium_acnes*	FALSE
238	NZ_CP013693.1	*Cutibacterium_acnes*	FALSE	**288**	NZ_VBYT01000001.1	*Cutibacterium_acnes*	FALSE
239	NZ_LZCJ01000001.1	*Cutibacterium_acnes*	FALSE	**289**	NZ_VBYX01000001.1	*Cutibacterium_acnes*	FALSE
240	NZ_LZCK01000001.1	*Cutibacterium_acnes*	FALSE	**290**	NZ_VCHU01000001.1	*Cutibacterium_acnes*	FALSE
241	NZ_LZCM01000001.1	*Cutibacterium_acnes*	FALSE	**291**	NZ_QJHT01000005.1	*Cutibacterium_acnes*	FALSE
242	NZ_LZCP01000007.1	*Cutibacterium_acnes*	FALSE	**292**	NZ_QJHQ01000001.1	*Cutibacterium_acnes*	FALSE
243	NZ_MDGS01000001.1	*Cutibacterium_acnes*	FALSE	**293**	NZ_QJHZ01000010.1	*Cutibacterium_acnes*	FALSE
244	NZ_CM008360.1	*Cutibacterium_acnes*	FALSE	**294**	NZ_QJHW01000011.1	*Cutibacterium_acnes*	FALSE
245	NZ_CM008362.1	*Cutibacterium_acnes*	FALSE	**295**	NZ_QJIE01000001.1	*Cutibacterium_acnes*	FALSE
246	NZ_MVCF01000010.1	*Cutibacterium_acnes*	FALSE	**296**	NZ_QJIN01000010.1	*Cutibacterium_acnes*	FALSE
247	NZ_CM008364.1	*Cutibacterium_acnes*	FALSE	**297**	NZ_QJHR01000001.1	*Cutibacterium_acnes*	FALSE
248	NZ_MVCJ01000001.1	*Cutibacterium_acnes*	FALSE	**298**	NZ_QJHP01000010.1	*Cutibacterium_acnes*	FALSE
249	NZ_CM008365.1	*Cutibacterium_acnes*	FALSE	**299**	NZ_QJHS01000001.1	*Cutibacterium_acnes*	FALSE
250	NZ_MVCL01000001.1	*Cutibacterium_acnes*	FALSE	**300**	NZ_QJHX01000001.1	*Cutibacterium_acnes*	FALSE

Supplementary Table 9. Continued.

301	NC_014121.1	*Enterobacter_cloacae*	TRUE	**351**	NZ_QKNF01000001.1	*Enterobacter_cloacae*	FALSE
302	NC_018079.1	*Enterobacter_cloacae*	TRUE	**352**	NZ_QMCS01000001.1	*Enterobacter_cloacae*	FALSE
303	NZ_JTBZ02000001.1	*Enterobacter_hormaechei*	FALSE	**353**	NZ_QMCV01000001.1	*Enterobacter_cloacae*	FALSE
304	NZ_JTCB02000001.1	*Enterobacter_hormaechei*	FALSE	**354**	NZ_QTTI01000002.1	*Enterobacter_cloacae*	FALSE
305	NZ_JTCD02000001.1	*Enterobacter_hormaechei*	FALSE	**355**	NZ_PXJT01000100.1	*Enterobacter_cloacae*	FALSE
306	NZ_JTCH02000001.1	*Enterobacter_hormaechei*	FALSE	**356**	NZ_QRKH01000001.1	*Enterobacter_cloacae*	FALSE
307	NZ_JUHY01000001.1	*Enterobacter_hormaechei*	FALSE	**357**	NZ_NRQN01000100.1	*Enterobacter_cloacae*	FALSE
308	NZ_JUHW01000001.1	*Enterobacter_hormaechei*	FALSE	**358**	NZ_RXRH01000001.1	*Enterobacter_cloacae*	FALSE
309	NZ_JUHX01000001.1	*Enterobacter_hormaechei*	FALSE	**359**	NZ_BIGD01000001.1	*Enterobacter_cloacae*	FALSE
310	NZ_JUIA01000001.1	*Enterobacter_hormaechei*	FALSE	**360**	NZ_QKPC01000100.1	*Enterobacter_cloacae*	FALSE
311	NZ_JUIB01000001.1	*Enterobacter_hormaechei*	FALSE	**361**	NZ_QBFC01000001.1	*Enterobacter_cloacae*	FALSE
312	NZ_JUIC01000001.1	*Enterobacter_hormaechei*	FALSE	**362**	NZ_QJSL01000100.1	*Enterobacter_cloacae*	FALSE
313	NZ_JXAE01000001.1	*Enterobacter_cloacae*	FALSE	**363**	NZ_CP026536.1	*Enterobacter_cloacae*	TRUE
314	NZ_JSWY01000001.1	*Enterobacter_cloacae*	FALSE	**364**	NZ_QFLN01000999.1	*Enterobacter_cloacae*	FALSE
315	NZ_LEDW02000001.1	*Enterobacter_ludwigii*	FALSE	**365**	NZ_CP039318.1	*Enterobacter_cloacae*	TRUE
316	NZ_LEDR01000001.1	*Enterobacter_cloacae*	FALSE	**366**	NZ_CP039303.1	*Enterobacter_cloacae*	TRUE
317	NZ_LEDL01000001.1	*Enterobacter_cloacae*	FALSE	**367**	NZ_VSZU01000045.1	*Enterobacter_cloacae*	FALSE
318	NZ_JWCB01000001.1	*Enterobacter_cloacae*	FALSE	**368**	NZ_VWXR01000001.1	*Enterobacter_cloacae*	FALSE
319	NZ_JWAU01000006.1	*Enterobacter_cloacae*	FALSE	**369**	NZ_CP046116.1	*Enterobacter_cloacae*	TRUE
320	NZ_JVIB01000001.1	*Enterobacter_cloacae*	FALSE	**370**	NZ_WNXG01000108.1	*Enterobacter_cloacae*	FALSE
321	NZ_JVCE01000010.1	*Enterobacter_cloacae*	FALSE	**371**	NZ_JAAATU010000001.1	*Enterobacter_cloacae*	FALSE
322	NZ_LPOY01000001.1	*Enterobacter_cloacae*	FALSE	**372**	NZ_JAAATM010000001.	*Enterobacter_cloacae*	FALSE
323	NZ_LPOT01000001.1	*Enterobacter_cloacae*	FALSE	**373**	NZ_VKVR01000001.1	*Enterobacter_cloacae*	FALSE
324	NZ_LPOK01000001.1	*Enterobacter_cloacae*	FALSE	**374**	NZ_JAALCJ010000010.1	*Enterobacter_cloacae*	FALSE
325	NZ_LRIO01000001.1	*Enterobacter_cloacae*	FALSE	**375**	NZ_NIVN01000001.1	*Enterobacter_cloacae*	FALSE
326	NZ_LVUS01000001.1	*Enterobacter_cloacae*	FALSE	**376**	NZ_CP056117.1	*Enterobacter_cloacae*	TRUE
327	NZ_MTFV01000001.1	*Enterobacter_cloacae*	FALSE	**377**	NZ_CP056779.1	*Enterobacter_cloacae*	TRUE
328	NZ_MWLI01000001.1	*Enterobacter_cloacae*	FALSE	**378**	NZ_AP022228.1	*Enterobacter_cloacae*	TRUE
329	NZ_MWMD01000001.1	*Enterobacter_cloacae*	FALSE	**379**	NZ_JACZMS010000001.1	*Enterobacter_cloacae*	FALSE
330	NZ_CP017990.1	*Enterobacter_cloacae*	TRUE	**380**	NZ_JADBPT010000001.1	*Enterobacter_cloacae*	FALSE
331	NZ_NJCP01000006.1	*Enterobacter_cloacae*	FALSE	**381**	NZ_JADBPU010000001.1	*Enterobacter_cloacae*	FALSE
332	NZ_NPNX01000039.1	*Enterobacter_cloacae*	FALSE	**382**	NZ_JADBQA010000001.1	*Enterobacter_cloacae*	FALSE
333	NZ_NPNI01000033.1	*Enterobacter_cloacae*	FALSE	**383**	NZ_JADBQJ010000001.1	*Enterobacter_cloacae*	FALSE
334	NZ_NPNK01000037.1	*Enterobacter_cloacae*	FALSE	**384**	NZ_JADBQK010000001.1	*Enterobacter_cloacae*	FALSE
335	NZ_NPNO01000034.1	*Enterobacter_cloacae*	FALSE	**385**	NZ_JADBQM010000001.	*Enterobacter_cloacae*	FALSE
336	NZ_NXHK01000010.1	*Enterobacter_cloacae*	FALSE	**386**	NZ_JADBQN010000001.	*Enterobacter_cloacae*	FALSE
337	NZ_NEEM01000001.1	*Enterobacter_cloacae*	FALSE	**387**	NZ_JADBQR010000001.1	*Enterobacter_cloacae*	FALSE
338	NZ_PGWQ01000001.1	*Enterobacter_cloacae*	FALSE	**388**	NZ_JADBQP010000001.1	*Enterobacter_cloacae*	FALSE
339	NZ_PJZE01000001.1	*Enterobacter_cloacae*	FALSE	**389**	NZ_JADBQX010000001.1	*Enterobacter_cloacae*	FALSE
340	NZ_PQKU01000001.1	*Enterobacter_cloacae*	FALSE	**390**	NZ_JADBRB010000001.1	*Enterobacter_cloacae*	FALSE
341	NZ_PQKW01000001.1	*Enterobacter_cloacae*	FALSE	**391**	NZ_JADBRA010000001.1	*Enterobacter_cloacae*	FALSE
342	NZ_PQLB01000001.1	*Enterobacter_cloacae*	FALSE	**392**	NZ_JACOJF010000001.1	*Enterobacter_cloacae*	FALSE
343	NZ_PQLC01000001.1	*Enterobacter_cloacae*	FALSE	**393**	NZ_JADBRC010000001.1	*Enterobacter_cloacae*	FALSE
344	NZ_CP020528.1	*Enterobacter_cloacae*	TRUE	**394**	NZ_JADBRI010000001.1	*Enterobacter_cloacae*	FALSE
345	NZ_PYGW01000010.1	*Enterobacter_cloacae*	FALSE	**395**	NZ_JADBRK010000001.1	*Enterobacter_cloacae*	FALSE
346	NZ_QBPF01000001.1	*Enterobacter_cloacae*	FALSE	**396**	NZ_JADBRP010000001.1	*Enterobacter_cloacae*	FALSE
347	NZ_CP028950.1	*Enterobacter_cloacae*	TRUE	**397**	NZ_JADBRQ010000001.1	*Enterobacter_cloacae*	FALSE
348	NZ_BGKJ01000001.1	*Enterobacter_cloacae*	FALSE	**398**	NZ_JADBRV010000001.1	*Enterobacter_cloacae*	FALSE
349	NZ_BGLL01000001.1	*Enterobacter_cloacae*	FALSE	**399**	NZ_JADCKW010000001.	*Enterobacter_cloacae*	FALSE
350	NZ_BGMD01000001.1	*Enterobacter_cloacae*	FALSE	**400**	NZ_JADCKZ010000001.1	*Enterobacter_cloacae*	FALSE

Supplementary Table 9. Continued.

401	NZ_GL454559.1	*Enterococcus_faecalis*	FALSE	**451**	NZ_RKMT01000001.1	*Enterococcus_faecalis*	FALSE
402	NZ_JH805465.1	*Enterococcus_faecalis*	FALSE	**452**	NZ_RKPT01000001.1	*Enterococcus_faecalis*	FALSE
403	NZ_KB932335.1	*Enterococcus_faecalis*	FALSE	**453**	NZ_AP018543.1	*Enterococcus_faecalis*	TRUE
404	NZ_KB932352.1	*Enterococcus_faecalis*	FALSE	**454**	NZ_PJWQ01000010.1	*Enterococcus_faecalis*	FALSE
405	NZ_KB932494.1	*Enterococcus_faecalis*	FALSE	**455**	NZ_PJXO01000010.1	*Enterococcus_faecalis*	FALSE
406	NZ_KB932555.1	*Enterococcus_faecalis*	FALSE	**456**	NZ_CP028720.1	*Enterococcus_faecalis*	TRUE
407	NZ_KB944513.1	*Enterococcus_faecalis*	FALSE	**457**	NZ_QPTI01000010.1	*Enterococcus_faecalis*	FALSE
408	NZ_KB944692.1	*Enterococcus_faecalis*	FALSE	**458**	NZ_QPUR01000010.1	*Enterococcus_faecalis*	FALSE
409	NZ_KB944757.1	*Enterococcus_faecalis*	FALSE	**459**	NZ_QPYI01000010.1	*Enterococcus_faecalis*	FALSE
410	NZ_KB944792.1	*Enterococcus_faecalis*	FALSE	**460**	NZ_QPZH01000010.1	*Enterococcus_faecalis*	FALSE
411	NZ_KB944831.1	*Enterococcus_faecalis*	FALSE	**461**	NZ_QPXA01000010.1	*Enterococcus_faecalis*	FALSE
412	NZ_KB944975.1	*Enterococcus_faecalis*	FALSE	**462**	NZ_QPYK01000010.1	*Enterococcus_faecalis*	FALSE
413	NZ_KB946405.1	*Enterococcus_faecalis*	FALSE	**463**	NZ_QPYN01000010.1	*Enterococcus_faecalis*	FALSE
414	NZ_KB947345.1	*Enterococcus_faecalis*	FALSE	**464**	NZ_QPZC01000010.1	*Enterococcus_faecalis*	FALSE
415	NZ_KB948475.1	*Enterococcus_faecalis*	FALSE	**465**	NZ_QPZF01000010.1	*Enterococcus_faecalis*	FALSE
416	NZ_KB948686.1	*Enterococcus_faecalis*	FALSE	**466**	NZ_VHSA01000010.1	*Enterococcus_faecalis*	FALSE
417	NZ_KB949329.1	*Enterococcus_faecalis*	FALSE	**467**	NZ_VHSD01000010.1	*Enterococcus_faecalis*	FALSE
418	NZ_KE350775.1	*Enterococcus_faecalis*	FALSE	**468**	NZ_VHRA01000099.1	*Enterococcus_faecalis*	FALSE
419	NZ_KE352476.1	*Enterococcus_faecalis*	FALSE	**469**	NZ_VHQZ01000010.1	*Enterococcus_faecalis*	FALSE
420	NZ_KE352767.1	*Enterococcus_faecalis*	FALSE	**470**	NZ_QYNZ01000010.1	*Enterococcus_faecalis*	FALSE
421	NZ_KI913051.1	*Enterococcus_faecalis*	FALSE	**471**	NZ_QYLV01000010.1	*Enterococcus_faecalis*	FALSE
422	NZ_KI913061.1	*Enterococcus_faecalis*	FALSE	**472**	NZ_WMGP01000001.1	*Enterococcus_faecalis*	FALSE
423	NZ_KI913085.1	*Enterococcus_faecalis*	FALSE	**473**	NZ_WMGA01000001.1	*Enterococcus_faecalis*	FALSE
424	NZ_KI913093.1	*Enterococcus_faecalis*	FALSE	**474**	NZ_VWNT01000010.1	*Enterococcus_faecalis*	FALSE
425	NZ_AYLN01000001.1	*Enterococcus_faecalis*	FALSE	**475**	NZ_VWNY01000010.1	*Enterococcus_faecalis*	FALSE
426	NZ_JVZS01000023.1	*Enterococcus_faecalis*	FALSE	**476**	NZ_VWOA01000010.1	*Enterococcus_faecalis*	FALSE
427	NZ_JVQF01000003.1	*Enterococcus_faecalis*	FALSE	**477**	NZ_WYAE01000001.1	*Enterococcus_faecalis*	FALSE
428	NZ_JUPP01000001.1	*Enterococcus_faecalis*	FALSE	**478**	NZ_WYAC01000001.1	*Enterococcus_faecalis*	FALSE
429	NZ_JUNN01000039.1	*Enterococcus_faecalis*	FALSE	**479**	NZ_WXSN01000010.1	*Enterococcus_faecalis*	FALSE
430	NZ_CP014949.1	*Enterococcus_faecalis*	TRUE	**480**	NZ_JABTDO010000001.1	*Enterococcus_faecalis*	FALSE
431	NZ_MSQS01000001.1	*Enterococcus_faecalis*	FALSE	**481**	NZ_JACIKJ010000010.1	*Enterococcus_faecalis*	FALSE
432	NZ_NGMZ01000001.1	*Enterococcus_faecalis*	FALSE	**482**	NZ_JACYZS010000085.1	*Enterococcus_faecalis*	FALSE
433	NZ_CP015883.1	*Enterococcus_faecalis*	TRUE	**483**	NZ_JACYZU010000010.1	*Enterococcus_faecalis*	FALSE
434	NZ_PTUE01000001.1	*Enterococcus_faecalis*	FALSE	**484**	NZ_JACZBY010000021.1	*Enterococcus_faecalis*	FALSE
435	NZ_PTUL01000001.1	*Enterococcus_faecalis*	FALSE	**485**	NZ_JADBPN010000001.1	*Enterococcus_faecalis*	FALSE
436	NZ_PTUN01000001.1	*Enterococcus_faecalis*	FALSE	**486**	NZ_JADBPE010000001.1	*Enterococcus_faecalis*	FALSE
437	NZ_PTVE01000001.1	*Enterococcus_faecalis*	FALSE	**487**	NZ_JAAMRT010000001.	*Enterococcus_faecalis*	FALSE
438	NZ_PTVR01000001.1	*Enterococcus_faecalis*	FALSE	**488**	NZ_FRXS01000224.1	*Enterococcus_faecalis*	FALSE
439	NZ_PTXI01000001.1	*Enterococcus_faecalis*	FALSE	**489**	NZ_CABGJT010000001.1	*Enterococcus_faecalis*	FALSE
440	NZ_PUBU01000001.1	*Enterococcus_faecalis*	FALSE	**490**	NZ_CABGKB010000001.	*Enterococcus_faecalis*	FALSE
441	NZ_PUAX01000001.1	*Enterococcus_faecalis*	FALSE	**491**	NZ_CABGKY010000001.1	*Enterococcus_faecalis*	FALSE
442	NZ_PUAQ01000001.1	*Enterococcus_faecalis*	FALSE	**492**	NZ_CABGLE010000001.1	*Enterococcus_faecalis*	FALSE
443	NZ_PTYM01000001.1	*Enterococcus_faecalis*	FALSE	**493**	NZ_CABGLU010000001.	*Enterococcus_faecalis*	FALSE
444	NZ_LMBT01000001.1	*Enterococcus_faecalis*	FALSE	**494**	NZ_CABGMH010000001	*Enterococcus_faecalis*	FALSE
445	NZ_QFYL01000001.1	*Enterococcus_faecalis*	FALSE	**495**	NZ_CABGNX010000001.	*Enterococcus_faecalis*	FALSE
446	NZ_KZ845811.1	*Enterococcus_faecalis*	FALSE	**496**	NZ_CABGOF010000001.	*Enterococcus_faecalis*	FALSE
447	NZ_KZ845861.1	*Enterococcus_faecalis*	FALSE	**497**	NZ_CABGPC010000001.1	*Enterococcus_faecalis*	FALSE
448	NZ_QSSZ01000001.1	*Enterococcus_faecalis*	FALSE	**498**	NZ_CABGST010000001.1	*Enterococcus_faecalis*	FALSE
449	NZ_LGBU01000001.1	*Enterococcus_faecalis*	FALSE	**499**	NZ_CABGTD010000001.	*Enterococcus_faecalis*	FALSE
450	NZ_LGBS01000001.1	*Enterococcus_faecalis*	FALSE	**500**	NZ_CABGXH010000001.	*Enterococcus_faecalis*	FALSE

Supplementary Table 9. Continued.

501	NZ_AERP01000070.1	*Escherichia_coli*	FALSE	**551**	NZ_QOXJ01000100.1	*Escherichia_coli*	FALSE
502	NZ_AFDR01000030.1	*Escherichia_coli*	FALSE	**552**	NZ_QOGT01000001.1	*Escherichia_coli*	FALSE
503	NZ_HE964769.1	*Escherichia_coli*	FALSE	**553**	NZ_QOFT01000001.1	*Escherichia_coli*	FALSE
504	NZ_AOQX01000001.1	*Escherichia_coli*	FALSE	**554**	NZ_QOOK01000031.1	*Escherichia_coli*	FALSE
505	NZ_KE136677.1	*Escherichia_coli*	FALSE	**555**	NZ_QGVL01000001.1	*Escherichia_coli*	FALSE
506	NZ_KE699111.1	*Escherichia_coli*	FALSE	**556**	NZ_QGZM01000001.1	*Escherichia_coli*	FALSE
507	NZ_KE700784.1	*Escherichia_coli*	FALSE	**557**	NZ_PNSN01000001.1	*Escherichia_coli*	FALSE
508	NZ_KE702728.1	*Escherichia_coli*	FALSE	**558**	NZ_POAK01000001.1	*Escherichia_coli*	FALSE
509	NZ_AVRJ01000001.1	*Escherichia_coli*	FALSE	**559**	NZ_NPIP01000001.1	*Escherichia_coli*	FALSE
510	NZ_KI534242.1	*Escherichia_coli*	FALSE	**560**	NZ_SHJC01000010.1	*Escherichia_coli*	FALSE
511	NZ_JOMW01000001.1	*Escherichia_coli*	FALSE	**561**	NZ_SNTL01000100.1	*Escherichia_coli*	FALSE
512	NZ_JORW01000001.1	*Escherichia_coli*	FALSE	**562**	NZ_SONW01000001.1	*Escherichia_coli*	FALSE
513	NZ_JSOO01000002.1	*Escherichia_coli*	FALSE	**563**	NZ_CM017186.1	*Escherichia_coli*	FALSE
514	NZ_JYIG01000001.1	*Escherichia_coli*	FALSE	**564**	NZ_SRWH01000100.1	*Escherichia_coli*	FALSE
515	NZ_JXVY01000001.1	*Escherichia_coli*	FALSE	**565**	NZ_SRWK01000100.1	*Escherichia_coli*	FALSE
516	NZ_CXYN01000001.1	*Escherichia_coli*	FALSE	**566**	NZ_RRRK01000161.1	*Escherichia_coli*	FALSE
517	NZ_LQWF01000001.1	*Escherichia_coli*	FALSE	**567**	NZ_RRJH01000111.1	*Escherichia_coli*	FALSE
518	NZ_LRXU01000001.1	*Escherichia_coli*	FALSE	**568**	NZ_RRJO01000133.1	*Escherichia_coli*	FALSE
519	NZ_LVSM01000001.1	*Escherichia_coli*	FALSE	**569**	NZ_BDJT01000001.1	*Escherichia_coli*	FALSE
520	NZ_CP015843.2	*Escherichia_coli*	TRUE	**570**	NZ_BFJE01000001.1	*Escherichia_coli*	FALSE
521	NZ_CP017249.1	*Escherichia_coli*	TRUE	**571**	NZ_BFVG01000001.1	*Escherichia_coli*	FALSE
522	NZ_MOJH01000001.1	*Escherichia_coli*	FALSE	**572**	NZ_BFWC01000001.1	*Escherichia_coli*	FALSE
523	NZ_MOKS01000001.1	*Escherichia_coli*	FALSE	**573**	NZ_BGCD01000001.1	*Escherichia_coli*	FALSE
524	NZ_MUIP01000036.1	*Escherichia_coli*	FALSE	**574**	NZ_BGDN01000001.1	*Escherichia_coli*	FALSE
525	NZ_NCSC01000001.1	*Escherichia_coli*	FALSE	**575**	NZ_BIES01000001.1	*Escherichia_coli*	FALSE
526	NZ_JZGQ01000001.1	*Escherichia_coli*	FALSE	**576**	NZ_SPVZ01000144.1	*Escherichia_coli*	FALSE
527	NZ_LYAQ01000001.1	*Escherichia_coli*	FALSE	**577**	NZ_SQAJ01000041.1	*Escherichia_coli*	FALSE
528	NZ_NHQF01000010.1	*Escherichia_coli*	FALSE	**578**	NZ_SQDY01000045.1	*Escherichia_coli*	FALSE
529	NZ_BDPA01000001.1	*Escherichia_coli*	FALSE	**579**	NZ_SQFV01000126.1	*Escherichia_coli*	FALSE
530	NZ_NTDL01000015.1	*Escherichia_coli*	FALSE	**580**	NZ_SQDI01000169.1	*Escherichia_coli*	FALSE
531	NZ_NJVD01000099.1	*Escherichia_coli*	FALSE	**581**	NZ_SQGX01000090.1	*Escherichia_coli*	FALSE
532	NZ_NJZJ01000100.1	*Escherichia_coli*	FALSE	**582**	NZ_SQIE01000090.1	*Escherichia_coli*	FALSE
533	NZ_NTPP01000010.1	*Escherichia_coli*	FALSE	**583**	NZ_QYJU01000001.1	*Escherichia_coli*	FALSE
534	NZ_NOCD01000100.1	*Escherichia_coli*	FALSE	**584**	NZ_BJPS01000001.1	*Escherichia_coli*	FALSE
535	NZ_NLWR01000099.1	*Escherichia_coli*	FALSE	**585**	NZ_WJFT01000001.1	*Escherichia_coli*	FALSE
536	NZ_NLVM01000100.1	*Escherichia_coli*	FALSE	**586**	NZ_WCEY01000001.1	*Escherichia_coli*	FALSE
537	NZ_NJSW01000100.1	*Escherichia_coli*	FALSE	**587**	NZ_WSWT01000100.1	*Escherichia_coli*	FALSE
538	NZ_NJRV01000100.1	*Escherichia_coli*	FALSE	**588**	NZ_WSJJ01000010.1	*Escherichia_coli*	FALSE
539	NZ_NONX01000100.1	*Escherichia_coli*	FALSE	**589**	NZ_JAAUVL010000001.1	*Escherichia_coli*	FALSE
540	NZ_NMBM01000098.1	*Escherichia_coli*	FALSE	**590**	NZ_WVVH01000001.1	*Escherichia_coli*	FALSE
541	NZ_LULC01000066.1	*Escherichia_coli*	FALSE	**591**	NZ_WVVS01000001.1	*Escherichia_coli*	FALSE
542	NZ_BEOS01000001.1	*Escherichia_coli*	FALSE	**592**	NZ_CP051727.1	*Escherichia_coli*	TRUE
543	NZ_PTNX01000010.1	*Escherichia_coli*	FALSE	**593**	NZ_JABBFK010000095.1	*Escherichia_coli*	FALSE
544	NZ_QBCU01000001.1	*Escherichia_coli*	FALSE	**594**	NZ_WEQV01000001.1	*Escherichia_coli*	FALSE
545	NZ_QFSO01000001.1	*Escherichia_coli*	FALSE	**595**	NZ_JABTYK010000078.1	*Escherichia_coli*	FALSE
546	NZ_QFRZ01000001.1	*Escherichia_coli*	FALSE	**596**	NZ_JACGZT010000001.1	*Escherichia_coli*	FALSE
547	NZ_NRUY01000065.1	*Escherichia_coli*	FALSE	**597**	NZ_CP059931.1	*Escherichia_coli*	TRUE
548	NZ_BGRX01000001.1	*Escherichia_coli*	FALSE	**598**	NZ_JACXWL010000100.1	*Escherichia_coli*	FALSE
549	NZ_NFTM01000087.1	*Escherichia_coli*	FALSE	**599**	NZ_JACZPA010000001.1	*Escherichia_coli*	FALSE
550	NZ_QOEB01000001.1	*Escherichia_coli*	FALSE	**600**	NZ_JADCMB010000010.	*Escherichia_coli*	FALSE

Supplementary Table 9. Continued.

601	NZ_AJZY01000001.1	*Klebsiella_pneumoniae*	FALSE	**651**	NZ_STAE01000001.1	*Klebsiella_pneumoniae*	FALSE
602	NZ_ARSQ01000001.1	*Klebsiella_pneumoniae*	FALSE	**652**	NZ_STAY01000001.1	*Klebsiella_pneumoniae*	FALSE
603	NZ_KK036582.1	*Klebsiella_pneumoniae*	FALSE	**653**	NZ_VIYC01000097.1	*Klebsiella_pneumoniae*	FALSE
604	NZ_JAQM01000001.1	*Klebsiella_pneumoniae*	FALSE	**654**	NZ_VTRN01000001.1	*Klebsiella_pneumoniae*	FALSE
605	NZ_CP010392.1	*Klebsiella_pneumoniae*	TRUE	**655**	NZ_WMJM01000100.1	*Klebsiella_pneumoniae*	FALSE
606	NZ_KQ089048.1	*Klebsiella_pneumoniae*	FALSE	**656**	NZ_VDJC01000079.1	*Klebsiella_pneumoniae*	FALSE
607	NZ_KQ236936.1	*Klebsiella_pneumoniae*	FALSE	**657**	NZ_CP039802.1	*Klebsiella_pneumoniae*	TRUE
608	NZ_LIPK01000096.1	*Klebsiella_pneumoniae*	FALSE	**658**	NZ_VKUM01000001.1	*Klebsiella_pneumoniae*	FALSE
609	NZ_CP012753.1	*Klebsiella_pneumoniae*	TRUE	**659**	NZ_CP053771.1	*Klebsiella_pneumoniae*	TRUE
610	NZ_JTFU01000001.1	*Klebsiella_pneumoniae*	FALSE	**660**	NZ_CP060421.1	*Klebsiella_pneumoniae*	TRUE
611	NZ_JXIY01000001.1	*Klebsiella_pneumoniae*	FALSE	**661**	NZ_CP050826.1	*Klebsiella_pneumoniae*	TRUE
612	NZ_LXMK01000001.1	*Klebsiella_pneumoniae*	FALSE	**662**	NZ_JAAWVU010000001	*Klebsiella_pneumoniae*	FALSE
613	NZ_LXMT01000001.1	*Klebsiella_pneumoniae*	FALSE	**663**	NZ_RQIX01000100.1	*Klebsiella_pneumoniae*	FALSE
614	NZ_LLWT01000001.1	*Klebsiella_pneumoniae*	FALSE	**664**	NZ_JABMMC010000001	*Klebsiella_pneumoniae*	FALSE
615	NZ_MPZH04000023.1	*Klebsiella_pneumoniae*	FALSE	**665**	NZ_JAANKQ010000010.1	*Klebsiella_pneumoniae*	FALSE
616	NZ_MPYK04000017.1	*Klebsiella_pneumoniae*	FALSE	**666**	NZ_JACEYW010000001.1	*Klebsiella_pneumoniae*	FALSE
617	NZ_CP018686.1	*Klebsiella_pneumoniae*	TRUE	**667**	NZ_CP061053.1	*Klebsiella_pneumoniae*	TRUE
618	NZ_CP015392.1	*Klebsiella_pneumoniae*	TRUE	**668**	NZ_JABMCK010000001.1	*Klebsiella_pneumoniae*	FALSE
619	NZ_MZZQ01000055.1	*Klebsiella_pneumoniae*	FALSE	**669**	NZ_JACYPJ010000001.1	*Klebsiella_pneumoniae*	FALSE
620	NZ_MOSN01000001.1	*Klebsiella_pneumoniae*	FALSE	**670**	NZ_JADAYE010000091.1	*Klebsiella_pneumoniae*	FALSE
621	NZ_NDEV01000008.1	*Klebsiella_pneumoniae*	FALSE	**671**	NZ_JADMWP010000100	*Klebsiella_pneumoniae*	FALSE
622	NZ_NFVU01000001.1	*Klebsiella_pneumoniae*	FALSE	**672**	NZ_JADODD010000100.1	*Klebsiella_pneumoniae*	FALSE
623	NZ_NPZK01000073.1	*Klebsiella_pneumoniae*	FALSE	**673**	NZ_CP063428.1	*Klebsiella_pneumoniae*	TRUE
624	NZ_NPZW01000011.1	*Klebsiella_pneumoniae*	FALSE	**674**	NZ_JADIGK010000010.1	*Klebsiella_pneumoniae*	FALSE
625	NZ_NLFC01000001.1	*Klebsiella_pneumoniae*	FALSE	**675**	NZ_FLDT01000001.1	*Klebsiella_pneumoniae*	FALSE
626	NZ_NLDJ01000001.1	*Klebsiella_pneumoniae*	FALSE	**676**	NZ_FWLQ01000065.1	*Klebsiella_pneumoniae*	FALSE
627	NZ_NIAT01000062.1	*Klebsiella_pneumoniae*	FALSE	**677**	NZ_FWML01000071.1	*Klebsiella_pneumoniae*	FALSE
628	NZ_NXDR01000067.1	*Klebsiella_pneumoniae*	FALSE	**678**	NZ_UGNL01000002.1	*Klebsiella_pneumoniae*	FALSE
629	NZ_NXHT01000010.1	*Klebsiella_pneumoniae*	FALSE	**679**	NZ_UFFA01000151.1	*Klebsiella_pneumoniae*	FALSE
630	NZ_NICX01000010.1	*Klebsiella_pneumoniae*	FALSE	**680**	NZ_UIEL01000083.1	*Klebsiella_pneumoniae*	FALSE
631	NZ_PJPR01000100.1	*Klebsiella_pneumoniae*	FALSE	**681**	NZ_UIRW01000116.1	*Klebsiella_pneumoniae*	FALSE
632	NZ_NWCJ01000099.1	*Klebsiella_pneumoniae*	FALSE	**682**	NZ_UIUR01000084.1	*Klebsiella_pneumoniae*	FALSE
633	NZ_NWEY01000100.1	*Klebsiella_pneumoniae*	FALSE	**683**	NZ_UIUY01000066.1	*Klebsiella_pneumoniae*	FALSE
634	NZ_CP037963.1	*Klebsiella_pneumoniae*	TRUE	**684**	NZ_UJCS01000118.1	*Klebsiella_pneumoniae*	FALSE
635	NZ_NCOK01000001.1	*Klebsiella_pneumoniae*	FALSE	**685**	NZ_UJFP01000102.1	*Klebsiella_pneumoniae*	FALSE
636	NZ_PIDV01000100.1	*Klebsiella_pneumoniae*	FALSE	**686**	NZ_UJHZ01000124.1	*Klebsiella_pneumoniae*	FALSE
637	NZ_PPZG01000001.1	*Klebsiella_pneumoniae*	FALSE	**687**	NZ_UJUX01000287.1	*Klebsiella_pneumoniae*	FALSE
638	NZ_PQAD01000001.1	*Klebsiella_pneumoniae*	FALSE	**688**	NZ_UJXA01000111.1	*Klebsiella_pneumoniae*	FALSE
639	NZ_PUZG02000118.1	*Klebsiella_pneumoniae*	FALSE	**689**	NZ_UKBA01000104.1	*Klebsiella_pneumoniae*	FALSE
640	NZ_PUZQ02000021.1	*Klebsiella_pneumoniae*	FALSE	**690**	NZ_UKDU01000066.1	*Klebsiella_pneumoniae*	FALSE
641	NZ_PVAC02000113.1	*Klebsiella_pneumoniae*	FALSE	**691**	NZ_UKEW01000530.1	*Klebsiella_pneumoniae*	FALSE
642	NZ_PZNK01000010.1	*Klebsiella_pneumoniae*	FALSE	**692**	NZ_UKHU01000064.1	*Klebsiella_pneumoniae*	FALSE
643	NZ_QIUI01000010.1	*Klebsiella_pneumoniae*	FALSE	**693**	NZ_UKKL01000055.1	*Klebsiella_pneumoniae*	FALSE
644	NZ_NELE01000098.1	*Klebsiella_pneumoniae*	FALSE	**694**	NZ_UKMM01000064.1	*Klebsiella_pneumoniae*	FALSE
645	NZ_QJNK01000001.1	*Klebsiella_pneumoniae*	FALSE	**695**	NZ_UKUJ01000181.1	*Klebsiella_pneumoniae*	FALSE
646	NZ_QRCG01000011.1	*Klebsiella_pneumoniae*	FALSE	**696**	NZ_UKUS01000072.1	*Klebsiella_pneumoniae*	FALSE
647	NZ_RHUI01000100.1	*Klebsiella_pneumoniae*	FALSE	**697**	NZ_UKVV01000148.1	*Klebsiella_pneumoniae*	FALSE
648	NZ_QFMA01000001.1	*Klebsiella_pneumoniae*	FALSE	**698**	NZ_UKZF01000063.1	*Klebsiella_pneumoniae*	FALSE
649	NZ_QFML01000070.1	*Klebsiella_pneumoniae*	FALSE	**699**	NZ_UKYI01000070.1	*Klebsiella_pneumoniae*	FALSE
650	NZ_SSZW01000001.1	*Klebsiella_pneumoniae*	FALSE	**700**	NZ_ULBA01000074.1	*Klebsiella_pneumoniae*	FALSE

Supplementary Table 9. Continued.

701	NC_012803.1	*Micrococcus_luteus*	TRUE	**751**	NZ_WBMF01000001.1	*Micrococcus_luteus*	FALSE
702	NZ_ADCD01000112.1	*Micrococcus_luteus*	FALSE	**752**	NZ_CP046319.1	*Micrococcus_luteus*	TRUE
703	NZ_CABC01000134.1	*Micrococcus_luteus*	FALSE	**753**	NZ_CP043842.1	*Micrococcus_luteus*	TRUE
704	NZ_JFYP01000001.1	*Micrococcus_luteus*	FALSE	**754**	NZ_CP043849.1	*Micrococcus_luteus*	TRUE
705	NZ_JFYU01000001.1	*Micrococcus_luteus*	FALSE	**755**	NZ_CP041689.1	*Micrococcus_luteus*	TRUE
706	NZ_JFYQ01000001.1	*Micrococcus_luteus*	FALSE	**756**	NZ_JAAAXP010000001.1	*Micrococcus_luteus*	FALSE
707	NZ_JTKE01000001.1	*Micrococcus_luteus*	FALSE	**757**	NZ_JAAOIP010000049.1	*Micrococcus_luteus*	FALSE
708	NZ_CP007437.1	*Micrococcus_luteus*	TRUE	**758**	NZ_JABAFF010000001.1	*Micrococcus_luteus*	FALSE
709	NZ_JWEM01000024.1	*Micrococcus_luteus*	FALSE	**759**	NZ_JABBWS010000001.	*Micrococcus_luteus*	FALSE
710	NZ_JUTN01000041.1	*Micrococcus_luteus*	FALSE	**760**	NZ_JABBWT010000001.	*Micrococcus_luteus*	FALSE
711	NZ_LQAC01000001.1	*Micrococcus_luteus*	FALSE	**761**	NZ_CP059046.1	*Micrococcus_luteus*	TRUE
712	NZ_LSKC01000001.1	*Micrococcus_luteus*	FALSE	**762**	NZ_CP058971.1	*Micrococcus_luteus*	TRUE
713	NZ_MAYP01000001.1	*Micrococcus_luteus*	FALSE	**763**	NZ_FRCE01000037.1	*Micrococcus_luteus*	FALSE
714	NZ_LWGN01000001.1	*Micrococcus_luteus*	FALSE	**764**	NZ_FUKT01000114.1	*Micrococcus_luteus*	FALSE
715	NZ_NAQZ01000001.1	*Micrococcus_luteus*	FALSE	**765**	NZ_UGPO01000002.1	*Micrococcus_luteus*	FALSE
716	NZ_NCWR01000010.1	*Micrococcus_luteus*	FALSE	**766**	NZ_UGPQ01000002.1	*Micrococcus_luteus*	FALSE
717	NZ_NCXE01000010.1	*Micrococcus_luteus*	FALSE	**767**	NZ_LS483396.1	*Micrococcus_luteus*	TRUE
718	NZ_NRSP01000001.1	*Micrococcus_luteus*	FALSE	**768**	NZ_CABIZL010000001.1	*Micrococcus_luteus*	FALSE
719	NZ_PDEF01000001.1	*Micrococcus_luteus*	FALSE	**769**	NC_010554.1	*Proteus_mirabilis*	TRUE
720	NZ_PKJT01000001.1	*Micrococcus_luteus*	FALSE	**770**	NZ_GG668629.1	*Proteus_mirabilis*	FALSE
721	NZ_PKJU01000001.1	*Micrococcus_luteus*	FALSE	**771**	NZ_JH815505.1	*Proteus_mirabilis*	FALSE
722	NZ_PKJV01000001.1	*Micrococcus_luteus*	FALSE	**772**	NZ_JH815533.1	*Proteus_mirabilis*	FALSE
723	NZ_PNGM01000001.1	*Micrococcus_luteus*	FALSE	**773**	NZ_AORN01000001.1	*Proteus_mirabilis*	FALSE
724	NZ_CP033200.1	*Micrococcus_luteus*	TRUE	**774**	NC_022000.1	*Proteus_mirabilis*	TRUE
725	NZ_CP026366.1	*Micrococcus_luteus*	FALSE	**775**	NZ_JTAW02000001.1	*Proteus_mirabilis*	FALSE
726	NZ_SEML01000001.1	*Micrococcus_luteus*	FALSE	**776**	NZ_CP026062.1	*Proteus_mirabilis*	TRUE
727	NZ_SWLA01000010.1	*Micrococcus_luteus*	FALSE	**777**	NZ_CP026051.1	*Proteus_mirabilis*	TRUE
728	NZ_SWFK01000010.1	*Micrococcus_luteus*	FALSE	**778**	NZ_LANL01000045.1	*Proteus_mirabilis*	FALSE
729	NZ_CP040019.1	*Micrococcus_luteus*	TRUE	**779**	NZ_JWCS01000021.1	*Proteus_mirabilis*	FALSE
730	NZ_VFQL01000098.1	*Micrococcus_luteus*	FALSE	**780**	NZ_JWBG01000030.1	*Proteus_mirabilis*	FALSE
731	NZ_VDOA01000010.1	*Micrococcus_luteus*	FALSE	**781**	NZ_JVWE01000001.1	*Proteus_mirabilis*	FALSE
732	NZ_VDOB01000100.1	*Micrococcus_luteus*	FALSE	**782**	NZ_JVVD01000013.1	*Proteus_mirabilis*	FALSE
733	NZ_VDOF01000010.1	*Micrococcus_luteus*	FALSE	**783**	NZ_JVXV01000028.1	*Proteus_mirabilis*	FALSE
734	NZ_VDOI01000100.1	*Micrococcus_luteus*	FALSE	**784**	NZ_JVTJ01000109.1	*Proteus_mirabilis*	FALSE
735	NZ_VDRP01000010.1	*Micrococcus_luteus*	FALSE	**785**	NZ_JVPB01000058.1	*Proteus_mirabilis*	FALSE
736	NZ_VDRY01000010.1	*Micrococcus_luteus*	FALSE	**786**	NZ_JVEH01000013.1	*Proteus_mirabilis*	FALSE
737	NZ_VDRZ01000100.1	*Micrococcus_luteus*	FALSE	**787**	NZ_JVMQ01000004.1	*Proteus_mirabilis*	FALSE
738	NZ_VDTK01000100.1	*Micrococcus_luteus*	FALSE	**788**	NZ_JUXR01000250.1	*Proteus_mirabilis*	FALSE
739	NZ_VDZA01000100.1	*Micrococcus_luteus*	FALSE	**789**	NZ_JVJQ01000012.1	*Proteus_mirabilis*	FALSE
740	NZ_VDZG01000100.1	*Micrococcus_luteus*	FALSE	**790**	NZ_JVHX01000018.1	*Proteus_mirabilis*	FALSE
741	NZ_VEAO01000010.1	*Micrococcus_luteus*	FALSE	**791**	NZ_JVHI01000004.1	*Proteus_mirabilis*	FALSE
742	NZ_VECJ01000010.1	*Micrococcus_luteus*	FALSE	**792**	NZ_JSUP01000001.1	*Proteus_mirabilis*	FALSE
743	NZ_VECL01000010.1	*Micrococcus_luteus*	FALSE	**793**	NZ_LNHT01000001.1	*Proteus_mirabilis*	FALSE
744	NZ_VECU01000100.1	*Micrococcus_luteus*	FALSE	**794**	NZ_KQ960957.1	*Proteus_mirabilis*	FALSE
745	NZ_VECV01000100.1	*Micrococcus_luteus*	FALSE	**795**	NZ_LQNO01000021.1	*Proteus_mirabilis*	FALSE
746	NZ_VEDL01000010.1	*Micrococcus_luteus*	FALSE	**796**	NZ_LWDB01000001.1	*Proteus_mirabilis*	FALSE
747	NZ_VEDR01000010.1	*Micrococcus_luteus*	FALSE	**797**	NZ_LTBK01000001.1	*Proteus_mirabilis*	FALSE
748	NZ_VEFX01000095.1	*Micrococcus_luteus*	FALSE	**798**	NZ_LWUL01000001.1	*Proteus_mirabilis*	FALSE
749	NZ_VEGF01000100.1	*Micrococcus_luteus*	FALSE	**799**	NZ_LWUM01000001.1	*Proteus_mirabilis*	FALSE
750	NZ_VLJZ01000001.1	*Micrococcus_luteus*	FALSE	**800**	NZ_MCOR01000001.1	*Proteus_mirabilis*	FALSE

Supplementary Table 9. Continued.

801	NZ_NJFA02000001.1	*Proteus_mirabilis*	FALSE	**851**	NZ_JAALCE010000010.1	*Proteus_mirabilis*	FALSE
802	NZ_PTIJ02000012.1	*Proteus_mirabilis*	FALSE	**852**	NZ_SCIO01000010.1	*Proteus_mirabilis*	FALSE
803	NZ_CP026581.1	*Proteus_mirabilis*	TRUE	**853**	NZ_CP049941.1	*Proteus_mirabilis*	TRUE
804	NZ_CP028356.1	*Proteus_mirabilis*	TRUE	**854**	NZ_JAAOMZ010000001.	*Proteus_mirabilis*	FALSE
805	NZ_CP028522.1	*Proteus_mirabilis*	TRUE	**855**	NZ_JAAONM010000001	*Proteus_mirabilis*	FALSE
806	NZ_QDAK01000001.1	*Proteus_mirabilis*	FALSE	**856**	NZ_JAAONH010000001.	*Proteus_mirabilis*	FALSE
807	NZ_QGGA01000095.1	*Proteus_mirabilis*	FALSE	**857**	NZ_JAAOND010000001.	*Proteus_mirabilis*	FALSE
808	NZ_BGKT01000001.1	*Proteus_mirabilis*	FALSE	**858**	NZ_JAAONE010000001.	*Proteus_mirabilis*	FALSE
809	NZ_BGKW01000001.1	*Proteus_mirabilis*	FALSE	**859**	NZ_JAAONI010000001.1	*Proteus_mirabilis*	FALSE
810	NZ_QHLF01000010.1	*Proteus_mirabilis*	FALSE	**860**	NZ_JAAUWF010000010.	*Proteus_mirabilis*	FALSE
811	NZ_CP029725.1	*Proteus_mirabilis*	TRUE	**861**	NZ_WNEI01000099.1	*Proteus_mirabilis*	FALSE
812	NZ_PURY01000044.1	*Proteus_mirabilis*	FALSE	**862**	NZ_CP051260.1	*Proteus_mirabilis*	TRUE
813	NZ_CP034091.1	*Proteus_mirabilis*	TRUE	**863**	NZ_JAAQRB010000001.	*Proteus_mirabilis*	FALSE
814	NZ_SELN01000001.1	*Proteus_mirabilis*	FALSE	**864**	NZ_JAAQQZ010000001.	*Proteus_mirabilis*	FALSE
815	NZ_SPSR01000001.1	*Proteus_mirabilis*	FALSE	**865**	NZ_JAAQQY010000001.	*Proteus_mirabilis*	FALSE
816	NZ_SPTD01000001.1	*Proteus_mirabilis*	FALSE	**866**	NZ_CP053894.1	*Proteus_mirabilis*	TRUE
817	NZ_SPSU01000001.1	*Proteus_mirabilis*	FALSE	**867**	NZ_CP053615.1	*Proteus_mirabilis*	TRUE
818	NZ_SPSV01000001.1	*Proteus_mirabilis*	FALSE	**868**	NZ_CP053682.1	*Proteus_mirabilis*	TRUE
819	NZ_SPSW01000001.1	*Proteus_mirabilis*	FALSE	**869**	NZ_KI519080.1	*Pseudomonas_aeruginosa*	FALSE
820	NZ_SPSX01000001.1	*Proteus_mirabilis*	FALSE	**870**	NZ_KI519305.1	*Pseudomonas_aeruginosa*	FALSE
821	NZ_SPOX01000010.1	*Proteus_mirabilis*	FALSE	**871**	NZ_AWYW01000001.1	*Pseudomonas_aeruginosa*	FALSE
822	NZ_SPTA01000001.1	*Proteus_mirabilis*	FALSE	**872**	NZ_JTTQ01000001.1	*Pseudomonas_aeruginosa*	FALSE
823	NZ_SPTH01000001.1	*Proteus_mirabilis*	FALSE	**873**	NZ_JTTB01000001.1	*Pseudomonas_aeruginosa*	FALSE
824	NZ_SPTI01000001.1	*Proteus_mirabilis*	FALSE	**874**	NZ_JTRQ01000001.1	*Pseudomonas_aeruginosa*	FALSE
825	NZ_QNXA01000070.1	*Proteus_mirabilis*	FALSE	**875**	NZ_JTQT01000001.1	*Pseudomonas_aeruginosa*	FALSE
826	NZ_VJOR01000032.1	*Proteus_mirabilis*	FALSE	**876**	NZ_JTOE01000001.1	*Pseudomonas_aeruginosa*	FALSE
827	NZ_CP041391.1	*Proteus_mirabilis*	TRUE	**877**	NZ_JTVD01000001.1	*Pseudomonas_aeruginosa*	FALSE
828	NZ_CP043332.1	*Proteus_mirabilis*	TRUE	**878**	NZ_JTXN01000001.1	*Pseudomonas_aeruginosa*	FALSE
829	NZ_CP043870.1	*Proteus_mirabilis*	FALSE	**879**	NZ_JTYF01000001.1	*Pseudomonas_aeruginosa*	FALSE
830	NZ_CP044028.1	*Proteus_mirabilis*	TRUE	**880**	NZ_LDKL01000001.1	*Pseudomonas_aeruginosa*	FALSE
831	NZ_WEKH01000001.1	*Proteus_mirabilis*	FALSE	**881**	NZ_LDLY01000001.1	*Pseudomonas_aeruginosa*	FALSE
832	NZ_WEKI01000001.1	*Proteus_mirabilis*	FALSE	**882**	NZ_LDKT01000001.1	*Pseudomonas_aeruginosa*	FALSE
833	NZ_WEKK01000001.1	*Proteus_mirabilis*	FALSE	**883**	NZ_LEJT01000001.1	*Pseudomonas_aeruginosa*	FALSE
834	NZ_CP045538.2	*Proteus_mirabilis*	TRUE	**884**	NZ_JYLC01000010.1	*Pseudomonas_aeruginosa*	FALSE
835	NZ_CP044134.1	*Proteus_mirabilis*	TRUE	**885**	NZ_JVPC01000001.1	*Pseudomonas_aeruginosa*	FALSE
836	NZ_CP044136.1	*Proteus_mirabilis*	TRUE	**886**	NZ_CWFR01000001.1	*Pseudomonas_aeruginosa*	FALSE
837	NZ_WNBF01000046.1	*Proteus_mirabilis*	FALSE	**887**	NZ_LFXR01000072.1	*Pseudomonas_aeruginosa*	FALSE
838	NZ_RCIK01000083.1	*Proteus_mirabilis*	FALSE	**888**	NZ_CTES01000001.1	*Pseudomonas_aeruginosa*	FALSE
839	NZ_RCIJ01000064.1	*Proteus_mirabilis*	FALSE	**889**	NZ_LLLT01000001.1	*Pseudomonas_aeruginosa*	FALSE
840	NZ_RCIL01000065.1	*Proteus_mirabilis*	FALSE	**890**	NZ_LLME01000001.1	*Pseudomonas_aeruginosa*	FALSE
841	NZ_RCIN01000062.1	*Proteus_mirabilis*	FALSE	**891**	NZ_LLOZ01000001.1	*Pseudomonas_aeruginosa*	FALSE
842	NZ_RCIM01000060.1	*Proteus_mirabilis*	FALSE	**892**	NZ_LLRL01000001.1	*Pseudomonas_aeruginosa*	FALSE
843	NZ_JAAATT010000001.1	*Proteus_mirabilis*	FALSE	**893**	NZ_LLRQ01000001.1	*Pseudomonas_aeruginosa*	FALSE
844	NZ_VKSC01000001.1	*Proteus_mirabilis*	FALSE	**894**	NZ_LLVE01000001.1	*Pseudomonas_aeruginosa*	FALSE
845	NZ_VKTS01000001.1	*Proteus_mirabilis*	FALSE	**895**	NZ_CP008857.1	*Pseudomonas_aeruginosa*	TRUE
846	NZ_VKWW01000001.1	*Proteus_mirabilis*	FALSE	**896**	NZ_LRSJ01000001.1	*Pseudomonas_aeruginosa*	FALSE
847	NZ_VKXD01000001.1	*Proteus_mirabilis*	FALSE	**897**	NZ_LSAE01000001.1	*Pseudomonas_aeruginosa*	FALSE
848	NZ_VKXK01000001.1	*Proteus_mirabilis*	FALSE	**898**	NZ_MAUQ01000231.1	*Pseudomonas_aeruginosa*	FALSE
849	NZ_CP047352.1	*Proteus_mirabilis*	TRUE	**899**	NZ_MVBS01000001.1	*Pseudomonas_aeruginosa*	FALSE
850	NZ_JAAMPE010000023.1	*Proteus_mirabilis*	FALSE	**900**	NZ_NFGX01000100.1	*Pseudomonas_aeruginosa*	FALSE

Supplementary Table 9. Continued.

901	NZ_MYFK01000046.1	*Pseudomonas_aeruginosa*	FALSE	**951**	NZ_JABGOB010000001.	*Pseudomonas_aeruginosa*	FALSE
902	NZ_MWXQ01000001.1	*Pseudomonas_aeruginosa*	FALSE	**952**	NZ_CP058332.1	*Pseudomonas_aeruginosa*	TRUE
903	NZ_MWYR01000001.1	*Pseudomonas_aeruginosa*	FALSE	**953**	NZ_JACEJW010000001.1	*Pseudomonas_aeruginosa*	FALSE
904	NZ_NRAX01000100.1	*Pseudomonas_aeruginosa*	FALSE	**954**	NZ_CP059063.1	*Pseudomonas_aeruginosa*	TRUE
905	NZ_NRAY01000010.1	*Pseudomonas_aeruginosa*	FALSE	**955**	NZ_JACWIG010000010.1	*Pseudomonas_aeruginosa*	FALSE
906	NZ_PKIS01000001.1	*Pseudomonas_aeruginosa*	FALSE	**956**	NZ_WODL01000001.1	*Pseudomonas_aeruginosa*	FALSE
907	NZ_QVBV02000100.1	*Pseudomonas_aeruginosa*	FALSE	**957**	NZ_JADPHX010000100.1	*Pseudomonas_aeruginosa*	FALSE
908	NZ_QVCZ02000100.1	*Pseudomonas_aeruginosa*	FALSE	**958**	NZ_FRGG01000058.1	*Pseudomonas_aeruginosa*	FALSE
909	NZ_QVBA02000099.1	*Pseudomonas_aeruginosa*	FALSE	**959**	NZ_FRHM01000090.1	*Pseudomonas_aeruginosa*	FALSE
910	NZ_RADB01000239.1	*Pseudomonas_aeruginosa*	FALSE	**960**	NZ_FRIH01000087.1	*Pseudomonas_aeruginosa*	FALSE
911	NZ_RACT01000020.1	*Pseudomonas_aeruginosa*	FALSE	**961**	NZ_FRIN01000188.1	*Pseudomonas_aeruginosa*	FALSE
912	NZ_RACH01000323.1	*Pseudomonas_aeruginosa*	FALSE	**962**	NZ_FRIS01000125.1	*Pseudomonas_aeruginosa*	FALSE
913	NZ_RABL01000011.1	*Pseudomonas_aeruginosa*	FALSE	**963**	NZ_FRIY01000227.1	*Pseudomonas_aeruginosa*	FALSE
914	NZ_QZZG01000344.1	*Pseudomonas_aeruginosa*	FALSE	**964**	NZ_FRNJ01000118.1	*Pseudomonas_aeruginosa*	FALSE
915	NZ_MCLY01000001.1	*Pseudomonas_aeruginosa*	FALSE	**965**	NZ_FRQU01000076.1	*Pseudomonas_aeruginosa*	FALSE
916	NZ_MCMV01000001.1	*Pseudomonas_aeruginosa*	FALSE	**966**	NZ_FRTT01000184.1	*Pseudomonas_aeruginosa*	FALSE
917	NZ_MCMQ01000001.1	*Pseudomonas_aeruginosa*	FALSE	**967**	NZ_FRUE01000212.1	*Pseudomonas_aeruginosa*	FALSE
918	NZ_NTGC01000004.1	*Pseudomonas_aeruginosa*	FALSE	**968**	NZ_FRVQ01000166.1	*Pseudomonas_aeruginosa*	FALSE
919	NZ_NSMX01000001.1	*Pseudomonas_aeruginosa*	FALSE	**969**	NZ_GG730158.1	*Staphylococcus_aureus*	FALSE
920	NZ_NSOA01000001.1	*Pseudomonas_aeruginosa*	FALSE	**970**	NZ_AHLO01000079.1	*Staphylococcus_aureus*	FALSE
921	NZ_NSRJ01000001.1	*Pseudomonas_aeruginosa*	FALSE	**971**	NZ_KB820894.1	*Staphylococcus_aureus*	FALSE
922	NZ_NSUB01000001.1	*Pseudomonas_aeruginosa*	FALSE	**972**	NZ_KI966573.1	*Staphylococcus_aureus*	FALSE
923	NZ_NSUG01000001.1	*Pseudomonas_aeruginosa*	FALSE	**973**	NZ_KI967784.1	*Staphylococcus_aureus*	FALSE
924	NZ_NSVM01000001.1	*Pseudomonas_aeruginosa*	FALSE	**974**	NZ_KI971688.1	*Staphylococcus_aureus*	FALSE
925	NZ_NSZO01000001.1	*Pseudomonas_aeruginosa*	FALSE	**975**	NZ_KI972776.1	*Staphylococcus_aureus*	FALSE
926	NZ_NSVR01000001.1	*Pseudomonas_aeruginosa*	FALSE	**976**	NZ_JBTX01000001.1	*Staphylococcus_aureus*	FALSE
927	NZ_RRBP01000010.1	*Pseudomonas_aeruginosa*	FALSE	**977**	NZ_JDCI01000001.1	*Staphylococcus_aureus*	FALSE
928	NZ_RRBL01000100.1	*Pseudomonas_aeruginosa*	FALSE	**978**	NZ_JDFC01000001.1	*Staphylococcus_aureus*	FALSE
929	NZ_RHSX01000010.1	*Pseudomonas_aeruginosa*	FALSE	**979**	NZ_KI976096.1	*Staphylococcus_aureus*	FALSE
930	NZ_RWWW01000040.1	*Pseudomonas_aeruginosa*	FALSE	**980**	NZ_KI976941.1	*Staphylococcus_aureus*	FALSE
931	NZ_RWWQ01000069.1	*Pseudomonas_aeruginosa*	FALSE	**981**	NZ_KI977779.1	*Staphylococcus_aureus*	FALSE
932	NZ_RXUI01000001.1	*Pseudomonas_aeruginosa*	FALSE	**982**	NZ_KI997674.1	*Staphylococcus_aureus*	FALSE
933	NZ_RXWC01000001.1	*Pseudomonas_aeruginosa*	FALSE	**983**	NZ_KK006024.1	*Staphylococcus_aureus*	FALSE
934	NZ_RWZY01000001.1	*Pseudomonas_aeruginosa*	FALSE	**984**	NZ_KK006722.1	*Staphylococcus_aureus*	FALSE
935	NZ_RWZZ01000001.1	*Pseudomonas_aeruginosa*	FALSE	**985**	NZ_KK015495.1	*Staphylococcus_aureus*	FALSE
936	NZ_RXAK01000001.1	*Pseudomonas_aeruginosa*	FALSE	**986**	NZ_KK017352.1	*Staphylococcus_aureus*	FALSE
937	NZ_RXCY01000001.1	*Pseudomonas_aeruginosa*	FALSE	**987**	NZ_KK023230.1	*Staphylococcus_aureus*	FALSE
938	NZ_SDVH01000010.1	*Pseudomonas_aeruginosa*	FALSE	**988**	NZ_KK023840.1	*Staphylococcus_aureus*	FALSE
939	NZ_SCYC01000001.1	*Pseudomonas_aeruginosa*	FALSE	**989**	NZ_KK031896.1	*Staphylococcus_aureus*	FALSE
940	NZ_SCZX01000001.1	*Pseudomonas_aeruginosa*	FALSE	**990**	NZ_KK047855.1	*Staphylococcus_aureus*	FALSE
941	NZ_RWPI01000001.1	*Pseudomonas_aeruginosa*	FALSE	**991**	NZ_KK068644.1	*Staphylococcus_aureus*	FALSE
942	NZ_RWQG01000001.1	*Pseudomonas_aeruginosa*	FALSE	**992**	NZ_KK070455.1	*Staphylococcus_aureus*	FALSE
943	NZ_RWTQ01000001.1	*Pseudomonas_aeruginosa*	FALSE	**993**	NZ_CAVU010000001.1	*Staphylococcus_aureus*	FALSE
944	NZ_RWVK01000001.1	*Pseudomonas_aeruginosa*	FALSE	**994**	NZ_KK074597.1	*Staphylococcus_aureus*	FALSE
945	NZ_RWVJ01000001.1	*Pseudomonas_aeruginosa*	FALSE	**995**	NZ_JDGT01000001.1	*Staphylococcus_aureus*	FALSE
946	NZ_VAUY01000001.1	*Pseudomonas_aeruginosa*	FALSE	**996**	NZ_JFPA01000001.1	*Staphylococcus_aureus*	FALSE
947	NZ_VIHG01000100.1	*Pseudomonas_aeruginosa*	FALSE	**997**	NZ_JFPV01000001.1	*Staphylococcus_aureus*	FALSE
948	NZ_WHVM01000010.1	*Pseudomonas_aeruginosa*	FALSE	**998**	NZ_KK095289.1	*Staphylococcus_aureus*	FALSE
949	NZ_WUDL01000100.1	*Pseudomonas_aeruginosa*	FALSE	**999**	NZ_KK096012.1	*Staphylococcus_aureus*	FALSE
950	NZ_JAACJB010000010.1	*Pseudomonas_aeruginosa*	FALSE	**1000**	NZ_KK097143.1	*Staphylococcus_aureus*	FALSE

Supplementary Table 9. Continued.

1001	NZ_KK098246.1	*Staphylococcus_aureus*	FALSE	**1051**	NZ_FFQE01000001.1	*Staphylococcus_aureus*	FALSE
1002	NZ_KK099682.1	*Staphylococcus_aureus*	FALSE	**1052**	NZ_FGAM01000001.1	*Staphylococcus_aureus*	FALSE
1003	NZ_KK099778.1	*Staphylococcus_aureus*	FALSE	**1053**	NZ_FGAV01000001.1	*Staphylococcus_aureus*	FALSE
1004	NZ_KK233903.1	*Staphylococcus_aureus*	FALSE	**1054**	NZ_FGND01000001.1	*Staphylococcus_aureus*	FALSE
1005	NZ_KK235057.1	*Staphylococcus_aureus*	FALSE	**1055**	NZ_FFPE01000001.1	*Staphylococcus_aureus*	FALSE
1006	NZ_JINN01000001.1	*Staphylococcus_aureus*	FALSE	**1056**	NZ_FGSR01000001.1	*Staphylococcus_aureus*	FALSE
1007	NZ_CSHX01000001.1	*Staphylococcus_aureus*	FALSE	**1057**	NZ_FGSX01000001.1	*Staphylococcus_aureus*	FALSE
1008	NZ_CTTO01000001.1	*Staphylococcus_aureus*	FALSE	**1058**	NZ_FHDC01000001.1	*Staphylococcus_aureus*	FALSE
1009	NZ_CTVB01000001.1	*Staphylococcus_aureus*	FALSE	**1059**	NZ_FHET01000001.1	*Staphylococcus_aureus*	FALSE
1010	NZ_CUBQ01000001.1	*Staphylococcus_aureus*	FALSE	**1060**	NZ_FHCJ01000001.1	*Staphylococcus_aureus*	FALSE
1011	NZ_CTYX01000001.1	*Staphylococcus_aureus*	FALSE	**1061**	NZ_FHEK01000001.1	*Staphylococcus_aureus*	FALSE
1012	NZ_CUBM01000001.1	*Staphylococcus_aureus*	FALSE	**1062**	NZ_FKMZ01000001.1	*Staphylococcus_aureus*	FALSE
1013	NZ_CTWW01000001.1	*Staphylococcus_aureus*	FALSE	**1063**	NZ_FKTC01000001.1	*Staphylococcus_aureus*	FALSE
1014	NZ_CTUT01000001.1	*Staphylococcus_aureus*	FALSE	**1064**	NZ_FKWV01000032.1	*Staphylococcus_aureus*	FALSE
1015	NZ_LFOH01000001.1	*Staphylococcus_aureus*	FALSE	**1065**	NZ_FKWX01000001.1	*Staphylococcus_aureus*	FALSE
1016	NZ_CVHN01000001.1	*Staphylococcus_aureus*	FALSE	**1066**	NZ_FQFU01000034.1	*Staphylococcus_aureus*	FALSE
1017	NZ_CFQI01000001.1	*Staphylococcus_aureus*	FALSE	**1067**	NZ_FQGH01000015.1	*Staphylococcus_aureus*	FALSE
1018	NZ_MAQR01000001.1	*Staphylococcus_aureus*	FALSE	**1068**	NZ_FQGX01000017.1	*Staphylococcus_aureus*	FALSE
1019	NZ_NDKM01000055.1	*Staphylococcus_aureus*	FALSE	**1069**	NZ_AFEH01000061.1	*Staphylococcus_epidermidis*	FALSE
1020	NZ_NDOH01000072.1	*Staphylococcus_aureus*	FALSE	**1070**	NZ_AKHJ01000001.1	*Staphylococcus_epidermidis*	FALSE
1021	NZ_NDQM01000009.1	*Staphylococcus_aureus*	FALSE	**1071**	NZ_AKHG01000001.1	*Staphylococcus_epidermidis*	FALSE
1022	NZ_NDPZ01000071.1	*Staphylococcus_aureus*	FALSE	**1072**	NZ_AKHE01000001.1	*Staphylococcus_epidermidis*	FALSE
1023	NZ_NMWL01000010.1	*Staphylococcus_aureus*	FALSE	**1073**	NZ_AKGW01000001.1	*Staphylococcus_epidermidis*	FALSE
1024	NZ_LGWE01000001.1	*Staphylococcus_aureus*	FALSE	**1074**	NZ_AMSJ01000001.1	*Staphylococcus_epidermidis*	FALSE
1025	NZ_LKVM01000057.1	*Staphylococcus_aureus*	FALSE	**1075**	NZ_ATCV02000263.1	*Staphylococcus_epidermidis*	FALSE
1026	NZ_LKVY01000139.1	*Staphylococcus_aureus*	FALSE	**1076**	NZ_ATCW02000148.1	*Staphylococcus_epidermidis*	FALSE
1027	NZ_PUUB01000010.1	*Staphylococcus_aureus*	FALSE	**1077**	NZ_ATDG02000114.1	*Staphylococcus_epidermidis*	FALSE
1028	NZ_PXUV01000010.1	*Staphylococcus_aureus*	FALSE	**1078**	NZ_ATCX02000246.1	*Staphylococcus_epidermidis*	FALSE
1029	NZ_BEYH01000001.1	*Staphylococcus_aureus*	FALSE	**1079**	NZ_JMIF01000001.1	*Staphylococcus_epidermidis*	FALSE
1030	NZ_BDXA01000001.1	*Staphylococcus_aureus*	FALSE	**1080**	NZ_HG813242.1	*Staphylococcus_epidermidis*	TRUE
1031	NZ_BEAZ01000001.1	*Staphylococcus_aureus*	FALSE	**1081**	NZ_CP009046.1	*Staphylococcus_epidermidis*	TRUE
1032	NZ_BHAQ01000001.1	*Staphylococcus_aureus*	FALSE	**1082**	NZ_JQHC01000001.1	*Staphylococcus_epidermidis*	FALSE
1033	NZ_BHBX01000001.1	*Staphylococcus_aureus*	FALSE	**1083**	NZ_JTAY02000001.1	*Staphylococcus_epidermidis*	FALSE
1034	NZ_RIXT01000010.1	*Staphylococcus_aureus*	FALSE	**1084**	NZ_JWFW01000009.1	*Staphylococcus_epidermidis*	FALSE
1035	NZ_RHFU01000001.1	*Staphylococcus_aureus*	FALSE	**1085**	NZ_JVVI01000017.1	*Staphylococcus_epidermidis*	FALSE
1036	NZ_SRPB01000100.1	*Staphylococcus_aureus*	FALSE	**1086**	NZ_JWCD01000015.1	*Staphylococcus_epidermidis*	FALSE
1037	NZ_VRQX01000010.1	*Staphylococcus_aureus*	FALSE	**1087**	NZ_JWBO01000020.1	*Staphylococcus_epidermidis*	FALSE
1038	NZ_AP019543.1	*Staphylococcus_aureus*	TRUE	**1088**	NZ_JVJD01000001.1	*Staphylococcus_epidermidis*	FALSE
1039	NZ_WKGQ01000010.1	*Staphylococcus_aureus*	FALSE	**1089**	NZ_JVEN01000004.1	*Staphylococcus_epidermidis*	FALSE
1040	NZ_WKIB01000100.1	*Staphylococcus_aureus*	FALSE	**1090**	NZ_JVAW01000004.1	*Staphylococcus_epidermidis*	FALSE
1041	NZ_WKGM01000093.1	*Staphylococcus_aureus*	FALSE	**1091**	NZ_JVAT01000132.1	*Staphylococcus_epidermidis*	FALSE
1042	NZ_JAANTX010000001.1	*Staphylococcus_aureus*	FALSE	**1092**	NZ_JVGH01000002.1	*Staphylococcus_epidermidis*	FALSE
1043	NZ_JAANBW010000010.1	*Staphylococcus_aureus*	FALSE	**1093**	NZ_JUXD01000016.1	*Staphylococcus_epidermidis*	FALSE
1044	NZ_WJBT01000010.1	*Staphylococcus_aureus*	FALSE	**1094**	NZ_JUWT01000011.1	*Staphylococcus_epidermidis*	FALSE
1045	NZ_JACEHP010000001.1	*Staphylococcus_aureus*	FALSE	**1095**	NZ_JUWI01000016.1	*Staphylococcus_epidermidis*	FALSE
1046	NZ_CP060608.1	*Staphylococcus_aureus*	TRUE	**1096**	NZ_JUVK01000006.1	*Staphylococcus_epidermidis*	FALSE
1047	NZ_JACJAL010000010.1	*Staphylococcus_aureus*	FALSE	**1097**	NZ_JUPY01000021.1	*Staphylococcus_epidermidis*	FALSE
1048	NZ_JACJAC010000010.1	*Staphylococcus_aureus*	FALSE	**1098**	NZ_JUKL01000014.1	*Staphylococcus_epidermidis*	FALSE
1049	NZ_FFXV01000001.1	*Staphylococcus_aureus*	FALSE	**1099**	NZ_BAZV01000001.1	*Staphylococcus_epidermidis*	FALSE
1050	NZ_FGHO01000001.1	*Staphylococcus_aureus*	FALSE	**1100**	NZ_LILE01000001.1	*Staphylococcus_epidermidis*	FALSE

Supplementary Table 9. Continued.

1101	NZ_LJIF01000001.1	*Staphylococcus_epidermidis*	FALSE	**1151**	NZ_WBRP01000010.1	*Staphylococcus_epidermidis*	FALSE
1102	NZ_LOAT01000001.1	*Staphylococcus_epidermidis*	FALSE	**1152**	NZ_WBRR01000010.1	*Staphylococcus_epidermidis*	FALSE
1103	NZ_LDTJ01000001.1	*Staphylococcus_epidermidis*	FALSE	**1153**	NZ_WBRT01000010.1	*Staphylococcus_epidermidis*	FALSE
1104	NZ_LWAU01000001.1	*Staphylococcus_epidermidis*	FALSE	**1154**	NZ_SHEH01000010.1	*Staphylococcus_epidermidis*	FALSE
1105	NZ_LWAV01000001.1	*Staphylococcus_epidermidis*	FALSE	**1155**	NZ_SHEU01000010.1	*Staphylococcus_epidermidis*	FALSE
1106	NZ_LWBB01000001.1	*Staphylococcus_epidermidis*	FALSE	**1156**	NZ_SHEV01000010.1	*Staphylococcus_epidermidis*	FALSE
1107	NZ_LWBF01000001.1	*Staphylococcus_epidermidis*	FALSE	**1157**	NZ_SHFF01000010.1	*Staphylococcus_epidermidis*	FALSE
1108	NZ_LYQA01000001.1	*Staphylococcus_epidermidis*	FALSE	**1158**	NZ_SHFQ01000010.1	*Staphylococcus_epidermidis*	FALSE
1109	NZ_LXRZ01000001.1	*Staphylococcus_epidermidis*	FALSE	**1159**	NZ_SHFS01000010.1	*Staphylococcus_epidermidis*	FALSE
1110	NZ_LYSC01000001.1	*Staphylococcus_epidermidis*	FALSE	**1160**	NZ_SHGA01000010.1	*Staphylococcus_epidermidis*	FALSE
1111	NZ_LYVT01000001.1	*Staphylococcus_epidermidis*	FALSE	**1161**	NZ_SHGE01000010.1	*Staphylococcus_epidermidis*	FALSE
1112	NZ_LYWH01000001.1	*Staphylococcus_epidermidis*	FALSE	**1162**	NZ_SHGK01000010.1	*Staphylococcus_epidermidis*	FALSE
1113	NZ_LOHL01000001.1	*Staphylococcus_epidermidis*	FALSE	**1163**	NZ_CP043784.1	*Staphylococcus_epidermidis*	TRUE
1114	NZ_CP018842.1	*Staphylococcus_epidermidis*	TRUE	**1164**	NZ_CP043804.1	*Staphylococcus_epidermidis*	TRUE
1115	NZ_MVFV01000010.1	*Staphylococcus_epidermidis*	FALSE	**1165**	NZ_JACMXP010000010.1	*Staphylococcus_epidermidis*	FALSE
1116	NZ_NPDA01000001.1	*Staphylococcus_epidermidis*	FALSE	**1166**	NZ_JACMXG010000010.	*Staphylococcus_epidermidis*	FALSE
1117	NZ_PEIQ01000001.1	*Staphylococcus_epidermidis*	FALSE	**1167**	NZ_JACMVB010000010.	*Staphylococcus_epidermidis*	FALSE
1118	NZ_CP013943.1	*Staphylococcus_epidermidis*	TRUE	**1168**	NZ_JACMVG010000010.	*Staphylococcus_epidermidis*	FALSE
1119	NZ_PPQG01000001.1	*Staphylococcus_epidermidis*	FALSE	**1169**	NZ_AJVA01000001.1	*Staphylococcus_haemolyticus*	FALSE
1120	NZ_QXSO01000001.1	*Staphylococcus_epidermidis*	FALSE	**1170**	NZ_JQHA01000001.1	*Staphylococcus_haemolyticus*	FALSE
1121	NZ_QXSQ01000001.1	*Staphylococcus_epidermidis*	FALSE	**1171**	NZ_JPRW01000001.1	*Staphylococcus_haemolyticus*	FALSE
1122	NZ_RDEK01000001.1	*Staphylococcus_epidermidis*	FALSE	**1172**	NZ_CVRV01000001.1	*Staphylococcus_haemolyticus*	FALSE
1123	NZ_MLXH01000001.1	*Staphylococcus_epidermidis*	FALSE	**1173**	NZ_JVMX01000026.1	*Staphylococcus_haemolyticus*	FALSE
1124	NZ_LGVL01000001.1	*Staphylococcus_epidermidis*	FALSE	**1174**	NZ_CUES01000001.1	*Staphylococcus_haemolyticus*	FALSE
1125	NZ_LGVK01000001.1	*Staphylococcus_epidermidis*	FALSE	**1175**	NZ_CUEL01000001.1	*Staphylococcus_haemolyticus*	FALSE
1126	NZ_LUBO01000001.1	*Staphylococcus_epidermidis*	FALSE	**1176**	NZ_CUEN01000001.1	*Staphylococcus_haemolyticus*	FALSE
1127	NZ_LUBT01000001.1	*Staphylococcus_epidermidis*	FALSE	**1177**	NZ_CUFL01000001.1	*Staphylococcus_haemolyticus*	FALSE
1128	NZ_SNNA01000001.1	*Staphylococcus_epidermidis*	FALSE	**1178**	NZ_CUDG01000001.1	*Staphylococcus_haemolyticus*	FALSE
1129	NZ_SNNN01000001.1	*Staphylococcus_epidermidis*	FALSE	**1179**	NZ_CUGL01000001.1	*Staphylococcus_haemolyticus*	FALSE
1130	NZ_SNNB01000001.1	*Staphylococcus_epidermidis*	FALSE	**1180**	NZ_CUHE01000001.1	*Staphylococcus_haemolyticus*	FALSE
1131	NZ_SNMZ01000001.1	*Staphylococcus_epidermidis*	FALSE	**1181**	NZ_CUDS01000001.1	*Staphylococcus_haemolyticus*	FALSE
1132	NZ_VDOM01000100.1	*Staphylococcus_epidermidis*	FALSE	**1182**	NZ_CUCS01000001.1	*Staphylococcus_haemolyticus*	FALSE
1133	NZ_VDPD01000010.1	*Staphylococcus_epidermidis*	FALSE	**1183**	NZ_CUDX01000001.1	*Staphylococcus_haemolyticus*	FALSE
1134	NZ_VDPK01000010.1	*Staphylococcus_epidermidis*	FALSE	**1184**	NZ_CUDF01000001.1	*Staphylococcus_haemolyticus*	FALSE
1135	NZ_VDPY01000010.1	*Staphylococcus_epidermidis*	FALSE	**1185**	NZ_CUCP01000001.1	*Staphylococcus_haemolyticus*	FALSE
1136	NZ_VDQE01000010.1	*Staphylococcus_epidermidis*	FALSE	**1186**	NZ_CUGD01000001.1	*Staphylococcus_haemolyticus*	FALSE
1137	NZ_VDRX01000010.1	*Staphylococcus_epidermidis*	FALSE	**1187**	NZ_CUEB01000001.1	*Staphylococcus_haemolyticus*	FALSE
1138	NZ_VDZZ01001000.1	*Staphylococcus_epidermidis*	FALSE	**1188**	NZ_CUDD01000001.1	*Staphylococcus_haemolyticus*	FALSE
1139	NZ_VECF01000100.1	*Staphylococcus_epidermidis*	FALSE	**1189**	NZ_CUFO01000001.1	*Staphylococcus_haemolyticus*	FALSE
1140	NZ_VECZ01000010.1	*Staphylococcus_epidermidis*	FALSE	**1190**	NZ_CUFC01000001.1	*Staphylococcus_haemolyticus*	FALSE
1141	NZ_VEEK01000010.1	*Staphylococcus_epidermidis*	FALSE	**1191**	NZ_CUEK01000001.1	*Staphylococcus_haemolyticus*	FALSE
1142	NZ_VEEO01000010.1	*Staphylococcus_epidermidis*	FALSE	**1192**	NZ_CUCR01000001.1	*Staphylococcus_haemolyticus*	FALSE
1143	NZ_VTQR01000100.1	*Staphylococcus_epidermidis*	FALSE	**1193**	NZ_CUDQ01000001.1	*Staphylococcus_haemolyticus*	FALSE
1144	NZ_WBME01000001.1	*Staphylococcus_epidermidis*	FALSE	**1194**	NZ_CUEX01000001.1	*Staphylococcus_haemolyticus*	FALSE
1145	NZ_WBPT01000100.1	*Staphylococcus_epidermidis*	FALSE	**1195**	NZ_CUDU01000001.1	*Staphylococcus_haemolyticus*	FALSE
1146	NZ_WBPW01000010.1	*Staphylococcus_epidermidis*	FALSE	**1196**	NZ_CUEF01000001.1	*Staphylococcus_haemolyticus*	FALSE
1147	NZ_WBQD01000010.1	*Staphylococcus_epidermidis*	FALSE	**1197**	NZ_CUGP01000001.1	*Staphylococcus_haemolyticus*	FALSE
1148	NZ_WBQV01000010.1	*Staphylococcus_epidermidis*	FALSE	**1198**	NZ_CUFX01000001.1	*Staphylococcus_haemolyticus*	FALSE
1149	NZ_WBRM01000010.1	*Staphylococcus_epidermidis*	FALSE	**1199**	NZ_CUFF01000001.1	*Staphylococcus_haemolyticus*	FALSE
1150	NZ_WBRN01000010.1	*Staphylococcus_epidermidis*	FALSE	**1200**	NZ_CUCK01000001.1	*Staphylococcus_haemolyticus*	FALSE

Supplementary Table 9. Continued.

1201	NZ_CUDR01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1251**	NZ_VEGW01000010.1	*Staphylococcus_haemolyticus*	FALSE
1202	NZ_CUHL01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1252**	NZ_WSYI01000010.1	*Staphylococcus_haemolyticus*	FALSE
1203	NZ_CUCN01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1253**	NZ_JRVS01000001.1	*Staphylococcus_haemolyticus*	FALSE
1204	NZ_CUDI01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1254**	NZ_JRVT01000001.1	*Staphylococcus_haemolyticus*	FALSE
1205	NZ_CUEI01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1255**	NZ_JACMWI010000010.	*Staphylococcus_haemolyticus*	FALSE
1206	NZ_CUHM01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1256**	NZ_JACMUM010000010	*Staphylococcus_haemolyticus*	FALSE
1207	NZ_CUGO01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1257**	NZ_JACMUN010000010.	*Staphylococcus_haemolyticus*	FALSE
1208	NZ_CUCT01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1258**	NZ_JADCSO010000001.1	*Staphylococcus_haemolyticus*	FALSE
1209	NZ_CUCL01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1259**	NZ_JADCSF010000001.1	*Staphylococcus_haemolyticus*	FALSE
1210	NZ_CUEO01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1260**	NZ_JACMRU010000001.	*Staphylococcus_haemolyticus*	FALSE
1211	NZ_LRHN01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1261**	NZ_JACGQO010000001.	*Staphylococcus_haemolyticus*	FALSE
1212	NZ_CP013911.1	*Staphylococcus_haemolyticus*	TRUE	**1262**	NZ_JACGQN010000001.	*Staphylococcus_haemolyticus*	FALSE
1213	NZ_JRAY01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1263**	NZ_JADICH010000001.1	*Staphylococcus_haemolyticus*	FALSE
1214	NZ_JRAZ01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1264**	NZ_JADICI010000001.1	*Staphylococcus_haemolyticus*	FALSE
1215	NZ_MIZW01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1265**	NZ_LT963441.1	*Staphylococcus_haemolyticus*	TRUE
1216	NZ_MRUZ01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1266**	NZ_CAJCGB010000001.1	*Staphylococcus_haemolyticus*	FALSE
1217	NZ_MRUY01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1267**	NZ_CAJCGF010000001.1	*Staphylococcus_haemolyticus*	FALSE
1218	NZ_NDWY01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1268**	NZ_CAJCGE010000001.1	*Staphylococcus_haemolyticus*	FALSE
1219	NZ_NCXH01000099.1	*Staphylococcus_haemolyticus*	FALSE	**1269**	NC_013853.1	*Streptococcus_mitis*	TRUE
1220	NZ_PPQA01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1270**	NZ_GL397179.1	*Streptococcus_mitis*	FALSE
1221	NZ_CP025396.1	*Staphylococcus_haemolyticus*	TRUE	**1271**	NZ_AEDT01000040.1	*Streptococcus_mitis*	FALSE
1222	NZ_PZIU01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1272**	NZ_AEDU01000035.1	*Streptococcus_mitis*	FALSE
1223	NZ_PZIV01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1273**	NZ_AEDV01000108.1	*Streptococcus_mitis*	FALSE
1224	NZ_PZIJ01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1274**	NZ_CP028414.1	*Streptococcus_mitis*	TRUE
1225	NZ_PZIH01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1275**	NZ_AFQT01000054.1	*Streptococcus_mitis*	FALSE
1226	NZ_PZIM01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1276**	NZ_AFQV01000034.1	*Streptococcus_mitis*	FALSE
1227	NZ_PZIT01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1277**	NZ_AFUO01000002.1	*Streptococcus_mitis*	FALSE
1228	NZ_PZIX01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1278**	NZ_AFUB01000066.1	*Streptococcus_mitis*	FALSE
1229	NZ_PZIY01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1279**	NZ_AFUF01000054.1	*Streptococcus_mitis*	FALSE
1230	NZ_QVPX01000092.1	*Staphylococcus_haemolyticus*	FALSE	**1280**	NZ_AICR01000001.1	*Streptococcus_mitis*	FALSE
1231	NZ_QVPW01000088.1	*Staphylococcus_haemolyticus*	FALSE	**1281**	NZ_AJJL01000001.1	*Streptococcus_mitis*	FALSE
1232	NZ_QYJJ01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1282**	NZ_AQTT01000001.1	*Streptococcus_mitis*	FALSE
1233	NZ_QXVT01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1283**	NZ_ATAA01000001.1	*Streptococcus_mitis*	FALSE
1234	NZ_QYJK01000001.1	*Staphylococcus_haemolyticus*	FALSE	**1284**	NZ_AYRR01000001.1	*Streptococcus_mitis*	FALSE
1235	NZ_CP033814.1	*Staphylococcus_haemolyticus*	TRUE	**1285**	NZ_AYRS01000001.1	*Streptococcus_mitis*	FALSE
1236	NZ_RXGD01000010.1	*Staphylococcus_haemolyticus*	FALSE	**1286**	NZ_JPFV01000001.1	*Streptococcus_mitis*	FALSE
1237	NZ_SUKI01000213.1	*Staphylococcus_haemolyticus*	FALSE	**1287**	NZ_JPFU01000001.1	*Streptococcus_mitis*	FALSE
1238	NZ_SUKZ01000063.1	*Staphylococcus_haemolyticus*	FALSE	**1288**	NZ_JPFW01000001.1	*Streptococcus_mitis*	FALSE
1239	NZ_CP035291.1	*Staphylococcus_haemolyticus*	TRUE	**1289**	NZ_JPFY01000001.1	*Streptococcus_mitis*	FALSE
1240	NZ_RIYU01000100.1	*Staphylococcus_haemolyticus*	FALSE	**1290**	NZ_CP028415.1	*Streptococcus_mitis*	TRUE
1241	NZ_VJMP01000010.1	*Staphylococcus_haemolyticus*	FALSE	**1291**	NZ_JPFZ01000001.1	*Streptococcus_mitis*	FALSE
1242	NZ_VDNK01000010.1	*Staphylococcus_haemolyticus*	FALSE	**1292**	NZ_JPFT01000001.1	*Streptococcus_mitis*	FALSE
1243	NZ_VDQP01000090.1	*Staphylococcus_haemolyticus*	FALSE	**1293**	NZ_JYGP01000001.1	*Streptococcus_mitis*	FALSE
1244	NZ_VDQU01000100.1	*Staphylococcus_haemolyticus*	FALSE	**1294**	NZ_JYGQ01000001.1	*Streptococcus_mitis*	FALSE
1245	NZ_VDSF01000100.1	*Staphylococcus_haemolyticus*	FALSE	**1295**	NZ_JWFA01000017.1	*Streptococcus_mitis*	FALSE
1246	NZ_VDZP01000100.1	*Staphylococcus_haemolyticus*	FALSE	**1296**	NZ_JWCV01000001.1	*Streptococcus_mitis*	FALSE
1247	NZ_VECW01000100.1	*Staphylococcus_haemolyticus*	FALSE	**1297**	NZ_JVYJ01000028.1	*Streptococcus_mitis*	FALSE
1248	NZ_VEDG01000010.1	*Staphylococcus_haemolyticus*	FALSE	**1298**	NZ_JVRN01000006.1	*Streptococcus_mitis*	FALSE
1249	NZ_VEGS01000100.1	*Staphylococcus_haemolyticus*	FALSE	**1299**	NZ_JUQO01000001.1	*Streptococcus_mitis*	FALSE
1250	NZ_VEHQ01000100.1	*Staphylococcus_haemolyticus*	FALSE	**1300**	NZ_KQ957840.1	*Streptococcus_mitis*	FALSE

Supplementary Table 9. Continued.

1301	NZ_CP014326.1	*Streptococcus_mitis*	TRUE	**1351**	NZ_WIJP01000010.1	*Streptococcus_mitis*	FALSE
1302	NZ_KQ970285.1	*Streptococcus_mitis*	FALSE	**1352**	NZ_WIJG01000100.1	*Streptococcus_mitis*	FALSE
1303	NZ_LROV01000001.1	*Streptococcus_mitis*	FALSE	**1353**	NZ_WIJO01000100.1	*Streptococcus_mitis*	FALSE
1304	NZ_LBMT01000001.1	*Streptococcus_mitis*	FALSE	**1354**	NZ_WIJF01000010.1	*Streptococcus_mitis*	FALSE
1305	NZ_MUYO01000001.1	*Streptococcus_mitis*	FALSE	**1355**	NZ_WIJH01000010.1	*Streptococcus_mitis*	FALSE
1306	NZ_MUXS01000001.1	*Streptococcus_mitis*	FALSE	**1356**	NZ_WIJK01000010.1	*Streptococcus_mitis*	FALSE
1307	NZ_NCVF01000026.1	*Streptococcus_mitis*	FALSE	**1357**	NZ_WIJL01000010.1	*Streptococcus_mitis*	FALSE
1308	NZ_NCVE01000029.1	*Streptococcus_mitis*	FALSE	**1358**	NZ_WIKD01000010.1	*Streptococcus_mitis*	FALSE
1309	NZ_NCVJ01000021.1	*Streptococcus_mitis*	FALSE	**1359**	NZ_WIJY01000010.1	*Streptococcus_mitis*	FALSE
1310	NZ_NCVG01000023.1	*Streptococcus_mitis*	FALSE	**1360**	NZ_WIKC01000010.1	*Streptococcus_mitis*	FALSE
1311	NZ_NCVK01000060.1	*Streptococcus_mitis*	FALSE	**1361**	NZ_WIJZ01000010.1	*Streptococcus_mitis*	FALSE
1312	NZ_NCVH01000033.1	*Streptococcus_mitis*	FALSE	**1362**	NZ_WIJS01000010.1	*Streptococcus_mitis*	FALSE
1313	NZ_NCVI01000020.1	*Streptococcus_mitis*	FALSE	**1363**	NZ_WIJM01000010.1	*Streptococcus_mitis*	FALSE
1314	NZ_NCVL01000056.1	*Streptococcus_mitis*	FALSE	**1364**	NZ_WIJR01000010.1	*Streptococcus_mitis*	FALSE
1315	NZ_NCVM01000026.1	*Streptococcus_mitis*	FALSE	**1365**	NZ_WIIT01000010.1	*Streptococcus_mitis*	FALSE
1316	NZ_PKID01000001.1	*Streptococcus_mitis*	FALSE	**1366**	NZ_WIIV01000010.1	*Streptococcus_mitis*	FALSE
1317	NZ_PKIE01000001.1	*Streptococcus_mitis*	FALSE	**1367**	NZ_WIIY01000010.1	*Streptococcus_mitis*	FALSE
1318	NZ_RJNH01000001.1	*Streptococcus_mitis*	FALSE	**1368**	NZ_WIIW01000010.1	*Streptococcus_mitis*	FALSE
1319	NZ_RJNQ01000001.1	*Streptococcus_mitis*	FALSE	**1369**	NC_011900.1	*Streptococcus_pneumoniae*	TRUE
1320	NZ_RJNR01000001.1	*Streptococcus_mitis*	FALSE	**1370**	NZ_JFJI01000001.1	*Streptococcus_pneumoniae*	FALSE
1321	NZ_RJNU01000001.1	*Streptococcus_mitis*	FALSE	**1371**	NZ_CLSQ01000001.1	*Streptococcus_pneumoniae*	FALSE
1322	NZ_RJNV01000001.1	*Streptococcus_mitis*	FALSE	**1372**	NZ_CKWW01000001.1	*Streptococcus_pneumoniae*	FALSE
1323	NZ_RJNY01000001.1	*Streptococcus_mitis*	FALSE	**1373**	NZ_CLLY01000001.1	*Streptococcus_pneumoniae*	FALSE
1324	NZ_RJNZ01000001.1	*Streptococcus_mitis*	FALSE	**1374**	NZ_CNUQ02000001.1	*Streptococcus_pneumoniae*	FALSE
1325	NZ_RJOC01000001.1	*Streptococcus_mitis*	FALSE	**1375**	NZ_CKCC01000001.1	*Streptococcus_pneumoniae*	FALSE
1326	NZ_RJOD01000001.1	*Streptococcus_mitis*	FALSE	**1376**	NZ_CKKU01000001.1	*Streptococcus_pneumoniae*	FALSE
1327	NZ_RJOE01000001.1	*Streptococcus_mitis*	FALSE	**1377**	NZ_CNXY02000001.1	*Streptococcus_pneumoniae*	FALSE
1328	NZ_RJOG01000001.1	*Streptococcus_mitis*	FALSE	**1378**	NZ_CLOI01000001.1	*Streptococcus_pneumoniae*	FALSE
1329	NZ_RJPW01000001.1	*Streptococcus_mitis*	FALSE	**1379**	NZ_CLTC01000001.1	*Streptococcus_pneumoniae*	FALSE
1330	NZ_RJPY01000001.1	*Streptococcus_mitis*	FALSE	**1380**	NZ_CLOH01000001.1	*Streptococcus_pneumoniae*	FALSE
1331	NZ_RJNS01000001.1	*Streptococcus_mitis*	FALSE	**1381**	NZ_CPLS01000001.1	*Streptococcus_pneumoniae*	FALSE
1332	NZ_RJNT01000001.1	*Streptococcus_mitis*	FALSE	**1382**	NZ_CPMD01000001.1	*Streptococcus_pneumoniae*	FALSE
1333	NZ_RJOB01000001.1	*Streptococcus_mitis*	FALSE	**1383**	NZ_CKRI01000001.1	*Streptococcus_pneumoniae*	FALSE
1334	NZ_RJOF01000001.1	*Streptococcus_mitis*	FALSE	**1384**	NZ_CKUS01000001.1	*Streptococcus_pneumoniae*	FALSE
1335	NZ_RJOA01000001.1	*Streptococcus_mitis*	FALSE	**1385**	NZ_CMWG01000001.1	*Streptococcus_pneumoniae*	FALSE
1336	NZ_RJPX01000001.1	*Streptococcus_mitis*	FALSE	**1386**	NZ_CPPQ01000001.1	*Streptococcus_pneumoniae*	FALSE
1337	NZ_SWFJ01000010.1	*Streptococcus_mitis*	FALSE	**1387**	NZ_CMHF01000001.1	*Streptococcus_pneumoniae*	FALSE
1338	NZ_SPGF01000100.1	*Streptococcus_mitis*	FALSE	**1388**	NZ_CNTS02000001.1	*Streptococcus_pneumoniae*	FALSE
1339	NZ_VTQF01000017.1	*Streptococcus_mitis*	FALSE	**1389**	NZ_CMUL01000001.1	*Streptococcus_pneumoniae*	FALSE
1340	NZ_WIJV01000100.1	*Streptococcus_mitis*	FALSE	**1390**	NZ_CNRC02000001.1	*Streptococcus_pneumoniae*	FALSE
1341	NZ_WIJU01000010.1	*Streptococcus_mitis*	FALSE	**1391**	NZ_CKHM01000001.1	*Streptococcus_pneumoniae*	FALSE
1342	NZ_WIIX01000010.1	*Streptococcus_mitis*	FALSE	**1392**	NZ_CMVM01000001.1	*Streptococcus_pneumoniae*	FALSE
1343	NZ_WIKB01000010.1	*Streptococcus_mitis*	FALSE	**1393**	NZ_COEN02000001.1	*Streptococcus_pneumoniae*	FALSE
1344	NZ_WIJD01000010.1	*Streptococcus_mitis*	FALSE	**1394**	NZ_CNTL02000001.1	*Streptococcus_pneumoniae*	FALSE
1345	NZ_WIJJ01000010.1	*Streptococcus_mitis*	FALSE	**1395**	NZ_CMKF01000001.1	*Streptococcus_pneumoniae*	FALSE
1346	NZ_WIJX01000010.1	*Streptococcus_mitis*	FALSE	**1396**	NZ_CMKI01000001.1	*Streptococcus_pneumoniae*	FALSE
1347	NZ_WIJI01000010.1	*Streptococcus_mitis*	FALSE	**1397**	NZ_CNPU02000001.1	*Streptococcus_pneumoniae*	FALSE
1348	NZ_WIJE01000010.1	*Streptococcus_mitis*	FALSE	**1398**	NZ_CLHS01000001.1	*Streptococcus_pneumoniae*	FALSE
1349	NZ_WIJN01000100.1	*Streptococcus_mitis*	FALSE	**1399**	NZ_CMDL01000001.1	*Streptococcus_pneumoniae*	FALSE
1350	NZ_WIJW01000010.1	*Streptococcus_mitis*	FALSE	**1400**	NZ_CNQU02000001.1	*Streptococcus_pneumoniae*	FALSE

Supplementary Table 9. Continued.

1401	NZ_CNOM02000001.1	*Streptococcus_pneumoniae*	FALSE	**1451**	NZ_FDBE01000001.1	*Streptococcus_pneumoniae*	FALSE
1402	NZ_CMWL01000001.1	*Streptococcus_pneumoniae*	FALSE	**1452**	NZ_FDXW01000001.1	*Streptococcus_pneumoniae*	FALSE
1403	NZ_CKSQ01000001.1	*Streptococcus_pneumoniae*	FALSE	**1453**	NZ_FDYK01000001.1	*Streptococcus_pneumoniae*	FALSE
1404	NZ_CKSS01000001.1	*Streptococcus_pneumoniae*	FALSE	**1454**	NZ_FDKG01000001.1	*Streptococcus_pneumoniae*	FALSE
1405	NZ_CKXK01000001.1	*Streptococcus_pneumoniae*	FALSE	**1455**	NZ_FHRA01000001.1	*Streptococcus_pneumoniae*	FALSE
1406	NZ_CMBZ01000001.1	*Streptococcus_pneumoniae*	FALSE	**1456**	NZ_FIBR01000001.1	*Streptococcus_pneumoniae*	FALSE
1407	NZ_CNOT02000001.1	*Streptococcus_pneumoniae*	FALSE	**1457**	NZ_FHRI01000001.1	*Streptococcus_pneumoniae*	FALSE
1408	NZ_CRQE01000001.1	*Streptococcus_pneumoniae*	FALSE	**1458**	NZ_FIXJ01000001.1	*Streptococcus_pneumoniae*	FALSE
1409	NZ_CRMP01000001.1	*Streptococcus_pneumoniae*	FALSE	**1459**	NZ_FHUI01000001.1	*Streptococcus_pneumoniae*	FALSE
1410	NZ_CRJD01000001.1	*Streptococcus_pneumoniae*	FALSE	**1460**	NZ_FHZH01000001.1	*Streptococcus_pneumoniae*	FALSE
1411	NZ_CRPT01000001.1	*Streptococcus_pneumoniae*	FALSE	**1461**	NZ_FLNB01000054.1	*Streptococcus_pneumoniae*	FALSE
1412	NZ_CRXK01000001.1	*Streptococcus_pneumoniae*	FALSE	**1462**	NZ_FWTI01000252.1	*Streptococcus_pneumoniae*	FALSE
1413	NZ_CHBN02000001.1	*Streptococcus_pneumoniae*	FALSE	**1463**	NZ_FWSR01000314.1	*Streptococcus_pneumoniae*	FALSE
1414	NZ_CFID02000001.1	*Streptococcus_pneumoniae*	FALSE	**1464**	NZ_FWUT01000333.1	*Streptococcus_pneumoniae*	FALSE
1415	NZ_CHXD02000001.1	*Streptococcus_pneumoniae*	FALSE	**1465**	NZ_FWRE01000243.1	*Streptococcus_pneumoniae*	FALSE
1416	NZ_CIKC02000001.1	*Streptococcus_pneumoniae*	FALSE	**1466**	NZ_FWSG01000284.1	*Streptococcus_pneumoniae*	FALSE
1417	NZ_CFKY01000001.1	*Streptococcus_pneumoniae*	FALSE	**1467**	NZ_FYIE01000060.1	*Streptococcus_pneumoniae*	FALSE
1418	NZ_CHYM01000001.1	*Streptococcus_pneumoniae*	FALSE	**1468**	NZ_FYJE01000062.1	*Streptococcus_pneumoniae*	FALSE
1419	NZ_CGFX01000001.1	*Streptococcus_pneumoniae*	FALSE	**1469**	NC_007297.2	*Streptococcus_pyogenes*	TRUE
1420	NZ_CFIY01000001.1	*Streptococcus_pneumoniae*	FALSE	**1470**	NZ_CAHN01000001.1	*Streptococcus_pyogenes*	FALSE
1421	NZ_CHGQ01000001.1	*Streptococcus_pneumoniae*	FALSE	**1471**	NZ_AFRY01000001.1	*Streptococcus_pyogenes*	TRUE
1422	NZ_CHUC01000001.1	*Streptococcus_pneumoniae*	FALSE	**1472**	NZ_ALKE01000001.1	*Streptococcus_pyogenes*	FALSE
1423	NZ_CGXF01000001.1	*Streptococcus_pneumoniae*	FALSE	**1473**	NC_020540.2	*Streptococcus_pyogenes*	TRUE
1424	NZ_CFBG01000001.1	*Streptococcus_pneumoniae*	FALSE	**1474**	NZ_JJKZ01000001.1	*Streptococcus_pyogenes*	FALSE
1425	NZ_CHQM01000001.1	*Streptococcus_pneumoniae*	FALSE	**1475**	NZ_KK213325.1	*Streptococcus_pyogenes*	FALSE
1426	NZ_CFBL01000001.1	*Streptococcus_pneumoniae*	FALSE	**1476**	NZ_KK213406.1	*Streptococcus_pyogenes*	FALSE
1427	NZ_CHWE01000001.1	*Streptococcus_pneumoniae*	FALSE	**1477**	NZ_KK213475.1	*Streptococcus_pyogenes*	FALSE
1428	NZ_LN831051.1	*Streptococcus_pneumoniae*	TRUE	**1478**	NZ_KK213617.1	*Streptococcus_pyogenes*	FALSE
1429	NZ_CHBJ02000001.1	*Streptococcus_pneumoniae*	FALSE	**1479**	NZ_KK213715.1	*Streptococcus_pyogenes*	FALSE
1430	NZ_LJLV01000001.1	*Streptococcus_pneumoniae*	FALSE	**1480**	NZ_KK213735.1	*Streptococcus_pyogenes*	FALSE
1431	NZ_LWKY01000001.1	*Streptococcus_pneumoniae*	FALSE	**1481**	NZ_KK213853.1	*Streptococcus_pyogenes*	FALSE
1432	NZ_MDYD01000001.1	*Streptococcus_pneumoniae*	FALSE	**1482**	NZ_JRLK01000001.1	*Streptococcus_pyogenes*	FALSE
1433	NZ_LXPM01000001.1	*Streptococcus_pneumoniae*	FALSE	**1483**	NZ_CP007561.1	*Streptococcus_pyogenes*	TRUE
1434	NZ_NOIL01000001.1	*Streptococcus_pneumoniae*	FALSE	**1484**	NZ_AP012335.1	*Streptococcus_pyogenes*	TRUE
1435	NZ_LTCL01000001.1	*Streptococcus_pneumoniae*	FALSE	**1485**	NZ_LRHJ01000001.1	*Streptococcus_pyogenes*	FALSE
1436	NZ_CQYV01000001.1	*Streptococcus_pneumoniae*	FALSE	**1486**	NZ_NCTM01000001.1	*Streptococcus_pyogenes*	FALSE
1437	NZ_JAAAQS010000100.1	*Streptococcus_pneumoniae*	FALSE	**1487**	NZ_NGQO01000010.1	*Streptococcus_pyogenes*	FALSE
1438	NZ_FEGQ01000001.1	*Streptococcus_pneumoniae*	FALSE	**1488**	NZ_NIYY01000010.1	*Streptococcus_pyogenes*	FALSE
1439	NZ_FDQP01000001.1	*Streptococcus_pneumoniae*	FALSE	**1489**	NZ_CP035430.1	*Streptococcus_pyogenes*	TRUE
1440	NZ_FEFF01000001.1	*Streptococcus_pneumoniae*	FALSE	**1490**	NZ_SSAK01000001.1	*Streptococcus_pyogenes*	FALSE
1441	NZ_FCSV01000001.1	*Streptococcus_pneumoniae*	FALSE	**1491**	NZ_CP040997.1	*Streptococcus_pyogenes*	TRUE
1442	NZ_FCRV01000001.1	*Streptococcus_pneumoniae*	FALSE	**1492**	NZ_SJLK01000001.1	*Streptococcus_pyogenes*	FALSE
1443	NZ_FDJQ01000001.1	*Streptococcus_pneumoniae*	FALSE	**1493**	NZ_SJLM01000001.1	*Streptococcus_pyogenes*	FALSE
1444	NZ_FEDP01000001.1	*Streptococcus_pneumoniae*	FALSE	**1494**	NZ_SJLG01000001.1	*Streptococcus_pyogenes*	FALSE
1445	NZ_FDCN01000001.1	*Streptococcus_pneumoniae*	FALSE	**1495**	NZ_SJLO01000001.1	*Streptococcus_pyogenes*	FALSE
1446	NZ_FCZA01000001.1	*Streptococcus_pneumoniae*	FALSE	**1496**	NZ_SJLN01000001.1	*Streptococcus_pyogenes*	FALSE
1447	NZ_FDHH01000001.1	*Streptococcus_pneumoniae*	FALSE	**1497**	NZ_WVHA01000010.1	*Streptococcus_pyogenes*	FALSE
1448	NZ_FDXS01000001.1	*Streptococcus_pneumoniae*	FALSE	**1498**	NZ_WVEW01000010.1	*Streptococcus_pyogenes*	FALSE
1449	NZ_FEJR01000001.1	*Streptococcus_pneumoniae*	FALSE	**1499**	NZ_JAAAMY010000001.	*Streptococcus_pyogenes*	FALSE
1450	NZ_FDAC01000001.1	*Streptococcus_pneumoniae*	FALSE	**1500**	NZ_JAAANA010000001.	*Streptococcus_pyogenes*	FALSE

Supplementary Table 9. Continued.

1501	NZ_JAAANC010000001.1	*Streptococcus_pyogenes*	FALSE	**1551**	NZ_CAAILU010000001.1	*Streptococcus_pyogenes*	FALSE
1502	NZ_JAAANK010000001.1	*Streptococcus_pyogenes*	FALSE	**1552**	NZ_CAAILZ010000001.1	*Streptococcus_pyogenes*	FALSE
1503	NZ_JAAACE010000010.1	*Streptococcus_pyogenes*	FALSE	**1553**	NZ_CAAIMH010000001.	*Streptococcus_pyogenes*	FALSE
1504	NZ_JABUAN010000001.1	*Streptococcus_pyogenes*	FALSE	**1554**	NZ_CAAIPD010000001.1	*Streptococcus_pyogenes*	FALSE
1505	NZ_JABUAQ010000001.1	*Streptococcus_pyogenes*	FALSE	**1555**	NZ_CAAIPG010000001.1	*Streptococcus_pyogenes*	FALSE
1506	NZ_JABUAZ010000001.1	*Streptococcus_pyogenes*	FALSE	**1556**	NZ_CAAIQI010000001.1	*Streptococcus_pyogenes*	FALSE
1507	NZ_AP023387.1	*Streptococcus_pyogenes*	TRUE	**1557**	NZ_CAAIQM010000001.	*Streptococcus_pyogenes*	FALSE
1508	NZ_UHHV01000002.1	*Streptococcus_pyogenes*	FALSE	**1558**	NZ_CAAISB010000001.1	*Streptococcus_pyogenes*	FALSE
1509	NZ_LS483323.1	*Streptococcus_pyogenes*	TRUE	**1559**	NZ_CAAISN010000001.1	*Streptococcus_pyogenes*	FALSE
1510	NZ_LS483437.1	*Streptococcus_pyogenes*	TRUE	**1560**	NZ_CAAITP010000001.1	*Streptococcus_pyogenes*	FALSE
1511	NZ_LR031521.1	*Streptococcus_pyogenes*	TRUE	**1561**	NZ_CAAITY010000001.1	*Streptococcus_pyogenes*	FALSE
1512	NZ_CAAHHS010000001.1	*Streptococcus_pyogenes*	FALSE	**1562**	NZ_CAAIVC010000001.1	*Streptococcus_pyogenes*	FALSE
1513	NZ_CAAHJH010000001.1	*Streptococcus_pyogenes*	FALSE	**1563**	NZ_CAAIVK010000001.1	*Streptococcus_pyogenes*	FALSE
1514	NZ_CAAHMR010000001.1	*Streptococcus_pyogenes*	FALSE	**1564**	NZ_CAAIWC010000001.	*Streptococcus_pyogenes*	FALSE
1515	NZ_CAAHNI010000001.1	*Streptococcus_pyogenes*	FALSE	**1565**	NZ_CAAIYI010000001.1	*Streptococcus_pyogenes*	FALSE
1516	NZ_CAAHNV010000001.1	*Streptococcus_pyogenes*	FALSE	**1566**	NZ_CAAIZH010000001.1	*Streptococcus_pyogenes*	FALSE
1517	NZ_CAAHNZ010000001.1	*Streptococcus_pyogenes*	FALSE	**1567**	NZ_CAAIZR010000001.1	*Streptococcus_pyogenes*	FALSE
1518	NZ_CAAHOJ010000001.1	*Streptococcus_pyogenes*	FALSE	**1568**	NZ_CAAIZZ010000001.1	*Streptococcus_pyogenes*	FALSE
1519	NZ_CAAHPQ010000001.1	*Streptococcus_pyogenes*	FALSE				
1520	NZ_CAAHSF010000001.1	*Streptococcus_pyogenes*	FALSE				
1521	NZ_CAAHUO010000001.1	*Streptococcus_pyogenes*	FALSE				
1522	NZ_CAAHUT010000001.1	*Streptococcus_pyogenes*	FALSE				
1523	NZ_CAAHVC010000001.1	*Streptococcus_pyogenes*	FALSE				
1524	NZ_CAAHVV010000001.1	*Streptococcus_pyogenes*	FALSE				
1525	NZ_CAAHVY010000001.1	*Streptococcus_pyogenes*	FALSE				
1526	NZ_CAAHVX010000001.1	*Streptococcus_pyogenes*	FALSE				
1527	NZ_CAAHWF010000001.1	*Streptococcus_pyogenes*	FALSE				
1528	NZ_CAAHWU010000001.1	*Streptococcus_pyogenes*	FALSE				
1529	NZ_CAAHWY010000001.1	*Streptococcus_pyogenes*	FALSE				
1530	NZ_CAAHXC010000001.1	*Streptococcus_pyogenes*	FALSE				
1531	NZ_CAAHXK010000001.1	*Streptococcus_pyogenes*	FALSE				
1532	NZ_CAAHXS010000001.1	*Streptococcus_pyogenes*	FALSE				
1533	NZ_CAAHXW010000001.1	*Streptococcus_pyogenes*	FALSE				
1534	NZ_CAAHZL010000001.1	*Streptococcus_pyogenes*	FALSE				
1535	NZ_CAAHZS010000001.1	*Streptococcus_pyogenes*	FALSE				
1536	NZ_CAAIAE010000001.1	*Streptococcus_pyogenes*	FALSE				
1537	NZ_CAAIAR010000001.1	*Streptococcus_pyogenes*	FALSE				
1538	NZ_CAAIBO010000001.1	*Streptococcus_pyogenes*	FALSE				
1539	NZ_CAAICW010000001.1	*Streptococcus_pyogenes*	FALSE				
1540	NZ_CAAIDZ010000001.1	*Streptococcus_pyogenes*	FALSE				
1541	NZ_CAAIEB010000001.1	*Streptococcus_pyogenes*	FALSE				
1542	NZ_CAAIEY010000001.1	*Streptococcus_pyogenes*	FALSE				
1543	NZ_CAAIFB010000001.1	*Streptococcus_pyogenes*	FALSE				
1544	NZ_CAAIGA010000001.1	*Streptococcus_pyogenes*	FALSE				
1545	NZ_CAAIGX010000001.1	*Streptococcus_pyogenes*	FALSE				
1546	NZ_CAAIIG010000001.1	*Streptococcus_pyogenes*	FALSE				
1547	NZ_CAAIJA010000001.1	*Streptococcus_pyogenes*	FALSE				
1548	NZ_CAAIKF010000001.1	*Streptococcus_pyogenes*	FALSE				
1549	NZ_CAAIKO010000001.1	*Streptococcus_pyogenes*	FALSE				
1550	NZ_CAAILQ010000001.1	*Streptococcus_pyogenes*	FALSE				

"What we observe is not nature itself, but nature

exposed to our method of questioning."

-Werner Heisenberg

6 SEQUENCING BLOOD CULTURE CLINICAL SAMPLES AND PREDICTING SPECIES ABUNDANCE

In Chapter 5, I demonstrated that bioinformatic contamination does exist and varies by the combination of sequencer, classifier and database used. I have also produced a misclassification catalogue that can be used to account for contamination from simulated bacterial reference genomes. I found that Bracken using a standard database had the best performance and filtering shorter reads (at least <1000bp) from error-prone Nanopore sequencing can provide performance either equivalent or superior to standard Illumina sequencing. This informs the bioinformatic pipeline

design that I will use to assess performance against the conventional laboratory. Additionally, it provides a method of accounting for bioinformatic contamination to produce more accurate results when using these tools. In this chapter, I process 211 positive blood cultures and 62 negative blood cultures using the laboratory and bioinformatic pipeline I have developed. I then demonstrate the diagnostic performance of this pipeline and compare it to the conventional laboratory in various aspects including performance and speed.

Bloodstream infections (BSIs) pose a significant global health challenge, given their high mortality rates and substantial strain on healthcare systems. The World Health Organization (WHO) estimates that BSI affects over 30 million individuals annually, with an alarming death toll of 6 million.[164] The associated mortality rate spans from 15% to 30%, underscoring the urgent need for efficacious diagnostic and therapeutic strategies.[164,165] Complicating the matter is the escalating prevalence of antibiotic-resistant bacteria. These multidrug-resistant organisms not only contribute to the rise in BSI rates but also limit treatment options, amplifying associated morbidity and mortality.[165] In severe BSI cases, the timely administration of appropriate antimicrobial therapy is pivotal, with studies demonstrating an average 7.6% decrease in survival for each hour of delay in antimicrobial administration;[166] a delay exceeding one day in initiating appropriate antimicrobial therapy correlates with a twofold increase in mortality risk.[87] These findings highlight the vital role of rapid and precise diagnostic approaches in identifying causative pathogens and informing timely, effective antimicrobial treatments.[234]

Recent advancements in technology have seen the rise of clinical metagenomics, emerging as a promising diagnostic tool for infectious diseases, including BSIs. Metagenomic sequencing of blood samples offers a fast and extensive method for identifying and characterizing pathogens. This innovative approach bypasses the limitations of traditional culture-based methodologies, often uncovering elusive organisms frequently overlooked by standard laboratory procedures.[167] However, clinical metagenomics is not without its set of obstacles, including contamination risks [64], a dearth of standardized workflows and protocols [118], and issues related to performance and sensitivity[142], which need addressing for broader application. Furthermore, the cost associated with initial setup and ongoing operational expenses, such as equipment maintenance and data storage, also requires consideration. [567]

Notwithstanding these hurdles, the transformative potential of metagenomic sequencing in BSIs is compelling. It offers several key benefits, among them being the identification of fastidious or unculturable organisms, providing a more comprehensive view of the microbial composition in bloodstream infections. This unique capability enables the detection of polymicrobial, and atypical or emerging pathogens, leading to improved diagnosis informing targeted treatment strategies.[2] Another significant advantage of metagenomic sequencing is its ability to predict antimicrobial resistance (AMR) profiles, facilitating the early selection of appropriate antimicrobial therapies in a more comprehensive manner compared to current molecular tests such as the BioFire® Blood Culture Identification (BCID)

Panel (BioFire Diagnostics, Salt Lake City, UT, USA). This information can guide clinicians in choosing the most effective treatment options and aid in the mitigation of AMR spread.[168] Beyond this, metagenomic sequencing has the potential to track outbreaks and monitor the transmission dynamics of BSI pathogens. By analysing the genetic relatedness of isolates, metagenomics can provide insights into the epidemiology and spread of BSI-causing pathogens. The identification of shared genetic markers and phylogenetic relationships among isolates can help determine the source of infection, trace transmission chains, and implement targeted infection control measures to prevent further dissemination.[84,169,170] Moreover, metagenomic sequencing can aid in the early detection of outbreaks by detecting clusters of related pathogens within a population. By implementing surveillance programs that incorporate metagenomics, healthcare facilities can proactively identify and respond to potential outbreaks, thus preventing further infections and reducing the overall disease burden.

Performing metagenomics directly on blood presents several challenges due to the high human to pathogen DNA ratio, which makes it difficult to detect and sequence the DNA of the causative pathogen. One approach taken is to rely on blood culture as a "pre-amplification step". This process involves incubating the blood sample to allow any bacteria or fungi present to multiply, thereby increasing the relative amount of pathogen DNA and making it easier to detect and sequence. While this approach has its limitations, such as the time required for pathogen growth and the potential for suboptimal growth of certain pathogens in liquid blood culture, it is nonetheless a standard procedure in microbiology laboratories. The routine nature

of this process, coupled with its proven effectiveness in enhancing the detectability of pathogen DNA, underscores its value in the diagnosis and management of bloodstream infections.

Building upon this established methodology, Luke Anson and colleagues propose a method for preparing bacterial DNA directly from positive blood culture bottles for whole-genome sequencing using the Illumina MiSeq and Nanopore MinION system.[3] They demonstrated the ability of their method to provide Gram-negative and *Staphylococcus sp.* species identification using both short-read Illumina sequencing and the emerging long-read Nanopore MinION system. Furthermore, they were able to predict limited antimicrobial resistance using Illumina sequencing data only. Another recent study by Patrick Harris and colleagues, demonstrated that direct-from-blood culture methods can accurately identify causative pathogens at the species level, particularly in monomicrobial infections approximately 9-17 hours after a positive blood culture result.[171] However, the study also highlights the limitations of predictive antimicrobial susceptibility testing, which showed varying levels of effectiveness across different bacterial species and antibiotic combinations. Another approach, the utilization of microbial cell-free DNA (cfDNA) sequencing, performed directly from blood samples and circumventing the necessity for blood culture, has demonstrated its capacity to accurately identify a diverse array of infectious species, but does not yield robust antibiotic susceptibility data.[95] The time frame required to obtain results from this method aligns with that of traditional laboratory procedures, underscoring its efficiency,

however, the prediction of antimicrobial resistance via cfDNA sequencing presents a significant hurdle.

Any approach that starts with direct from patient blood needs an amplification step, as often only very few bacteria are present, even as low as 1-10 organisms per ml of blood. Growth in culture provides one form of amplification time, with doubling times of around 20 minutes, that selectively amplifies bacteria, as human cells to do not divide this rapidly or in these conditions. For a molecular technique to better this requires use of PCR or similar technologies that, for example, using extreme PCR, may offer shorter doubling times where cycles last less than a minute, however specifically amplifying bacterial DNA, while still amplifying the whole genome is an unmet challenge.

Given these findings, the proposed workflow in this chapter will aim to build on the strengths of these methods and address their limitations. By doing so, it is hoped that this workflow will contribute to the advancement of infectious disease diagnostics in blood stream infections, ultimately improving patient outcomes.

6.1 Methods

I developed a workflow to deliver direct-from-positive-blood culture speciation results (**Figure 36**). Detailed sample processing time was recorded to facilitate comparisons with the conventional microbiology workflow.

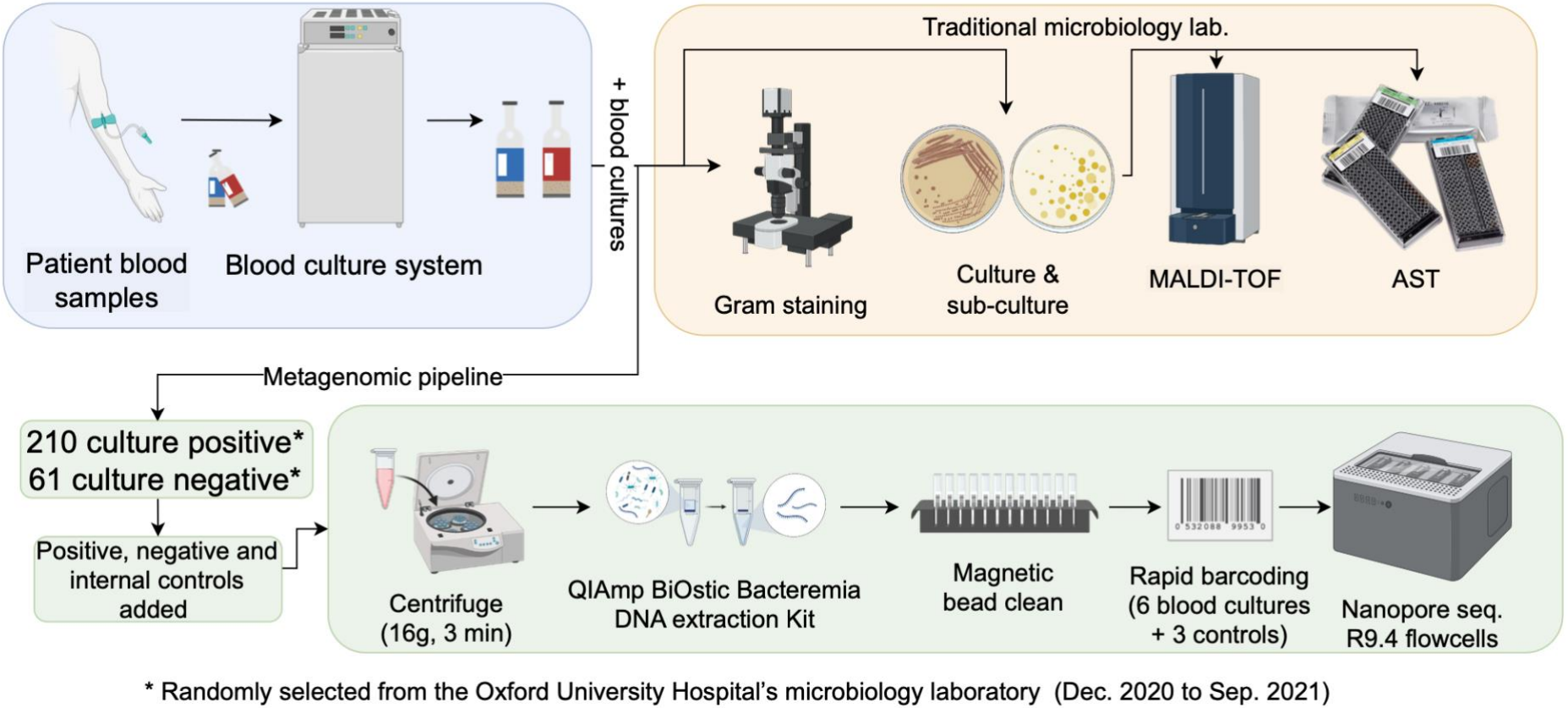

Figure 36. Traditional microbiology vs. metagenomic pipeline. Positive blood cultures are identified by monitoring metabolic changes, indicating bacterial growth, by the incubator.

6.1.1 Sample collection and processing

I processed a total of 211 positive blood cultures and 62 negative blood cultures (de-duplicated to one BSI isolate per 30-day period per patient) between November 9, 2020, and November 3, 2022. These samples were submitted to Oxford University Hospitals' (OUH) microbiology laboratory at the John Radcliffe Hospital, Oxford, UK, as part of routine clinical care. Six samples were consecutively selected to be processed each day on which sequencing was performed from those received by the OUH laboratory. Samples were processed with no pre-selection, that is, no selection according to speciation, *in vitro* antibiotic susceptibility, or patient factors, and hence I refer to this as a completely "unselected" or "random" sample population.

Blood cultures were collected using the BD BACTEC™ system (BD, USA). into aerobic and anaerobic bottles. I did not process clinical samples that were collected using paediatric blood culture bottles, the BD BACTEC™ Peds Plus™ media. Samples that was due to be discarded were taken immediately after routine tests were completed. Therefore, individual patient consent was not required. This study was done with Research Ethics Committee approval (REC reference 17/LO/1420). Although aerobic and anaerobic cultures are drawn together, only one was processed as this was the bottle that flagged positive during the sample collection period. Collected blood culture bottles included BD BACTEC^{TM} Plus Aerobic, and Lytic Anaerobic (BD, USA). The former contains resin, and the latter lytic agents

for the release of intracellular pathogens. The BD BACTEC™ FX blood culture system stored liquid blood culture bottles at 35 °C until positivity or negativity (defined as not flagging positive after five days). For the purposes of this research, 1.5ml of liquid blood culture was removed and immediately processed.

6.1.2 DNA extraction from blood culture

Samples were extracted using the BiOstic Bacteraemia kit (MoBio, Qiagen, USA) according to the manufacturers' protocol with no pre-treatment steps (Appendix 2). Prior to extraction, two additional negative controls, and one positive control was added to each batch of 6 clinical samples for processing, totalling 9 samples being processed in each batch. As a positive control for each batch, I used a single 5ul loop of *Arcanobacterium haemolyticum* grown from blood agar plate cultures. The culture broth from uninoculated BD BACTEC™ Plus Aerobic, and Lytic Anaerobic bottles were used as negative controls. Extracted DNA from clinical samples were normalised prior to library preparation, where possible, and DNA from positive controls were normalised to one-fourth of this concentration to ensure greater sequencing coverage depth of clinical samples. A spiked internal control of *Thermus thermophilus* DNA at a concentration of 2% (with an absolute minimum amount of 0.5ng) of the normalised DNA concentration was added to every clinical sample, positive and one negative control, leaving the other negative control completely free from any DNA input. CFU counts were determined for every sample at the point of extraction using chocolate, blood and chromes agar plates.

6.1.3 Library preparation and Nanopore sequencing

Samples were prepared for GridION sequencing using the Rapid Sequencing Kit (SQK-RBK004) following the manufacturer's protocol. Sequencing was performed using the R9.4.1 flow cells.

6.1.4 Bioinformatic analysis

Metagenomic bioinformatic pipelines for species prediction is shown in (**Figure 37**). Resistance prediction methodology is further detailed in Chapter 7.

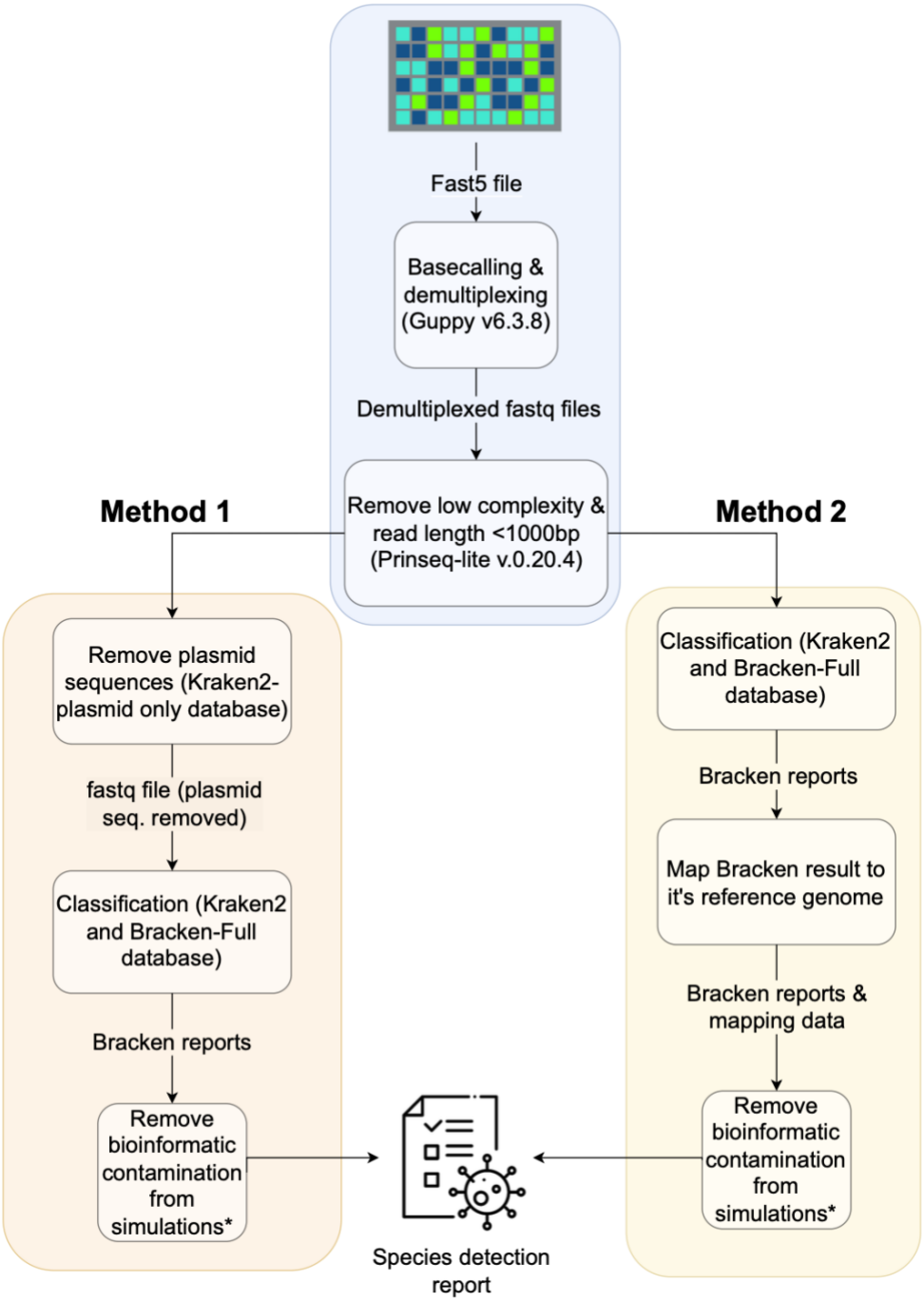

Figure 37. Bioinformatic pipeline: Speciation

6.1.4.1 Species prediction

GridION sequencing data was pre-processed using Guppy (version 6.4.8) to demultiplex barcodes and prinseq-lite (v0.20.4) to remove reads with low complexity and lengths <1000bp (threshold based on previous work in Chapter 5).[172] Two bioinformatic workflow methods were then tested. The first method removed plasmid reads prior to classification using a plasmid only Kraken2 database that was built specifically for the analysis (sequences downloaded from NCBI on the 16/3/23), as plasmids, due to horizontal gene transfer, may be classified into multiple species erroneously. Species abundance including polymicrobial infections were identified using Bracken (V2.5; specific settings available in Appendix 1) using a full standard database built from archaeal, bacterial, plasmid, viral and human reference genomes downloaded from NCBI RefSeq on the 13/2/2022. The second method did not remove plasmid reads, but instead used Minimap2 (v2.24-r1122; using the map-ont function) to map sequencing reads to each species identified by Bracken to a species-specific reference genome (downloaded randomly from NBCI) after minimal filtering that excluded only the least plausible species before mapping. Samtools (v 1.17) was used to determine coverage breadth, depth of coverage, and percentage of reads mapped from extracted species reads. In both methods, simulations were performed to control for contamination arising from bioinformatic misclassification using methods previously described and published (See Chapter 5).[172] Databases were built according to each tool's default setting.

6.1.4.2 Statistical analysis

The performance of metagenomic sequencing was assessed by comparing the species identified from sequencing data with the species from pure culture isolates grown from blood culture samples [species from Bruker microflex MALDI-TOF MS (Bruker, USA)].

6.1.4.2.1 Threshold setting

To correct for samples which may contain small numbers of contaminating and nonspecific bacterial reads, a threshold was determined to identify the presence of true infection, using 60% of the culture bottles processed as a derivation set (and the remaining 40% as the testing set) selected at random using the sample_frac() function in RStudio (V 2023.03.0+386). The same split was used throughout all approaches tested to ensure comparable results.

The following filters were used: read- and percentage cut-off, that is, the total number of reads classified as a specific species, and the percentage of all bacterial reads classified as a specific species. Additionally in the second approach tested, mapping breadth (i.e., proportion of the reference genome called), depth (mean depth across the whole reference genome) and the percentage of extracted reads mapped were also used as filters. Parameter values were selected by numerical optimization, using RStudio comparing sequencing results to blood culture results and maximizing the value of the Youden index (sensitivity + specificity - 1).[173] Additionally and reported separately, parameter values were subsequently updated

to maximise sensitivity, after adjusting results for false positive results that were in fact plausible infections missed by the routine laboratory.[19]

Sensitivity was calculated taking each species identified from each culture-positive sample as a separate data point contributing to the numerator and denominator; therefore culture-positive samples with a single species contributed once, and culture-positive samples with two or more species contributed two or more times. Specificity was determined both by considering the proportion of samples deemed negative by blood culture that were identified as negative by metagenomic sequencing, and in culture-positive samples. In the latter samples false-positive results are still possible where an additional species not found by culture was reported. In culture positive results, specificity was defined on a per sample basis, where any sample with >1 additional species found from sequencing was considered to contribute as 1 to the false negative count.

6.1.4.2.2 Assessing false positive results for plausibility

False positive results, that is, additional species detected from sequencing not found on routine culture, were assessed for bioinformatic plausibility using two approaches. Firstly, 1000 reads were extracted from additional species detected and queried using NCBI BLAST (v2.13.0+) with the ref_prok_rep database (downloaded 12/3/2023) to see if the same species was found. Hits were filtered using percent identity>=95, coverage >=80, E-value <1e-50 and the -max_target_seqs option set to 1. Secondly, for each additional species identified by

Bracken, extracted reads were mapped using Minimap2 (v2.24-r1122; using the map-ont function) to a randomly selected refseq genome of the same species (downloaded from NCBI 29/4/23, accession numbers available in appendix). Samtools (v 1.17) was used to calculate genome coverage breadth and depth.

To investigate the extent to which bioinformatic misclassification might arise between genomically similar species, the Average Nucleotide Index (ANI) was calculated for additional bioinformatic species found against the MALDI result using ANI.pl (downloaded reference genomes from NCBI 29/4/23).[174] Additional microbiology data was captured from other blood culture samples, if any were available, within 5 days of the initial collected blood sample to look for evidence of the additional species in these samples. Clinical notes from laboratory requests were also reviewed for relevance to additional infections found by sequencing. Provision for this is included in the ethical approvals cited above.[20]

6.1.4.2.3 Machine learning using random forests

In addition to the numerically optimised threshold approach, a machine learning approach using a random forest model was applied to classify Bracken results after correction with simulation data using five input features (i.e. percentage of reads assigned to the species [i.e. covered by the clade rooted at this taxon], number of reads covered by the clade rooted at this taxon, mapped breadth [i.e. percentage of reference genome covered by at least 1 read], mean mapped depth, and percentage of extracted reads mapped to the species reference genome). The randomForest package in R developed by A. Liaw et al was used.[175]

Hyperparameter selection was made using a repeated cross-validation approach with 5 folds and 40 repeats. The data was split into a training set (60% of the data) and a testing set (40%). A random forest model was trained using the training set, and its variable importance was assessed. The model was evaluated on the testing set using a confusion matrix, and performance metrics including sensitivity and specificity, were calculated. Predictions on the testing set were analysed, and instances of agreement and disagreement between predicted and actual values were identified. Thereafter I re-train the model using a "corrected" training dataset, correcting for plausible false positive and false negative errors that were identified using the previous heuristic method. I then assess the updated model performance on the testing dataset. All code used in this analysis is available in Appendix 1.

6.2 Results: Method 1 – optimised threshold approach (plasmid sequences removed)

The dataset (n=205 blood culture samples [216 individual species]) was randomly split into 60% (n=123 [128 individual species]) for the training dataset, and 40% (n=82 [88 individual species]) for the testing dataset. Optimal thresholds for determining if samples contained low-level contamination or true infection from sequencing were determined by numerical optimization in the training set, choosing thresholds that maximized the Youden's Index (Sensitivity + specificity - 1) (**Figure 38**). The final, most sensitive, threshold chosen to determine the presence of true species was 750 reads assigned to that species and having a total percentage cut-off of 10% of the total bacterial reads assigned to that species (**Table 12**). The Youden's

index after using the optimised threshold in the training dataset, testing and entire dataset was 94.4%, 87.1% and 91.4% respectively (**Table 13**). In the testing dataset, 81/88 species were detected [i.e., sensitivity of 94% (95% CI, 85-97%)] while 4/82 culture-positive samples were found to have an additional false positive result [i.e., Specificity: 95% (88-99%)] (**Table 26**).

When applying the model to the full dataset, 201/216 species were detected [i.e., sensitivity of 94% (95% CI, 91-97%)] while 7/205 culture-positive samples were found to have a false positive result [i.e., Specificity: 97% (95% CI, 93-99%)].

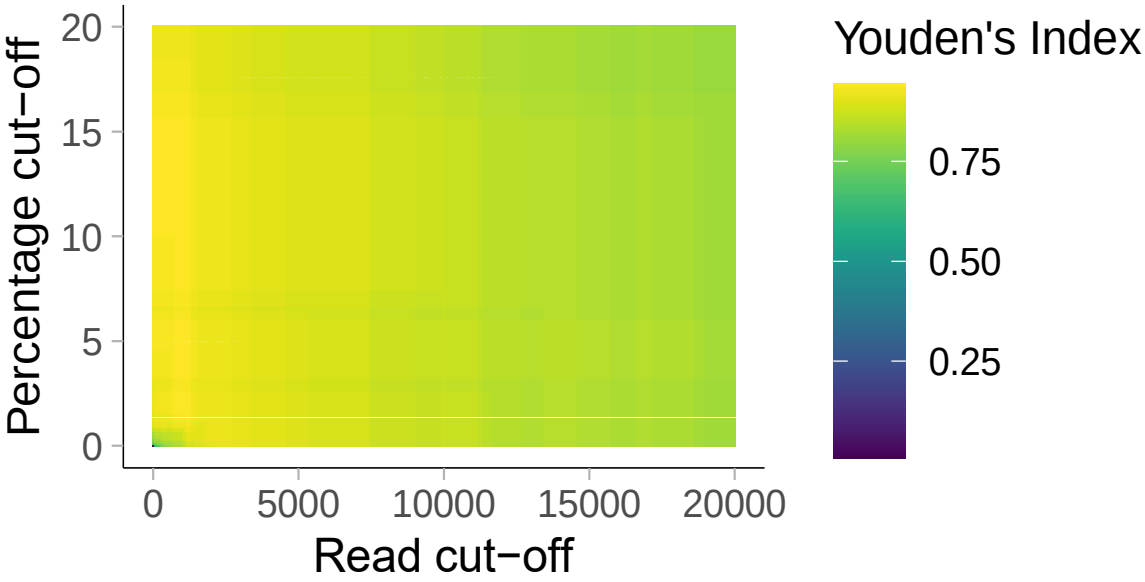

Figure 38. Filtering calibration heat map from training dataset in Method 1.

Table 12. Method 1 applied to the training dataset: Read and percentage cut-off maximising the Youden's Index.

Percentage cut-off	Read cut-off	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index
10	750	97% (92-99)	98% (93-99)	94.4%
10	1250	97% (92-99)	98% (93-99)	94.4%
10	1300	96% (91-99)	98% (93-99)	93.6%
10.1	0	97% (92-99)	98% (93-99)	94.4%
10.1	1250	97% (92-99)	98% (93-99)	94.4%
10.1	1300	96% (91-99)	98% (93-99)	93.6%
10.1 to 15.6	0 to 1250	97% (92-99)	98% (93-99)	94.4%
15.6	0	97% (92-99)	98% (93-99)	94.4%
15.6	1250	97% (92-99)	98% (93-99)	94.4%
15.6	1300	96% (91-99)	98% (93-99)	93.6%

Table 13. Method 1 applied to the training, testing and full dataset: Read and percentage cut-off maximising the Youden's Index.

Dataset	Percentage cut-off	Read cut-off	Sensitivity	Specificity	Youden's Index
Training	10	750	97% (92-99)	98% (93-99)	94.4%
Testing	10	750	92% (85-97)	95% (88-99)	87.1%
Full set	10	750	94% (91-97)	97% (93-99)	91.4%

6.3 Results: Method 2 – optimised threshold approach (with mapping)

Initial Bracken results were liberally filtered, to reduce computational resources required at the next step for mapping, by requiring the percentage of reads covered by the clade rooted at this taxon to be >0.1% and number of reads covered by the clade rooted at this taxon to be >50. The dataset used the same training/test split as above in Method 1 for the testing dataset. Mapping thresholds was set to maximise sensitivity enabling the detection of bacteria confirmed by MALDI result (including any polymicrobial phenotypic results). These were set as a minimum breadth, depth and mapped percentage of extracted reads as 32%, 0.4, and 75% respectively, which were the lowest thresholds found in true sample in the training dataset.

The optimal threshold chosen to determine the presence of a true species was a cut-off of 0 reads (i.e., disabling this filter) and having a total percentage cut-off of 3.3% of the total bacterial reads after removing misclassification using the simulation data (**Figure 39, Table 14**). The Youden's index after using the optimising thresholds in the training dataset, testing and entire dataset was 95.1%, 88.2% and 91.3% respectively (**Table 14, Table 15**). Compared to Method 1, Method 2 had overall better accuracy for a lower percentage cut-off threshold. In the testing dataset, 83/88 species were detected [i.e., sensitivity of 94% (95%CI, 87-98%)] while 5/82 samples were found to have a false positive result [i.e., Specificity: 94% (86-98%)] (**Table 26**). When applying the model to the full dataset, 210/216 species were detected [i.e., sensitivity of 97% (95% CI, 94-99%)]

while 12/205 samples were found to have a false positive result [i.e., Specificity: 94% (95% CI, 94-97%)].

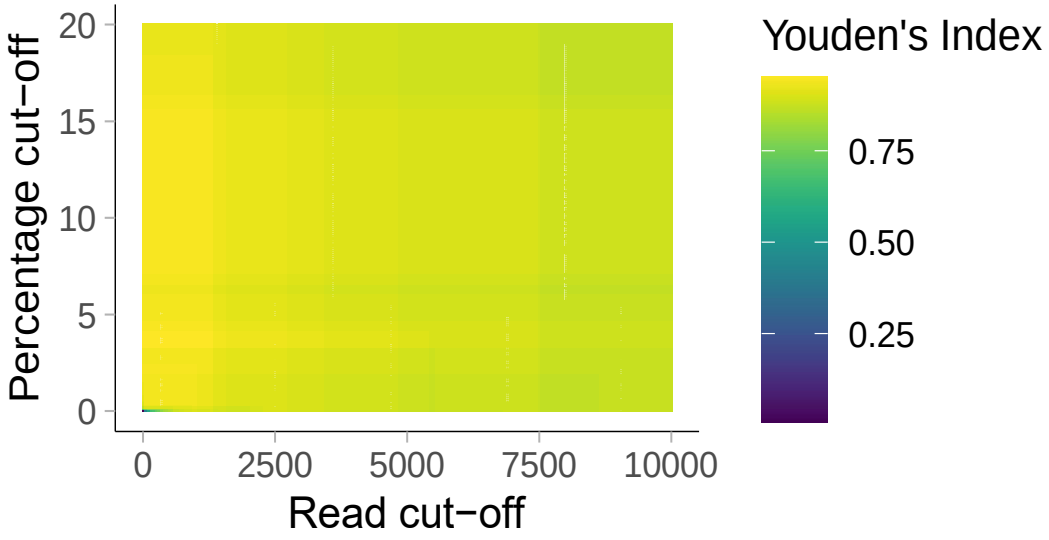

Figure 39. Filtering calibration heat map from training dataset in Method 2.

Table 14. Method 2 applied to the training dataset: Read and percentage cut-off maximising the Youden's Index.

Percentage cut-off	Read cut-off	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index
3.2	0	99% (96-100)	94% (89-98)	93.5%
3.2	1300	99% (96-100)	94% (89-98)	93.5%
3.2	1350	98% (93-100)	94% (89-98)	91.9%
3.3	0	99% (96-100)	96% (91-99)	95.1%
3.3	1300	99% (96-100)	96% (91-99)	95.1%
3.3	1350	98% (93-100)	96% (91-99)	93.5%
3.3 to 4.1	0 to 1300	99% (96-100)	96% (91-99)	95.1%
4.1	0	99% (96-100)	96% (91-99)	95.1%
4.1	1300	99% (96-100)	96% (91-99)	95.1%
4.1	1350	99% (96-100)	96% (91-99)	93.5%
4.2	0	98% (94-100)	96% (91-99)	94.4%
4.2	1300	98% (94-100)	96% (91-99)	94.4%
4.2	1350	97% (92-99)	96% (91-99)	92.8%

Table 15. Method 2 applied to the training, testing and full dataset: Read and percentage cut-off maximising the Youden's Index.

Dataset	Percentage cut-off	Read cut-off	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index
Training	3.3	0	99% (96-100)	96% (91-99)	95.1%
Testing	3.3	0	94% (87-98)	94% (86-98)	88.2%
Full set	3.3	0	97% (94-99)	94% (90-97)	91.3%

6.3.1 Assessing false negative results for plausibility in Method 2 (i.e., potential inaccurate sequencing results or inaccurate MALDI-TOF results)

A total of 6 false negative results were identified across the training and test datasets using Method 2, that is, where sequencing failed to detect a species found by MALDI-TOF. In four false negative results, the sequencing results included a species that was discordant with MALDI-TOF species, but closely related (**Table 16**). In a further two species (samples 143 and 147) the percentage threshold of 3.3 was too high to call these samples positive; notably both species were part of a cultured polymicrobial growth.

To interpret MALDI results, probability scores are used: a value of >2.3 indicates "highly probable species identification", a score >2 and <2.3 indicates "secure genus identification, probable species identification", a score >1.7 and <2 indicates "probable genus identification", and a score <1.7 indicates "unreliable identification".[176] In samples 22, 23, 46 and 49, MALDI-TOF predicted *Corynebacterium lipophiloflavum, Staphylococcus epidermidis, Klebsiella pneumoniae* and *Klebsiella pneumoniae (MALDI scores of 2.08, 2.07, 2.35, and 2.35 respectively)* while sequencing results predicted *Corynebacterium sanguinis*, *Staphylococcus hominis, Klebsiella quasipneumoniae* and *Klebsiella quasipneumoniae* respectively. Samples were repeat tested grown from saved stock cultures and yield the same MALDI-TOF results as described. In all four samples, there was higher coverage breadth of the sequencing reference genome species than

the MALDI-TOF reference genome species (93% vs 92.7%, 95% vs 63%, 62% vs 59% and 68% vs 62%). Furthermore, almost all blast hits were classified to the species identified by sequencing (99.2%, 97%, 96% and 96% respectively) rather than the MALDI-TOF species. Prior studies have shown that *Corynebacterium sanguinis* and *Corynebacterium lipophiloflavum,* and *Staphylococcus epidermidis* and *Staphylococcus hominis* are genotypically similar respectively and are often misidentified by MALDI-TOF.[177–179] Furthermore, the MALDI-TOF probability scores for these two organisms were 2.08 and 2.07 respectively, suggesting a reliable genus identification, but only probable species identification.

Klebsiella quasipneumoniae is a newly defined species, not readily distinguished from *Klebsiella pneumoniae* with current MALDI-TOF identification techniques, which has resulted in its true prevalence being unknown.[180] This is likely due to a lack of discrimination between spectra patterns being produced by these species. *Klebsiella quasipneumoniae* has been shown to be capable of nosocomial transmission and acquiring and maintaining relevant resistance elements, placing an increased necessity for accurate identification of this organism.[180] Biomerieux API (Analytical Profile Index) 20E testing was performed on samples 46 and 49 which were both sequenced as *Klebsiella quasipneumoniae,* as well as sample 60 which was sequenced as *Klebsiella pneumoniae* as a control for phenotype. Sample 60 was identified by the API test to have a high probability (i.e., % ID: 82) as being *Klebsiella pneumoniae*, while both samples 46 and 49 were identified with a much lower probability (i.e., % ID: 21) as being identified as *Klebsiella pneumoniae*. This difference in probability score was due to the two samples sequenced as *Klebsiella*

quasipneumoniae having a negative LDC result, which tests for decarboxylations of the amino acid lysine by lysine decarboxylase (**Figure 40**). Furthermore, *Klebsiella quasipneumoniae* is not in the API 20E database, and therefore a much lower probability is expected. In summary, all four discordant organisms identified though sequencing are likely to be similar or a more accurate representation of true species compared to the phenotypic MALDI-TOF results.

Table 16. Assessing false negative species results for plausibility (i.e., potential inaccurate MALDI-TOF results).

Sample no.	Phenotypic result		Sequencing result				ANI
	MALDI species	MALDI probability score	Bracken species (%)	Blast result (1000 extracted reads)	Bracken sp. reads mapped to Bracken sp. reference genome: Coverage; mean depth	Bracken sp. reads mapped to MALDI sp. reference genome: Coverage; mean depth	
22	*Corynebacterium lipophiloflavum*	2.08	*Corynebacterium sanguinis (76%)*	*Corynebacterium sanguinis (99.2%)* *Corynebacterium lipophiloflavum (0.5%)*	93%, 128x	92.7%, 129x	91
23	*Staphylococcus epidermidis*	2.07	*Staphylococcus hominis (90%)*	*Staphylococcus hominis (97%)* *Staphylococcus haemolyticus (1.5%)*	95% 19x	63%, 9x	76
46	*Klebsiella pneumoniae*	2.35	*Klebsiella quasipneumoniae (84%)*	*Klebsiella quasipneumoniae (96%)* *Klebsiella pneumoniae (2%)* *Klebsiella oxytoca (2%)*	62%, 2.6x	59%, 2.1x	94
49	*Klebsiella pneumoniae*	2.35	*Klebsiella quasipneumoniae (84%)*	*Klebsiella quasipneumoniae (96%)* *Klebsiella oxytoca (2%)* *Citrobacter telavivensis (2%)*	68%, 3.6x	62%, 2.4x	94



Figure 40. Biomerieux API (Analytical Profile Index) 20E testing for *Klebsiella pneumoniae* (Sample 60) vs *Klebsiella quasipneumoniae* (Sample 46 & 49).

6.3.2 Assessing false positives in Method 2 by maximizing sensitivity (i.e., potential contaminants or potential missed infections by the routine laboratory)

As the 'false positive' results found by the initial thresholds chosen were in fact plausible results, I considered relaxing the stringency of filtering to optimise sensitivity and identify the 2 species missed. I then investigated if this would lead to more potentially plausible additional species being identified. Repeat numerical optimisation was used to optimise for sensitivity (i.e., restricting sensitivity in the training data to be 100%) and then specificity, rather than using the Youden's Index. The optimal thresholds chosen to determine the presence of true single species was 50 reads and having a total percentage cut-off of 0.5% of the total bacterial reads after removing misclassification using the simulation data (**Table 17**). Using these thresholds, sensitivity and specificity in the test dataset were 100% (95% CI, 97-100%) [128/128] and 94% (88-97%) [115/123] respectively, and 98% (95-99%) [212/216] and 92% (88-95%) [189/205] respectively in the whole dataset (**Table 18**). The four remaining false negative results are explained by the MALDI-TOF/sequencing discrepancies described above.

Table 17. Method 2 applied to the entire dataset: Read and percentage cut-off maximising for Sensitivity (rather than for Youden's Index).

Percentage cut-off	Read cut-off	Sensitivity (95% CI)	Specificity (95% CI)
0.4	50	98% (95-99%)	92% (87-97%)
0.4	1000	98% (95-99%)	92% (87-97%)
0.5	0	98% (95-99%)	91% (86-94%)
0.5	50	98% (95-99%)	92% (88-95%)
0.5	1000	98% (95-99%)	92% (88-95%)
0.5	1050	98% (95-99%)	92% (87-97%)
0.5 to 1.3	50 to 1000	98% (95-99%)	92% (88-95%)
1.2	0	98% (95-99%)	92% (87-97%)
1.3	50	98% (95-99%)	92% (88-95%)
1.3	1000	98% (95-99%)	92% (88-95%)
1.3	1050	98% (95-99%)	92% (87-97%)
1.4	50	98% (95-99%)	92% (87-97%)
1.4	1000	98% (95-99%)	92% (87-97%)

Table 18. Method 2 applied to the training, testing and full dataset: Read and percentage cut-off maximising Sensitivity and then Specificity.

Dataset	Percentage cut-off	Read cut-off	Sensitivity (95% CI)	Specificity (95% CI)
Training	0.5	50	100% (97-100%) [128/128]	94% (88-97%) [115/123]
Testing	0.5	50	95% (89-99%) [84/88]	93% (85-97%) [76/82]
Full set	0.5	50	98% (95-100%) [212/216]	92% (88-95%) [189/205]

6.3.3 Assessing false positive results for plausibility:

Plausibility of additional 'false positive' results generated by the new thresholds was assessed by various methods including blasting 1000 randomly selected reads from each species, mapping to the reference and other polymicrobial genomes of species present, average nucleotide identity and associated clinical and microbiology factors. Further methodological details are available in *6.1.4.2.2 Assessing false positive results for plausibility*.

In total, there were 16 blood culture samples that was assessed to have false positive results. Four blood culture samples (each containing an individual species each) are described in the section above on false negative results, as the species found on sequencing differed to the MALDI-TOF species likely due to MALDI-TOF inaccuracy (**Table 16**).

In a further 12 blood culture samples, 13 additional species were found using sequencing and the sensitivity optimized thresholds. These potential polymicrobial infections were assessed and found to be bioinformatically and biologically plausible and were likely to be true additional infections (**Table 19**). Of these, 11/13 additional species results found had a minimum genome coverage breadth of >75%, with a depth of coverage ranging from >3x to 143x. One species, *Collinsella aerofaciens*, had a 48% coverage breadth and 3x depth, and was identified in the other culture bottle. Furthermore, this was supported by all reads blasting to this

specific species. Another species, *Bacteroides nordii*, was found at a 68% coverage breadth and 1.5x depth together with *Bacteroides fragilis*, and *Bacteroides thetaiotaomicron*. The presence of this species was also supported by all reads blasting to this specific species. In the gut, different species of bacteria occupy different ecological niches and perform specialized functions. *Bacteroides fragilis*, *Bacteroides thetaiotaomicron*, and *Bacteroides nordii* have distinct metabolic capabilities and can utilize different nutrients, which reduces competition for resources and may lead multiple anaerobic species to coexist.[181] Poor overall coverage in these two samples of 1.5x and 3x may have led to a reduced coverage breadth in comparison to other additional polymicrobial species identified.

Six out of the 13 additional species identified were anaerobes, suggesting that current microbiological diagnostic techniques and conditions may not be fully optimized to grow and detect these organisms. For instance, sample 78 was found to additionally have *Bacteroides fragilis*, which is known to play a role in abscess formation and may have led to the formation of the liver abscess seen in this patient. Additionally in sample 154, *Haemophilus influenzae* was found as an additional infection in a patient who had squamous cell carcinoma of the lung and a pleural effusion. None of the additional polymicrobial species identified had high ANIs (>95%) with the species found phenotypically suggesting that additional species identified were distinct. In sample 33, *Pseudomonas aeruginosa* was additionally found and is a significant finding in the light of the presence of an old central venous catheter suggesting a central line-associated bloodstream infection. Sample 39 and 129, both detected low levels of *Collinsella aerofaciens* and

Staphylococcus haemolyticus respectively though sequencing, which was interestingly also detected by MALDI-TOF only from the other culture bottle taken. This suggests that some organisms exists in low levels in liquid blood culture bottles, that may often not be identified by the routine microbiology laboratory. In sample 135, *Enterococcus faecium* was found at low levels as an additional infection, and interestingly only was found by the routine microbiological laboratory from blood culture samples taken from the patient the following day.

Table 19. Assessing false positive species results for plausibility.

Sample no.	Phenotypic result	Sequencing result				ANI	Clinical or microbiology factors
	MALDI species	Bracken species (%)	Blast Bracken sp. reads (1000 extracted reads)	Bracken sp. reads mapped to Bracken sp. reference genome: Coverage; mean depth	Bracken sp. reads mapped to other Bracken sp. reference genome: Coverage; mean depth		
5	*Klebsiella pneumoniae*	*Klebsiella pneumoniae (85%)*	*Klebsiella pneumoniae* (97%) *Klebsiella quasivariicola (2%)*	92%, 167x	63%, 86x	78%	
		Escherichia coli (15%)	*Escherichia coli (30%)* *Shigella dysenteriae (40%)* *Shigella sonnei (20%)*	90%, 22x	54%, 10x		
16	*Clostridium perfringens*	*Clostridium perfringens (75%)*	*Clostridium perfringens (100%)*	88%, 30x	11%, 1.2x	68%	
		Paeniclostridium sordellii (25%)	*Paeniclostridium sordellii (100%)*	93%, 14x	11%, 0.8x		
33	*Lacticaseibacillus rhamnosus* *Klebsiella pneumoniae*	*Lacticaseibacillus rhamnosus (73%)*	*Lacticaseibacillus rhamnosus (96%)*	90%, 47x	0.3%, 0.15x	64%	Old central venous catheter left subclavian.
		Pseudomonas aeruginosa (19%)	*Pseudomonas aeruginosa (100%)*	94%, 5x	0.6%, 0.02x		
		Klebsiella pneumoniae (7%)	*Klebsiella pneumoniae (92%)* *Klebsiella quasivariicola (7%)*	78%, 2x	4%, 0.1x	68%	
39	*Bacteroides caccae*	*Bacteroides caccae (96%)*	*Bacteroides caccae (99%)*	79%, 196x	1%, 0.8x	63%	*Collinsella aerofaciens* found in other blood culture bottle taken
		Collinsella aerofaciens (3%)	*Collinsella aerofaciens (100%)*	48%, 3x	0.28%, 0.01x		
78	*Streptococcus anginosus*	*Bacteroides fragilis (92%)*	*Bacteroides fragilis (100%)*	88%, 143x	1%, 1.2x	63%	Liver abscess
		Streptococcus anginosus (6%)	*Streptococcus anginosus (100%)*	77%, 25x	0.5%, 0.1x		

Table 19. Continued.

93	*Staphylococcus haemolyticus*	*Staphylococcus haemolyticus (56%)*	*Staphylococcus haemolyticus (98%)*	92%, 18x	83%, 13x	77%	
		Staphylococcus hominis (43%)	*Staphylococcus hominis (100%)*	92%, 13x	73%, 8x		
121	*Bacteroides fragilis*	*Bacteroides fragilis (86%)*	*Bacteroides fragilis (100%)*	86%, 40x	32%, 9x	76%	
		Bacteroides thetaiotaomicron (9%)	*Bacteroides thetaiotaomicron (96%)* *Bacteroides faecis (4%)*	78%, 5x	35%, 2x		
		Bacteroides nordii (2%)	*Bacteroides nordii (100%)*	68%, 1.5x	24%, 0.4x	76%	
129	*Enterococcus faecium*	*Enterococcus faecium (98%)*	*Enterococcus faecium (94%)*	92%, 167x	4%, 4x	66%	*Staphylococcus haemolyticus grew in other culture bottle*
		Staphylococcus haemolyticus (2%)	*Staphylococcus haemolyticus (99%)*	79%, 3x	2.4%, 0.2x		
134	*Staphylococcus capitis*	*Staphylococcus capitis (69%)*	*Staphylococcus capitis (99%)*	95%, 17x	76%, 11x	78%	
		Staphylococcus epidermidis (30%)	*Staphylococcus epidermidis (98%)*	97%, 10x	79%, 7x		
135	*Enterobacter cloacae complex*	*Enterobacter cloacae complex (96%)*	*Enterobacter hormaechei* (89%) *Enterobacter quasihormaechei (11%)*	90%, 141x	1.2%, 1.3x	65%	*Enterococcus faecium* isolated from additional blood culture
		Enterococcus faecium (2%)	*Enterococcus faecium (94%)* *Enterococcus avium (2%)*	85%, 5x	1%, 0.1x		
154	*Staphylococcus hominis*	*Haemophilus influenzae* (82%)	*Haemophilus influenzae (55%)* *Haemophilus aegyptius (44%)*	86%,101x	2%,2x	63%	Squamous cell carcinoma of the lung. Pleural effusion.
		Staphylococcus hominis (17%)	*Staphylococcus hominis (100%)*	92%,18x	2%,0.3x		
191	*Klebsiella variicola*	*Klebsiella variicola (96%)*	*Klebsiella variicola* (96%) *Klebsiella africana (1%)*	94%, 61x	1%, 0.3x	66%	
		Enterococcus faecium (4%)	*Enterococcus faecium (95%)* *Enterococcus avium (2%)*	89%, 5x	0.6%, 0.1x		

6.3.4 Performance after accounting for plausible false positive and false negative results

After correcting for plausible false positive and false negative infections, the optimal threshold chosen to determine the presence of a species was 50 reads and having a total percentage cut-off of 0.4% of the total bacterial reads (**Table 20**). The Youden's index after using the optimising threshold for sensitivity in the training dataset, testing and entire dataset was 100, 100 and 100 respectively (**Table 21).** In the testing dataset, 90/90 species were detected [i.e., sensitivity of 100% (95 CI, 96-100)] while 0/82 samples were found to have a false negative result [i.e., Specificity: 100% (95 CI, 95-100)]. When accounting for plausible results and applying the model to the full dataset, 229/229 species were detected [i.e., sensitivity of 100% (95 CI, 98-100)] while 0/205 samples were found to have a false negative result [i.e., Specificity: 100% (95 CI, 98-100)] (**Table 26**).

Table 20. Method 2 applied to the training dataset after adjusting for plausible infections: Read and percentage cut-off maximising for Youden's Index.

Percentage cut-off	Read cut-off	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index
0.4	50	100% (98-100%)	100% (97-100%)	100%
0.4	1000	100% (98-100%)	100% (97-100%)	100%
0.4 to 1.2	50 to 100	100% (98-100%)	100% (97-100%)	100%
1.2	50	100% (98-100%)	100% (97-100%)	100%
1.2	1000	100% (98-100%)	100% (97-100%)	100%

Table 21. Method 2 applied to the training, testing and full dataset after adjusting for plausible infections: Read and percentage cut-off maximising for Youden's Index.

Dataset	Percentage cut-off	Read cut-off	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index
Training	0.4	50	100% (98-100%)	100% (97-100%)	100%
Testing	0.4	50	100% (96-100%)	100% (95-100%)	100%
Full set	0.4	50	100% (98-100%)	100% (98-100%)	100%

6.3.5 Performance with even lower threshold cut-offs (after accounting for plausible false positives and false negatives)

To explore if there was a lower limit on which additional species might be plausible, I further relaxed the thresholds. The percentage cut-off was set to >0.1 and read cut-off to 50 as simulation studies did not account for bioinformatic contamination lower than these thresholds. Performance from this lower threshold yielded a sensitivity and specificity of 100% (98-100%) and 99.5% (97-100%) respectively up until a 0.4% percentage cut-off where performance improved to 100% (98-100%) and 100% (98-100%) respectively (**Table 22**). One false positive result accounted for the reduced performance at a lower threshold, which is likely to be a bioinformatic contaminant with a poor reference genome coverage and inconsistent blast results (**Table 23**). *Staphylococcus aureus* is known to be a bioinformatic contaminant (**Supplementary Table 10**) in the simulation experiments of *Enterococcus faecium*, *Staphylococcus haemolyticus*, and *Streptococcus pyogenes* which were all polymicrobial species in this specific sample. It is likely that bioinformatic simulation experiments have not captured the extent of the genomic diversity across these specific species during simulations, accounting for some of the *Staphylococcus aureus* bioinformatic contamination in the sample, but not enough to completely exclude this contaminant. There was no other sample containing *Staphylococcus aureus* that was sequenced in this particular batch.

Table 22. Method 2 applied to the entire dataset after adjusting for plausible infections: Read and percentage cut-off with lower threshold cut-offs.

Percentage cut-off	Read cut-off	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index
>0.1-0.3	50	100% (98-100%)	99.5% (97-100%)	99.5%
>0.1-0.3	1000	100% (98-100%)	99.5% (97-100%)	99.5%
0.4	50	100% (98-100%)	100% (98-100%)	100%
0.4	1000	100% (98-100%)	100% (98-100%)	100%

Table 23. Assessing false positive species results for plausibility with lower threshold cut-offs.

Sample no.	Phenotypic result	Sequencing result				ANI	Clinical or microbiology factors
	MALDI species	Bracken species (%)	Blast Bracken sp. reads (1000 extracted reads)	Bracken sp. reads mapped to Bracken sp. reference genome: Coverage; mean depth	Bracken sp. reads mapped to other Bracken sp. reference genome: Coverage; mean depth		
143	*Enterococcus faecium* *Staphylococcus haemolyticus* *Streptococcus pyogenes*	*Enterococcus faecium (75%)*[1]	*Enterococcus faecium (89%)*	100%, 94x	4%, 1.6x	64%	
		Staphylococcus haemolyticus (13%)[2]	*Staphylococcus haemolyticus (99%)*	92%, 14x	57%, 6x	74%	
		Streptococcus pyogenes (1.3%)[3]	*Streptococcus pyogenes (99%)*	77%, 1.8x	1%, 0.03x	66%	
		Staphylococcus aureus (0.36%)	*Staphylococcus aureus (65%)* *Staphylococcus haemolyticus (5%)*	54%,1x	59%, 3x[1] 28%, 0.5x[2] 2%, 0.04x[3]		

6.4 Culture negative samples

Overall unadjusted species-level specificity in culture-negative samples was 61/62 (98%; 95% CI, 91 to 100%), and after adjusting for plausible infections was 61/61 (100%; 95% CI, 94 to 100%). Filtering thresholds were used from the best performing method: Threshold Method 2 (Mapping thresholds: breadth >32%, depth >0.4 and mapped percentage extracted reads >75%. Percentage cut-off 0.4%, Read cut-off 50). Sample N47 was negative by the routine laboratory (did not flag after 5 days of incubation in the Phoenix BD), however, sequencing confirmed the presence of *Pseudomonas aeruginosa*, which was also present in another blood culture obtained within a few hours from the same patient. It is likely that this sample contained growth below the threshold required to flag positive in the Phoenix BD. Furthermore, this Sample N47 was an anaerobic culture bottle, which may have resulted in a slower growth of *Pseudomonas aeruginosa*, which is an aerobic (and at times facultatively anaerobic) bacterium. **Figure 41** shows the number of reads from pre-filtered culture negative clinical samples where no threshold had been applied. Minor cross-contamination which fell well below the thresholds for filtering was evident particularly in samples that contained *Arcanobacterium haemolyticum*, which was the organisms used as a positive control in this study.



Figure 41. Pre-filtered culture negative clinical samples (filter thresholds not applied). DFB_N47 was shown to be a true infection (*Pseudomonas aeruginosa*) by sequencing and repeat culture.

6.5 Method 3 - Random Forest approach

6.5.1 Random forest model 1: Not accounting for previously identified plausible false negative and false positive species results.

As above, the dataset used the previous split for the training and testing dataset. The trained model was also applied to a test set of 62 culture negative samples. Fitting a random forest model, as described above, without adjusting for plausibility, resulted in an 81/88 [92% (95 CI, 84-97) sensitivity] and 75/82 [95% (95 CI, 88-99) specificity] in the testing dataset (**Table 26**). The seven false negative results included the 4 MALDI-TOF misclassifications described above, and 3 other missed species (*Klebsiella pneumonia* from sample DFB_33, and *Streptococcus pyogenes* and *Staphylococcus haemolyticus* from sample DFB_143). Within the seven false positive results in the testing data, four were likely due to MALDI-TOF inaccuracy (**Table 16**) and three had been previously assessed as being plausible (**Table 24**). However, this approach failed to identify one additional polymicrobial results judged plausible in the threshold-based analysis (**Table 25**). Overall unadjusted species-level specificity in culture-negative samples was 61/62 (98%; 95% CI, 91 to 100%). The area under the ROC curve was 99% (**Figure 42**). Using the in-built feature importance ranking, the most important variables used for prediction in the random forest were the percentage of reads assigned to the species (71%), the total reads assigned (58%) while coverage breadth from mapping (15%), percentage reads mapped (15%) and mean depth of coverage (7%) had less predictive power (**Figure 43**).

Table 24. False positive results in random forest model 1 testing dataset (excluding four false positive results due to MALDI inaccuracy).

Random forest result	Sample ID	Bacteria	Reads	Percentage reads	Coverage breadth	Mean depth	Reads mapped	MALDI result	Model prediction	Plausible
False positive	DFB_16	*Paeniclostridium sordellii*	21993	24.9	93.2	14.3	97.9	absent	present	yes
False positive	DFB_5	*Escherichia coli*	71042	8.2	92.7	22.5	83.4	absent	present	yes
False positive	DFB_33	*Pseudomonas aeruginosa*	14282	15.0	94.1	5.3	91.8	absent	present	yes

Table 25. Additional infections identified in the threshold-based method but not in the random forest model 1 testing dataset.

Random forest result	Sample ID	Bacteria	Reads	Percentage reads	Coverage breadth	Mean depth	Reads mapped	MALDI result	Model prediction	Plausible
Not identified	DFB_134	*Staphylococcus epidermidis*	10366	7.1	95.3	10.2	95.6	absent	absent	yes

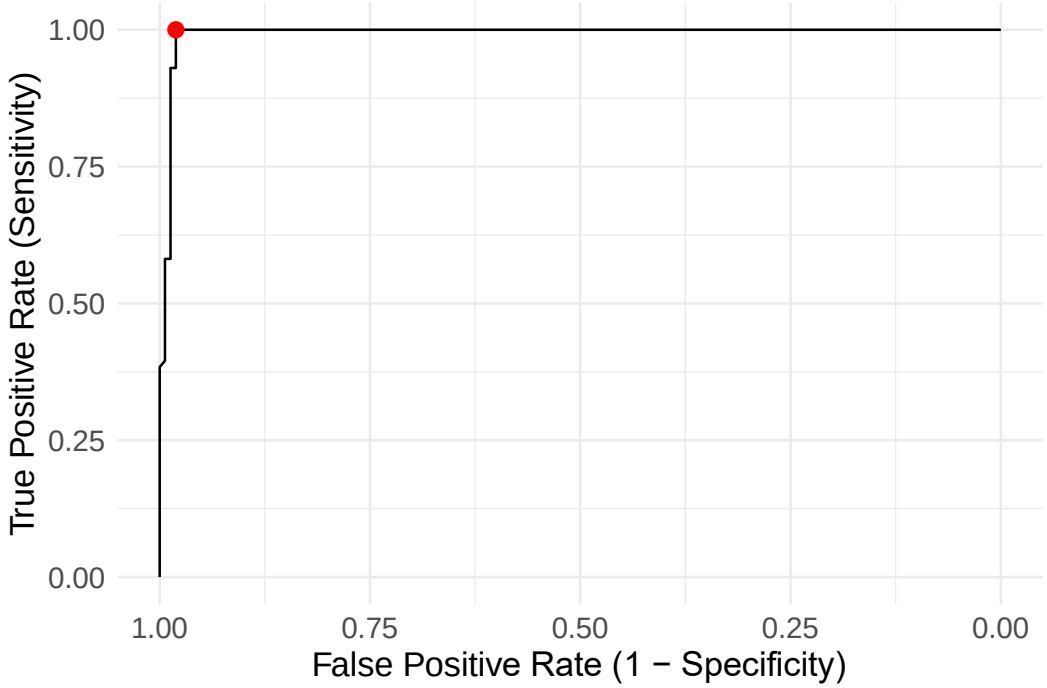

Figure 42. ROC curve of test dataset without adjusting for plausibility in both training and test datasets.

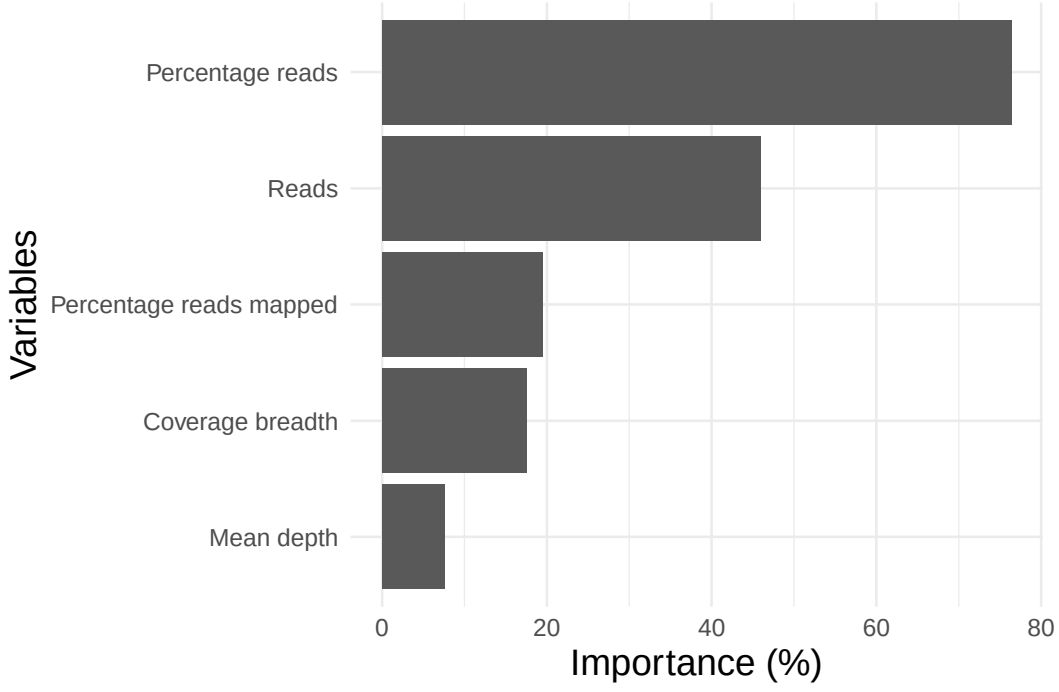

Figure 43. Random forest model 1 variable importance plot.

6.5.2 Random forest model 2: Accounting for previously identified plausible false negatives and false positives species in the training and testing dataset.

As above, the dataset (n=229 species, 216 MALDI-TOF species + 13 additional plausible species detected [205 blood culture samples]) was randomly split into 60% (n=137, 128 MALDI-TOF species + 9 additional plausible species detected [from 123 blood culture samples]) for the training dataset, and 40% (n=92, 88 MALDI-TOF species + 4 additional plausible species detected [from 82 blood culture samples]) for the testing dataset.

Adjusted performance was 100% sensitive (95 CI, 97-100, 92/92) and 100% specific (95 CI, 96-100, 82/82) in the testing dataset (**Table 26**). All additional species considered to be polymicrobial identified from the previous numerically optimised analysis that were present in the testing dataset were identified. Adjusted species-level specificity for plausible infections in culture-negative samples was 61/61 (100%; 95% CI, 94 to 100%). Similar to above, the most important variables used for prediction in the random forest were percentage reads (63%), total reads (63%) while coverage breadth (16%), percentage reads mapped (17%) and mean depth of coverage (11%) had less predictive power (**Figure 44**). In this model, more importance on predictive power was placed on total reads than in model 1.

In summary, random forest models applied here perform similarly well to numerically optimised threshold-based methods. These models improve with higher quality training datasets as seen in this study after adjusting for previously identified plausible species. An advantage of using random forest models is their robustness to overfitting, accuracy and feature important analysis, however, their lack of interpretability for individual predictions in random forest models must be considered when applied in clinical settings.

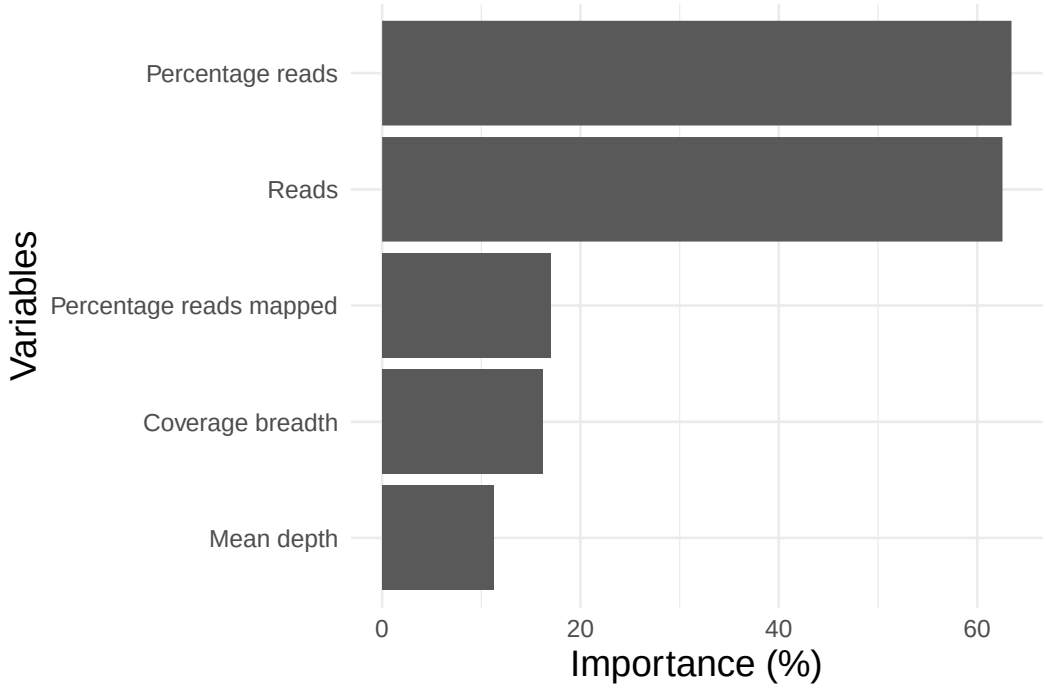

Figure 44. Random forest model 2 variable importance plot.

Table 26. Species performance results across all models.

	Training		**Testing**		**Whole dataset**		
	Culture positive		Culture positive		Culture positive		Culture negative
	Sensitivity (95% CI) [sample #]	Specificity (95% CI) [sample #]	Sensitivity (95% CI) [sample #]	Specificity (95% CI) [sample #]	Sensitivity (95% CI) [sample #]	Specificity (95% CI) [sample #]	Specificity (95% CI) [sample #]
Plasmid removal (Method 1): naive	97% (92-99) [124/128]	98% (93-99) [120/123]	92% (85-97) [81/88]	95% (88-99) [78/82]	94% (91-97 [201/216]	97% (93-99) [198/205]	98% (91-100%) [61/62]
Mapping (Method 2): naive	99% (96-100) [127/128]	96% (91-99) [118/123]	94% (87-98) [83/88]	94% (86-98) [77/82]	97% (94-99) [210/216]	94% (90-97) [193/205]	98% (91-100%) [61/62]
Mapping (Method 2): adjusted	100% (97-100) [137/137]	100% (97-100) [123/123]	100% (96-100) [92/92]	100% (97-100) [82/82]	100% (95 CI, 98-100) [229/229]	100% (95 CI, 98-100) [205/205]	100% (94-100%) [61/61]
Random forest (Method 3): naive	100% (97-100) [128/128]	100% (97-100) [123/123]	92% (84-97%) [81/88]	91% (83-96%) [75/82]	97% (93-99%) [209/216]	97% (93-99%) [198/205]	98% (91-100%) [61/62]
Random forest (Method 3): adjusted	100% (97-100) [137/137]	100% (97-100) [123/123]	100% (96-100) [92/92]	100% (96-100) [82/82]	100% (95 CI, 98-100) [229/229]	100% (95 CI, 98-100) [205/205]	100% (94-100%) [61/61]

*Specificity was calculated on a per sample basis, where any sample with >1 additional species found from sequencing was considered to contribute as 1 to the false negative count.

6.6 Samples unidentifiable by the routine lab in positive blood cultures

A total of six blood culture specimens flagged positive, suggesting metabolic activity consistent with bacteria growth, however, species identification remained elusive through routine microbiology laboratory techniques (**Table 27**). In sample 44, both conventional methodologies and sequencing failed to yield a precise species-level identification, suggesting the possibility of a true false positive blood culture flagging event. This is likely due to changes in cell metabolism as a result of the patient having AML where a substantial amount of white blood cells may have been highly metabolically active resulting in a false positive. MALDI-TOF only provided a genus-level identification in sample 91, while sequencing facilitated a higher resolution of bacterial taxonomy, identifying the organism as *Acinetobacter variabilis*. A remarkable highlight of this study was evident in samples 2, 122, 131, and 181, where the routine microbiology laboratory failed to identify the organisms present, while sequencing revealed organisms previously hidden from conventional detection methods. They were respectively identified as *Turicibacter sanguinis*, *Faecalibacterium prausnitzii*, *Corynebacterium kroppenstedtii,* and *Capnocytophaga canimorsus*. The first two species were not found the in MALDI species database, however, the latter two species were present yet still not detected. Therefore, the lack of identification of these species, besides being absent from the MALDI database, might be due to several other reasons such as the absence of growth on standard media or the stringent culture conditions required by these organisms.

Table 27. Sequencing results for fastidious unidentifiable culture positive results by the routine laboratory.

Sample	Laboratory result	Sequencing result	Clinical and microbiology notes
DFB_2	Gram-positive rods- unable to identify further.	*Turicibacter sanguinis*	Query covid.
DFB_44_0	No organism seen	No organism found	AML.
DFB_91	*Acinetobacter sp.* (Only up to Genus level)	*Acinetobacter variabilis*	Very ill.
DFB_122	No organism seen	*Faecalibacterium prausnitzii*	Query sepsis.
DFB_131	Failure to thrive. Gram-positive rods- unable to identify further.	*Corynebacterium kroppenstedtii*	-
DFB_181	No organism seen	*Capnocytophaga canimorsus*	This patient presented with evidence of bite marks and a swollen left hand.

6.7 Sequencing depth and coverage

In monomicrobial samples, the median sequencing depth and coverage was 69x (IQR, 29-147x) [Range, 1.2-1087x] and 93% (88-95%) [42-100%] respectively, and in polymicrobial samples (including newly identified pathogens in this analysis) the median sequencing depth and coverage was 25x (7-113x) [0.4-298x] and 90% (84-95%) [32-100%].

6.8 Controls

6.8.1 Negative control: No spike

In all negative control samples with no spike added, total reads did not exceed 125. **Figure 45** shows these reads prior to applying any filtering threshold.

6.8.2 Negative control: *Thermus thermophilus* spike

When *Thermus thermophilus* spikes was added to negative controls in each sample, >96% of the classified reads were from this species, except for sample DFB_R_22 which had a low read number of 150 (**Figure 46**). After applying filtering thresholds (reads>50, percent reads>0.4%), all samples were classified as having >98% of the spike added, while filtering by reads>150 and percent reads>0.4% resulted in 100% of the spiked in species in all samples. The negative control from batch 1 (DFB_R_1) has been removed in the figures, as lambda DNA was added as the spike instead of *T. thermophilus* in this initial run, before the possibility of mis-identification of lambda DNA as *E. coli* was identified.

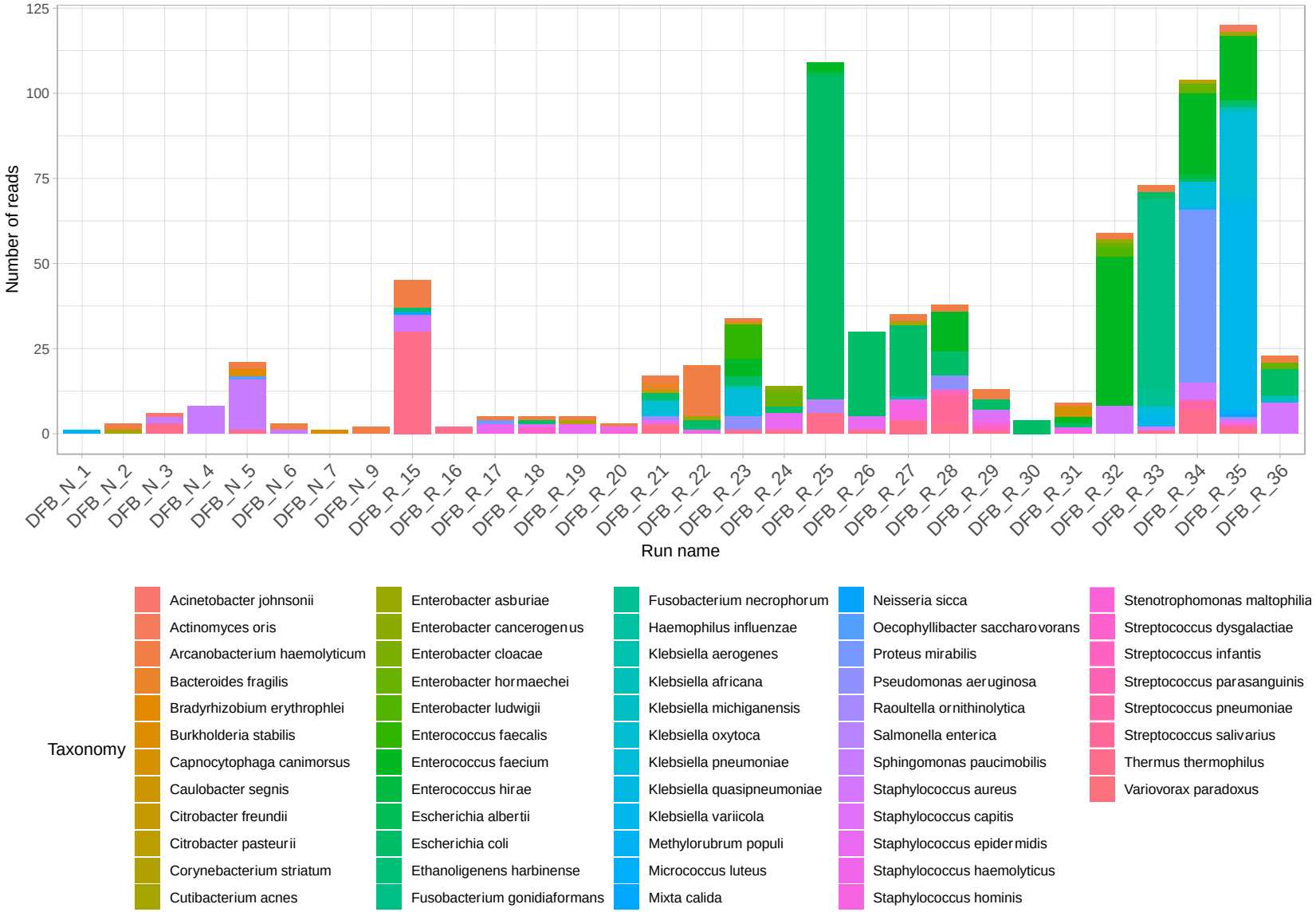

Figure 45. Pre-filtered negative control with no spike (no filter thresholds applied).

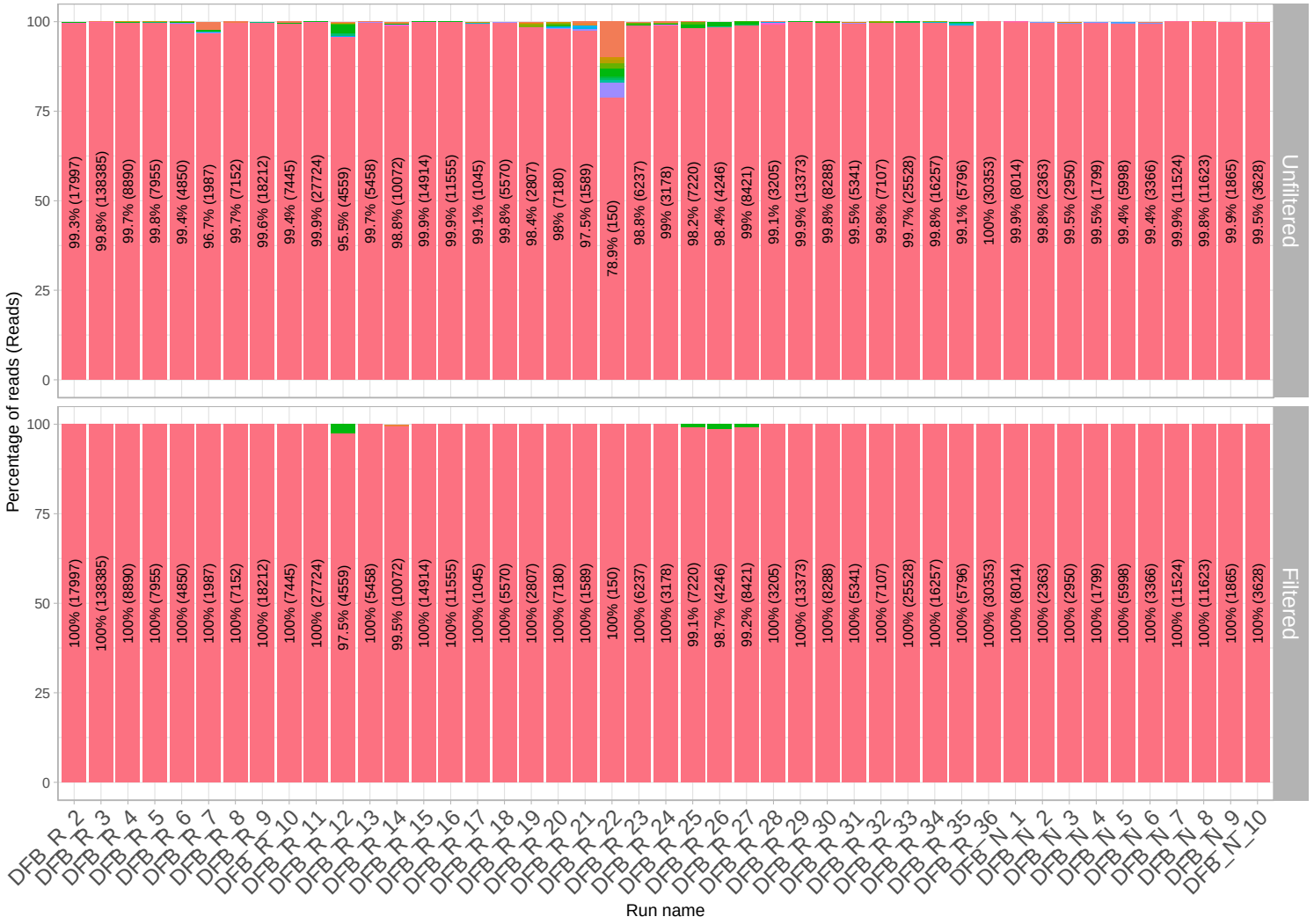

Figure 46. Unfiltered and filtered spiked negative control. Red bars correspond to *Thermus thermophilus.* The percentage of internal control is shown with absolute read numbers in brackets.

6.8.3 Negative control: Increasing *Thermus thermophilus* spike levels

In samples 15 and 16, an additional negative control of *Thermus thermophilus* spiked at 20% of the individual sample DNA concentration was added. This was to assess the effect of increasing levels of spiked *T. thermophilus* concentration on contamination. Prior to any filtering, increasing the concentration of *T. thermophilus* DNA from 0% to 0.5% and to 20% improved contamination control and resulted in 66.7%, 99.87% and 99.92% and 100%, 99.89% and 99.97% of reads classified as *T. thermophilus* in both samples respectively (**Figure 47**). Although *T. thermophilus* reads were not expected to be found in samples with a 0% spike, the low numbers of reads may be explained by potential barcode or laboratory cross-contamination, however, these contaminating reads of 2 and 30 respectively are well below filtering thresholds set for optimal performance.



Figure 47. Increasing levels of spiked *Thermus thermophilus (Tth)* concentration improves contamination control. Percentage of spike shown with absolute read numbers shown in brackets.

6.8.4 Positive control

Positive controls which consisted of *Arcanobacterium haemolyticum* were present in every run. Unfiltered positive controls had >99.8% of reads across all samples classified as *A. haemolyticum*. After applying filtering thresholds (percent>0.4, reads>1000, coverage_breadth>32, reads_mapped>75, mean_depth>0.4), all positive controls had 100% of reads classified to *A. haemolyticum* in all samples (**Figure 48**).

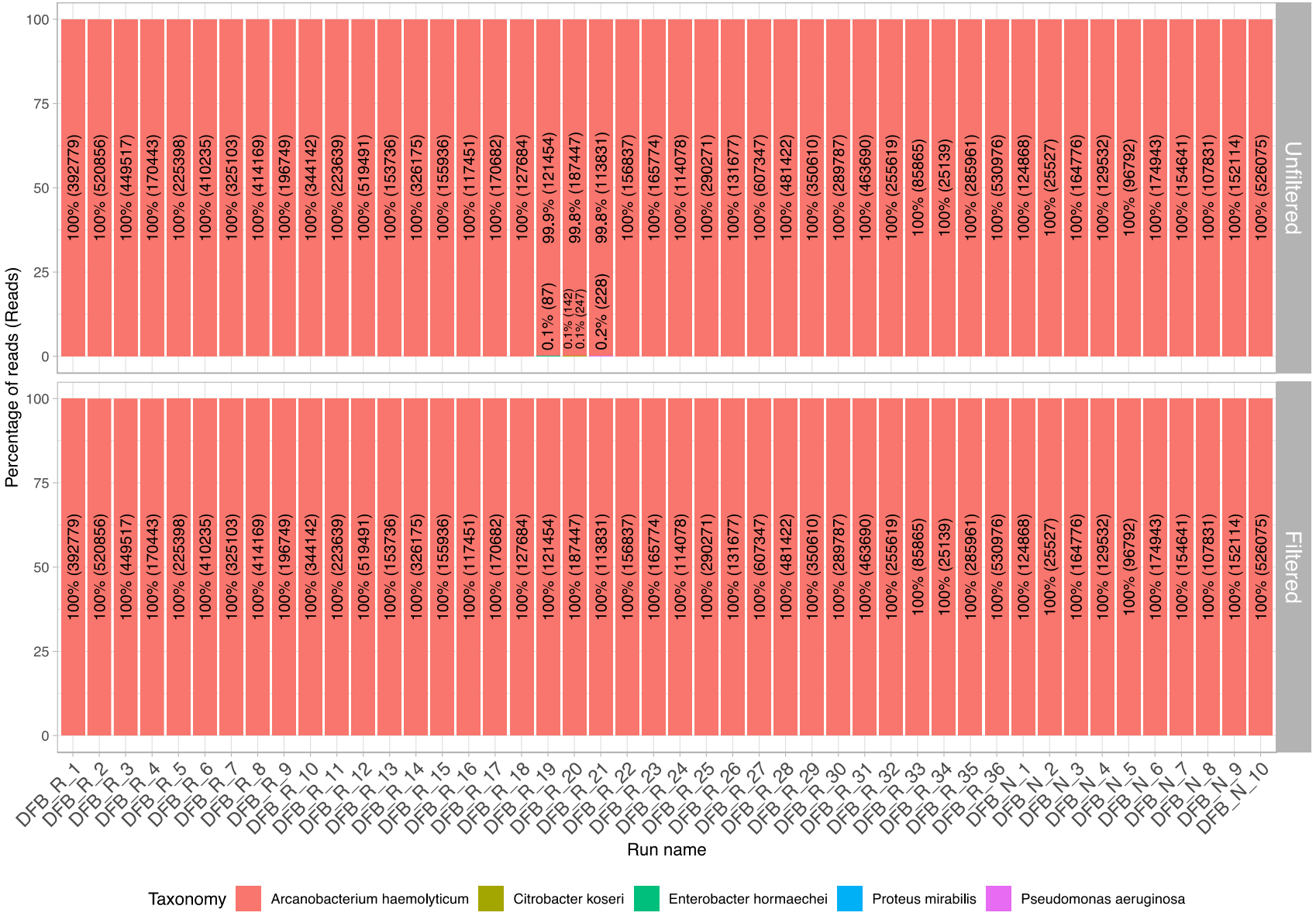

Figure 48. Unfiltered and filtered positive controls. The percentage of positive control is shown with absolute read numbers in brackets.

6.8.5 Internal spike control present in positive controls

The internal spike of *Thermus thermophilus* was present in all positive controls. The median percent of *T. thermophilus* reads was 2.4% (1.7-4) [0.9-14] in positive control samples (**Figure 49**). The internal control was spiked at 2% of total DNA after DNA extraction using the Qubit 2.0 fluorometer (broad range kit; Thermo Fisher Scientific, USA). In cases where the total positive control DNA amount was less than 25ng, a fixed amount of 0.5ng *T. thermophilus* DNA was added as the internal control, as lower amounts of *T. thermophilus* DNA become difficult to reliably quantify. Therefore, in these select samples *T. thermophilus* would make up >2% of total reads.

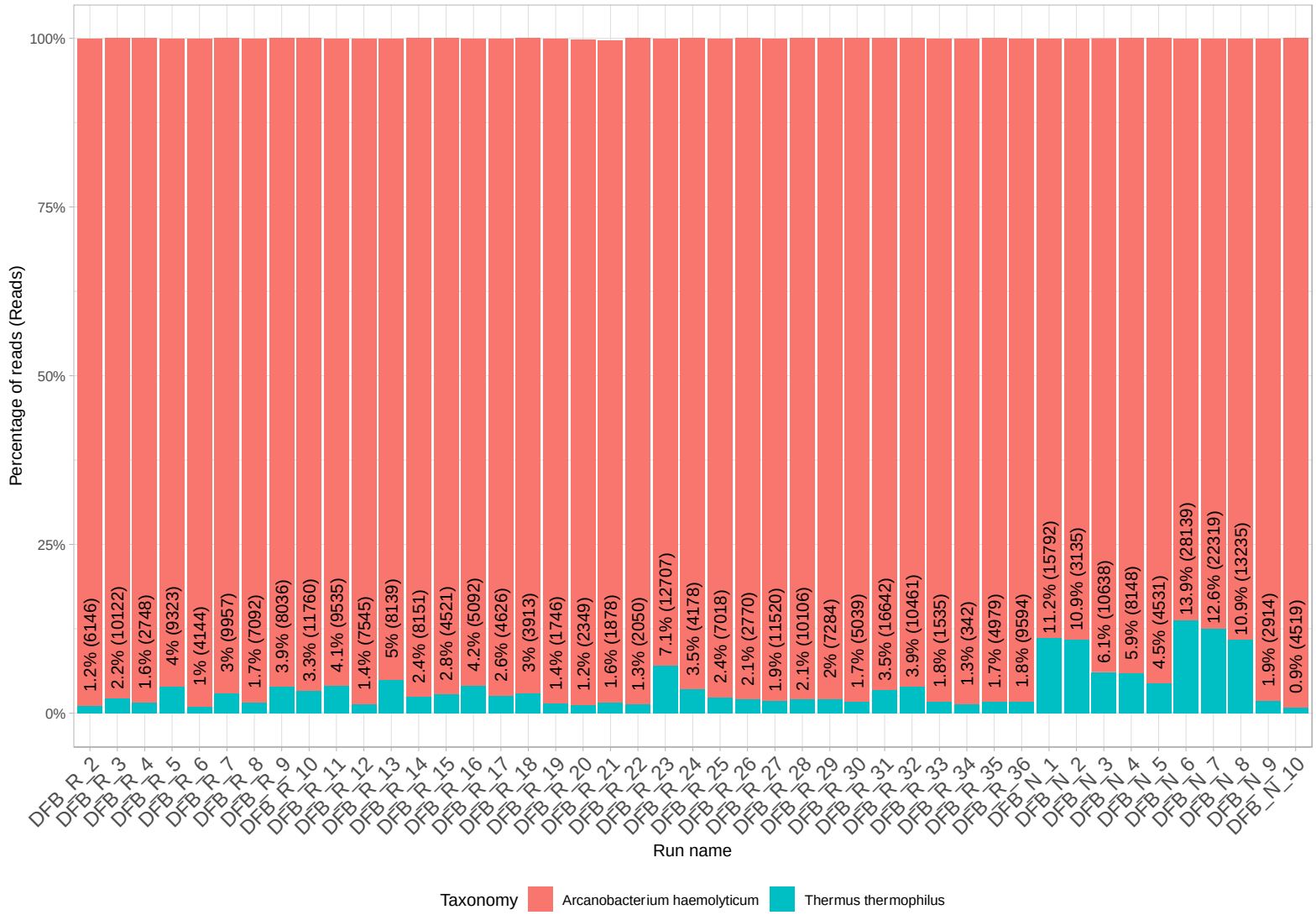

Figure 49. Unfiltered and filtered positive controls. The percentage of positive control is shown with absolute read numbers in brackets.

6.8.6 Internal spike control present in culture positive and culture negative clinical samples

The internal spike of *Thermus thermophilus* was present in all culture positive and culture negative clinical samples. The median percentage of *T. thermophilus* reads was 3.6% (2-8) [0.7-58] in culture positive clinical samples (**Figure 50**), and 100% (100-100) [4.5-100] in culture negative samples (**Figure 51**). As in the positive controls, the internal control was spiked at 2% of total DNA after DNA extraction, except where the total sample DNA amount was less than 50ng or below the limit of detection (i.e., 2ng/mL), when a fixed amount of 0.5ng *T. thermophilus* DNA was added as the internal control. This might explain the range in percentage reads of *T. thermophilus*. For instance, sample 45 which contained 58% *T. thermophilus* had the lowest DNA concentration in the run of 28 ng/mL while sample 49 had the highest DNA concentration of 642 ng/mL in the same run which after normalisation contained 3.6% of *T. thermophilus*. Clinical negative cultures contained generally 100% of the internal control *T. thermophilus*, except for sample N47 which found to contain *Pseudomonas aeruginosa* by sequencing and re-culture.



Figure 50. Internal spike control shown in culture-positive clinical samples. The percentage of internal control is shown with absolute read numbers in brackets.

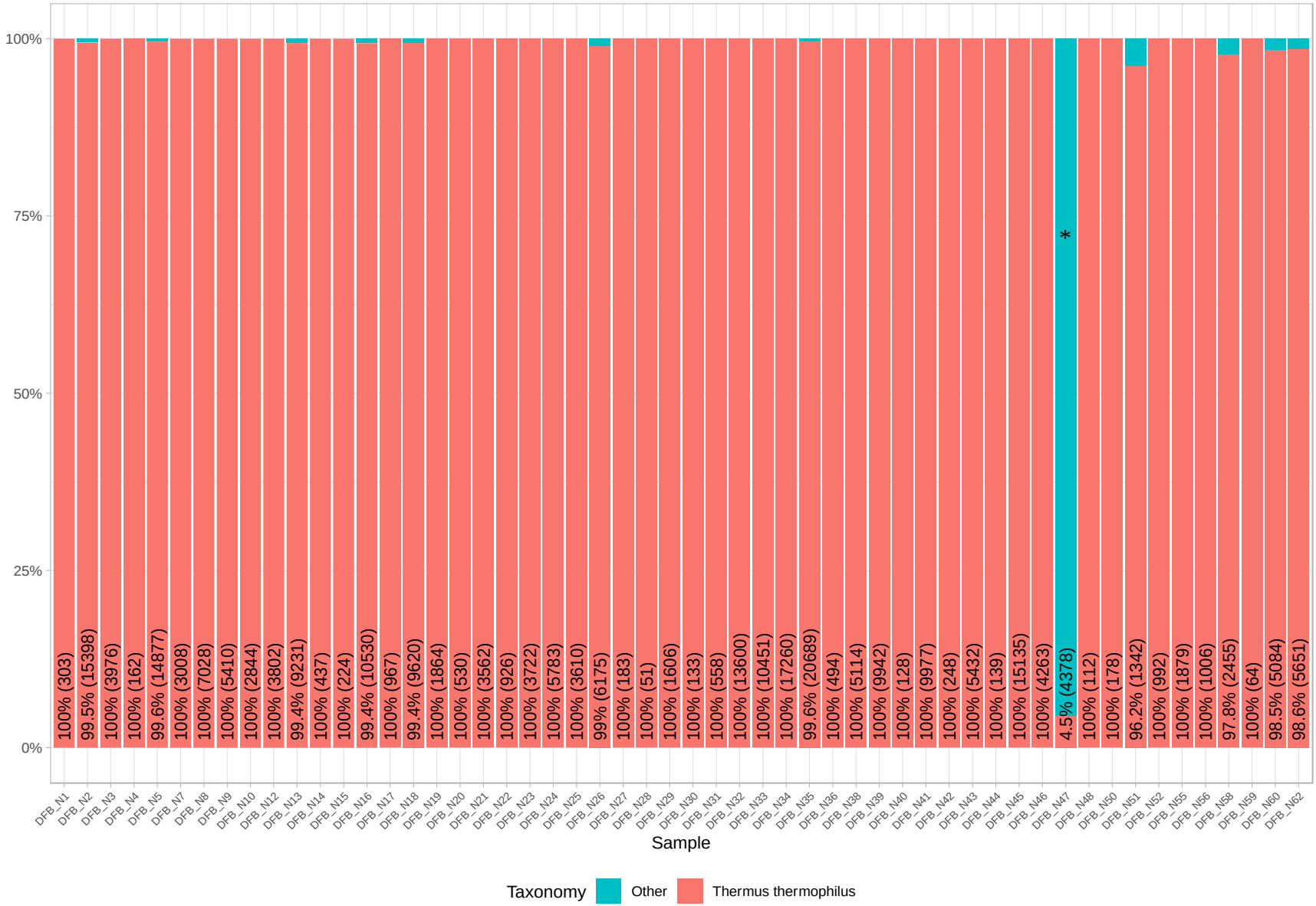

Figure 51. Internal spike control shown in culture-negative clinical samples. The percentage of internal control is shown with absolute read numbers in brackets. * classified and re-cultured as *Pseudomonas aeruginosa*.

6.9 Limit of detection

Although this study did not explicitly test the limit of detection, the sample with the lowest concentration of culturable bacteria had 10^4 CFUs/mL (Sample 98).

6.10 Barcode contamination

To allow for sequencing multiple samples in the same sequencing run, short recognisable sequences (barcodes) are added to DNA fragments during library preparation to tag samples separately. These barcodes must be read and used to separate out the resulting sequencing data. However within the multiplex sequencing data, reads were assigned to barcodes not included in the experiments (defined as mis-assigned reads), and reads were assigned to the incorrect barcode bins in-use (referred to as cross-assigned).[128] In this study, I assessed both mis-assigned reads by Guppy demultiplexer (previously described in methods) and cross-assigned reads using the positive control *Arcanobacterium haemolyticum* as a proxy. *A. haemolyticum* was used as this was added only to a single positive control sample per run sequenced and this species has not been present in any of the MALDI-TOF results from this dataset.

6.10.1 Mis-assigned reads

If there are 12 barcodes available, and 11 are in use with barcode number 3 not in use, $number(barcode\ 3\ reads) = \big(1 - p(t)\big) \times \left[\frac{1}{11}\right] \times z$, where p(t): probability of a read being classified to the correct barcode and z: total number of reads. The term 1/11 reflects that when mis-assigned there are 11 possible barcodes that a read can be mis-assigned to. Therefore,

$$p(t) = 1 - [\frac{number(barcode\ 3\ reads)}{z} \times 11\ ].$$

Now assuming there are more barcodes not in use (a1, a2, …). The equation then becomes:

$$p(t) = 1 - [\frac{number(bar.\ a1\ reads)}{z} + \frac{number(bar.\ a2\ reads)}{z} + \ldots] \times \frac{(number\ of\ total\ barcodes\ in\ use - 1)}{(number\ of\ barcodes\ not\ in\ use)}.$$

The median probability across all runs for a read to be mis-assigned was =0.000054 (IQR, 0.000008-0.000129) [Range, 0-0.000379] for unfiltered reads, and =0.000050 (0.000008-0.000119) [0- 0.000372] for filtered reads. Therefore, I estimate that for every 18,519 unfiltered and 20,000 filtered reads, one read would be misassigned, but this varied between runs with up to 1 in 2639 reads mis-

assigned in the run with the greatest mis-assignment and up to no mis-assignment of other runs (**Figure 52**). The reasons for this variation in misassignment rates between runs are not entirely clear. However, it could be due to a number of factors. For instance, the quality and purity of the DNA used can significantly influence the results, with degraded or contaminated input material potentially leading to higher misassignment rates. The sequencing depth, or the number of times a particular region is sequenced, can also impact the accuracy of read assignment. Additionally, variations in library preparation methods can introduce bias and affect the accuracy of read assignment. The inherent error rate of Oxford Nanopore technology can also contribute to misassignment of reads.

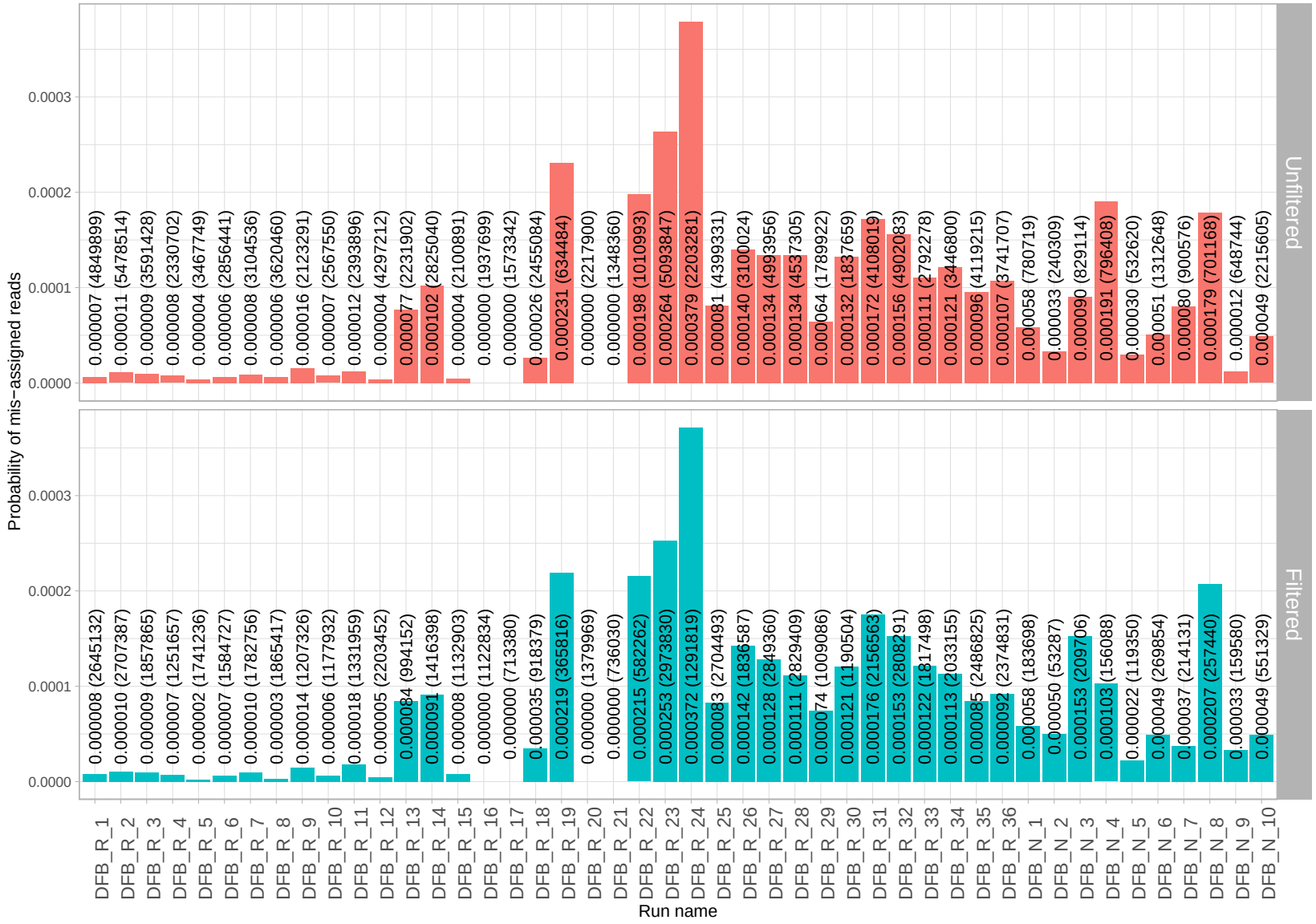

Figure 52. Probability of reads mis-assigned to other barcodes per run. The mean probability of mis-assigned reads is shown with the absolute number of reads in brackets.

6.10.2 Cross-assigned reads

The probability of reads cross-assigned was calculated by the number of *Arcanobacterium haemolyticum* reads not in the positive control barcode, but assigned to a barcode in use, divided by the total *Arcanobacterium haemolyticum* reads present in a run assigned to any barcode in use. The median (IQR) [range] probability was 0.000028 (0.000019-0.000039) [0.000009-0.000199] (**Figure 53**), while in culture negative samples it was 0.000021 (0.000010-0.000026) [0.000008-0.000052] (**Figure 54**). Therefore, for every 35,714 correctly assigned reads in positive samples, a median 1 read was cross-assigned, while for negative samples for every 47,619 correctly assigned reads, 1 read was cross-assigned.

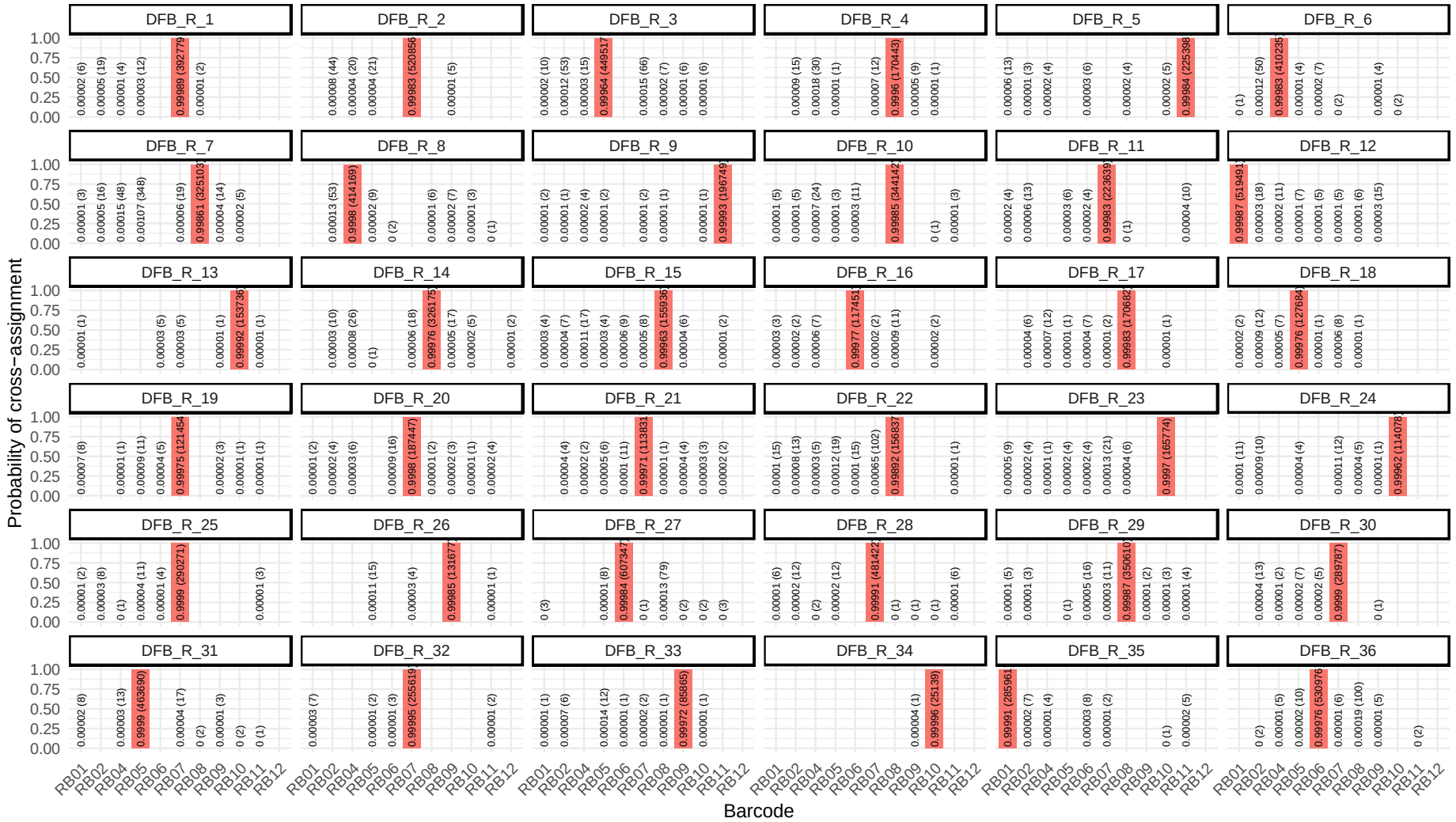

Figure 53. Probability of cross-assignment in culture positive samples (estimated using *Arcanobacterium haemolyticum*). The largest red bar in every run corresponds to the correct positive control sample. The mean probability of cross-assigned reads is shown with the absolute number of reads in brackets.

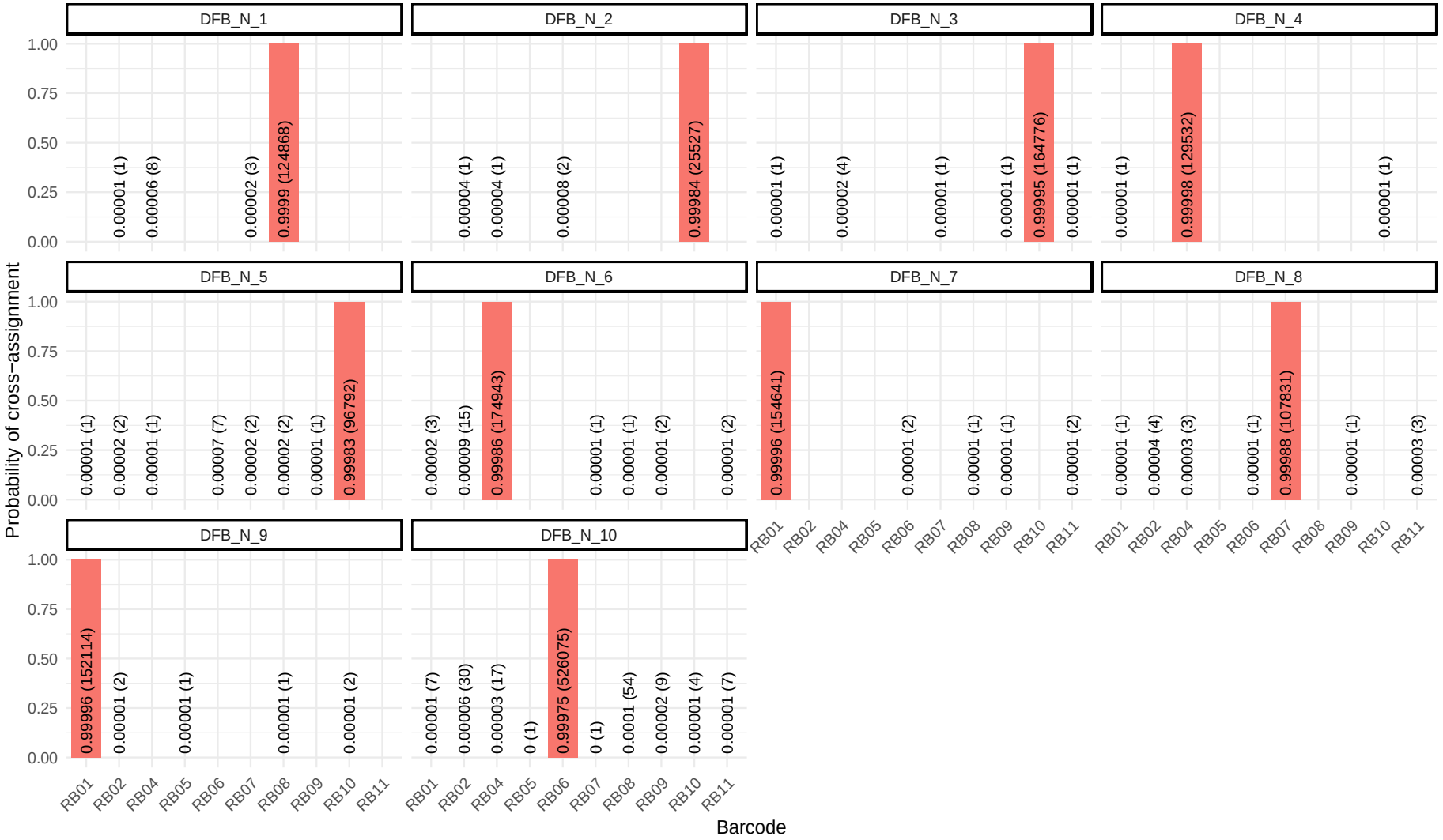

Figure 54. Probability of cross-assignment in culture negative samples (estimated using *Arcanobacterium haemolyticum*). The largest red bar in every run corresponds to the correct positive control sample. The mean probability of cross-assigned reads is shown with the absolute number of reads in brackets.

6.10.3 Cross-assigned contamination by off-target barcode

To explore the impact of specific Nanopore barcodes on the proportion of off-target sequencing reads sequencing data was categorized according to the barcode used, followed by the corresponding mean off-target read proportions. The mean proportions of off-target reads varied significantly across barcodes, with the lowest mean proportion at 0.000013 observed for barcode 01 and the highest at 0.000062 for barcode 08 (**Figure 55**).

A linear regression analysis was conducted to ascertain the association between the mean off-target read proportions and the specific barcode used. All barcodes, beside barcode 08, demonstrated negligible associations with the mean off-target read proportion ($p>0.05$ when compared to barcode 01). In contrast, barcode 08 exhibited a significant positive association with the mean off-target read proportion (p-value= 0.004).

Certain barcodes, such as barcode number 8, may be more susceptible to misreading than others, yet they might still be considered within acceptable limits by Oxford Nanopore Technologies (ONT). Sequencing or alignment errors could potentially misinterpret parts of other barcodes as being barcode number 8, thereby inflating its read count. Alternatively, if barcode number 8 possesses specific characteristics that make it more prone to sequencing or alignment errors, this could also lead to a higher rate of misassignment. These observations underscore the significance of

meticulous barcode design and selection to minimize sequence overlap to ensure precise read assignment. These results highlights the distinctive role of barcode 08 in augmenting the proportion of off-target sequencing reads, which necessitates further research into the design and selection of barcodes. **Figure 56** presents a further detailed graphical representation of the mean probability of a read being misattributed across off-target barcodes, juxtaposed with the absolute mean reads that are attributable to each barcode.

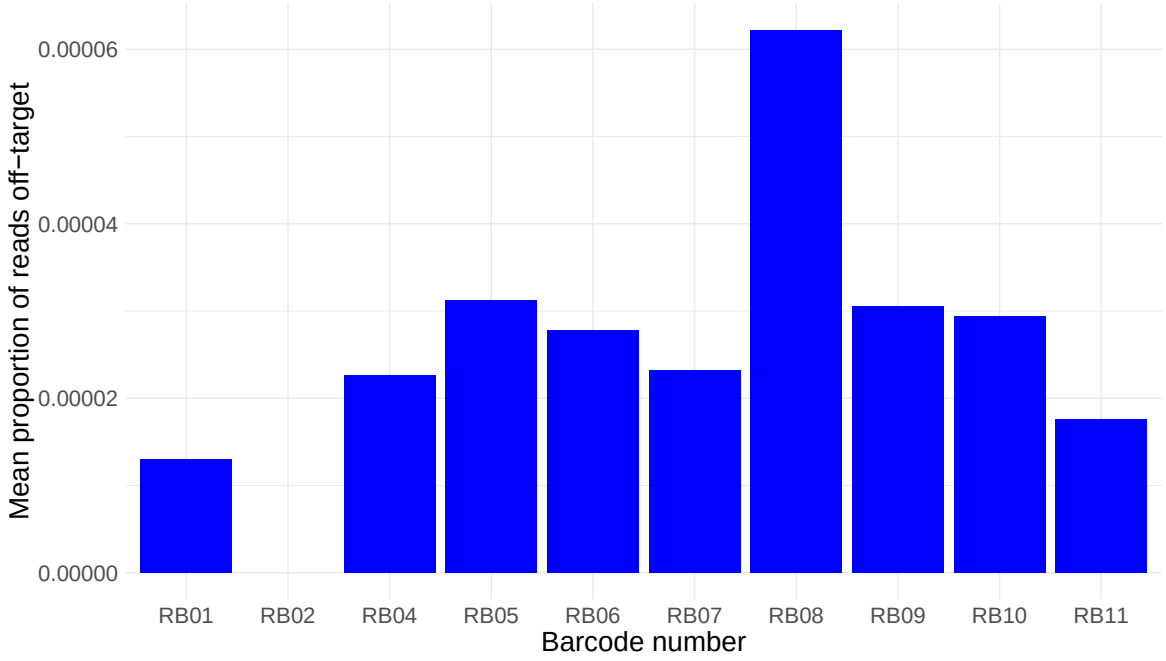

Figure 55. Mean proportion of off target reads by barcode.



Figure 56. Cross-assigned *Arcanobacterium haemolyticum* reads shown across off-target barcodes. The mean probability of cross-assigned reads is shown with the absolute number of reads in brackets.

6.10.4 Probability of assigning a correct read

The probability of reads correctly assigned was defined as 1 minus the reads misassigned and the reads cross-assigned: . Therefore, the probability of (unfiltered) reads being correctly assigned in the study was 0.9999215 (1-0.000054-0.0000245). Therefore, for every 12,739 correctly assigned reads, 1 read would be incorrectly assigned.

These calculations assume that all barcodes have an equal chance of being misassigned. If this is not the case (for example, if some barcodes are more prone to misreads due to sequencing errors), these calculations might need to be adjusted. Cross-assignment probability could be influenced by various factors such as barcode similarities or sequencing errors, which are not considered in this calculation. The study assumes that *A. haemolyticum* is a suitable proxy for all species. If this species behaves differently than others in terms of barcode assignment, this might impact these results, however as the barcode sequence is separate to the sample sequence and designed not to be naturally occurring, I expect my findings to be generalisable.

6.11 Time to speciation (Routine laboratory vs metagenomic approach)

6.11.1 Time to speciation by the routine laboratory

From positive blood culture, the routine laboratory took a median time of 10hr 33min (IQR, 8hr 45min-15hr 49min) [Range, 2hr 33min-13hr 50min] to the first MALDI species result.

Additional time between sampling and results was accounted for by the median time from sample collection to the laboratory receiving the sample of 106 min (IQR, 66-186 min) [Range, 28-735 min] and the median time for blood culture incubation until positivity was 17hrs (IQR, 11-26hrs) [Range, 3-122hrs].

6.11.2 Time to speciation by the metagenomic pipeline

The time from a positive blood culture to the start of sequencing had a median duration of 3hr20min (IQR, 3hr7min-3hr27min) [Range, 2hr45min-3hr48min]. Specifically, the median time from blood culture extraction to eluted DNA was 1hr 35min (IQR, 1hr30min-1hr38min) [Range, 1hr20min-1hr47min]. The process of DNA quantification, DNA normalization, library preparation to the start of sequencing had a median duration of 1hr46min (IQR, 1hr37min-1hr49min) [Range, 1hr19min-2hr3min].

Using the threshold filtering method derived from the training dataset after sequencing end (24 hours from sequencing start) and after adjustment for 'false positives' resulted in a sensitivity and specificity of 100% (95 CI, 98-100) [229/229] and 100% (95 CI, 98-100) [205/205] respectively in the complete dataset. Applying the same derived thresholds to earlier sequencing time intervals, for instance, at 10 minutes after sequencing start resulted in a sensitivity and specificity of 78% (95 CI, 72-83) [178/229] and 100% (95 CI, 98-100) [205/205] respectively, while sensitivity improved to sequencing end (**Figure 57**). Filtering thresholds (with mapping data) were re-trained according to time interval training datasets that resulted in further optimised thresholds (**Table 28**). Thereafter the metagenomics pipeline achieved a sensitivity and specificity of 98% (95 CI, 95-99) [224/229] and 100% (95 CI, 98-100) [205/205] respectively within a sequencing duration of 10 minutes using the dynamic threshold filtering method after adjustment for 'false positives'. This improved to a near perfect score of 100% (95 CI, 98-100) [229/229] and 100% (95 CI, 97-100) [204/205] for both sensitivity and specificity respectively after 120 minutes of sequencing, even in the presence of polymicrobial infections with a low read abundance of 2% (**Figure 58**).

In conclusion, the metagenomics pipeline exhibited a sensitivity and specificity of 98% and 100% respectively after 3hr 30min from the point a blood culture flagged positive, and a near perfect score of 100% for both after 4hr 40min. Remarkably, the metagenomics pipeline was significantly faster, taking nearly half the time compared to the routine laboratory, which had a median time of 10hr 33min to identify species.

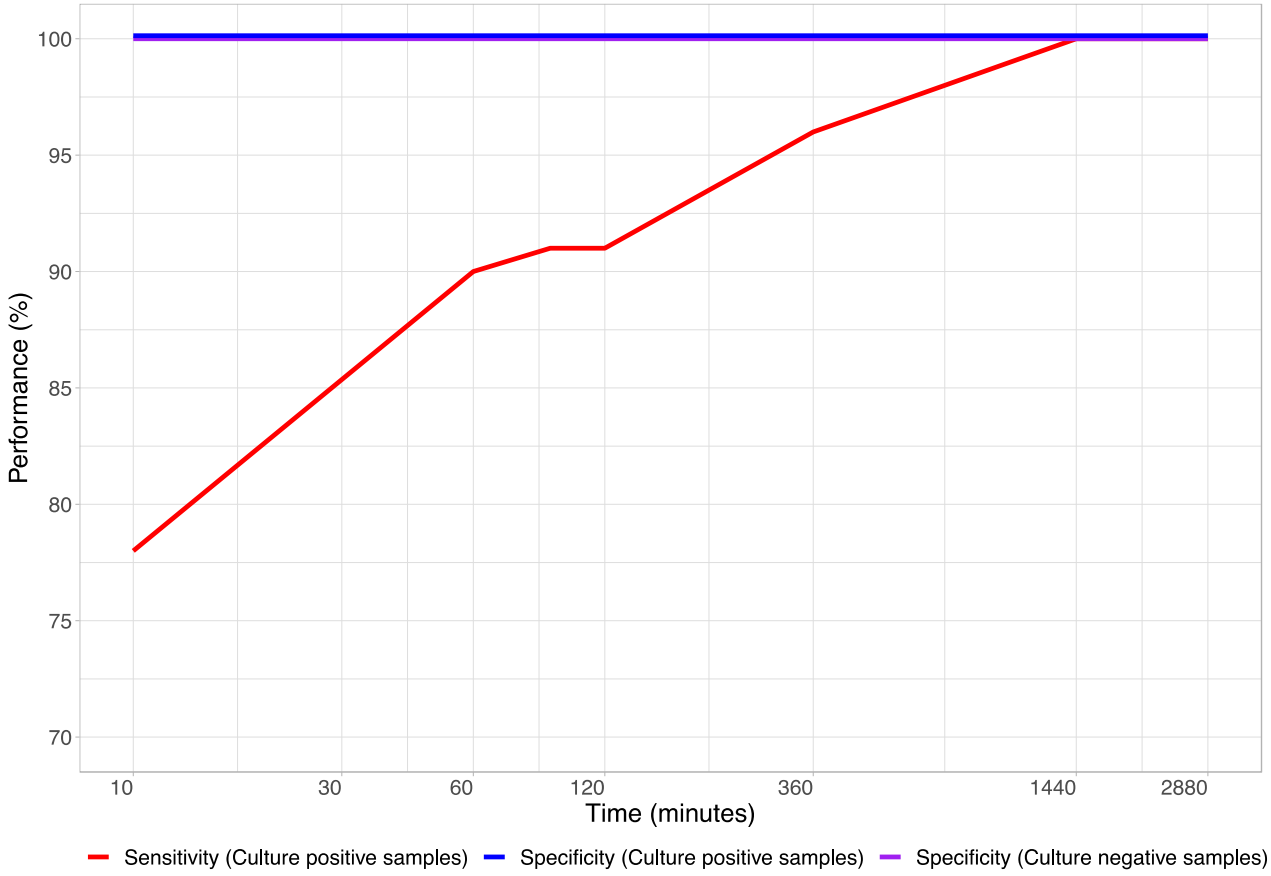

Figure 57. Species performance over sequencing time (filtering thresholds based on training dataset at the end of sequencing).

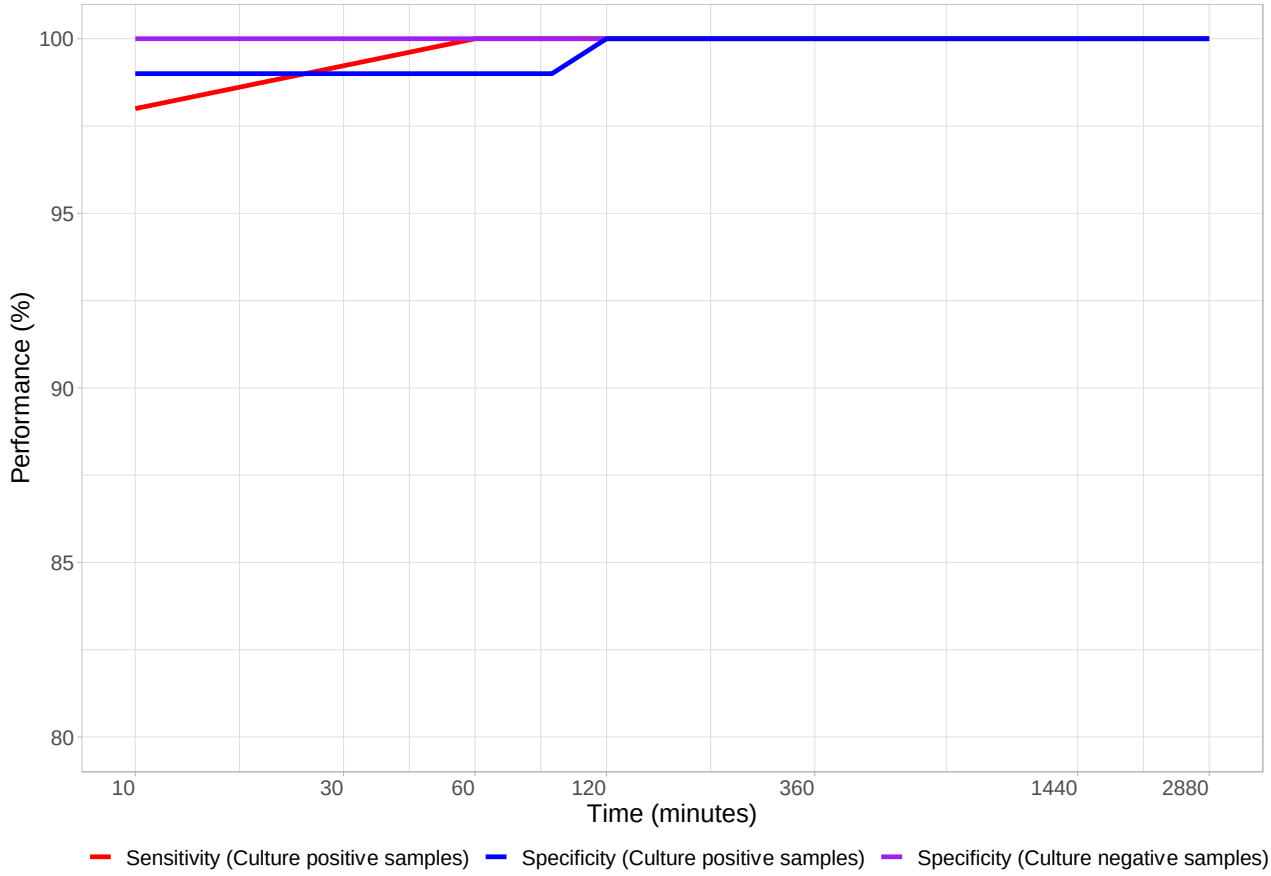

Figure 58. Species performance over sequencing time (dynamic filtering thresholds based on training dataset at given time points from sequencing start).

Table 28. Dynamic filtering thresholds derived from the training dataset at given intervals from sequencing start.

Minutes from sequencing start	Coverage breadth threshold	Mean depth threshold	% reads mapped threshold
10	>2	>0.02	>75
60	>3.4	>0.03	>75
90	>4.3	>0.045	>75
120	>5.2	>0.055	>75
360	>12	>0.13	>75
1440	>27.8	>0.33	>75
2880	>32.7	>0.41	>75

6.12 CFU counts, DNA extraction yield and time to positivity.

6.12.1 Does positive blood culture bottle weight (a proxy for inoculated blood volume) affect CFU count?

Aerobic positive blood culture bottles after inoculation had a median weight of 65 grams (IQR, 63-69) [Range, 57-78] while anaerobic culture bottles had a median weight of 71 grams (68-75) [65-78]. To ensure that new blood culture bottles did not have differing weight, ten new aerobic and anaerobic blood culture bottles each were sampled for weight prior to use. For the aerobic bottles the median weight was 60.61 grams (IQR, 60.45-60.76) [Range, 60.12-60.83] and for the anaerobic bottles the median weight was 65.39 grams (IQR, 65.26-65.46) [Range, 64.89-65.83]. This implies the median volume of blood inoculated in aerobic bottles is 4.4 ml (assuming a density of ~ 1g / ml) and in anaerobic bottles is 5.6 ml. A limitation is that weight was only measured after the routine lab aliquoted an amount for routine culture, which may lead to an underestimation of the median volume of blood reported here. In **Figure 59**, we see a lack of association between blood culture bottle weight and final CFU count trends in aerobic and anaerobic culture bottles (R^2=0.016, p=0.17 and R^2=0.013, p=0.36 respectively).

In the BD BACTEC™ Automated Blood Culture System, microbial growth is detected through a CO_2-reactive dye in a sensor at the culture bottle's base. Microbes' metabolic CO_2 output interacts with this dye, changing the light absorption of a sensor-based fluorescent material, thereby indicating growth.

Initial blood volume in culture bottles can affect starting CFU counts, given a uniform bacterial distribution in the patient's blood. However, once the culture is incubated and bacterial growth begins, the BD BACTEC™ System flags a bottle as positive typically during the exponential growth phase, when bacteria are replicating most rapidly. The rate of bacterial growth during this phase is not dependent on the initial volume of blood or the initial CFU count, but rather on the bacterial species growth characteristic.

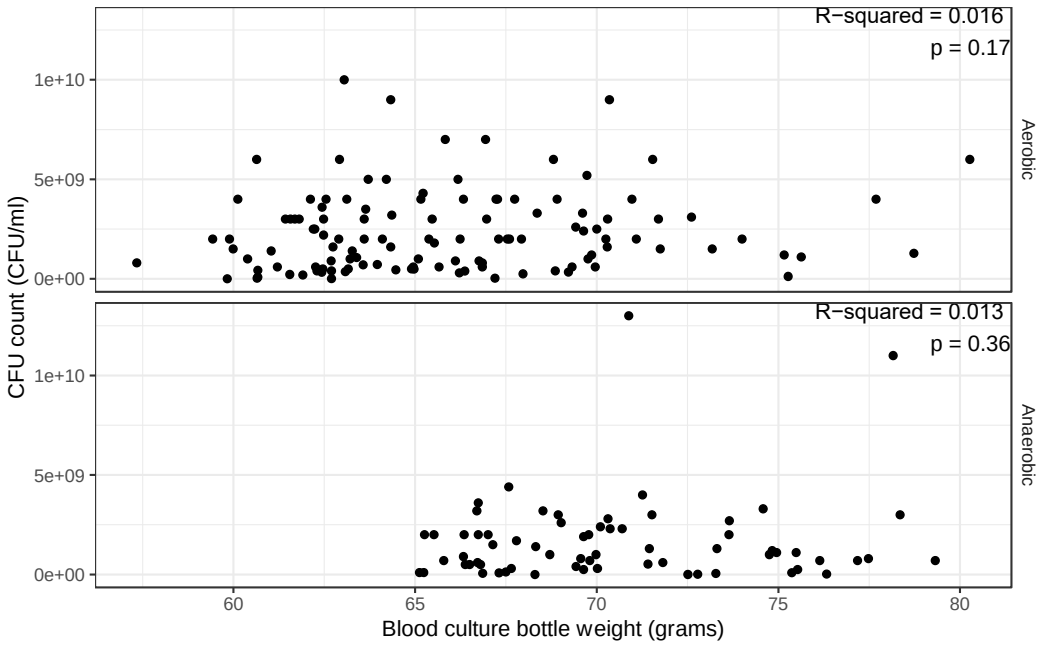

Figure 59. Scatter plot of blood culture bottle weight vs. CFU count.

6.12.2 Does blood culture bottle weight (a proxy for inoculated blood volume) affect incubation time, that is, blood culture time to positivity?

The lack of a relationship between blood culture bottle weight and time to positivity, as observed in **Figure 60**, is somewhat unexpected given the previous literature on higher inoculation volume being associated with a greater probability of positive blood cultures and a faster time to positivity.[182] In an analysis stratified by culture bottle type for the two predominant isolates, *Escherichia coli* demonstrated no correlation between culture weight and time to positivity (**Figure 61**). However, a relationship was observed for *Staphylococcus epidermidis* (**Figure 61**). Specifically, in aerobic conditions, an inverse relationship between culture weight and time to positivity was noted as expected (R^2=0.49, p=0.001). However, unexpectedly, under anaerobic conditions, a direct relationship was observed, indicating contrasting growth dynamics in response to oxygen availability (R^2=0.6, p=0.001). Several factors could contribute to the observed discordance. Firstly, the growth characteristics and metabolic activity of the different species (or even specific strain types within species) of microorganisms may have a more significant impact than bottle weight on the duration of incubation. In this study we include a large range of bacterial species, including contaminants, whereas other studies have typically focused on pathogens. Secondly, variations in composition of the culture medium between batches and changes in ambient temperature during the course of the study may have added noise to the growth rates. Additionally, the presence of inhibitory substances, such as antibiotics, could have varied over time.

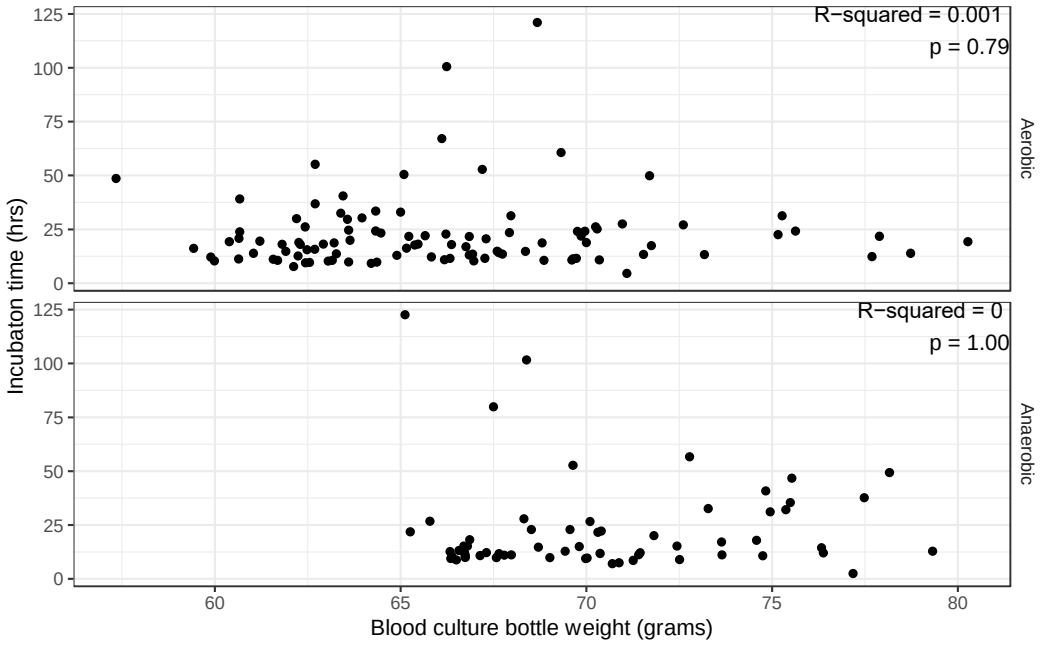

Figure 60. Scatter plot of blood culture incubation time vs. blood culture bottle weight.

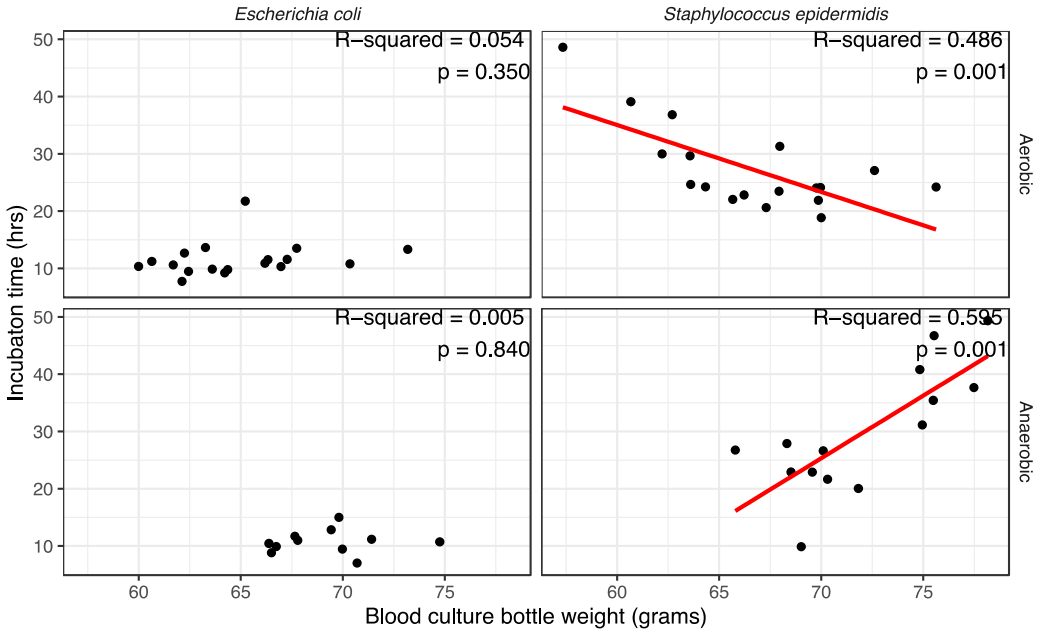

Figure 61. Scatter plot of blood culture incubation time vs. blood culture bottle weight for *E. coli* and *S. epidermidis*.

6.12.3 Is CFU count associated with blood culture time to positivity?

In both aerobic and anaerobic culture bottles, the analysis reveals a negative association between CFU count and blood culture time to positivity (R^2=0.158, p<0.001 and R^2=0.079, p=0.038 respectively; **Figure 62**). The finding that a high CFU count results in a shorter blood culture time to positivity is not unexpected, as the faster the rate of growth the more bacteria there are likely to be present, once a given cumulative amount of metabolic activity has occurred. It is also likely that faster growth means fewer bacteria will have died at the time the blood culture flags positive, also contributing to higher CFU counts. This finding may be relevant for diagnosing and monitoring infections, as a higher CFU count could potentially indicate a more severe or rapidly progressing infection, if any impact of the initial blood culture volume is accounted for.

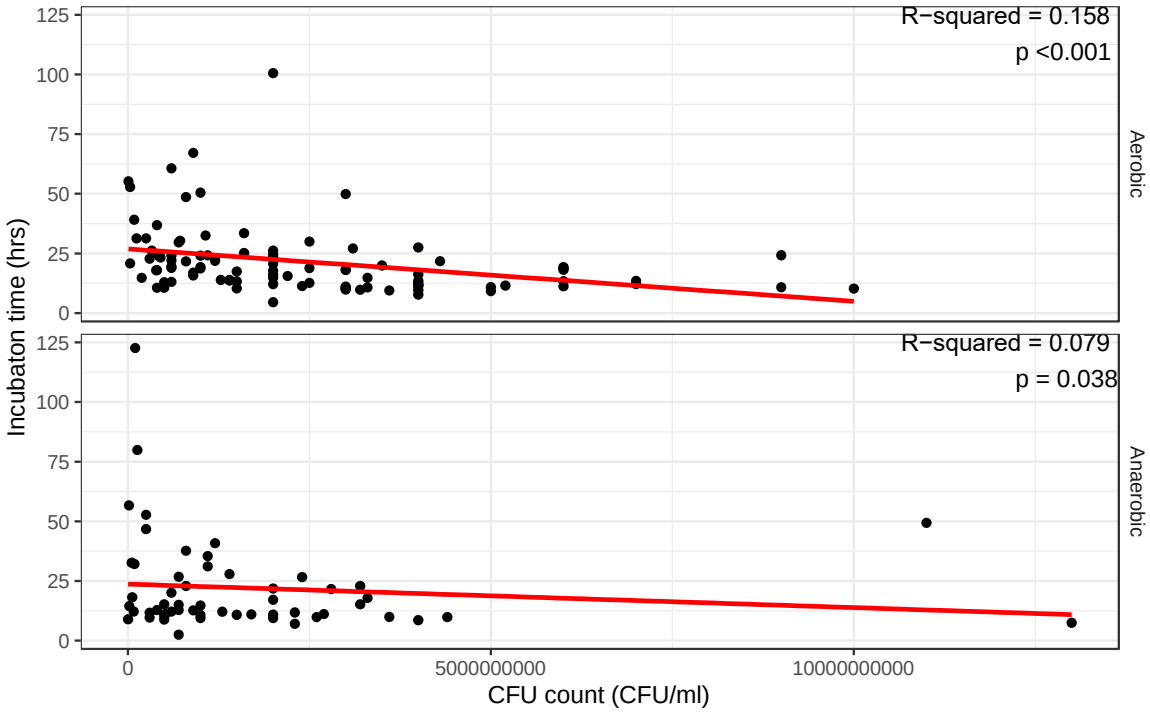

Figure 62. Scatter plot of CFU count vs. time to positive blood culture.

6.12.4 Does CFU count relate to DNA extraction yield?

Higher CFU counts were associated with greater DNA extraction yields in aerobic cultures (**Figure 63**, R^2=0.14, p<0.001), however, did not have an association in anaerobic cultures (R^2=0.004, p=0.62). This indicates an expected link between microbial abundance and the amount of DNA obtained from the sample, but surprisingly the extent of the DNA yields explained by CFU counts is still relatively low, for example, because of variations between bacterial species in the efficiency of DNA extraction. However, the association still has practical implications. The DNA extraction yield is a critical factor in many molecular biology applications,

such as genetic analysis or diagnostic testing. Understanding the relationship between CFU count and DNA extraction yield can help researchers and practitioners optimize protocols and interpret results accurately. A higher DNA extraction yield from samples with higher CFU counts can enhance the sensitivity and reliability of downstream molecular analyses. There was a distinct lack of association between CFU counts and DNA extraction yields in anaerobic cultures, a contrast to the significant association identified in aerobic cultures. The explanation for this observed phenomenon may lie in the unique attributes of anaerobic bacterial growth. First, anaerobic bacteria often necessitate longer incubation periods compared to their aerobic counterparts. This protracted growth timeline may introduce additional variability into the CFU measurements, making it challenging to discern a direct correlation with DNA extraction yields. Second, the cultivation of anaerobic organisms mandates stringent, oxygen-free conditions - a level of environmental control that can be challenging to maintain consistently. Fluctuations or inaccuracies in these conditions could potentially influence the CFU counts. Finally, the requirement for specific growth media to foster optimal growth of diverse anaerobic organisms might also contribute to the observed lack of association. Each of these factors points to the inherent complexity of studying anaerobic cultures and suggests avenues for future, more nuanced investigation.

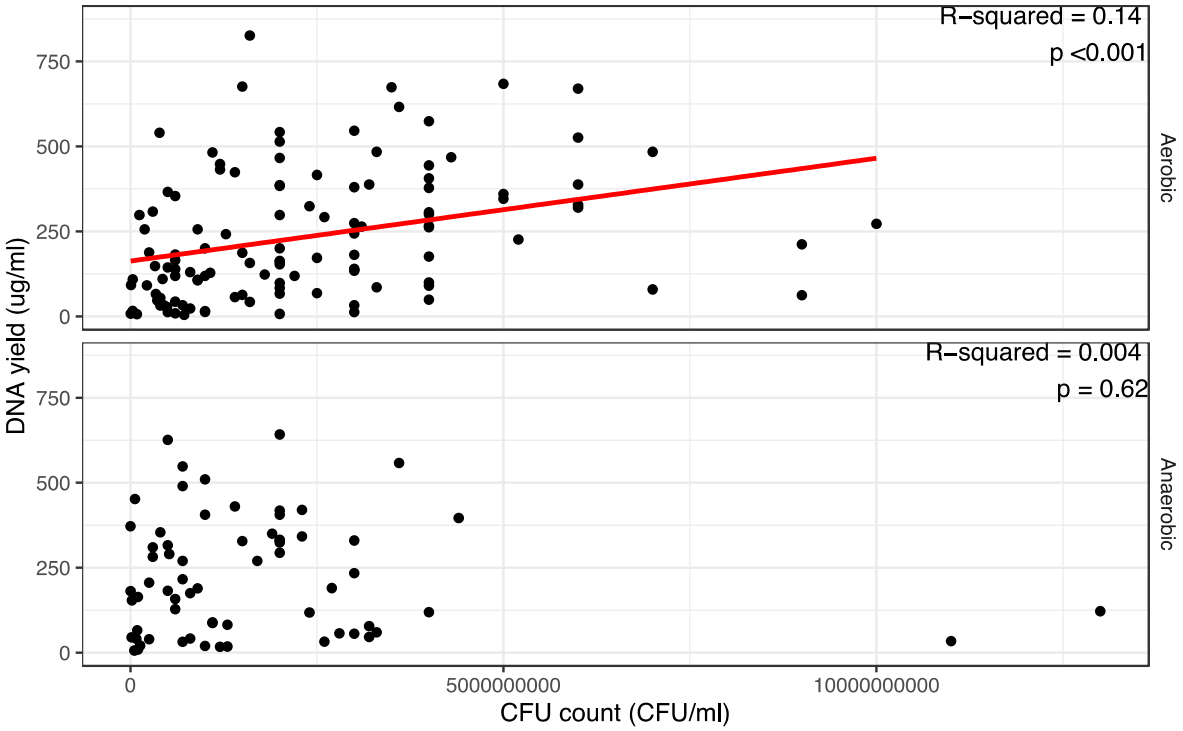

Figure 63. Scatter plot of CFU count vs. DNA extraction yield.

6.12.5 Does DNA extraction yield relate to the number of bases sequenced?

Higher DNA extraction yields were associated with a greater number of bases sequenced (**Figure 64**, R^2=0.08, p<0.001). This relationship implies that a higher DNA extraction yield is associated with a larger amount of DNA available for sequencing, resulting in a greater number of bases that can be accurately determined. The significance of this association has important implications in as the number of bases sequenced is a key measure of the amount of genetic information obtained in sequencing experiments. A higher DNA extraction yield corresponds to an increased quantity of DNA available for sequencing, which can

enhance the depth and coverage of genomic analysis. It allows for more comprehensive exploration of genetic variations, gene expression, and other genomic features.

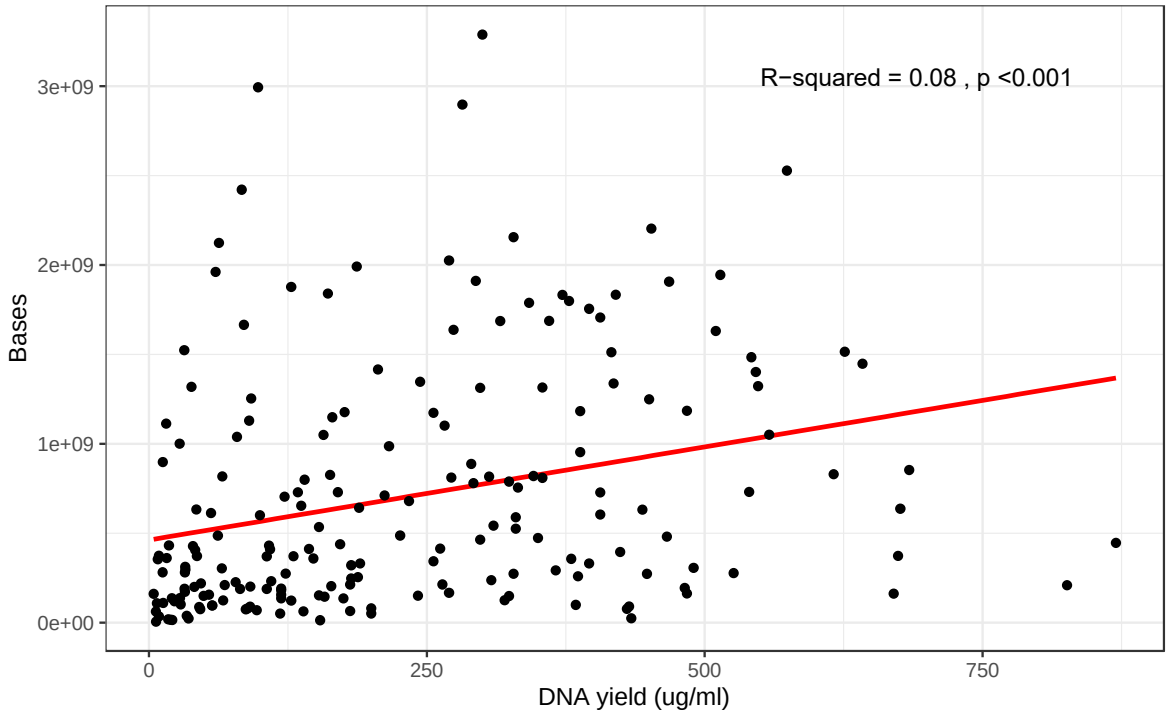

Figure 64. Scatter plot of DNA extraction yield vs. number of bases sequenced.

6.12.6 Does available nanopore sequencing pores relate to the number of bases sequenced?

Not all Nanopore flow cells are equal in terms of the number of sequencing pores available at the start of a sequencing run. 'Good' flowcells tend to have more pores available. I found the numbers of available pores is positively associated with sequencing yields (**Figure 65**, R^2=0.14, p<0.001). As the number of available

nanopores increases, the number of bases generated also tends to increase. This relationship implies that a higher number of nanopores available for sequencing allows for a greater capacity to generate bases accurately during the sequencing process. By maximizing the number of available nanopores at sequencing start, researchers can achieve higher sequencing throughput and increased data yield. Moreover, the correlation between the number of nanopores and the number of bases generated has additional implications for the time to result in nanopore sequencing. As the number of available nanopores increases, the sequencing speed improves, resulting in reduced time to obtain results. This is particularly relevant in time-sensitive applications, such as infectious disease diagnostics, where prompt identification of pathogens and timely treatment decisions are critical. The findings highlight the importance of optimizing nanopore sequencing protocols to maximize the number of available nanopores. Strategies aimed at enhancing the efficiency and capacity of nanopores, including improvements in library preparation methods, flow cell design, and sample loading techniques, can further expedite the time to result.

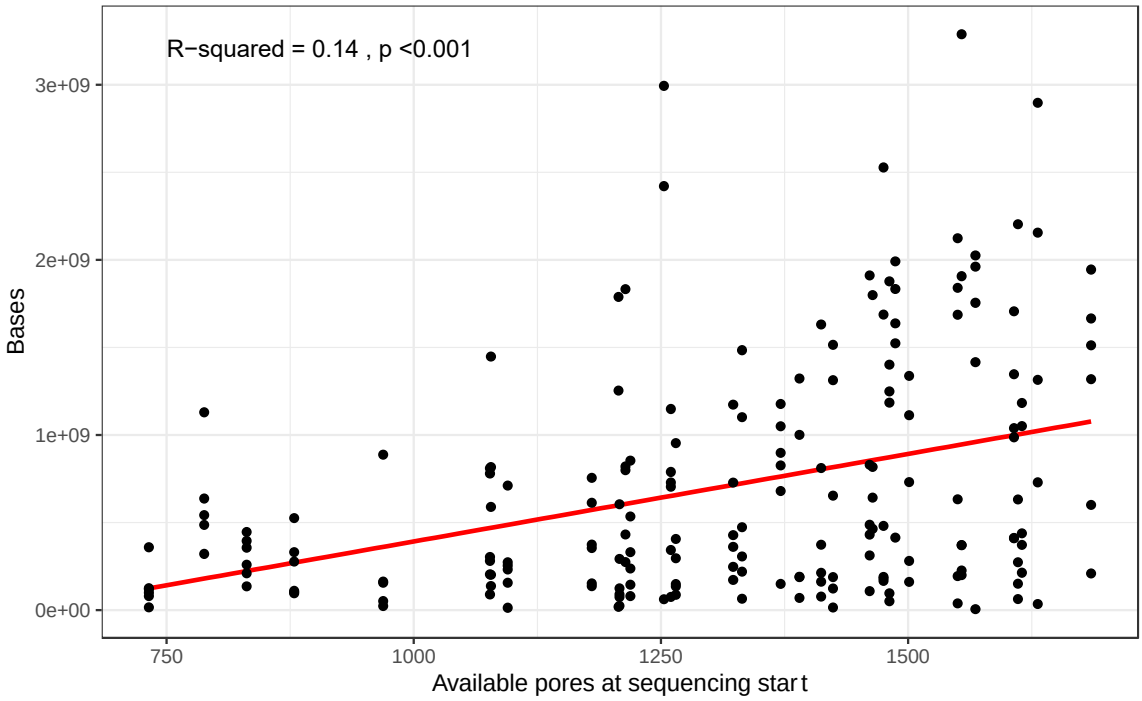

Figure 65. Scatter plot of available nanopores at sequencing start vs. bases generated.

6.13 Cost per sample

The total consumable cost estimated per sample processed was $105 (**Table 29**). This is based on running six samples on each flow cell, and including a positive and negative control in each sequencing run. It is important to note that more than half of the consumable cost was the Nanopore flow cell itself. Further multiplexing of >6 samples may lower the costs as well as using techniques such as adaptive sampling to optimize sequencing where only a defined number of reads is captured per barcode to make an accurate prediction. Furthermore, as Nanopore flow cells are still manufactured by manual assembly, it is hoped the cost of flow cells will be reduced with increasing automation in the manufacturing process. Multiplexing 22 samples, which would still result in a >20x coverage for a species identified, would cost $40 per sample, while multiplexing by 94, that might provide a species only result, would cost $20 (**Table 30**). However, if should be remembered that these costs do not include capital and serving costs of the sequencing machinery or allowances for the staff time and laboratory space required.

Table 29. Itemized Costs per Single Metagenomic Sample in This Study (Based on 6 Sample Multiplexing).

Items	Cost per item* (quantity)	Cost per sample (multiplexing 6 sample)
Extraction		
BD Vacutainer Holders with Luer Adapter	$49 (250)	$0.20
Sterile syringe	$20 (100)	$0.20
QIAamp BiOstic Bacteremia DNA Kit	$200 (50)	$4
Qubit		
dsDNA Quantitation, broad range	$108 (100)	$1.08
Library preparation		
Internal control DNA (*Thermus thermophilus HB8* Genomic 2 ug DNA)	$495	$1.98
Rapid barcoding kit + flow cell priming kit	$650 (6 reactions with 6 multiplexed samples each)	$18
Agencourt AMPure XP beads	$1432 (60ml)	$0.24
Sequencing		
R9.1 flow cell	$450 (1)	$75
Miscellaneous		
1.5 ml Eppendorf DNA LoBind tubes	$25 (500)	$0.2
Pipette tips	$100 (960)	$1.4
Total		$102
Cost of 2x controls per run	$18.6/6#	$3
Grand total		$105

*Prices as of June 2023.

#excluding sequencing and barcoding costs which have already been considered for controls per run

Table 30. Predicted Itemized Costs per Single Metagenomic Sample (Based on 24, 96 Sample Multiplexing).

Items	**Cost per item* (quantity)**	**Cost per sample (multiplexing 22 sample)**	**Cost per sample (multiplexing 94 sample)**
Extraction			
BD Vacutainer Holders with Luer Adapter	$49 (250)	$0.20	$0.20
Sterile syringe	$20 (100)	$0.20	$0.20
QIAamp BiOstic Bacteremia DNA Kit	$200 (50)	$4	$4
Qubit			
dsDNA Quantitation, broad range	$108 (100)	$1.08	$1.08
Library preparation			
Internal control DNA (*Thermus thermophilus HB8* Genomic 2 ug DNA)	$495	$1.98	$1.98
Rapid barcoding kit + flow cell priming kit	$990 (12 reactions)	$8	$4
Agencourt AMPure XP beads	$1432 (60ml)	$0.24	$0.24
Sequencing			
R9.1 flow cell	$450 (1)	$20	$5
Miscellaneous			
1.5 ml Eppendorf DNA LoBind tubes	$25 (500)	$0.2	$0.2
Pipette tips	$100 (960)	$1.4	$1.4
Total		$37	$17
Cost of 2x controls per run	$18.6/6[#]	$3	$3
Grand total		$40	$20

*Prices as of June 2023.

#excluding sequencing and barcoding costs which have already been considered for controls per run

6.14 Results summary

A total of 211 positive blood cultures and 62 negative blood cultures were selected randomly for metagenomic processing in real time. In five positive blood cultures, no species result could be determined by the routine laboratory, and in one positive blood culture sample, MALDI-TOF could only detect a genus. Therefore, 205 positive blood culture samples [216 individual species] was left for the analysis (**Table 31**). The resulting data was analysed using both a threshold-based and random forest machine learning approach using a training and testing dataset to predict true species from sequencing data. Additionally, I tested two approaches, of which the latter was superior, the first involved removing plasmid reads prior to classification, while the second did not remove plasmid reads but instead mapped all species found in Bracken results to the Bracken species reference genome.

Using the threshold method applied to the testing dataset in culture positive samples, unadjusted performance was 94% sensitive (95% CI, 87-98%) and 94% specific (95% CI, 86-98%). In culture-negative samples, specificity was 98% (95% CI, 91-100%). However, after accounting for plausible coinfections (using clinical, microbiological, reference genome mapping and blast techniques), perfect adjusted performance in the testing dataset was achieved (100% sensitivity [95% CI, 97-100], 100% specificity [98-99%]), and all negative controls were negative (100% specific [94-100%]).

Similar performance was achieved prior to adjustment for plausible additional species using a random forest model, 92% sensitivity (95% CI, 84-97%) and 91% specificity (83-96%) in the testing dataset. In culture-negative samples, specificity was 98% (91-100%). After adjusting both the training and testing datasets for plausible coinfections resulted in the same perfect performance as in the threshold-based approach with 100% sensitivity and specificity in the test dataset, and 100% specificity in the negative controls.

This study substantiated the capacity of metagenomic testing to recognize an additional 13 co-infections with high probability, amounting to 6% of the processed samples, which eluded detection by standard laboratory procedures. This finding was supported by clinical and microbiology data, along with reference genome mapping and BLAST data. Moreover, routine laboratory techniques failed to identify an additional six positive samples amounting to 3% of the processed samples, wherein metagenomics succeeded in confidently ascertaining a species in five instances. The sample where metagenomics did not detect any pathogens, might indeed have been a genuine false positive, supported by the patient's medical condition of an acute leukaemia which is characterized by the presence of hypermetabolic cells.

Table 31. Breakdown of culture positive samples and results.

	Culture positive clinical samples	**Species by culture result**	**Species by sequencing result**
	Results that had sufficient routine laboratory result		
Monomicrobial	195	195	195
Polymicrobial	9+1=10	18 (9x 2 species) + 3 (1 x 3 species) = 21	21+13
Total	205	216	229
	Results that had insufficient routine laboratory results		
No routine microbiology result (in flagged positive samples)	5	0	4
Monomicrobial (genus level only)	1	0	1
Total	6	0	5
Summary	211	216	234

6.15 Discussion

This study provides compelling evidence that metagenomic sequencing directly from blood culture samples offers a rapid and accurate method for species identification. The results demonstrate high sensitivity and specificity, with the ability to identify species within approximately 3 hours and 30 minutes from sample to result, representing a significant improvement compared to the routine laboratory turnaround time for species identification by almost half. This approach offers the ability to comprehensively understand infections, guiding effective treatment strategies by identifying additional pathogens beyond traditional methods and predicting antimicrobial resistance. With consumable costs comparable to molecular tests like the BioFire system, metagenomic sequencing demonstrates cost-effectiveness while offering the advantages of pathogen agnosticism and more comprehensive AMR prediction. A notable strength of this study lies in its prospective and blinded collection of samples, ensuring unbiased representation of the routine microbiology laboratory. Moreover, this study implemented rigorous quality control measures, including the use of negative, positive, and internal controls, to ensure the reliability and validity of the sequencing results. These measures enhance the confidence in the accuracy and reproducibility of the findings. In bloodstream infection, each additional hour of delay in treatment increases the odds of mortality by 4%.[183] Thus, the potential of metagenomic sequencing to reduce the time to species identification could play a pivotal role in improving patient outcomes by reducing mortality associated with treatment delays.

6.15.1 Performance

In this study, after accounting for plausible additional pathogens found on sequencing (using clinical, microbiological, reference genome mapping and blast data), the performance in culture positive samples was 100% sensitive and 100% specific, and in culture negative samples was 100% specific. I was able to achieve this by using a combination of taxonomic read classification and confirmation by mapping. Although I attempted removing plasmid reads prior to kraken2 classification to reduce misclassification, I found the performance to be better when mapping to reference genomes and filtering by set thresholds. It is likely that removal of plasmid reads was imperfectly specific such that it also reduced the amount of non-plasmid sequencing data thereby reducing performance and may in some cases excessively removed similar sequences that resemble plasmids and in other cases miss removing plasmids when not defined in the plasmid database.

Sequencing identified 13 co-infections in blood culture samples amounting to 6% of the processed samples, which eluded detection by standard laboratory procedures. This is a significant finding, as it suggests that metagenomic sequencing could reveal a broader spectrum of pathogens involved in bloodstream infections than currently recognized by conventional diagnostic methods. For instance, *Bacteroides fragilis* was found from sequencing in a patient with a liver abscess that was not identified by the routine laboratory. In another case, metagenomics identified *Entercoccus facium* to be a co-infection at a low level, however, it was only cultured from samples taken a day later. Six out of the 13

additional species identified were anaerobes, suggesting that current microbiological diagnostic techniques and conditions may not be fully optimized to grow and detect these organisms. Several infections can be challenging to diagnose using traditional methods, as some pathogens may be overlooked or outcompeted in culture. In addition, metagenomic sequencing can identify pathogens that cannot be identified by MALDI-TOF mass spectrometry and other phenotypic methods. These methods rely on the physical and biochemical characteristics of organisms, which can vary and may not be distinctive for all pathogens. In contrast, metagenomic sequencing identifies organisms based on their DNA, which can provide a more accurate and definitive identification. This could be particularly useful for identifying rare, novel, or atypical pathogens that may not be recognized by phenotypic methods. For instance, 5/6 samples were identified by sequencing where the routine lab failed. One sample was likely a true false positive supported by the patient having AML leading to hypermetabolic cells. In one sample, sequencing found *Capnocytophaga canimorsus*, and interestingly this patient had evidence of a bite mark in keeping with transmission from an animal source. These findings suggest that metagenomic sequencing could lead to a paradigm shift in our understanding of bloodstream infections and their management. By providing a more comprehensive and accurate diagnosis, metagenomic sequencing could guide more targeted and effective treatment strategies, potentially improving patient outcomes. However, it remains to be seen whether additional information would ultimately improve outcomes; additional species may already be treated by current antibiotics, for example anaerobes found in blood stream infection associated with intra-abdominal infections may already

be covered by empirical therapies. However, the case of *Capnocytophaga canimorsus* identification is an example where this information was shared with the clinical team during the study given its significance, and this did change antibiotic treatment. Ultimately, randomised trials of the impact of metagenomics will be needed to evaluate if it does change outcomes.

Compared to routine culture methods, metagenomic sequencing offers several potential advantages. First, while routine culture methods can take several days to yield results, metagenomic sequencing could potentially provide results within hours. This could significantly reduce the time to diagnosis and treatment. Second, routine culture methods can miss fastidious or slow-growing organisms, whereas metagenomic sequencing can potentially detect any pathogen, as it sequences all DNA present in a sample. When compared to multiplex PCR tests like the BioFire, metagenomic sequencing also offers potential advantages. While multiplex PCR tests are limited to detecting a predefined set of pathogens, metagenomic sequencing can potentially detect any pathogen. This means that metagenomic sequencing could identify pathogens far beyond what is included in a standard BioFire panel. Furthermore, metagenomic sequencing can also provide information about antimicrobial resistance, which is not typically provided by multiplex PCR tests, or at least just partially covered with only a handful of key resistance genes included. In comparison to cell-free DNA (cfDNA) metagenomics, metagenomic sequencing of whole blood offers a more comprehensive approach to microbial detection.[184] Whole blood metagenomics captures both intracellular and extracellular microbes, providing a more accurate representation of the active

microbial community in the body. In contrast, cfDNA metagenomics, which primarily captures DNA from microbes present in the extracellular space or those undergoing cell lysis, may miss intracellular microbes and can contain DNA from dead or dying microbes, complicating result interpretation.

In conclusion, the comprehensive microbial detection capability of whole blood metagenomics could guide clinicians in choosing the most effective treatment options. This could reduce the use of broad-spectrum antibiotics and help mitigate the spread of antimicrobial resistance, highlighting the potential of metagenomic sequencing in revolutionizing infectious disease diagnostics and management.

6.15.2 Quality control

In this study, a strong emphasis was placed on the implementation of rigorous quality control measures to ensure the accuracy and reliability of the results, and to minimise contamination. These measures were carefully designed and integrated into different stages of the study, from sample collection and processing to bioinformatic analysis. Meticulous laboratory work was conducted in a controlled environment, including the use of a sterile laminar flow hood and strict adherence to established aseptic techniques. Negative controls to identify any contamination had few reads, all well below filtering thresholds that were set to optimise performance. Positive controls consisted of a known bacterial species, which allowed to verify the accuracy of the sequencing and classification methods of every run. Additionally, internal controls were present in every sample, serving as

an internal reference to ensure the quality and integrity of the sequencing data. These internal controls allowed for the monitoring of various parameters, such as sequencing depth and coverage, to assess the overall performance of the sequencing run. Barcode contamination was meticulously assessed and found to be minimal, aligning with established literature on the topic.[128] In this study, the rate of misassignment or cross-assignment was extremely low, with only 1 incorrectly assigned read observed for every 12,739 correctly assigned reads. Overall, the quality control measures implemented in this study were integral to its success, ensuring the reliability and accuracy of the findings. These measures not only increased confidence in the results but also provided a solid foundation for future research in this area.

6.15.3 CFU counts, DNA extraction yield and time to positivity.

In the chain between growth, DNA extraction and sequencing I identified several relevant relationships. There was no clear correlation between blood culture bottle weight and CFU count, as the rate of bacterial growth in the exponential phase (that flags as positive) is likely not dependent on the initial volume of blood or the initial CFU count, but rather on the bacterial species growth characteristics. Although I unexpectedly did not observe a relationship between blood culture bottle weight and time to positivity, there was a relationship when stratified by specific species such as *Staphylococcus epidermidis.* The analysis found that CFU count is negatively associated with blood culture time to positivity as expected. More growth, that is, higher CFU counts, resulted in more DNA being extracted. In turn more DNA being extracted resulted in more sequencing data being generated.

Furthermore, Nanopore flow cells with more available sequencing pores also generated more sequencing data.

6.15.4 Cost

Cost is a critical factor in the implementation of any diagnostic tool in a clinical setting. The MALDI-TOF MS method, which is commonly used in clinical microbiology laboratories, has a reported reagent cost of approximately $0.50 per isolate for organism identification.[185,186] However, this cost does not include the initial setup costs for the instrument, which is priced at $270k for a complete system, nor does it include the annual maintenance costs of $30k.[186] These additional costs can significantly increase the overall cost of using MALDI-TOF MS, particularly for smaller laboratories or those with a lower throughput of samples. Excluding the initial purchase price of MALDI-TOF, studies have estimated the cost per sample to be $3.14 (incl. reagents, technologist time and maintenance).[186]

The BioFire test, a multiplex PCR test, is reported to cost $45,000 for a single main unit+ $32,000 for every additional unit up to a maximum of 12 (each unit can process a single sample at a given time).[187] Maintenance costs is ±$4500 per year and raw reagent costs per sample range from $120-180. The BioFire test has a short hands-on time of 2 minutes, and total reported costs vary from $200-500.[187] While this is significantly higher than the per-isolate cost of MALDI-TOF MS, it's important to note that the BioFire test includes both organism identification and limited antimicrobial resistance testing, which could potentially provide more

comprehensive information for clinical decision-making. In comparison, the cost of metagenomic sequencing is becoming increasingly affordable. The cost of the GridION, which can run five flow cells simultaneously, is $50,000 (Aug 2023), while annual servicing costs is $10,000. In my calculations, I found that the consumables cost per sample for metagenomic sequencing in this study was approximately $105, however, this can be further reduced to $40 if multiplexing more samples. Furthermore, if we are only interested in species identification (i.e., a low genome coverage) and multiplex 94 samples in a run, the cost per sample would then go down to $20. The costs of metagenomic sequencing can be very similar to the cost of the BioFire test, making metagenomic sequencing a realistic replacement for the BioFire test in the clinical setting. Moreover, metagenomic sequencing holds promise in predicting antimicrobial resistance, which I will return to in the next chapter. This potential is particularly exciting as it could guide clinicians in choosing the most effective treatment options, reducing the use of broad-spectrum antibiotics and helping to mitigate the spread of antimicrobial resistance. Significant progress has already been made in using genetic information to predict resistance in several pathogens, such as *Mycobacterium tuberculosis*[60], *Staphylococcus aureus*, *Escherichia coli and Klebsiella pneumoniae* among others.[61,62] This suggests that with further research and development, metagenomic sequencing could become a powerful tool for AMR prediction in a wide range of pathogens. As the cost of sequencing continues to decrease, it is likely to become an increasingly viable option for routine clinical use. However, it's important to note that the cost-effectiveness of metagenomic sequencing will also depend on

other factors, such as the potential for improved patient outcomes through more accurate and timely diagnosis.

6.15.5 Limitations of this study

This study, while providing valuable insights, has several limitations. Firstly, the use of blood cultures rather than direct patient samples may introduce biases or inaccuracies, as the process of liquid blood culturing can selectively favour certain organisms over others. Crucially this means my approach also depends on a period of pre-culture rather than working directly on samples from patients, introducing a median 17-hour delay, compared to approaches that process cell free DNA in initial blood samples. Secondly, the study did not include a comparison to multiplex panels, which are commonly used in clinical practice for infectious disease diagnostics. Such a comparison could provide a more comprehensive evaluation of the performance and utility of metagenomic sequencing relative to existing diagnostic methods. Lastly, the high cost per sample for metagenomic sequencing is a significant limitation. This high cost may limit the feasibility of implementing this method in routine clinical practice, particularly in resource-limited settings. Future research should explore strategies to reduce the cost of metagenomic sequencing without compromising its performance.

6.15.6 Challenges and future direction

The results of this study are promising, demonstrating the potential of metagenomic sequencing as a powerful tool for the rapid and accurate diagnosis of bloodstream infections. However, these results also highlight several challenges that need to be addressed before this technology can be widely implemented in clinical practice. The biggest challenge remains how to move to being able to sequence samples direct from patients, to substantially decrease the time to clinically impactful results. Culture is an effective amplification technique, however, potentially more sensitive detection methods for growth might allow sequencing to be started sooner. Enrichment and isolation techniques, for instance, serve to enhance the proportion of bacterial cells relative to human cells within a blood specimen. A variety of current techniques have been tested such as differential lysis, saponin treatment, adaptive sampling (in Nanopore flow cells), however, currently none emerge as the clear solution.[188] Newer techniques such as microfluidics cell sorting, magnetic-activated cell sorting (MACS), and CRISPR-Cas based approaches to deplete human DNA may hold promise. In addition to isolation, techniques that facilitate the amplification or capture of bacterial DNA within the sample are of critical importance. Polymerase chain reaction (PCR) has been instrumental in selectively amplifying bacterial DNA, and rapid PCR techniques may hold promise for clinical use. Alternatively, isothermal amplification techniques, typified by loop-mediated isothermal amplification (LAMP), provide a valuable option, especially in scenarios where resources are constrained. Bait capture sequencing and hybridization capture represent two captures strategies. Further work is needed to

establish the lower limits of the number of bacteria required for successful sequencing.

Another important challenge is contamination, which can introduce foreign DNA into the samples and lead to false-positive results. This can complicate the interpretation of sequencing data and potentially lead to incorrect diagnoses. Therefore, it's crucial to develop and implement strategies to minimize contamination. This could involve refining sample collection and processing methods, implementing stringent quality control measures, and developing bioinformatic tools to identify and remove contaminant sequences from the data. Another significant challenge is the lack of standardized workflows for metagenomic sequencing and data analysis. Currently, different laboratories and studies may use different methods for sample preparation, sequencing, and data analysis, which can make it difficult to compare results across different studies. The development of standardized, validated workflows could help to address this issue, improving the reproducibility and comparability of results.

The challenge of human DNA depletion is another key consideration. In metagenomic sequencing of clinical samples, the vast majority of the DNA present is often human, which can interfere with the detection of microbial DNA. Therefore, effective methods for depleting human DNA are needed to increase the sensitivity of metagenomic sequencing for pathogen detection. The challenge of automation is another key consideration. Currently, many steps in the metagenomic

sequencing workflow require manual intervention, which can be time-consuming and prone to error. Developing automated workflows could increase the efficiency and consistency of the process, making it more feasible for routine clinical use. Reliance on a single sequencing provider is another potential challenge. This could limit the availability and affordability of sequencing services, and could also pose a risk if the provider encounters technical issues or other problems. Therefore, it's important to establish relationships with multiple sequencing providers, or to develop in-house sequencing capabilities, to ensure the reliability and sustainability of the sequencing services. Cost considerations are another important factor that may limit the widespread adoption of metagenomic sequencing. While the cost of sequencing has been decreasing, it is still relatively high compared to other diagnostic methods. Furthermore, the cost of data storage and analysis can also be significant, particularly for whole-genome sequencing. Therefore, efforts to reduce costs, such as the development of more efficient sequencing technologies and bioinformatic tools, could facilitate the wider adoption of this technology.

In the interim, while these challenges are being addressed and further cost-effectiveness studies are ongoing, metagenomic sequencing could be used for reference laboratory testing. This would allow for the continued development and refinement of the technology in a controlled setting, while also providing valuable diagnostic information for complex or unresolved cases. The ultimate goal would be to perform metagenomic sequencing directly from a patient's blood, without the need for blood culture. This would significantly reduce the time to diagnosis, potentially improving patient outcomes. However, this approach presents its own

set of challenges, such as the need for extremely sensitive methods to detect the typically low levels of pathogen DNA in blood.

Looking forward, future research should focus on addressing these challenges and exploring the full potential of metagenomic sequencing in clinical practice. This could involve further clinical studies to validate the accuracy and utility of this method, as well as research to refine and optimize the sequencing and data analysis workflows. With these advancements, metagenomic sequencing holds the potential to revolutionize the diagnosis and management of infectious diseases.

Supplementary Table 10. Bioinformatic contaminants (samples with median >1% shown for brevity).

Bacteria	Bioinformatic contaminant	Median	Q1	Q3	Max
Bacteroides caccae	*Akkermansia muciniphila*	3.635	1.058	3.812	4.02
Bacteroides caccae	*Bacteroides fragilis*	1.05	0.7	1.335	2.66
Bacteroides caccae	*Bacteroides thetaiotaomicron*	1.08	0.64	1.53	2.54
Bacteroides caccae	*Bacteroides xylanisolvens*	1.72	0.87	2.37	3.75
Bacteroides caccae	*Phascolarctobacterium faecium*	1.715	1.692	1.738	1.76
Brevibacterium casei	*Kocuria flava*	1.27	0.16	1.46	1.5
Brevibacterium casei	*Microbacterium paraoxydans*	1.06	0.953	1.205	1.34
Brevibacterium casei	*Staphylococcus warneri*	1.08	1	1.16	1.24
Clostridium perfringens	*Staphylococcus hominis*	2.79	2.79	2.79	2.79
Corynebacterium lipophiloflavum	*Corynebacterium sanguinis*	95.6	95.6	95.6	95.6
Corynebacterium minutissimum	*Corynebacterium aurimucosum*	1.285	0.942	64.088	90.9
Corynebacterium sanguinis	*Corynebacterium lipophiloflavum*	12.28	8.555	14.383	18.04
Corynebacterium sanguinis	*Corynebacterium qintianiae*	1.095	0.917	1.752	2.67
Corynebacterium sanguinis	*Microbacterium paraoxydans*	1.78	1.735	1.845	1.91
Corynebacterium striatum	*Corynebacterium resistens*	1.01	0.74	1.055	1.18
Enterobacter hormaechei	*Enterobacter asburiae*	1.13	0.715	1.907	2.58

Enterobacter hormaechei	*Enterobacter roggenkampii*	1.25	0.755	1.97	3.59
Enterobacter hormaechei	*Ralstonia pickettii*	1.02	1.02	1.02	1.02
Enterobacter ludwigii	*Enterococcus faecalis*	1.68	1.57	1.695	1.71
Enterococcus faecalis	*Klebsiella pneumoniae*	1.05	1.05	1.05	1.05
Enterococcus faecium	*Acinetobacter baylyi*	3.68	3.68	3.68	3.68
Enterococcus faecium	*Anaerobutyricum hallii*	1.44	1.44	1.44	1.44
Escherichia coli	*Vibrio cholerae*	1.75	1.75	1.75	1.75
Fusobacterium gonidiaformans	*Fusobacterium necrophorum*	6.67	6.67	7.74	7.74
Fusobacterium gonidiaformans	*Streptococcus anginosus*	1.66	0.94	1.7	1.74
Gordonia sputi	*Gordonia otitidis*	2.905	1.688	4.122	5.34
Klebsiella aerogenes	*Klebsiella pneumoniae*	1.715	0.853	2.322	8.38
Klebsiella grimontii	*Enterococcus casseliflavus*	1.38	1.01	1.665	1.95
Klebsiella grimontii	*Klebsiella michiganensis*	6.165	5.155	7.155	39.38
Klebsiella grimontii	*Klebsiella oxytoca*	5.475	3.965	6.425	19.21
Klebsiella grimontii	*Klebsiella pneumoniae*	5.575	3.758	6.93	14.29
Klebsiella michiganensis	*Enterococcus faecalis*	1.28	1.28	1.28	1.28
Klebsiella michiganensis	*Klebsiella oxytoca*	4.57	2.365	9.145	52.36
Klebsiella michiganensis	*Klebsiella pneumoniae*	3.975	2.697	6.15	15.73
Klebsiella michiganensis	*Pluralibacter gergoviae*	1.02	1.02	1.02	1.02
Klebsiella oxytoca	*Klebsiella pneumoniae*	1.85	1.065	3.412	14.37

Klebsiella pneumoniae	*Klebsiella quasipneumoniae*	2.825	1.407	4.585	6.05
Klebsiella pneumoniae	*Klebsiella variicola*	1.92	1.155	2.89	3.93
Klebsiella quasipneumoniae	*Klebsiella pneumoniae*	9.095	6.63	11.778	42.17
Klebsiella quasipneumoniae	*Klebsiella variicola*	1.11	0.65	1.63	6.6
Klebsiella variicola	*Citrobacter amalonaticus*	1.25	1.25	1.25	1.25
Klebsiella variicola	*Klebsiella pneumoniae*	18.255	13.49	23.732	48.83
Klebsiella variicola	*Klebsiella quasipneumoniae*	3.425	2.27	4.57	7.17
Klebsiella variicola	*Mycobacterium avium*	1.285	1.178	1.393	1.5
Klebsiella variicola	*Pseudomonas aeruginosa*	1.14	0.16	1.31	1.34
Lacticaseibacillus rhamnosus	*Bifidobacterium longum*	3.84	3.695	3.985	4.13
Lacticaseibacillus rhamnosus	*Ligilactobacillus ruminis*	2.21	2.21	2.21	2.21
Leuconostoc mesenteroides	*Mycolicibacterium fortuitum*	1.31	1.31	1.31	1.31
Leuconostoc mesenteroides	*Staphylococcus epidermidis*	1.145	1.072	1.218	1.29
Micrococcus luteus	*Microbacterium paraoxydans*	1.7	1.045	1.845	1.99
Micrococcus luteus	*Streptomyces fradiae*	1.82	1.82	1.82	1.82
Mixta calida	*Klebsiella pneumoniae*	1.715	0.318	3.212	3.64
Mixta calida	*Raoultella ornithinolytica*	1.88	1.84	1.92	1.96
Proteus vulgaris	*Proteus mirabilis*	2.275	1.557	3.49	8.82
Providencia rettgeri	*Proteus vulgaris*	1.81	1.41	2.26	4.41
Pseudomonas aeruginosa	*Pluralibacter gergoviae*	1.37	1.37	1.37	1.37

Pseudomonas aeruginosa	*Pseudomonas juntendi*	64.14	64.14	64.14	64.14
Schaalia odontolytica	*Klebsiella pneumoniae*	1.07	1.06	1.08	1.09
Schaalia odontolytica	*Pauljensenia hongkongensis*	1.17	0.5	1.46	2.88
Schaalia odontolytica	*Schaalia meyeri*	1.69	0.63	2.58	25.17
Staphylococcus aureus	*Escherichia coli*	3.05	1.89	4.21	5.37
Staphylococcus capitis	*Staphylococcus epidermidis*	1.01	0.56	1.71	15.1
Staphylococcus epidermidis	*Enterococcus faecalis*	4.24	4.24	4.24	4.24
Staphylococcus epidermidis	*Pseudomonas protegens*	1.75	1.75	1.75	1.75
Staphylococcus epidermidis	*Staphylococcus carnosus*	1.61	1.61	1.61	1.61
Staphylococcus epidermidis	*Staphylococcus condimenti*	10.99	10.99	10.99	10.99
Staphylococcus haemolyticus	*Cronobacter virus GAP31*	3.81	3.78	3.84	3.87
Staphylococcus haemolyticus	*Lentilactobacillus kefiri*	2.36	2.36	2.36	2.36
Staphylococcus hominis	*Micrococcus luteus*	1.7	1.7	1.7	1.7
Staphylococcus pasteuri	*Kocuria palustris*	92.595	92.558	92.632	92.67
Streptococcus anginosus	*Streptococcus agalactiae*	2.09	0.568	2.757	5.48
Streptococcus anginosus	*Streptococcus constellatus*	1.855	0.835	2.48	8.77
Streptococcus anginosus	*Streptococcus intermedius*	1.5	0.832	2.49	4.44
Streptococcus anginosus	*Streptococcus milleri*	1.425	0.96	1.848	76.76
Streptococcus anginosus	*Streptococcus periodonticum*	5.24	1.515	6.355	31.65

Streptococcus dysgalactiae	*Erysipelothrix rhusiopathiae*	1.745	0.52	2.01	2.16
Streptococcus dysgalactiae	*Streptococcus pyogenes*	4.23	0.675	6.04	12.64
Streptococcus gordonii	*Homo sapiens*	2.83	1.565	3.135	3.44
Streptococcus gordonii	*Phocaeicola vulgatus*	1.71	1.665	1.755	1.8
Streptococcus gordonii	*Streptococcus lactarius*	1.245	1.11	1.515	1.58
Streptococcus oralis	*Homo sapiens*	1.84	1.01	2.065	2.29
Streptococcus oralis	*Streptococcus uberis*	1.01	1.01	1.01	1.01
Streptococcus parasanguinis	*Enterococcus durans*	1.92	1.875	1.965	2.01
Streptococcus parasanguinis	*Streptococcus australis*	1.14	0.58	1.598	4.97
Streptococcus parasanguinis	*Streptococcus lactarius*	1.45	0.98	2.14	59.61
Streptococcus pneumoniae	*Bacillus virus DeepPurple*	1.64	1.64	1.64	1.64
Streptococcus pneumoniae	*Escherichia coli*	1.5	1.5	1.5	1.5
Streptococcus salivarius	*Acidaminococcus intestini*	2.255	2.228	2.282	2.31
Streptococcus sanguinis	*Streptococcus lactarius*	1.29	1.04	1.46	1.6

APPENDIX 1: CODE

Guppy demultiplexing code:

(guppy_barcoder --input_path basecalled_fastq --save_path output_fastq --barcode_kits SQK-RBK004 -q 0 -t 10 --detect_mid_strand_barcodes --min_score_barcode_front 60 --num_extra_bases_trim 50)

Prinseq code to filter low quality reads, and reads <1000bp:

(prinseq-lite.pl -fastq input_reads -lc_method dust -lc_threshold 7 -out_bad null -out_good output_reads -min_len 1000)

Building kraken2 database:

kraken2-build --db . --download-library archaea, bacteria, viral, human, plasmid

Building kraken2 database (plasmid only):

kraken2-build --db . --download-library plasmid

Extracting plasmid sequences from fastq files:

kraken2 --db kraken2_plasmid_db sample.fastq --report output.txt --unclassified-output.fastq --confidence 0.4

Classifying sequences using Kraken2:

kraken2 --db standard_13_2_22 sample.fastq --report output.txt --output standard_kraken_report.txt

Converting Kraken2 report to Bracken report:

bracken -d standard_13_2_22 -i standard_kraken_report.txt -o standard_bracken_report.txt

APPENDIX 2: DNA EXTRACTION PROTOCOL

DNA Extraction protocol using the QIAmp BiOstic Bacteremia DNA Kit.

1. Spin 1.5ml blood in a tube at max speed (16g) for 3min. Discard supernatant.
2. Add **450µl of solution MBL** to the pellet and resuspend by pipetting. Transfer the lysate into a 2ml powerbead tube and close. Vortex for 10s to mix and place in a 70°C heat block for 15min.
3. Secure the 2ml microbead tube horizontally on the Fast Prep instrument at 5.5m/s for 30 sec X1.
4. Centrifuge the 2ml MicroBead Tube to pellet debris at 10,000g for 1min. Transfer the supernatant to a new 2ml collection tube (provided).
5. Add **100µl of Solution IRS** and vortex to mix. Incubate for 5 min at room temperature.
6. Centrifuge at 10,000g for 1 min to pellet debris. Transfer the supernatant to a new 2ml collection tube (provided). Discard pellet.
7. Add **800ul** of Solution BB. Pipette or pulse vortex to mix. Briefly centrifuge to collect any liquid from the top of the lid.

8. Load **650µl of lysate** onto a Spin Column and centrifuge at 10,000g for 1min.
9. Discard the flow-through liquid and place the Spin Column back into the 2ml collection tube. Repeat until all the lysate has been loaded onto the Spin Column.
10. Transfer the Spin Column to a new 2ml collection tube (provided) and wash by adding **500µl of solution CB.** Centrifuge 10, 000g for 1 min. Discard the flow-through and put the Spin Column back into the 2ml collection tube.
11. Wash with another **500µl of solution CB** and spin at 10, 000g for 1 min. Discard the flow-through and put the Spin Column back into the 2ml collection tube.
12. Centrifuge at 13,000g for 2min to dry the Spin Column membrane.
13. Transfer the Spin Column to a new 2ml collection tube (provided).
14. Elute by adding **30µl of solution EB** directly in the center of the membrane. Allow the Spin Column to sit at room temperature for up to 5min to maximise the elution. (Do not heat elution buffer).
15. Centrifuge at 10, 000g for 1min.
16. Discard the Spin Column and cap the 2 ml collection tube containing the genomic DNA.

"One microbe at a time, the physician discerns

the patterns of resistance as a scientist."

- Gaurav Agarwal

7 METAGENOMIC RESISTANCE PREDICTION IN BLOOD STREAM INFECTIONS

7.1.1 Introduction

In Chapter 6, I presented a metagenomic pipeline with the capability to precisely identify a wide range of species, including previously uncultivable bacteria and polymicrobial infections, which conventional microbiology labs often fail to detect. This pipeline delivers species results in a mere 3hrs30min following the detection of a positive blood culture, cutting the time almost by half in comparison to the 10hrs33min average by the routine microbiology laboratory. While early identification of species can potentially modify empirical treatments due to our knowledge of inherent species characteristics, the game-changer lies in targeted and

comprehensive antimicrobial resistance (AMR) prediction. This chapter delves deeper, exploring real-time AMR prediction through various resistance predictive software and databases. Early AMR prediction becomes especially salient when considering the ramifications of delayed or imprecise treatments in conditions such as bacterial bloodstream infections (BSIs) where early targeted therapy is of utmost importance.

BSIs are a major cause of morbidity and mortality with nearly 2 million episodes and 250,000 deaths annually in North America and Europe combined.[189] Furthermore, the total burden of disease is predicted to rise due to increasing antibiotic resistance.[190] Early appropriate antibiotic treatment is vital to reduce the mortality and morbidity, length of stay and healthcare costs associated with BSIs.[191] In patients with severe bloodstream infection, each hour delay in antimicrobial administration is associated with an average decrease in survival of 7.6%, and a time interval of >1 day before initiation of appropriate antimicrobial therapy is associated with a 2-fold increase in the risk of mortality.[87,166] As empirical treatment may be unsuitable in up to 35% of BSIs cases, the need for earlier targeted treatment is becoming increasingly warranted.[192]

Antimicrobial susceptibility testing (AST) remains critical both for guiding appropriate clinical use of antimicrobial agents and supporting surveillance programs that aim to monitor the prevalence of antimicrobial resistance. At present, in vitro phenotypic AST methodologies, such as broth microdilution and disc diffusion — based on principles that date back to the pioneering work by Alexander Fleming in the 1920s — are viewed as the gold standard for evaluating AMR in

bacteria. Automated phenotypic methods such as microbroth dilution (e.g. BD Phoenix; BD, USA) have significantly reduced turnaround time, improved reliability and have demonstrated the benefits associated with earlier antibiotic therapy.[193] However, this method is still reliant on pure cultures that typically take from 24 hours to several days to complete.[194] Despite their conceptual simplicity, phenotypic methods are not without their drawbacks, including challenges in achieving consistent results even when adhering to international standards. Variability can be introduced by factors ranging from the skill and judgement of the individual executing the test and interpreting the results, to the nuances of bacterial growth and stability of reagents. Further limitations stem from a lack of consensus on which panels of antimicrobial agents to test and the criteria for interpretation, as prescribed by different organizations. Beyond these issues, it has been argued that standard laboratory conditions used for AST may not accurately predict clinical efficacy for certain antimicrobial agent-bacterium combinations due to the differing bacterial growth dynamics at infection sites versus in vitro conditions. This, alongside inducible resistance genes and so-called 'silent' genes that may become active during clinical antimicrobial treatment, among other factors, further complicates the picture. Additionally, phenotypic methods offer limited resolution for epidemiological conclusions on the spread of resistance genes and bacteria due to their limited discriminatory power.

Genotypic methods have been suggested as viable alternatives to phenotypic AST since the early 1990s.[195] Microarrays and PCR-based molecular tests have been developed to target drug-resistance markers in primary cultures, however, these tests are unable to identify the full diversity of drug-resistance markers due to limited sized panels. Furthermore, the spectrum of sequences that exists may

diverge from primer regions affecting the sensitivity of these methods, whereas DNA contamination from human and other bacterial cells may affect specificity. Thus, molecular methods continue to present a challenge.[196] The increasing affordability and accessibility of next-generation sequencing (NGS) has positioned whole-genome sequencing (WGS) as a practical methodology for routine bacterial characterization. WGS offers a solution to the limitations of molecular methods as it is unrestricted by a target panel and has the advantage of linage and relatedness prediction. However, traditionally most WGS-based approaches have required pure isolates which take at least 6 to 48 hours to sub-cultivate, resulting in a major bottleneck in the process. Due to recent advancements in sequencing throughput and bioinformatic techniques, sequencing using a metagenomic approach obviates the need for subculture and could provide resistance predictions within a clinically meaningful timeframe. NGS-based methods surpass other genotypic approaches in AMR detection, as they can identify virtually any known AMR gene/mutation and uncover new variants of existing AMR determinants. Furthermore, sequence data can be stored indefinitely, allowing for re-analysis when new AMR determinants are phenotypically discovered. However, genotypic AST methods are not without their limitations. The main caveat related to any genotypic AST method is represented by the fact that only known AMR mechanisms can be detected, whereas resistance caused by new mechanisms and/or modulation of gene expression (increased expression of efflux pumps, heteroresistance, etc.) might be overlooked. The identification of antimicrobial resistance in bacterial genomes is conducted using several bioinformatics tools, each with unique features that detect specific resistance mechanisms.

ResFinder 4.0, an advanced version of a popular web-based platform, provides comprehensive analysis of acquired AMR genes in bacterial genomes.[29] It contains four databases including AMR genes (ResFinder), chromosomal gene mutations mediating AMR (PointFinder), translation of genotypes into phenotypes and species-specific panels for *in silico* antibiograms. ResFinder is a critical tool in identifying an extensive array of acquired antimicrobial resistance genes associated with various resistance mechanisms including detecting genes causing drug target modifications, such as the *mecA* gene in *Staphylococcus aureus* that induces methicillin resistance by coding for an altered penicillin-binding protein with a reduced affinity for beta-lactam antibiotics. In terms of drug inactivation, ResFinder efficiently identifies numerous associated resistance genes, such as various types of beta-lactamase genes that code for enzymes capable of inactivating beta-lactam antibiotics. Similarly, for the drug efflux mechanism, ResFinder can detect resistance genes encoding efflux pumps, like the *tet*(A) and *tet*(B) genes, which confer tetracycline resistance. However, while ResFinder and its PointFinder component represent robust tools for detecting a wide range of AMR genes and certain mutations, they may not capture all possible resistance mechanisms, especially those driven by complex regulatory changes or yet unidentified genetic factors. PointFinder, part of ResFinder, can detect point mutations associated with resistance, which commonly involve target modification. The PointFinder database was extended to include chromosomal gene mutations leading to ampicillin resistance in *Enterococcus faecium*, ciprofloxacin resistance in *E. faecium* and *Enterococcus faecalis*, and resistance to cefoxitin, chloramphenicol, ciprofloxacin, fusidic acid, linezolid, mupirocin, quinupristin–dalfopristin, rifampicin and trimethoprim in *Staphylococcus aureus*. The Comprehensive Antibiotic Resistance Database (CARD) additionally offers a crucial resource in

understanding and combating antimicrobial resistance.[197] It includes a feature called the Resistance Gene Identifier (RGI) software, which predicts antimicrobial resistance genes in new bacterial genomes. RGI utilizes two main methods: homology and SNP (Single Nucleotide Polymorphism) model predictions. Homology involves comparing the sequence of the bacterial genome being tested to known antimicrobial resistance genes in the CARD database, whereas SNP model predictions identify specific single nucleotide changes that can cause antibiotic resistance. This comprehensive approach aids in effective treatment strategies and in understanding the evolution of antimicrobial resistance. NCBI's AMRFinderPlus is another powerful tool designed to identify AMR genes in prokaryotic genome sequences.[198] Its algorithm incorporates a curated database that highlights both acquired AMR genes and resistance-associated mutations. In summary, these tools collectively offer an inclusive detection platform for AMR genes, ranging from acquired resistance genes to specific chromosomal mutations. In this chapter, I benchmark these AMR predictive tools and databases. I then report a method to predict antimicrobial resistance from sequencing of pathogen DNA directly-from-blood culture, and demonstrate the ability to provide faster targeted antibiotic therapy than conventional laboratory approaches in a clinically meaningful timeframe.

7.2 Methods

In this study 205 positive and 62 negative blood culture specimens were randomly selected for metagenomic sequencing from the Oxford University Hospital's microbiology laboratory between December 2020 and October 2022. DNA, extracted using the BiOstic Bacteraemia kit, was sequenced using the Oxford

Nanopore GridION (R9.4.1 flowcell; SQK-RBK004) with 6 libraries multiplexed per flowcell. Positive, negative, and internal controls were used for every batch as described previously in Chapter 6. To ensure high quality predictions and to keep plasmid sequences, sequencing data was pre-processed by filtering out human and internal control *(Thermus thermophilus)* reads prior to resistance prediction using the Kraken Software Suite, rather than extracting only identified species reads.[199] Genotypic resistance prediction was performed with fasta assemblies [Flye v. 2.9-b1768 with '-meta' and '-nano-raw'] and raw reads when possible using ResFinder 4.0 (ResFinder & PointFinder db downloaded 3/7/22), Resistance Gene Identifier (CARD downloaded 18/8/22), and NCBI AMRFinderPlus 3.10.40 (db downloaded 30/7/22). Additionally, a curated mapping database (that maps genes found to genotypic resistance) from ResPipe, an open-source software pipeline, was also applied to RGI (CARD) hits to determine genotypic resistance predictions which is referred to as ResPipe (modified CARD, mCARD).[200] Phenotype organisms-drug panels where a genotypic prediction could be made are listed in **Supplementary Table 13**.

In the process of configuring RGI(CARD), I have excluded hits associated with the "antibiotic efflux" mechanism and the "protein variant model". The former, a nonspecific mechanism influenced by factors such as efflux pump expression levels, and the latter, a context-dependent model susceptible to variability among different bacteria and conditions, were filtered out to focus on more specific and reliable resistance mechanisms. Concurrently, I set stringent thresholds for best identities (>60), best hit bitscore (>33), and percentage length of reference sequence (>99) to ensure significant sequence similarity, high-quality alignment, and comprehensive coverage of the reference sequence, respectively.[201,202] These

measures collectively aim to minimize false positives, thereby increasing the confidence in the RGI results.

For AMRFinder, I implemented a similar approach. I set a high percentage coverage of the reference sequence (>85) to ensure that the identified sequences likely represent substantial portions or complete genes, rather than mere fragments, reducing the risk of false positives from partial gene matches. I also established a high percentage identity to the reference sequence (>95), ensuring a significant degree of similarity between the identified and reference sequences, thereby increasing the confidence in the identification.[201,202] To further refine the search, I filtered for the "AMR" (Antimicrobial Resistance) subtype, focusing the analysis on elements specifically associated with antimicrobial resistance and excluding irrelevant elements. Lastly, I set a minimum target length (>120bp) to ensure that the identified sequences are of sufficient length to likely represent complete, functional resistance genes. This length threshold further reduces the likelihood of false positives from short sequences that might represent gene fragments or spurious matches.

Of note, I used the *Staphylococcus aureus* rules across all *Staphylococcus* species to ensure a broader coverage and robust detection of resistance elements. A sample was considered to be genotypically sensitive if no resistant genes were found, in samples with a genome sequencing depth of at least 10x at the end of sequencing.

Phenotypic antimicrobial susceptibility testing (AST) was used as a reference standard to assess the performance of genotypic resistance predictions. Categorical

AST results (sensitive, intermediate, resistant) were generated from pure culture isolates obtained in the Oxford University Hospitals (OUH) microbiology laboratory during each patient's admission [using the BD Phoenix microbroth dilution (BD, USA)]. Results that was determined to be "intermediate" was excluded from the analysis. Results of comparisons are presented using categorical agreement (CA; genotypic result classified in the same susceptibility category as the phenotypic result), very major error (VME; predicted susceptible but phenotypically resistant) and major error (ME; predicted resistant but phenotypically susceptible) rates. Only ResFinder provided drug-specific resistance predictions while the other tools provided drug class predictions. When summarising phenotypic results at a drug class level, any resistant result within a class was considered as the sample containing a resistant isolate (**Supplementary Table 12**). Monomicrobial and polymicrobial infections (including those identified in Chapter 6) were analysed separately.

Based on historic data on the most common organisms found in blood cultures, the following organisms was considered to be the top ten most clinically relevant species within this dataset: *Escherichia coli*, *Staphylococcus aureus*, *Enterococcus* sp. (*Enterococcus faecalis* and *Enterococcus faecium*), *Klebsiella pneumoniae*, *Enterobacter cloacae*, *Bacteroides fragilis*, *Proteus mirabilis*, *Streptococcus pneumoniae*, *Pseudomonas aeruginosa* and Coagulase-negative staphylococci (*Staphylococcus hominis*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, and *Staphylococcus capitis*).[203]

7.3 Results

7.3.1 Study characteristics

A total of 205 positive and 62 negative blood culture specimens were successfully sequenced. The median total reads and bases generated per flow cell were 6 million (IQR:4.7-6.4; range:2.4-10) and 11 gigabases (IQR:10-12; range:5-20) respectively. All positive and internal controls were identified, and the negative controls appropriately yielded negative results, as detailed in Chapter 6.

Amongst the ten most clinically relevant species, phenotypic resistance was present in 18% (224 resistant, 1029 sensitive) of drug-organism combinations in monomicrobial samples (**Table 32**). The highest phenotypic resistance was seen in *Coagulase-negative staphylococci* followed by *Enterococcus sp.* (43% and 38% respectively), and the lowest was seen in *Pseudomonas aeruginosa* and *Streptococcus pneumoniae* (both had no recorded phenotypic resistance).

Table 32. Phenotypic resistance and sensitivity by species.

Species	Resistant phenotype	Sensitive phenotype	Total phenotype
Escherichia coli	58 (10%)	539 (90%)	597
Staphylococcus aureus	12 (8%)	139 (92%)	151
Enterococcus sp.	24 (38%)	40 (62%)	64
Klebsiella pneumoniae	25 (27%)	66 (73%)	91
Enterobacter cloacae	10 (24%)	32 (76%)	42
Proteus mirabilis	6 (11%)	50 (89%)	56
Streptococcus pneumoniae	0 (0%)	11 (100%)	11
Pseudomonas aeruginosa	0 (0%)	34 (100%)	34
Coagulase-negative staphylococci	89 (43%)	118 (57%)	207
Total	**224 (18%)**	**1029 (82%)**	**1253**

Phenotypic resistance by drug was most commonly seen for penicillin and amoxicillin (85% and 59% respectively, **Table 33**). For *Escherichia coli* phenotypic resistance was most seen in amoxicillin and ampicillin+clavulanic acid (42% and 37% respectively, **Table 34**) and for *Staphylococcus aureus* in penicillin, clindamycin, erythromycin (89%, 18% and 18% respectively). Other species-drug combination detailing phenotypic resistance and sensitivity can be found in **Supplementary Table 11**.

Table 33. Phenotypic resistance and sensitivity by drug.

Drug	Resistant phenotype	Sensitive phenotype	Total phenotype
Tigecycline	1 (100%)	0 (0%)	1
Penicillin	11 (85%)	2 (15%)	13
Amoxicillin	46 (59%)	32 (41%)	78
Erythromycin	18 (58%)	13 (42%)	31
Clindamycin	15 (52%)	14 (48%)	29
Fusidic acid	13 (43%)	17 (57%)	30
Ampicillin+clavulanic acid	22 (39%)	35 (61%)	57
Gentamicin	34 (32%)	73 (68%)	107
Trimethoprim	13 (22%)	45 (78%)	58
Tetracycline	5 (20%)	20 (80%)	25
Teicoplanin	6 (14%)	38 (86%)	44
Vancomycin	6 (13%)	40 (87%)	46
Tobramycin	7 (11%)	56 (89%)	63
Ciprofloxacin	6 (10%)	55 (90%)	61
Piperacillin+tazobactam	7 (9%)	69 (91%)	76
Temocillin	4 (7%)	52 (93%)	56
Ceftriaxone	4 (7%)	53 (93%)	57
Ceftazidime	3 (5%)	57 (95%)	60
Chloramphenicol	1 (4%)	26 (96%)	27
Rifampicin	1 (3%)	29 (97%)	30
Linezolid	1 (2%)	46 (98%)	47
Amikacin	0 (0%)	63 (100%)	63
Ertapenem	0 (0%)	57 (100%)	57
Fosfomycin	0 (0%)	63 (100%)	63
Meropenem	0 (0%)	62 (100%)	62
Mupirocin	0 (0%)	12 (100%)	12
Total	**224 (18%)**	**1029 (82%)**	**1253**

Table 34. Phenotypic resistance and sensitivity by species-drug combination.

Species	Drug	Resistant phenotype	Sensitive phenotype	Total phenotype
Escherichia coli	Amoxicillin	18 (42%)	25 (58%)	43
Escherichia coli	Ampicillin+clavulanic acid	16 (37%)	27 (63%)	43
Escherichia coli	Trimethoprim	8 (19%)	35 (81%)	43
Escherichia coli	Piperacillin+tazobactam	3 (7%)	39 (93%)	42
Escherichia coli	Ciprofloxacin	3 (7%)	40 (93%)	43
Escherichia coli	Gentamicin	3 (7%)	40 (93%)	43
Escherichia coli	Tobramycin	3 (7%)	40 (93%)	43
Escherichia coli	Temocillin	2 (5%)	40 (95%)	42
Escherichia coli	Ceftazidime	1 (2%)	40 (98%)	41
Escherichia coli	Ceftriaxone	1 (2%)	41 (98%)	42
Escherichia coli	Amikacin	0 (0%)	43 (100%)	43
Escherichia coli	Ertapenem	0 (0%)	43 (100%)	43
Escherichia coli	Fosfomycin	0 (0%)	43 (100%)	43
Escherichia coli	Meropenem	0 (0%)	43 (100%)	43
Staphylococcus aureus	Penicillin	7 (88%)	1 (12%)	8
Staphylococcus aureus	Clindamycin	2 (18%)	9 (82%)	11
Staphylococcus aureus	Erythromycin	2 (18%)	9 (82%)	11
Staphylococcus aureus	Fusidic acid	1 (9%)	10 (91%)	11
Staphylococcus aureus	Tetracycline	0 (0%)	11 (100%)	11
Staphylococcus aureus	Chloramphenicol	0 (0%)	11 (100%)	11
Staphylococcus aureus	Ciprofloxacin	0 (0%)	11 (100%)	11
Staphylococcus aureus	Fosfomycin	0 (0%)	11 (100%)	11
Staphylococcus aureus	Gentamicin	0 (0%)	11 (100%)	11

Staphylococcus aureus	Linezolid	0 (0%)	11 (100%)	11
Staphylococcus aureus	Mupirocin	0 (0%)	11 (100%)	11
Staphylococcus aureus	Rifampicin	0 (0%)	11 (100%)	11
Staphylococcus aureus	Teicoplanin	0 (0%)	11 (100%)	11
Staphylococcus aureus	Vancomycin	0 (0%)	11 (100%)	11

* For other organisms, refer to **Supplementary Table 11**.

7.3.2 Performance in monomicrobial infections

7.3.2.1 Overall classifier performance

The predictive capacity of drug resistance was evaluated through a comparative analysis with four software and database tools on the top ten species of clinical relevance, inclusive of *Coagulase-negative staphylococci*. Given that ResFinder was the only tool capable of predicting resistance at the individual drug level, I adopted the approach of interpreting resistance at the class level. This approach assumes that if a gene within a particular class is identified as resistant, all drugs within that same class are presumed to exhibit resistance.

Using assembled genomes, ResFinder [Input: fastq reads at 24hrs] had the highest overall sensitivity and specificity of 84% (95% CI: 78-88%, 188/224) and 96% (95% CI: 94-97%, 983/1029) respectively, while mCARD [Input: fasta from assembly at 24hrs] had the lowest sensitivity and specificity of 76% (70-81%, 170/224) and 76% (74-79%, 786/1029) respectively (**Table 35**).

In older versions of ResFinder, only assembled genomes in FASTA format were accepted as input. However, with ResFinder 4.0 a new alignment method was developed, which could handle highly similar gene content in the ResFinder databases which allowed for the analysis of raw read data directly.[204] Using this new feature (which is not available on other tools), I tested performance using raw fastq reads [all reads at 24hours] which yielded a higher sensitivity and specificity of 91% (95% CI: 87-94%, 204/224) and 90% (95% CI: 88-91%, 923/1029) respectively (**Table 36**). I then tested performance of raw reads at earlier time intervals. Using ResFinder with raw reads one hour after sequencing resulted in a sensitivity and specificity of 82% (95% CI: 76-87%, 184/224) and 94% (95% CI: 92-95%, 963/1029), while raw reads after two hours of sequencing had a sensitivity and specificity of 88% (82-92%, 196/224) and 93% (91-94%, 953/1029) respectively (**Table 37**).

Table 35. Genotypic performance of top 10 clinically relevant species by tool after genome assembly.

Database	Software	Input format	Sequencing start time to result	Sensitivity	Specificity	Major Error rate	Very Major Error rate
ResFinder and PointFinder database	ResFinder	fasta	24 hours	84% (95% CI: 78-88%, 188/224)	96% (95% CI: 94-97%, 983/1029)	4%	16%
NCBI's AMRFinderPlus database*	AMRFinderPlus	fasta	24 hours	79% (74-85%, 178/224)	79% (77-82%, 816/1029)	21%	21%
Comprehensive Antibiotic Resistance Database (CARD)*	Resistance Gene Identifier (RGI)	fasta	24 hours	79% (73-84%, 177/224)	77% (74-79%, 788/1029)	23%	21%
modified Comprehensive Antibiotic Resistance Database (CARD)*	Resistance Gene Identifier (RGI)	fasta	24 hours	76% (70-81%, 170/224)	76% (74-79%, 786/1029)	24%	24%

*These tools predicted genotypic resistance at class level only, which was interpreted as resistance at all drug levels within that specific class.

Table 36. Genotypic performance of top 10 clinically relevant species using raw reads.

Database	Software	Input format	Sequencing start time to result	Sensitivity	Specificity	Major Error rate	Very Major Error rate
ResFinder and PointFinder database	ResFinder	fastq	24 hours	91% (95% CI: 87-94%, 204/224)	90% (95% CI: 88-91%, 923/1029)	10%	9%

Table 37. Genotypic performance of top 10 clinically relevant species using raw reads at earlier time intervals.

Database	Software	Input format	Sequencing start time to result	Sensitivity	Specificity	Major Error rate	Very Major Error rate
ResFinder and PointFinder database	ResFinder	fastq	1 hour	82% (95% CI: 76-87%, 184/224)	94% (95% CI: 92-95%, 963/1029)	6%	18%
ResFinder and PointFinder database	ResFinder	fastq	2 hours	88% (82-92%, 196/224)	93% (91-94%, 953/1029)	7%	12%

7.3.2.2 Species performance

For *Escherichia coli,* sensitivity and specificity with ResFinder [Input fastq at 2hrs] was 91% [95% CI, 80-97%] (53/58 drug-organism combinations) and 94% [95% CI, 92-96%] (505/539) respectively, and for *Staphylococcus aureus* was 100% [95% CI, 77-100%] (12/12) and 99% [95% CI, 96-100%] (138/139) respectively (**Figure 66**).

Sensitivity using ResFinder [Input fastq at 2hrs] was highest among *Staphylococcus aureus, Escherichia coli and Coagulase−negative staphylococci* across all tools (sensitivity & specificity, 100% [95% CI, 77-100%] & 99% [95% CI, 96-100%], 91% [95% CI, 76-99%] & 94% [95% CI, 80-99%] and 90% [95% CI, 88-98%] & 94% [95% CI, 89-98%] respectively).

ResFinder-based tools generally performed better than the other tools and databases. However, RGI (CARD) [Input fasta at 24hrs] performed notably better for *Klebsiella pneumoniae* and *Enterobacter cloacae* with sensitivities and specificities of 100% and 95% (vs 88% and 69% for ResFinder fastq at 24hrs) and 100% and 100% (vs 80% and 69% for ResFinder fastq at 24hrs), likely due to a different database structure compared to ResFinder-based tools. AMRFinder [Input fasta at 24hrs] performed similarly superior for *Klebsiella pneumoniae* and *Enterobacter cloacae.*

Poor sensitivity for detecting phenotypic resistance was seen globally with *Proteus mirabilis* (17% across tools), while AMRFinder and RGI (CARD) has poor

sensitivity for *coagulase-negative staphylococcus* (Ranging between 58-65%). Poor specificity for detecting phenotypic was seen with *Pseudomonas aeruginosa* (15% to 74% across tools). There were no resistant phenotypic drug-organism combinations for *Streptococcus pneumoniae* and *Pseudomonas aeruginosa*, and therefore sensitivity could not be evaluated in these organisms.

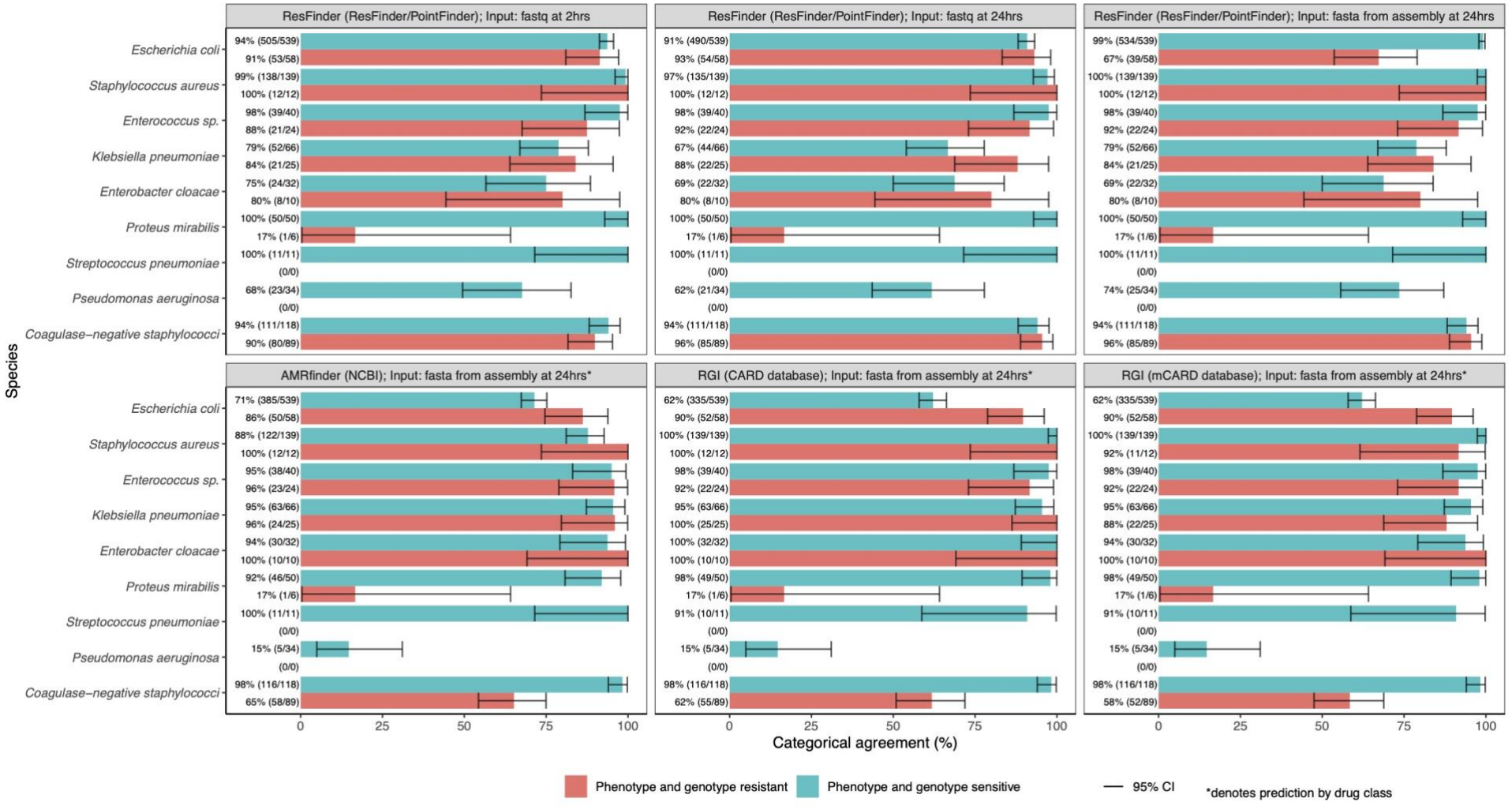

Figure 66. Genotypic performance by species across tools.

7.3.2.3 Drug performance

7.3.2.3.1 Resistance phenotype prediction (sensitivity)

ResFinder [Input: fastq reads at 24hrs] had the highest overall sensitivity, that is, resistance detection rate, by drug class ranging from 100% within the glycopeptide (12/12), amphenicol (1/1), macrolide (33/33) and steroid antibacterials (13/13) to 83% and 71% within the tetracyclines (5/6 correct resistant predictions) and quinolones (5/7 correct resistant predictions) respectively. AMRFinder and RGI (CARD) performed better for beta-lactams and folate pathway antagonist, however performed similarly or worse for other drug classes (**Figure 67**).

Although only ResFinder was able to predict resistance at the drug level, I considered a class level resistance prediction by other tools to be interpreted as resistance within all specific drug classes. By specific drug, ResFinder [Input: fastq reads at 24hrs] performed similarly or better than all other tools except for ampicillin+clavulanic acid (where CARD-based tools performed better, 100% (22/22) vs 95% (21/22)), ceftriaxone (where CARD and AMRFinder tools performed better, 75% (3/4) vs 100% (4/4)) and trimethoprim (where AMRFinder performed better, 92% (12/13) vs 100% (13/13)) (**Figure 68** and **Figure 69**).

7.3.2.3.2 Sensitive phenotype prediction (specificity)

ResFinder [Input: fastq reads at 24hrs] drug class level specificity, i..e correct detection of drug sensitivity, ranged from 97% in glycopeptides (76/78) and 96% (25/26) in amphenicols to 83% (336/405) in beta-lactams and 82% (37/45) in folate pathway antagonists (**Figure 67**). ResFinder [Input: fastq reads at 2hrs] performed similarly. AMRFinder and RGI (CARD) performed more poorly at 65% (263/405)

and 51% (207/405) in beta-lactams and 76% (145/192) and 81% (155/192) in aminoglycosides respectively. AMRFinder performed particularly poorly in tetracyclines with a performance of 45% (9/20).

By specific drugs, ResFinder [Input: fastq reads at 24hrs] had the highest specificity of 100% (2/2) in penicillin and 100% (56/56) in tobramycin to the lowest of 66% (35/53) in ceftriaxone and 56% (32/57) in ceftazidime (**Figure 68** and **Figure 69**). ResFinder [Input: fastq reads at 2hrs] performed similarly. Both RGI (CARD) and AMRFinder had a poor performance of 28% (9/32) and 50% (16/32) in amoxicillin respectively. RGI (CARD) had a particularly poor performance of 34% (12/35) in ampicillin+clavulanic acid, while AMRFinder had a particularly poor performance of 45% (9/20) in tetracyclines.

RGI (mCARD) performed either similar or worse than RGI (CARD) across all drugs and drug classes. Sensitive prediction performance for drugs with only sensitive phenotypes (and no resistant phenotypes) which included amikacin, ertapenem, fosfomycin, meropenem, mupirocin, and their respective drug classes are shown in **Figure 70**.

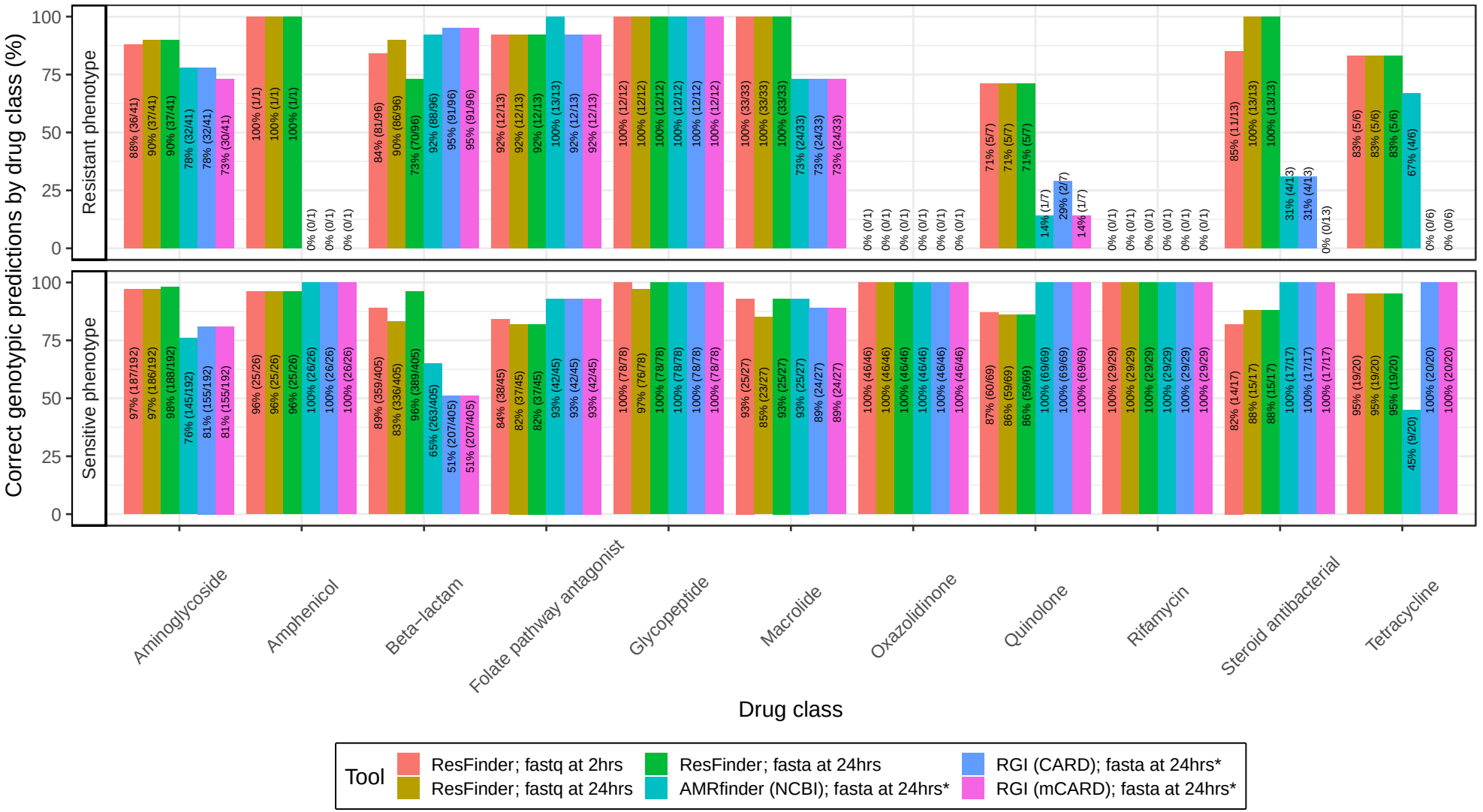

Figure 67. Genotypic performance by drug class.

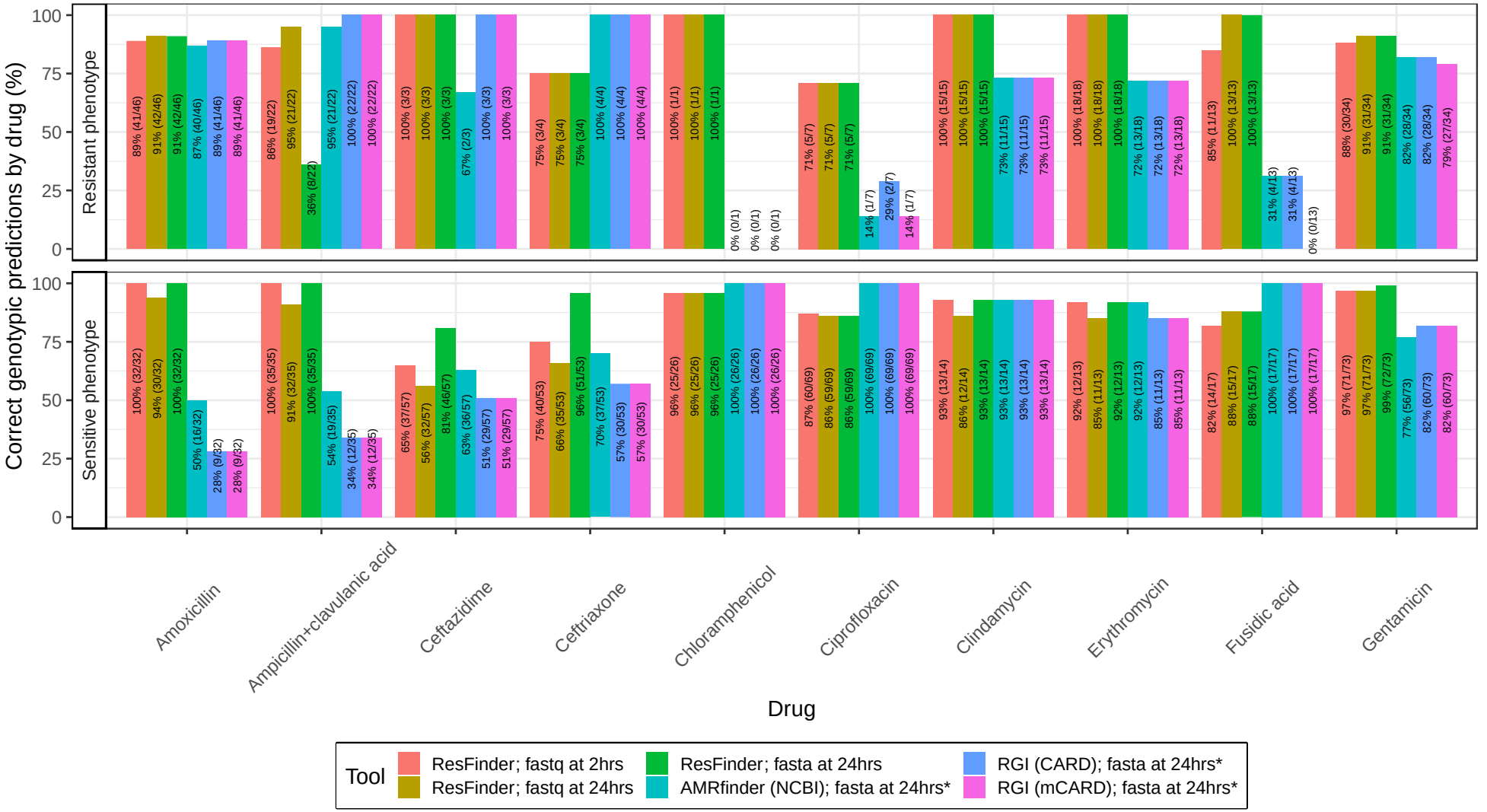

Figure 68. Genotypic performance by specific drug (1/2).

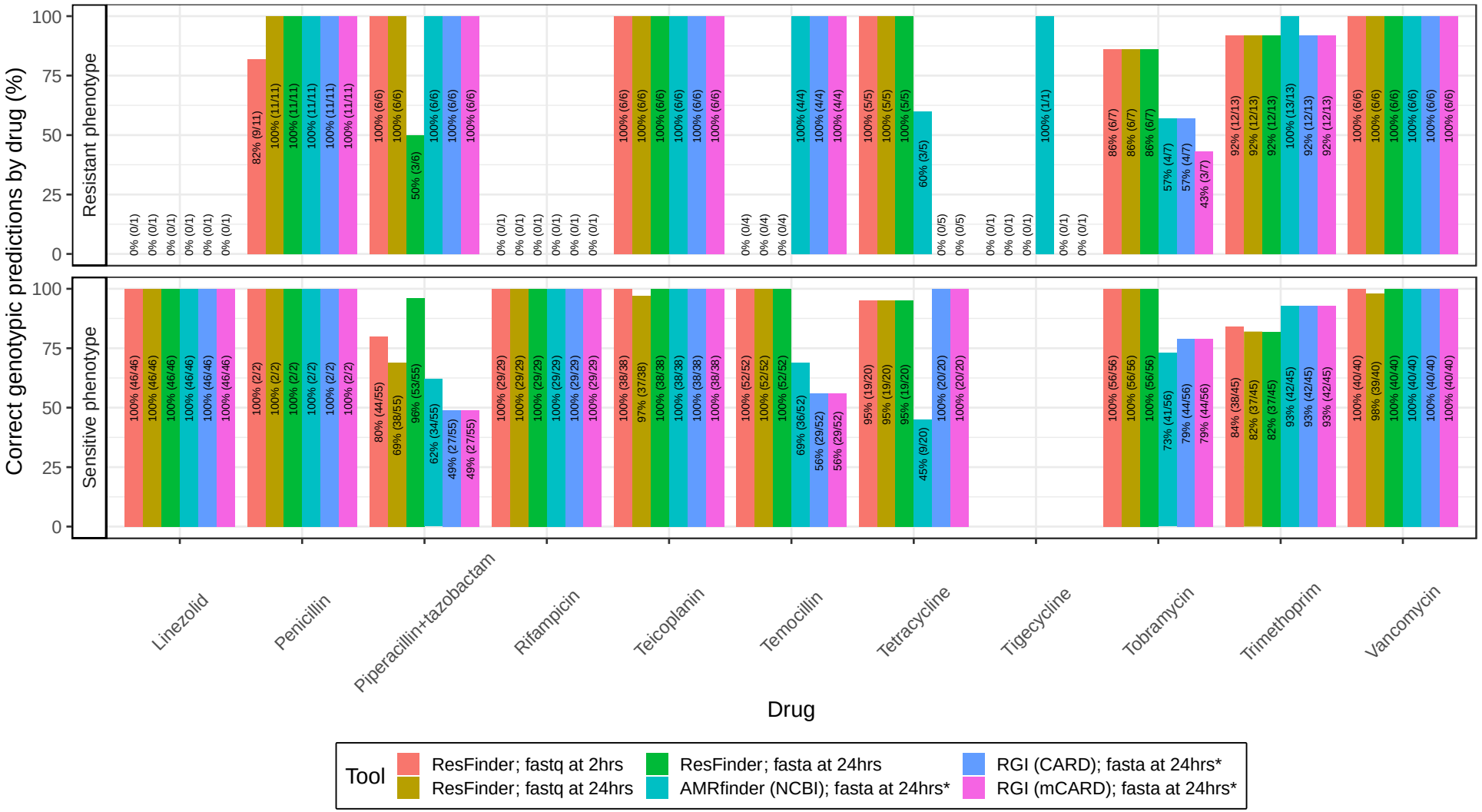

Figure 69. Genotypic performance by specific drug (2/2).

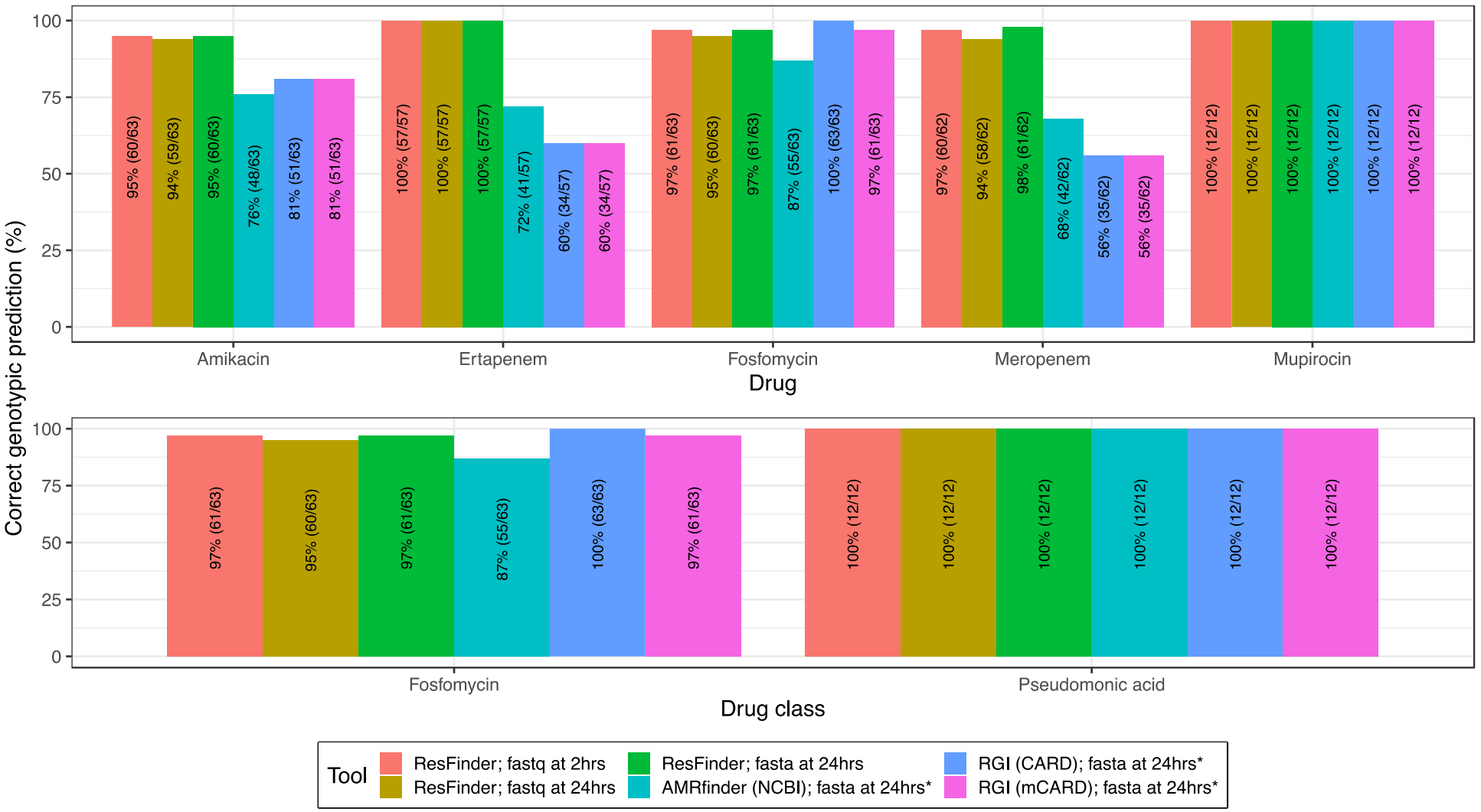

Figure 70. Genotypic performance for sensitive only drug/drug class.

7.3.2.4 Patterns in resistance and sensitivity prediction across tools

By all tools used, 59% (132/224) of phenotypic resistant drug-organism combinations (for aminoglycosides, beta−lactams, folate pathway antagonist, glycopeptides, macrolides and oxazolidinone) were identified (**Figure 71**). 13% (30/224) were identified by Resfinder-based tools alone (including resistance to aminoglycosides, amphenicols, beta−lactams, macrolides, quinolones and tetracyclines), while 5% (10/224) were identified by AMRFinder and RGI (CARD) alone (beta-lactam and folate pathway antagonist resistance). 4% (9/224) had no genotypic calls across all tools (for aminoglycosides, beta−lactams, glycopeptides, oxazolidinones, quinolones and rifamycins). RGI (modified CARD) [Input fasta at 24hrs] called 3% (6/224) less compared to other tools (it missed aminoglycosides and steroid antibacterials), while RGI (CARD)-based tools in general called 2% (4/224) less compared to other tools. ResFinder [Input: fasta reads at 24hrs] and ResFinder [Input: fastq reads at 1hr] were the mostly likely tools to miss beta-lactam resistance detected, calling 6% (14/224) less compared to other tools.

53% (544/1029) of phenotypic sensitive drug-organism combinations were correctly called as sensitive by all tools used (**Figure 72**). 29% (294/1029) were correctly called sensitive by Resfinder-based tools only (aminoglycosides, beta−lactams, folate pathway antagonist and macrolides), and 2% (17/1029) by AMRFinder and RGI (CARD) alone (amphenicols, folate pathway antagonists, quinolones, steroid antibacterials and tetracyclines). 6% (57/1029) were called sensitive by Resfinder-based tools and AMRFinder excluding RGI (CARD)-based tools (for beta-lactams and macrolides). 3% (27/1029) were called sensitive by RGI (CARD)-based tools and Resfinder-based tools excluding AMRFinder (for

aminoglycosides, fosfomycins and tetracyclines). 2% (23/1029) had false genotypic calls across all tools (for aminoglycosides, beta−lactams, folate pathway antagonists and macrolides).

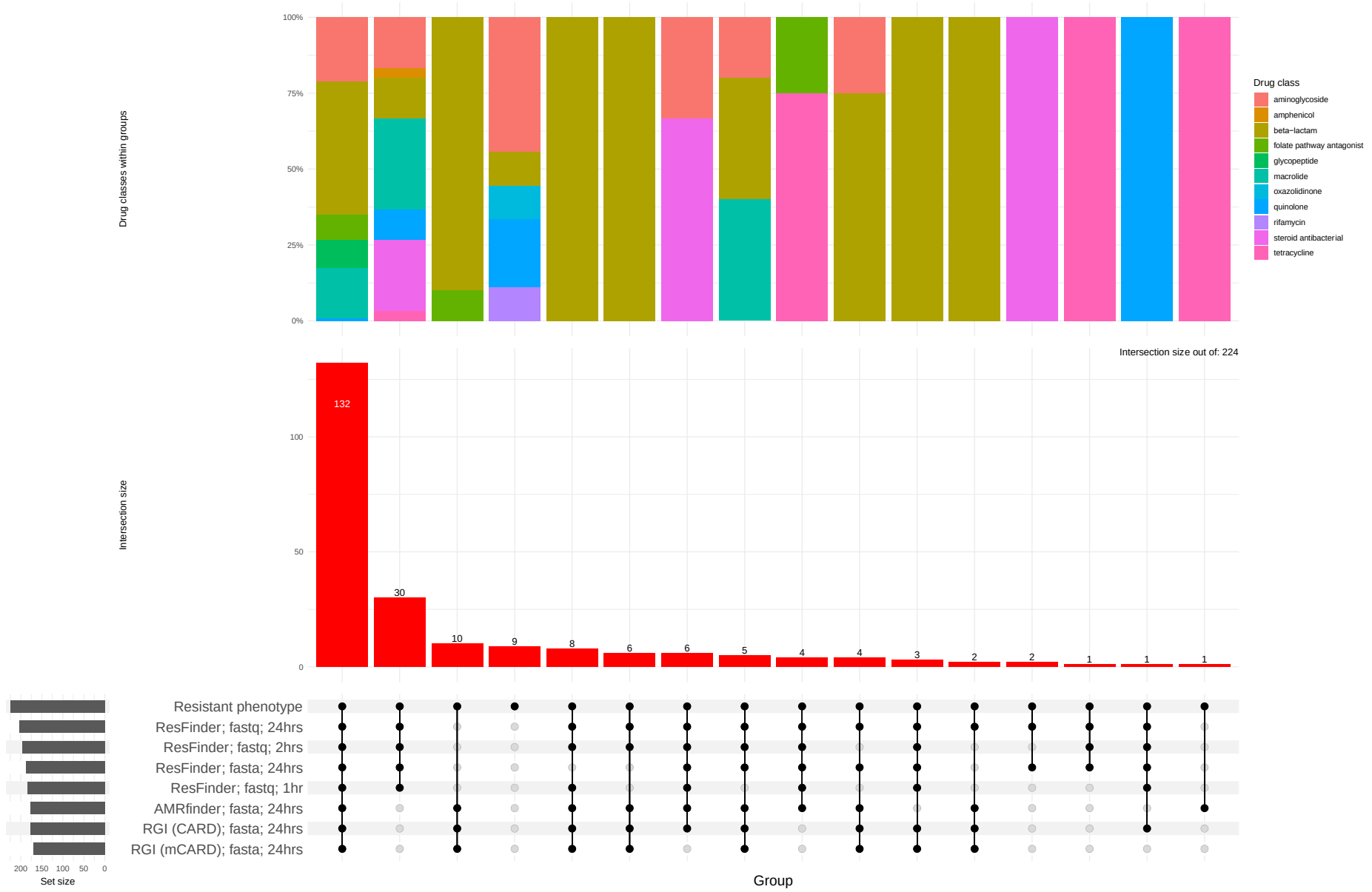

Figure 71. Upset plot showing various groups within resistant phenotype calls.

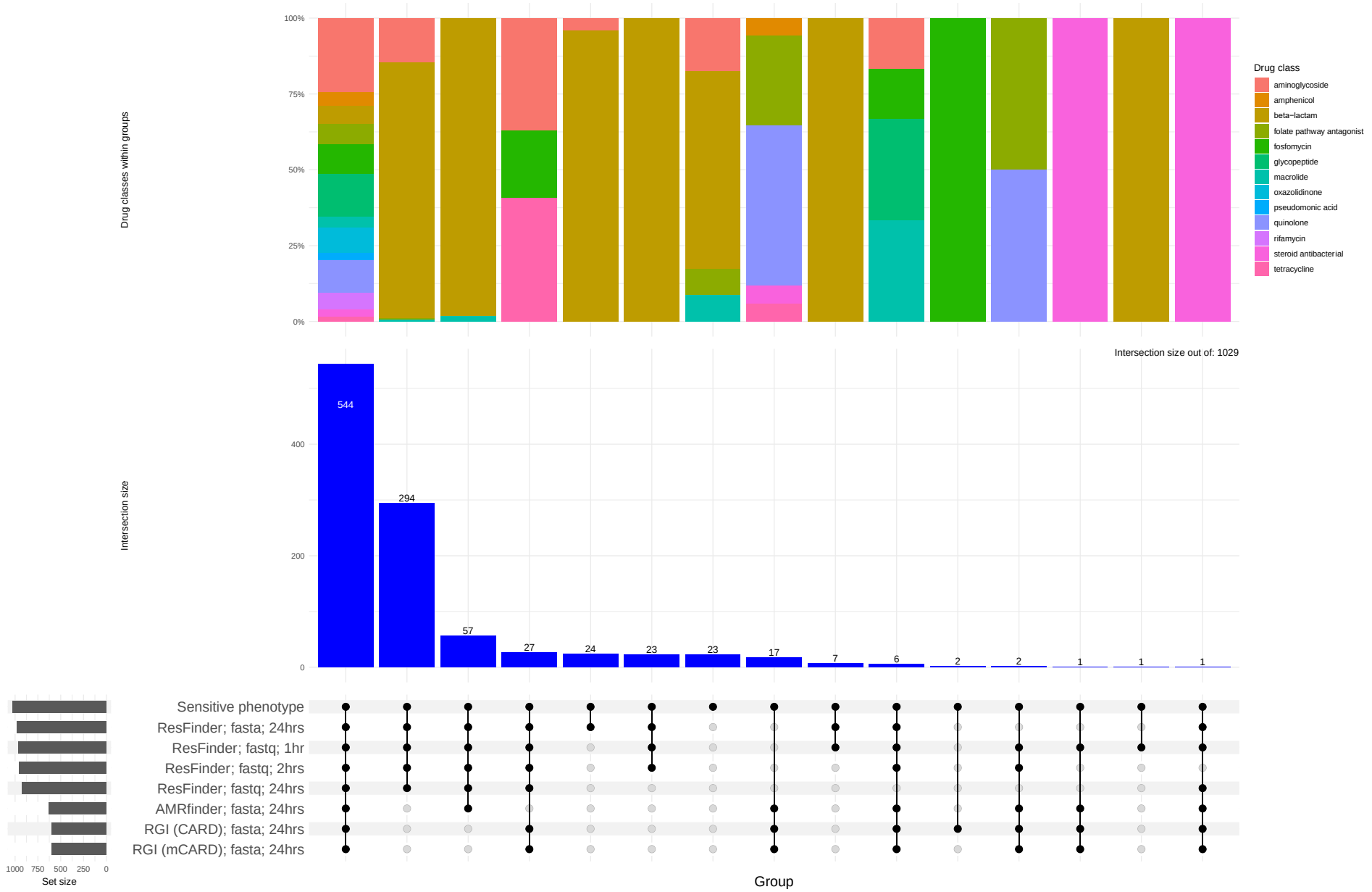

Figure 72. Upset plot showing various groups within sensitive phenotypes.

7.3.2.5 Genes implicated in true calls.

In the analysis, ResFinder, when provided with fastq reads at a 24-hour interval, exhibited superior performance. Consequently, I present the genes identified in association with these calls here. As multiple genes may be detected in a single sample, I present aggregated data across all genes implicated in true calls.

In the analysis, the most frequently observed calls for each drug were as follows:

- **Amoxicillin:** *blaTEM* (n=324), *blaZ* (n=91), *blaCTX* (n=90), *blaSHV* (n=70), and *mecA* (n=44).
- **Ampicillin+clavulanic acid:** *blaTEM* (n=77).
- **Ceftazidime:** *blaCTX* (n=45).
- **Ceftriaxone:** *blaCTX* (n=90).
- **Chloramphenicol:** *cat* (n=1).
- **Ciprofloxacin:** *aac(6')-Ib-cr* (n=3), and *gyrA* (n=2).
- **Clindamycin:** *erm(C)* (n=62), and *erm(A)* (n=5).
- **Erythromycin:** *erm(C)* (n=62) and *erm(A)* (n=6).
- **Fusidic acid:** *fusB* (n=18).
- **Gentamicin:** *aac(6')-aph(2'')* (n=26), and *aph(2'')-Ia* (n=23).
- **Penicillin:** *blaZ* (n=46).
- **Piperacillin+tazobactam:** *blaTEM* (n=24).
- **Teicoplanin:** *VanHAX* (n=12).

- **Tetracycline:** *tet(K)* (n=10).
- **Tobramycin:** *aac(3)-IIa* (n=9).
- **Trimethoprim:** *dfrA1* (n=21).
- **Vancomycin:** *VanHAX* (n=12).

Further detailed true calls can be found in **Figure 73**.

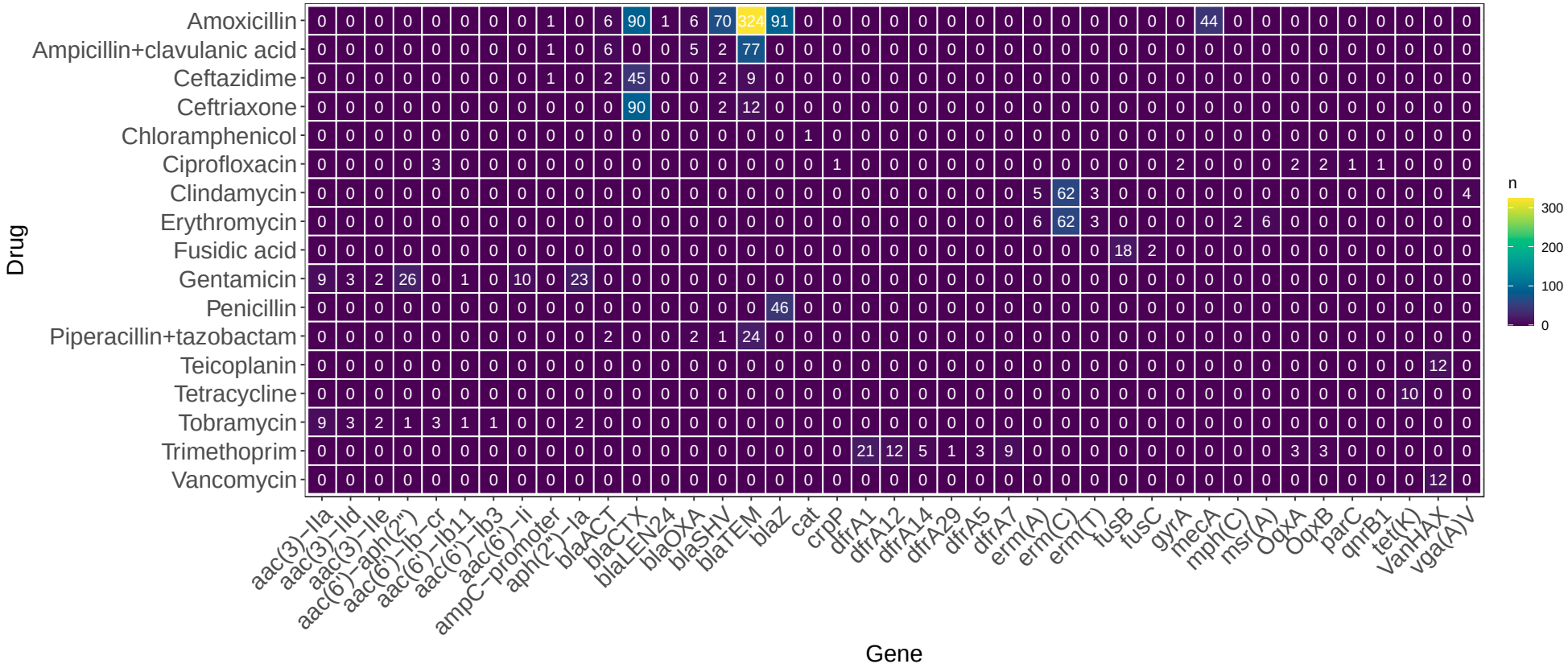

Figure 73. Genes implicated in true genotypic calls (ResFinder with fastq reads at 24hr).

*For conciseness, variants of the genes *blaSHV*, *blaTEM*, *blaCTX*, *blaOXA*, *blaACT*, and *cat* have been consolidated.

7.3.2.6 Genes implicated in major error (false resistance) calls.

As ResFinder [Input: fastq reads at 24hrs] had the best overall performance, understanding what genes led to false positive genotypic results are important to improve predictions and understand which genes are clinically significant. It is important to note that multiple genes may be encoded by the ResFinder mapping data that confer resistance to the same drug, and therefore aggregate implicated genes are presented.

Reported detection of *bla*$_{\text{TEM}}$ genes was responsible for the most major error (ME) or false resistance calls (295x incorrect gene calls[IGC]) followed by *bla*$_{\text{SHV}}$ (52x IGC), *bla*$_{\text{PAO}}$ (19x IGC), *bla*$_{\text{CTX}}$ (15x IGC) and *bla*$_{\text{OXA}}$ (9x IGC) which were particularly seen in ceftazidime, ceftriaxone and piperacillin+tazobactam (**Figure 74**). Detection of *OqxA* (11x IGC) and *OqxB* (13x IGC) were responsible for ME calls in ciprofloxacin and trimethoprim, *erm(C)* (10x IGC) in clindamycin and erythromycin and *fosA* (4x IGC) in fosfomycin.

To further investigate *bla* genes implicated in ME calls, grouped *bla* gene sequences are shown in **Figure 75** and **Figure 76**. In ceftazidime and ceftriaxone, *bla*$_{\text{TEM-143}}$ (25x IGC), *bla*$_{\text{TEM-15}}$ (24x IGC), *bla*$_{\text{TEM-207}}$ (24x IGC), *bla*$_{\text{TEM-213}}$ (18x IGC), and *bla*$_{\text{TEM-47}}$ (12x IGC) were the most frequent genes responsible for a false call. However, in true positive calls very few of the same genes were responsible for correct resistant calls. For instance, *bla*$_{\text{TEM-183}}$ was able to differentiate and call the correct resistance in amoxicillin without any false resistance calls in ceftazidime

and ceftriaxone. Only *bla*$_{\text{TEM-102}}$ and most *bla*$_{\text{CTX-M}}$ genes was able to discriminate resistance in ceftazidime and ceftriaxone without any false positive calls.

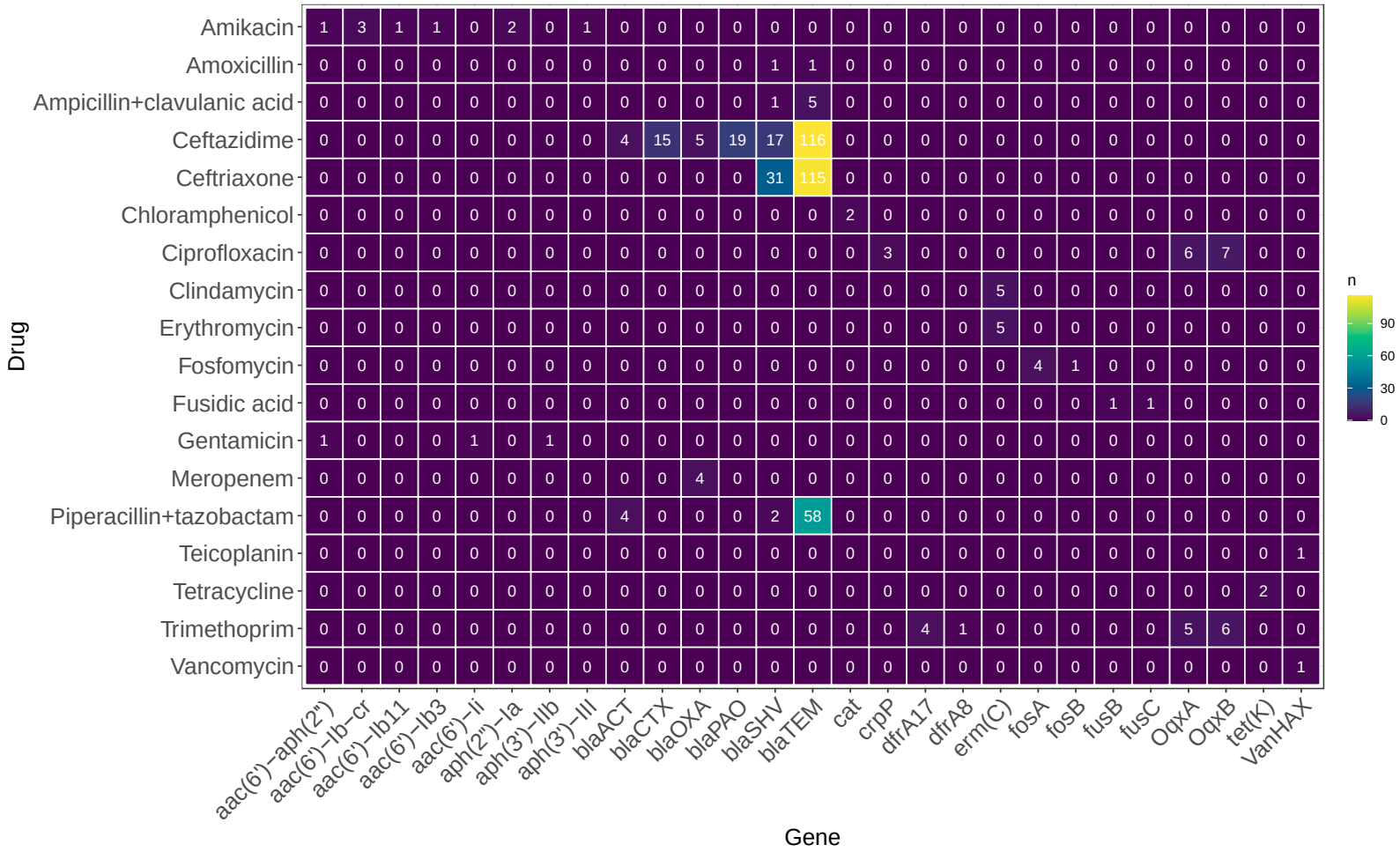

Figure 74. Genes implicated in major error genotypic calls (ResFinder with fastq reads at 24hr).

*For conciseness, variants of the genes *blaSHV*, *blaTEM*, *blaCTX*, *blaOXA*, *blaACT*, and *cat* have been consolidated.

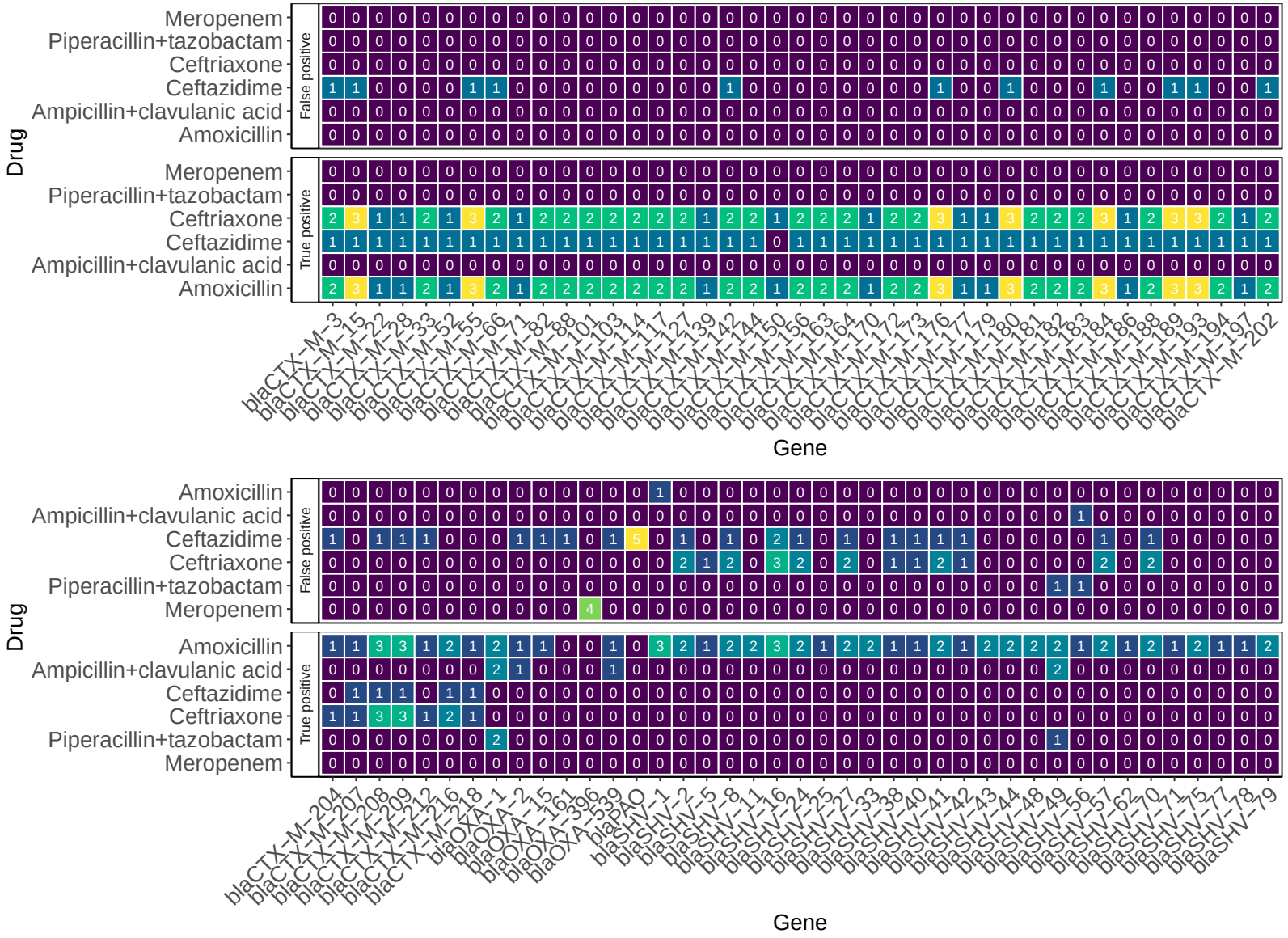

Figure 75. Specific *bla* genes implicated in false and true positive genotypic calls (1/2).

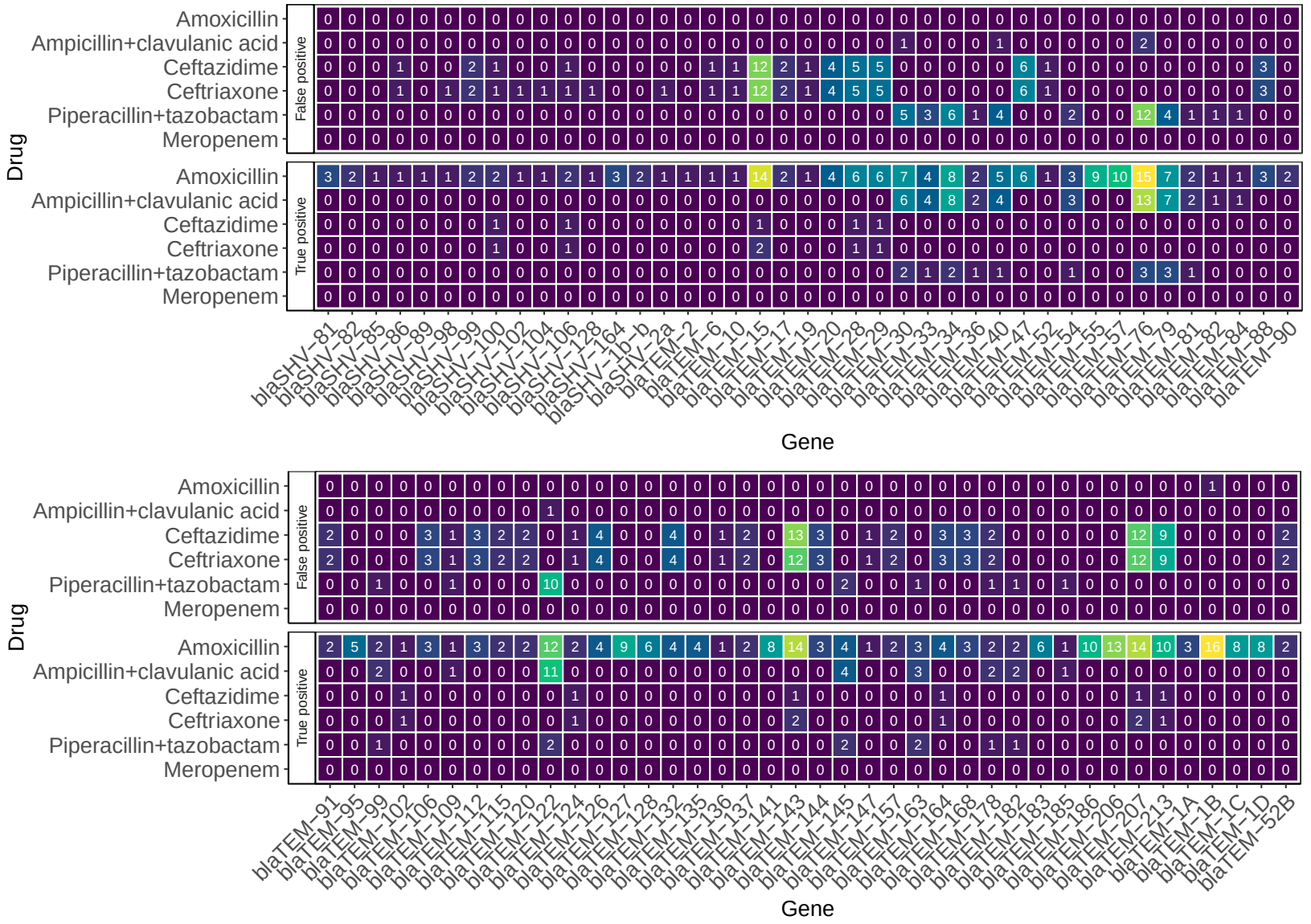

Figure 76. Specific *bla* genes implicated in false and true positive genotypic calls (2/2).

7.3.2.7 Genes to improve performance of ResFinder to reduce Very Major Errors (VMEs).

As ResFinder [Input: fastq reads at 24hrs] had the overall best performance, results from other tools where a resistant call was made (by specific genes) were compared to false negative results by ResFinder to identify possibly missed or unidentified genes (**Table 38**). AMRFinder and RGI (CARD) found numerous genes not found in the ResFinder database including *blaEC*, *blaACT*, *Klebsiella pneumoniae OmpK37* for several beta-lactams and *dfrS1* and *dfrC* for trimethoprim. Although *blaOKP-A-3, tet(L), blaSHV-28, blaCTX-M-15* and *cepA* was present in the ResFinder database, the tool failed to pick up resistance in specific drugs as these genes were encoded to confer resistance at a class level, explaining why AMRFinder and RGI (CARD) both correctly identified this. No tool was able to predict phenotypic resistance in 9 species-drug combinations especially in aminoglycosides and quinolones (**Table 39**).

Table 38. Genes identified in other tools to improve VME call in ResFinder.

Species	Drug	Drug class	Phenotypic result	Resfinder fastq (24h)	AMRFinder	RGI (CARD)
Escherichia coli	Temocillin	Beta-lactam	R	S	blaEC	Klebsiella pneumoniae OmpK37
Escherichia coli	Amoxicillin	Beta-lactam	R	S	blaEC	Klebsiella pneumoniae OmpK37
Escherichia coli	Ampicillin+clavulanic acid	Beta-lactam	R	S	blaEC	Klebsiella pneumoniae OmpK37
Klebsiella pneumoniae	Amoxicillin	Beta-lactam	R	S	blaOKP-A-3	Klebsiella pneumoniae OmpK37
Klebsiella pneumoniae	Amoxicillin	Beta-lactam	R	S	blaOKP-A-3	OKP-A-3
Enterobacter cloacae	Ceftriaxone	Beta-lactam	R	S	blaACT	Klebsiella pneumoniae OmpK37
Enterobacter cloacae	Temocillin	Beta-lactam	R	S	blaACT	Klebsiella pneumoniae OmpK37
Klebsiella pneumoniae	Temocillin	Beta-lactam	R	S	blaSHV-28, blaCTX-M-15	Klebsiella pneumoniae OmpK37
Klebsiella pneumoniae	Temocillin	Beta-lactam	R	S	blaSHV-28, blaCTX-M-15	SHV-28

Klebsiella pneumoniae	Temocillin	Beta-lactam	R	S	blaSHV-28, blaCTX-M-15	CTX-M-15
Staphylococcus epidermidis	Trimethoprim	Folate pathway antagonist	R	S	dfrS1	dfrC
Enterococcus faecium	Tigecycline	Tetracycline	R	S	tet(L)	NA
Escherichia coli	Temocillin	Beta-lactam	R	S	blaEC	Klebsiella pneumoniae OmpK37
Escherichia coli	Amoxicillin	Beta-lactam	R	S	blaEC	Klebsiella pneumoniae OmpK37
Escherichia coli	Ampicillin+clavulanic acid	Beta-lactam	R	S	blaEC	Klebsiella pneumoniae OmpK37
Klebsiella pneumoniae	Amoxicillin	Beta-lactam	R	S	blaOKP-A-3	Klebsiella pneumoniae OmpK37

Table 39. Organism-drug combinations where no tool predicted resistance.

Species	Drug	Drug class	Phenotypic result	Genotypic result (across all tools)
Proteus mirabilis	Gentamicin	Aminoglycoside	R	S
Proteus mirabilis	Gentamicin	Aminoglycoside	R	S
Proteus mirabilis	Gentamicin	Aminoglycoside	R	S
Proteus mirabilis	Tobramycin	Aminoglycoside	R	S
Proteus mirabilis	Amoxicillin	Beta-lactam	R	S
Enterococcus faecium	Linezolid	Oxazolidinone	R	S
Staphylococcus haemolyticus	Ciprofloxacin	Quinolone	R	S
Staphylococcus capitis	Ciprofloxacin	Quinolone	R	S
Staphylococcus haemolyticus	Rifampicin	Rifamycin	R	S

7.3.2.8 Time to result

Time from blood culture positivity to sequencing start and speciation has previously been reported in Chapter 6 (6.11 Time to Speciation).

7.3.2.8.1 Laboratory time to AMR result

From positive blood culture, the routine microbiology lab took a median time of 27hr20min (IQR, 24-34hrs) [Range, 9-119hrs] to produce an AMR result.

7.3.2.8.2 Metagenomic time to AMR result

From positive blood culture, the metagenomic pipeline took a median duration of 3hr20min (IQR, 3hr7min-3hr27min) [Range, 2hr45min-3hr48min] to the start of sequencing. An hour after sequencing (including analysis time) yielded AMR results with a sensitivity and specificity of 82% (95% CI: 76-87%) and 94% (95%

CI: 92-95%) respectively, and two hours after sequencing (including analysis time) 88% (95% CI, 82-92%) and 93% (91-94%) respectively across the 10 most clinically relevant pathogens.

7.3.2.9 Performance in polymicrobial infections

The overall sensitivity and specificity at the drug level using ResFinder [Input: fastq reads at 24hrs] for clinically relevant polymicrobial isolates was 96% (95% CI, 79-100%, 26/27) and 74% (56-89%, 26/35) respectively. Drugs only relevant to specific organisms were tested for depending on the ResFinder database as previously detailed. Performance by individual sample is shown in **Table 40**. In cases where discordant phenotypic susceptibility results existed in polymicrobial species, a genotypic result of resistant was required to correctly predict this phenotype, counting as a single drug prediction. The polymicrobial sample *Enterococcus faecium* and *Staphylococcus haemolyticus* had the highest sensitivity and specificity performance of 100% (10/10) and 100% (4/4) respectively, while the *Escherichia coli* and *Providencia rettgeri* sample had the lowest of 0% (0/5) and 100% (9/9) respectively.

Table 40. Phenotypic sensitivity and specificity for polymicrobial isolates.

MALDI species	MALDI additional species	Sensitivity (Resistant agreement)	Specificity (Sensitive agreement)	Major Error rate	Very Major Error rate
Samples that include only top 10 clinically relevant species					
Klebsiella pneumoniae	*Escherichia coli*	100% (95% CI: 29-100%, 3/3)	54.5% (95% CI: 23-83%, 6/11)	45.5%	0%
Staphylococcus haemolyticus	*Staphylococcus capitis**	66.7% (9-99%, 2/3)	90.9% (59-100%, 10/11)	9.1%	33.3%
Enterococcus faecium	*Staphylococcus haemolyticus*	100% (10/10)	100% (4/4)	0%	0%
Enterococcus faecalis	*Pseudomonas aeruginosa*	100% (3-100%, 1/1)	66.7% (30-93%, 6/9)	33.3%	0%
Overall performance		96% (79-100%, 26/27)	74% (56%-89%, 26/35)	26%	4%
Samples including any additional species					
Escherichia coli	*Klebsiella aerogenes*	75% (19-99%, 3/4)	54.5% (23-83%, 6/11)	45.5%	25%
Lactobacillus rhamnosus	*Klebsiella pneumoniae*	50% (1-99%, 1/2)	75% (43-95%, 9/12)	25%	50%
Staphylococcus epidermidis	*Leuconostoc mesenteroides*	100% (40-100%, 4/4)	60% (26-88%, 6/10)	40%	0%
Escherichia coli	*Providencia rettgeri*	0% (0-52%, 0/5)	100% (66-100%, 9/9)	0%	100%
Proteus mirabilis	*Citrobacter koseri*	66.7% (9-99%, 2/3)	81.8% (48-98%, 9/11)	18.2%	33.3%

* *Staphylococcus capitis* was not tested for phenotypic resistance.

7.4 Discussion

This study demonstrates that clinical metagenomic sequencing can provide rapid, comprehensive, and actionable targeted antimicrobial therapy direct-from-positive blood culture to treat bloodstream infections within 5hr30min, with a sensitivity and specificity of 88% (95% CI, 82-92%) and 93% (91-94%) respectively (in the top 10 most clinically relevant bloodstream infections). Notably, this method delivered results 22 hours earlier than conventional microbiology laboratory techniques for analogous samples.

Across the top 10 clinically relevant species (including *Coagulase-negative staphylococci*), ResFinder [Input: fastq reads at 24hrs] had the highest overall AMR sensitivity and specificity (91% and 90% respectively). Similar performance could be obtained from raw reads after only two hours of sequencing, using ResFinder [Input: fastq reads at 2hrs] with an overall sensitivity of 88% and specificity of 93%. Poor sensitivity for detecting phenotypic resistance was seen in *Proteus mirabilis* and *Enterobacter cloacae*, while poor specificity was seen in *Pseudomonas aeruginosa*, *Enterobacter cloacae*, and *Klebsiella pneumoniae*.

Across all the classifiers used, only *Staphylococcus aureus*, *Escherichia coli*, *Enterococcus sp.*, *Klebsiella pneumonia*, *Enterobacter cloacae,* and *Coagulase-negative staphylococci* from this study met the FDA recommended standards with a ME rate of ≤3% and VME of ≤1.5% with an upper limit of ≤7.5%.[205] Therefore, further research is required for many of these major pathogens to better understand

and characterise the mechanism of resistance to improve the ability to predict this genotypically.

Sensitivity and specificity for antimicrobial resistance prediction using ResFinder at the two-hour interval for *Staphylococcus aureus* was 100% [95% CI, 77-100%] (12/12) and 99% [95% CI, 96-100%] (138/139), and at the 24-hour interval was 100% [95% CI, 77-100%] (12/12) and 97% [95% CI, 93-99%] (135/139). One major error was seen each in clindamycin, erythromycin, teicoplanin and vancomycin finding the *erm(C)* or *VanHAX* gene, perhaps suggesting intermediate resistance.

For *Escherichia coli,* sensitivity and specificity using ResFinder at the two-hour interval was 91% [95% CI, 80-97%] (53/58) and 94% [95% CI, 92-96%] (505/539), and at the 24-hour interval was 93% [95% CI, 83-98%] (54/58) and 91% [95% CI, 88-93%] (490/539). Four very major errors were seen in beta-lactams (temocillin, amoxicillin, ampicillin + clavulanic acid). Two were possibly due to the incomplete encoding of the *ampC-promoter* gene to temocillin resistance by the ResFinder database. The other two major errors were from the same isolate, where either genes that might be implicated in beta-lactam resistance have not been characterized or the phenotype result might be inaccurate. There were 49 major errors, of which 46 were in beta-lactams. Improved characterization and mapping of genes to specific drugs might reduce these errors, as often any phenotypic beta-lactam resistance in a sample would be mapped to multiple specific beta-lactam drugs as genotypically resistant resulting in major errors. This study found specific genes such as $bla_{\text{TEM-183}}$ was able to differentiate and call the correct resistance in amoxicillin without

any false resistance calls in ceftazidime and ceftriaxone. Similarly, *bla*$_{\text{TEM-102}}$ and most *bla*$_{\text{CTX-M}}$ genes was able to discriminate resistance in ceftazidime and ceftriaxone without any major errors.

A principal biological impediment resides in our depth of understanding concerning the genotype-to-phenotype association. It is imperative that the genotype serves as a robust predictor of resistance. Notably, this association is becoming increasingly conspicuous in select bacterium-drug pairings. Both rule- and model-based approaches have been applied for AMR prediction from pure isolates across several species including *Staphylococcus aureus, Escherichia coli, Mycobacterium tuberculosis, Salmonella enterica, Streptococcus pneumoniae, Enterococcus faecalis, Enterococcus faecium, Pseudomonas aeruginosa* and *Klebsiella pneumoniae* with various levels of performance (sensitivities of 88-100%, and specificities of 90-100%).[206] Specifically, previous studies have demonstrated high performance for *Staphylococcus aureus* with sensitivities of 97-99% and specificities of >99% for at least 12 antibiotics, which was consistent with the finding in this study.[207,208] Although a previous study using WGS from pure isolates demonstrated slightly higher performance for *Escherichia coli* resistance prediction with a sensitivity and specificity of 99% and 96% respectively, the study was limited to seven antibiotics whereas 26 antibiotics are considered in this study.[209]

Variation in AMR predictive performance across species may be due to several factors. Principal among these is the extent to which resistance mechanisms within certain species are characterized, and the wealth of correlating genomic and phenotypic data collected. Additionally, the propensity for clonal spread, the

complexity of resistance pathways, and the prevalence of horizontal gene transfer within a species can modulate the ease of genotypic predictions.[210] Furthermore, the inherent genomic diversity and evolutionary rate of a species can either facilitate or complicate the identification of consistent resistance markers. Genotypic predictions tend to be more straightforward for Gram-positive bacteria, such as *Staphylococcus aureus* or *Mycobacterium tuberculosis* because the number of resistance mechanisms in these species are limited.[211] By contrast, resistant Gram-negative bacteria, such as species in the Enterobacteriaceae, *Acinetobacter* spp. and *Pseudomonas* spp., have many different resistance determinants (resistance genes and mutations), and often the phenotypic resistance results from a combination of determinants that encode resistance mechanisms acting together at different sites in the bacterial cell. Therefore, as resistance is so complex, only a subset of resistance mechanisms may be picked up.[211]

The poorest performing resistance prediction in this study was seen in quinolones (71%), tetracyclines (83%) and beta-lactams (84%) particularly in ceftriaxone (75%), ciprofloxacin (86%), tobramycin (86%) and amoxicillin (89%). The difficulty of curating point mutations for a broad number of species is evident and might explain the poor performance seen in quinolones. Predicting drug resistance varies in complexity depending on the drug in question. Drugs with clear resistance mechanisms, like rifampicin where mutations in the *rpoB* gene indicate resistance, simplify predictions. Stability of resistance mechanisms also influences predictability; consistent genetic markers allow for easier predictions, whereas transient resistance complicates them. Additionally, drugs targeted by multiple resistance methods, such as efflux pumps and enzyme modifications, present a greater challenge for genomic-based predictions.

Although 59% of phenotypic resistance calls were called by all databases used, an additional 13% was called in ResFinder-based tools, while an additional 5% was called in AMRFinder and RGI (CARD) tools suggesting that there is potential to integrate databases to expand the resistance catalogue and improve prediction. For example, there were several genes found in the AMRFinder and RGI (CARD) databases that conferred resistance, however, were not present in the Resfinder-based database such as *blaEC*, *blaACT* and *blaOKP-A-33*. Variations in the performance of tools primarily are due to differences in their database composition and algorithmic approaches. For instance, the ResFinder database is highly curated for predicting genes associated with antimicrobial resistance, especially those of clinical relevance, whereas the CARD database is primarily geared to serve as an extensive catalogue for resistance genes.

Overall sensitivity and specificity for clinically relevant polymicrobial isolates was 96% (95% CI, 79-100%) and 74% (56%-89%) respectively. A lower specificity was seen in these samples due to numerous false positive results within the beta-lactam class. Where a single drug within the class may contain phenotypic resistance, the genotypic resistance appears to be overestimated within the same class, which might explain poor performance within the beta-lactam class.

Further work in this study might involve limiting genes identified that contribute to MEs, and reducing false negative predictions by enriching the database with relevant genes that may be present in other databases to reduce VMEs. Using personalized database thresholds for specific species rather than a uniform

threshold applied to the whole database may also improve performance, however, would require a much higher organism-drug sample size in order to validate. Furthermore, larger training and testing datasets are ultimately required to obtain regulatory approval for clinical use.

There were several limitations to this study. Only categorical results were considered in this analysis, while MIC results might better represent the phenotype, especially for samples that may have borderline breakpoints. Furthermore, routine clinical laboratory results from the BD Phoenix system using EUCAST breakpoints[212] was used rather than disc diffusion resistance testing. Bioinformatically, the plasmid copy number could not be determined to infer the contribution of multiple gene copies on resistance, for example, bla_{TEM} might lead to hyperproduction of enzymes leading to phenotypic resistance by a gene dosage effect. A significant constraint was the limited power of samples for resistance prediction, especially considering drug-bug combinations which resulted in varied confidence intervals. Future research should consider larger samples and a more extensive range of drug-bug combinations to enhance the robustness of findings. Additionally, as this study was a DNA-based analysis, epigenetic and gene expression mechanisms could not be determined.

In summary, predicting AMR using a direct-from-positive blood culture metagenomic approach is feasible and achieved high sensitivity and specificity in a record time of <6 hours from sample receipt, producing comprehensive and actionable AMR results 22hrs prior to the conventional laboratory. Clinical metagenomics may well become part of routine microbiological workflows in the

future allowing for unbiased (including unculturable and polymicrobial) pathogen identification, fast AMR prediction reliably and bacterial strain typing for infection prevention and control or epidemiological reasons.

Supplementary Table 11. Phenotypic resistance and sensitivity by species-drug combination.

Species	Drug	Resistant phenotype	Sensitive phenotype	Total phenotype
Coagulase-negative staphylococci	Penicillin	4 (100%)	0 (0%)	4
Coagulase-negative staphylococci	Tobramycin	1 (100%)	0 (0%)	1
Coagulase-negative staphylococci	Trimethoprim	1 (100%)	0 (0%)	1
Coagulase-negative staphylococci	Amoxicillin	17 (100%)	0 (0%)	17
Coagulase-negative staphylococci	Erythromycin	16 (84%)	3 (16%)	19
Coagulase-negative staphylococci	Gentamicin	16 (84%)	3 (16%)	19
Coagulase-negative staphylococci	Clindamycin	13 (72%)	5 (28%)	18
Coagulase-negative staphylococci	Ciprofloxacin	2 (67%)	1 (33%)	3
Coagulase-negative staphylococci	Fusidic acid	12 (63%)	7 (37%)	19
Coagulase-negative staphylococci	Tetracycline	5 (38%)	8 (62%)	13
Coagulase-negative staphylococci	Chloramphenicol	1 (7%)	14 (93%)	15
Coagulase-negative staphylococci	Rifampicin	1 (5%)	18 (95%)	19
Coagulase-negative staphylococci	Amikacin	0 (0%)	1 (100%)	1
Coagulase-negative staphylococci	Fosfomycin	0 (0%)	2 (100%)	2
Coagulase-negative staphylococci	Linezolid	0 (0%)	19 (100%)	19
Coagulase-negative staphylococci	Mupirocin	0 (0%)	1 (100%)	1
Coagulase-negative staphylococci	Teicoplanin	0 (0%)	17 (100%)	17
Coagulase-negative staphylococci	Vancomycin	0 (0%)	19 (100%)	19

Enterobacter cloacae	Amoxicillin	3 (100%)	0 (0%)	3
Enterobacter cloacae	Ampicillin + clavulanic acid	3 (100%)	0 (0%)	3
Enterobacter cloacae	Ceftazidime	1 (33%)	2 (67%)	3
Enterobacter cloacae	Ceftriaxone	1 (33%)	2 (67%)	3
Enterobacter cloacae	Piperacillin + tazobactam	1 (33%)	2 (67%)	3
Enterobacter cloacae	Temocillin	1 (33%)	2 (67%)	3
Enterobacter cloacae	Amikacin	0 (0%)	3 (100%)	3
Enterobacter cloacae	Ciprofloxacin	0 (0%)	3 (100%)	3
Enterobacter cloacae	Ertapenem	0 (0%)	3 (100%)	3
Enterobacter cloacae	Fosfomycin	0 (0%)	3 (100%)	3
Enterobacter cloacae	Gentamicin	0 (0%)	3 (100%)	3
Enterobacter cloacae	Meropenem	0 (0%)	3 (100%)	3
Enterobacter cloacae	Tobramycin	0 (0%)	3 (100%)	3
Enterobacter cloacae	Trimethoprim	0 (0%)	3 (100%)	3
Enterococcus sp.	Tigecycline	1 (100%)	0 (0%)	1
Enterococcus sp.	Gentamicin	10 (71%)	4 (29%)	14
Enterococcus sp.	Teicoplanin	6 (40%)	9 (60%)	15
Enterococcus sp.	Vancomycin	6 (40%)	9 (60%)	15
Enterococcus sp.	Linezolid	1 (6%)	15 (94%)	16
Enterococcus sp.	Amoxicillin	0 (0%)	3 (100%)	3
Klebsiella pneumoniae	Amoxicillin	7 (100%)	0 (0%)	7
Klebsiella pneumoniae	Ampicillin + clavulanic acid	3 (43%)	4 (57%)	7
Klebsiella pneumoniae	Trimethoprim	3 (43%)	4 (57%)	7
Klebsiella pneumoniae	Ceftriaxone	2 (29%)	5 (71%)	7
Klebsiella pneumoniae	Ciprofloxacin	2 (29%)	5 (71%)	7
Klebsiella pneumoniae	Gentamicin	2 (29%)	5 (71%)	7
Klebsiella pneumoniae	Piperacillin + tazobactam	2 (29%)	5 (71%)	7
Klebsiella pneumoniae	Tobramycin	2 (29%)	5 (71%)	7

Klebsiella pneumoniae	Ceftazidime	1 (14%)	6 (86%)	7
Klebsiella pneumoniae	Temocillin	1 (14%)	6 (86%)	7
Klebsiella pneumoniae	Amikacin	0 (0%)	7 (100%)	7
Klebsiella pneumoniae	Ertapenem	0 (0%)	7 (100%)	7
Klebsiella pneumoniae	Meropenem	0 (0%)	7 (100%)	7
Proteus mirabilis	Gentamicin	3 (75%)	1 (25%)	4
Proteus mirabilis	Amoxicillin	1 (25%)	3 (75%)	4
Proteus mirabilis	Tobramycin	1 (25%)	3 (75%)	4
Proteus mirabilis	Trimethoprim	1 (25%)	3 (75%)	4
Proteus mirabilis	Amikacin	0 (0%)	4 (100%)	4
Proteus mirabilis	Ampicillin + clavulanic acid	0 (0%)	4 (100%)	4
Proteus mirabilis	Ceftazidime	0 (0%)	4 (100%)	4
Proteus mirabilis	Ceftriaxone	0 (0%)	4 (100%)	4
Proteus mirabilis	Ciprofloxacin	0 (0%)	4 (100%)	4
Proteus mirabilis	Ertapenem	0 (0%)	4 (100%)	4
Proteus mirabilis	Fosfomycin	0 (0%)	4 (100%)	4
Proteus mirabilis	Meropenem	0 (0%)	4 (100%)	4
Proteus mirabilis	Piperacillin + tazobactam	0 (0%)	4 (100%)	4
Proteus mirabilis	Temocillin	0 (0%)	4 (100%)	4
Pseudomonas aeruginosa	Amikacin	0 (0%)	5 (100%)	5
Pseudomonas aeruginosa	Ceftazidime	0 (0%)	5 (100%)	5
Pseudomonas aeruginosa	Ciprofloxacin	0 (0%)	5 (100%)	5
Pseudomonas aeruginosa	Gentamicin	0 (0%)	5 (100%)	5
Pseudomonas aeruginosa	Meropenem	0 (0%)	4 (100%)	4
Pseudomonas aeruginosa	Piperacillin + tazobactam	0 (0%)	5 (100%)	5

Pseudomonas aeruginosa	Tobramycin	0 (0%)	5 (100%)	5
Streptococcus pneumoniae	Amoxicillin	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Ceftriaxone	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Chloramphenicol	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Erythromycin	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Gentamicin	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Linezolid	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Meropenem	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Penicillin	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Teicoplanin	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Tetracycline	0 (0%)	1 (100%)	1
Streptococcus pneumoniae	Vancomycin	0 (0%)	1 (100%)	1

Supplementary Table 12. Antibiotic classes and specific drugs.

No.	# Class	Antibiotics	No.	# Class	Antibiotics
1	Aminocyclitol	Spectinomycin	51	Beta-lactam	Ceftazidime
2	Aminocyclitol	Unknown Aminocyclitol	52	Beta-lactam	Ceftazidime+Avibactam
3	Aminoglycoside	Amikacin	53	Beta-lactam	ceftiofur
4	Aminoglycoside	Apramycin	54	Beta-lactam	Ceftriaxone
5	Aminoglycoside	Arbekacin	55	Beta-lactam	Cefuroxime
6	Aminoglycoside	Astromicin	56	Beta-lactam	Cephalothin
7	Aminoglycoside	Bleomycin	57	Beta-lactam	Cephalotin
8	Aminoglycoside	Butiromycin	58	Beta-lactam	pivmecillinam
9	Aminoglycoside	Butirosin	59	Beta-lactam	piptaz
10	Aminoglycoside	Capreomycin	60	Beta-lactam	methicillin
11	Aminoglycoside	Dibekacin	61	Beta-lactam	Ertapenem
12	Aminoglycoside	Fortimicin	62	Beta-lactam	Imipenem
13	Aminoglycoside	G418	63	Beta-lactam	Meropenem
14	Aminoglycoside	Gentamicin	64	Beta-lactam	Penicillin
15	Aminoglycoside	gentamycin	65	Beta-lactam	flucloxacillin
16	Aminoglycoside	Gentamicin c	66	Beta-lactam	Piperacillin
17	Aminoglycoside	Hygromycin	67	Beta-lactam	Piperacillin+Clavulanic acid
18	Aminoglycoside	Isepamicin	68	Beta-lactam	Piperacillin+Tazobactam
19	Aminoglycoside	Kanamycin	69	Beta-lactam	Temocillin
20	Aminoglycoside	Kanamycin a	70	Beta-lactam	Ticarcillin
21	Aminoglycoside	Kasugamycin	71	Beta-lactam	Ticarcillin+Clavulanic acid
22	Aminoglycoside	Lividomycin	72	Beta-lactam	Unknown Beta-lactam
23	Aminoglycoside	Neomycin	73	cephalosporin	cephalosporins
24	Aminoglycoside	Netilmicin	74	cephalosporin	cephalosporin
25	Aminoglycoside	Paromomycin	75	Diarylquinoline	Bedaquiline
26	Aminoglycoside	Ribostamycin	76	Diarylquinoline	Unknown Diarylquinoline
27	Aminoglycoside	Sisomicin	77	Folate pathway antagonist	Sulfamethoxazole
28	Aminoglycoside	Streptomycin	78	Folate pathway antagonist	Trimethoprim
29	Aminoglycoside	Tobramycin	79	Folate pathway antagonist	trimeth−sulfamethoxaz
30	Aminoglycoside	Unknown Aminoglycoside	80	Folate pathway antagonist	diaminopyrimidine
31	Amphenicol	Chloramphenicol	81	Folate pathway antagonist	sulfonamide
32	Amphenicol	Florfenicol	82	Folate pathway antagonist	Unknown Folate pathway antagonist
33	Amphenicol	Unknown Amphenicol	83	Fosfomycin	Fosfomycin
34	Analog of d-alanine	D-Cycloserine	84	Fosfomycin	Unknown Fosfomycin
35	Analog of d-alanine	Unknown Analog of d-alanine	85	Glycopeptide	Teicoplanin
36	Beta-lactam	Amoxicillin	86	Glycopeptide	Unknown Glycopeptide
37	Beta-lactam	Amoxicillin+Clavulanic acid	87	Glycopeptide	Vancomycin
38	Beta-lactam	Ampicillin	88	Iminophenazine	Clofazimine
39	Beta-lactam	ampicillin+clavulanic acid	89	Iminophenazine	Unknown Iminophenazine
40	Beta-lactam	penam	90	Isonicotinic acid hydrazide	Isoniazid
41	Beta-lactam	penem	91	Isonicotinic acid hydrazide	Unknown Isonicotinic acid hydrazide
42	Beta-lactam	Aztreonam	92	Lincosamide	Lincomycin
43	Beta-lactam	carbapenem	93	Lincosamide	Unknown Lincosamide
44	Beta-lactam	Cefazolin	94	lipopeptide	daptomycin
45	Beta-lactam	Cefepime	95	Macrolide	Azithromycin
46	Beta-lactam	Cefixime	96	Macrolide	Carbomycin
47	Beta-lactam	Cefotaxime	97	Macrolide	clindamycin
48	Beta-lactam	Cefotaxime+Clavulanic acid	98	Macrolide	Erythromycin
49	Beta-lactam	Cefoxitin	99	Macrolide	Oleandomycin
50	Beta-lactam	Ceftaroline	100	Macrolide	Spiramycin

Supplementary Table 12. Continued.

No.	# Class	Antibiotics
101	Macrolide	Telithromycin
102	Macrolide	Tylosin
103	Macrolide	Unknown Macrolide
104	monobactam	azteonam
105	Nitroimidazole	Metronidazole
106	Nitroimidazole	Unknown Nitroimidazole
107	Oxazolidinone	Linezolid
108	Oxazolidinone	Unknown Oxazolidinone
109	Pleuromutilin	Tiamulin
110	Pleuromutilin	Unknown Pleuromutilin
111	Polymyxin	Colistin
112	Polymyxin	Unknown Polymyxin
113	Pseudomonic acid	Mupirocin
114	Pseudomonic acid	Unknown Pseudomonic acid
115	Quinolone	Ciprofloxacin
116	Quinolone	Levofloxacin
117	Quinolone	moxifloxacin
118	Quinolone	Fluoroquinolone
119	Quinolone	Nalidixic acid
120	Quinolone	Unknown Quinolone
121	Rifamycin	Rifampicin
122	Rifamycin	Unknown Rifamycin
123	Salicylic acid - anti-folate	Para-aminosalicyclic acid
124	Salicylic acid - anti-folate	Unknown Salicylic acid - anti-folate
125	Steroid antibacterial	Fusidic acid
126	Steroid antibacterial	fusidane
127	Steroid antibacterial	Unknown Steroid antibacterial
128	Streptogramin A	Dalfopristin
129	Streptogramin A	Pristinamycin IIA
130	Streptogramin A	Quinupristin+Dalfopristin
131	Streptogramin A	Unknown Streptogramin A
132	Streptogramin A	Virginiamycin M
133	Streptogramin B	Pristinamycin IA
134	Streptogramin B	Quinupristin
135	Streptogramin B	Unknown Streptogramin B
136	Streptogramin B	Virginiamycin S
137	streptogramins	synercid
138	Tetracycline	Doxycycline
139	Tetracycline	Minocycline
140	Tetracycline	Tetracycline
141	Tetracycline	Tigecycline
142	Tetracycline	Unknown Tetracycline
143	Thioamide	Ethionamide
144	Thioamide	Unknown Thioamide
145	Unspecified	Ethambutol

Supplementary Table 13. Phenotype panel where a genetic resistance call could be made.

No.	Species	Drug	No.	Species	Drug
1	*Escherichia coli*	amikacin	51	*Staphylococcus aureus*	quinupristin+dalfopristin
2	*Escherichia coli*	ampicillin	52	*Staphylococcus aureus*	rifampicin
3	*Escherichia coli*	ampicillin+clavulanic acid	53	*Staphylococcus aureus*	spiramycin
4	*Escherichia coli*	amoxicillin	54	*Staphylococcus aureus*	sulfamethoxazole
5	*Escherichia coli*	azithromycin	55	*Staphylococcus aureus*	teicoplanin
6	*Escherichia coli*	cefepime	56	*Staphylococcus aureus*	telithromycin
7	*Escherichia coli*	cefotaxime	57	*Staphylococcus aureus*	tetracycline
8	*Escherichia coli*	cefoxitin	58	*Staphylococcus aureus*	tiamulin
9	*Escherichia coli*	ceftazidime	59	*Staphylococcus aureus*	tobramycin
10	*Escherichia coli*	ceftriaxone	60	*Staphylococcus aureus*	trimethoprim
11	*Escherichia coli*	chloramphenicol	61	*Staphylococcus aureus*	vancomycin
12	*Escherichia coli*	ciprofloxacin	62	*Klebsiella pneumoniae*	amikacin
13	*Escherichia coli*	colistin	63	*Klebsiella pneumoniae*	amoxicillin
14	*Escherichia coli*	ertapenem	64	*Klebsiella pneumoniae*	amoxicillin+clavulanic acid
15	*Escherichia coli*	fosfomycin	65	*Klebsiella pneumoniae*	ampicillin
16	*Escherichia coli*	gentamicin	66	*Klebsiella pneumoniae*	ampicillin+clavulanic acid
17	*Escherichia coli*	imipenem	67	*Klebsiella pneumoniae*	apramycin
18	*Escherichia coli*	meropenem	68	*Klebsiella pneumoniae*	arbekacin
19	*Escherichia coli*	nalidixic acid	69	*Klebsiella pneumoniae*	astromicin
20	*Escherichia coli*	piperacillin+tazobactam	70	*Klebsiella pneumoniae*	azithromycin
21	*Escherichia coli*	sulfamethoxazole	71	*Klebsiella pneumoniae*	aztreonam
22	*Escherichia coli*	temocillin	72	*Klebsiella pneumoniae*	benzylkonium chloride
23	*Escherichia coli*	tetracycline	73	*Klebsiella pneumoniae*	bleomycin
24	*Escherichia coli*	tigecycline	74	*Klebsiella pneumoniae*	butiromycin
25	*Escherichia coli*	tobramycin	75	*Klebsiella pneumoniae*	butirosin
26	*Escherichia coli*	trimethoprim	76	*Klebsiella pneumoniae*	carbapenem
27	*Enterococcus faecium*	ampicillin	77	*Klebsiella pneumoniae*	carbomycin
28	*Enterococcus faecium*	chloramphenicol	78	*Klebsiella pneumoniae*	cefepime
29	*Enterococcus faecium*	ciprofloxacin	79	*Klebsiella pneumoniae*	cefixime
30	*Enterococcus faecium*	erythromycin	80	*Klebsiella pneumoniae*	cefotaxime
31	*Enterococcus faecium*	fosfomycin	81	*Klebsiella pneumoniae*	cefotaxime+clavulanic acid
32	*Enterococcus faecium*	gentamicin	82	*Klebsiella pneumoniae*	cefoxitin
33	*Enterococcus faecium*	linezolid	83	*Klebsiella pneumoniae*	ceftazidime
34	*Enterococcus faecium*	quinupristin+dalfopristin	84	*Klebsiella pneumoniae*	ceftazidime+avibactam
35	*Enterococcus faecium*	teicoplanin	85	*Klebsiella pneumoniae*	ceftiofur
36	*Enterococcus faecium*	tetracycline	86	*Klebsiella pneumoniae*	ceftriaxone
37	*Enterococcus faecium*	tigecycline	87	*Klebsiella pneumoniae*	cephalosporins
38	*Enterococcus faecium*	vancomycin	88	*Klebsiella pneumoniae*	cephalothin
39	*Staphylococcus aureus*	amikacin	89	*Klebsiella pneumoniae*	cephalotin
40	*Staphylococcus aureus*	cefoxitin	90	*Klebsiella pneumoniae*	cetylpyridinium chloride
41	*Staphylococcus aureus*	chloramphenicol	91	*Klebsiella pneumoniae*	chloramphenicol
42	*Staphylococcus aureus*	ciprofloxacin	92	*Klebsiella pneumoniae*	chlorhexidine
43	*Staphylococcus aureus*	clindamycin	93	*Klebsiella pneumoniae*	ciprofloxacin
44	*Staphylococcus aureus*	erythromycin	94	*Klebsiella pneumoniae*	clindamycin
45	*Staphylococcus aureus*	fosfomycin	95	*Klebsiella pneumoniae*	colistin
46	*Staphylococcus aureus*	fusidic acid	96	*Klebsiella pneumoniae*	dalfopristin
47	*Staphylococcus aureus*	gentamicin	97	*Klebsiella pneumoniae*	dibekacin
48	*Staphylococcus aureus*	linezolid	98	*Klebsiella pneumoniae*	doxycycline
49	*Staphylococcus aureus*	mupirocin	99	*Klebsiella pneumoniae*	ertapenem
50	*Staphylococcus aureus*	penicillin	100	*Klebsiella pneumoniae*	erythromycin

Supplementary Table 13. Continued.

101	*Klebsiella pneumoniae*	ethidium bromide	151	*Klebsiella pneumoniae*	tobramycin
102	*Klebsiella pneumoniae*	florfenicol	152	*Klebsiella pneumoniae*	trimethoprim
103	*Klebsiella pneumoniae*	fluoroquinolone	153	*Klebsiella pneumoniae*	tylosin
104	*Klebsiella pneumoniae*	formaldehyde	154	*Klebsiella pneumoniae*	unknown aminoglycoside
105	*Klebsiella pneumoniae*	fortimicin	155	*Klebsiella pneumoniae*	unknown beta-lactam
106	*Klebsiella pneumoniae*	fusidic acid	156	*Klebsiella pneumoniae*	unknown quinolone
107	*Klebsiella pneumoniae*	gentamicin	157	*Klebsiella pneumoniae*	vancomycin
108	*Klebsiella pneumoniae*	hydrogen peroxide	158	*Klebsiella pneumoniae*	virginiamycin m
109	*Klebsiella pneumoniae*	hygromycin	159	*Klebsiella pneumoniae*	virginiamycin s
110	*Klebsiella pneumoniae*	imipenem	160	*Streptococcus pneumoniae*	amikacin
111	*Klebsiella pneumoniae*	isepamicin	161	*Streptococcus pneumoniae*	amoxicillin
112	*Klebsiella pneumoniae*	kanamycin	162	*Streptococcus pneumoniae*	amoxicillin+clavulanic acid
113	*Klebsiella pneumoniae*	kasugamycin	163	*Streptococcus pneumoniae*	ampicillin
114	*Klebsiella pneumoniae*	lincomycin	164	*Streptococcus pneumoniae*	ampicillin+clavulanic acid
115	*Klebsiella pneumoniae*	linezolid	165	*Streptococcus pneumoniae*	apramycin
116	*Klebsiella pneumoniae*	lividomycin	166	*Streptococcus pneumoniae*	arbekacin
117	*Klebsiella pneumoniae*	meropenem	167	*Streptococcus pneumoniae*	astromicin
118	*Klebsiella pneumoniae*	metronidazole	168	*Streptococcus pneumoniae*	azithromycin
119	*Klebsiella pneumoniae*	minocycline	169	*Streptococcus pneumoniae*	aztreonam
120	*Klebsiella pneumoniae*	mupirocin	170	*Streptococcus pneumoniae*	benzylkonium chloride
121	*Klebsiella pneumoniae*	nalidixic acid	171	*Streptococcus pneumoniae*	bleomycin
122	*Klebsiella pneumoniae*	neomycin	172	*Streptococcus pneumoniae*	butiromycin
123	*Klebsiella pneumoniae*	netilmicin	173	*Streptococcus pneumoniae*	butirosin
124	*Klebsiella pneumoniae*	norfloxacin	174	*Streptococcus pneumoniae*	carbapenem
125	*Klebsiella pneumoniae*	oleandomycin	175	*Streptococcus pneumoniae*	carbomycin
126	*Klebsiella pneumoniae*	paromomycin	176	*Streptococcus pneumoniae*	cefepime
127	*Klebsiella pneumoniae*	penicillin	177	*Streptococcus pneumoniae*	cefixime
128	*Klebsiella pneumoniae*	piperacillin	178	*Streptococcus pneumoniae*	cefotaxime
129	*Klebsiella pneumoniae*	piperacillin+clavulanic acid	179	*Streptococcus pneumoniae*	cefotaxime+clavulanic acid
130	*Klebsiella pneumoniae*	piperacillin+tazobactam	180	*Streptococcus pneumoniae*	cefoxitin
131	*Klebsiella pneumoniae*	pristinamycin ia	181	*Streptococcus pneumoniae*	ceftazidime
132	*Klebsiella pneumoniae*	pristinamycin iia	182	*Streptococcus pneumoniae*	ceftazidime+avibactam
133	*Klebsiella pneumoniae*	quinupristin	183	*Streptococcus pneumoniae*	ceftiofur
134	*Klebsiella pneumoniae*	quinupristin+dalfopristin	184	*Streptococcus pneumoniae*	ceftriaxone
135	*Klebsiella pneumoniae*	ribostamycin	185	*Streptococcus pneumoniae*	cephalosporins
136	*Klebsiella pneumoniae*	rifampicin	186	*Streptococcus pneumoniae*	cephalothin
137	*Klebsiella pneumoniae*	sisomicin	187	*Streptococcus pneumoniae*	cephalotin
138	*Klebsiella pneumoniae*	spectinomycin	188	*Streptococcus pneumoniae*	cetylpyridinium chloride
139	*Klebsiella pneumoniae*	spiramycin	189	*Streptococcus pneumoniae*	chloramphenicol
140	*Klebsiella pneumoniae*	streptomycin	190	*Streptococcus pneumoniae*	chlorhexidine
141	*Klebsiella pneumoniae*	sulfamethoxazole	191	*Streptococcus pneumoniae*	ciprofloxacin
142	*Klebsiella pneumoniae*	teicoplanin	192	*Streptococcus pneumoniae*	clindamycin
143	*Klebsiella pneumoniae*	telithromycin	193	*Streptococcus pneumoniae*	colistin
144	*Klebsiella pneumoniae*	temocillin	194	*Streptococcus pneumoniae*	dalfopristin
145	*Klebsiella pneumoniae*	temperature	195	*Streptococcus pneumoniae*	dibekacin
146	*Klebsiella pneumoniae*	tetracycline	196	*Streptococcus pneumoniae*	doxycycline
147	*Klebsiella pneumoniae*	tiamulin	197	*Streptococcus pneumoniae*	ertapenem
148	*Klebsiella pneumoniae*	ticarcillin	198	*Streptococcus pneumoniae*	erythromycin
149	*Klebsiella pneumoniae*	ticarcillin+clavulanic acid	199	*Streptococcus pneumoniae*	ethidium bromide
150	*Klebsiella pneumoniae*	tigecycline	200	*Streptococcus pneumoniae*	florfenicol

Supplementary Table 13. Continued.

201	*Streptococcus pneumoniae*	fluoroquinolone	251	*Streptococcus pneumoniae*	trimethoprim
202	*Streptococcus pneumoniae*	formaldehyde	252	*Streptococcus pneumoniae*	tylosin
203	*Streptococcus pneumoniae*	fortimicin	253	*Streptococcus pneumoniae*	unknown aminoglycoside
204	*Streptococcus pneumoniae*	fosfomycin	254	*Streptococcus pneumoniae*	unknown beta-lactam
205	*Streptococcus pneumoniae*	fusidic acid	255	*Streptococcus pneumoniae*	unknown quinolone
206	*Streptococcus pneumoniae*	gentamicin	256	*Streptococcus pneumoniae*	vancomycin
207	*Streptococcus pneumoniae*	hydrogen peroxide	257	*Streptococcus pneumoniae*	virginiamycin m
208	*Streptococcus pneumoniae*	hygromycin	258	*Streptococcus pneumoniae*	virginiamycin s
209	*Streptococcus pneumoniae*	imipenem	259	*Enterococcus faecalis*	amikacin
210	*Streptococcus pneumoniae*	isepamicin	260	*Enterococcus faecalis*	amoxicillin
211	*Streptococcus pneumoniae*	kanamycin	261	*Enterococcus faecalis*	amoxicillin+clavulanic acid
212	*Streptococcus pneumoniae*	kasugamycin	262	*Enterococcus faecalis*	ampicillin
213	*Streptococcus pneumoniae*	lincomycin	263	*Enterococcus faecalis*	ampicillin+clavulanic acid
214	*Streptococcus pneumoniae*	linezolid	264	*Enterococcus faecalis*	apramycin
215	*Streptococcus pneumoniae*	lividomycin	265	*Enterococcus faecalis*	arbekacin
216	*Streptococcus pneumoniae*	meropenem	266	*Enterococcus faecalis*	astromicin
217	*Streptococcus pneumoniae*	metronidazole	267	*Enterococcus faecalis*	azithromycin
218	*Streptococcus pneumoniae*	minocycline	268	*Enterococcus faecalis*	aztreonam
219	*Streptococcus pneumoniae*	mupirocin	269	*Enterococcus faecalis*	benzylkonium chloride
220	*Streptococcus pneumoniae*	nalidixic acid	270	*Enterococcus faecalis*	bleomycin
221	*Streptococcus pneumoniae*	neomycin	271	*Enterococcus faecalis*	butiromycin
222	*Streptococcus pneumoniae*	netilmicin	272	*Enterococcus faecalis*	butirosin
223	*Streptococcus pneumoniae*	norfloxacin	273	*Enterococcus faecalis*	carbapenem
224	*Streptococcus pneumoniae*	oleandomycin	274	*Enterococcus faecalis*	carbomycin
225	*Streptococcus pneumoniae*	paromomycin	275	*Enterococcus faecalis*	cefepime
226	*Streptococcus pneumoniae*	penicillin	276	*Enterococcus faecalis*	cefixime
227	*Streptococcus pneumoniae*	piperacillin	277	*Enterococcus faecalis*	cefotaxime
228	*Streptococcus pneumoniae*	piperacillin+clavulanic acid	278	*Enterococcus faecalis*	cefotaxime+clavulanic acid
229	*Streptococcus pneumoniae*	piperacillin+tazobactam	279	*Enterococcus faecalis*	cefoxitin
230	*Streptococcus pneumoniae*	pristinamycin ia	280	*Enterococcus faecalis*	ceftazidime
231	*Streptococcus pneumoniae*	pristinamycin iia	281	*Enterococcus faecalis*	ceftazidime+avibactam
232	*Streptococcus pneumoniae*	quinupristin	282	*Enterococcus faecalis*	ceftiofur
233	*Streptococcus pneumoniae*	quinupristin+dalfopristin	283	*Enterococcus faecalis*	ceftriaxone
234	*Streptococcus pneumoniae*	ribostamycin	284	*Enterococcus faecalis*	cephalosporins
235	*Streptococcus pneumoniae*	rifampicin	285	*Enterococcus faecalis*	cephalothin
236	*Streptococcus pneumoniae*	sisomicin	286	*Enterococcus faecalis*	cephalotin
237	*Streptococcus pneumoniae*	spectinomycin	287	*Enterococcus faecalis*	cetylpyridinium chloride
238	*Streptococcus pneumoniae*	spiramycin	288	*Enterococcus faecalis*	chloramphenicol
239	*Streptococcus pneumoniae*	streptomycin	289	*Enterococcus faecalis*	chlorhexidine
240	*Streptococcus pneumoniae*	sulfamethoxazole	290	*Enterococcus faecalis*	ciprofloxacin
241	*Streptococcus pneumoniae*	teicoplanin	291	*Enterococcus faecalis*	clindamycin
242	*Streptococcus pneumoniae*	telithromycin	292	*Enterococcus faecalis*	colistin
243	*Streptococcus pneumoniae*	temocillin	293	*Enterococcus faecalis*	dalfopristin
244	*Streptococcus pneumoniae*	temperature	294	*Enterococcus faecalis*	dibekacin
245	*Streptococcus pneumoniae*	tetracycline	295	*Enterococcus faecalis*	doxycycline
246	*Streptococcus pneumoniae*	tiamulin	296	*Enterococcus faecalis*	ertapenem
247	*Streptococcus pneumoniae*	ticarcillin	297	*Enterococcus faecalis*	erythromycin
248	*Streptococcus pneumoniae*	ticarcillin+clavulanic acid	298	*Enterococcus faecalis*	ethidium bromide
249	*Streptococcus pneumoniae*	tigecycline	299	*Enterococcus faecalis*	florfenicol
250	*Streptococcus pneumoniae*	tobramycin	300	*Enterococcus faecalis*	fluoroquinolone

Supplementary Table 13. Continued.

301	*Enterococcus faecalis*	formaldehyde	351	*Enterococcus faecalis*	tylosin
302	*Enterococcus faecalis*	fortimicin	352	*Enterococcus faecalis*	unknown aminoglycoside
303	*Enterococcus faecalis*	fosfomycin	353	*Enterococcus faecalis*	unknown beta-lactam
304	*Enterococcus faecalis*	fusidic acid	354	*Enterococcus faecalis*	unknown quinolone
305	*Enterococcus faecalis*	gentamicin	355	*Enterococcus faecalis*	vancomycin
306	*Enterococcus faecalis*	hydrogen peroxide	356	*Enterococcus faecalis*	virginiamycin m
307	*Enterococcus faecalis*	hygromycin	357	*Enterococcus faecalis*	virginiamycin s
308	*Enterococcus faecalis*	imipenem	358	*Enterococcus faecium*	amikacin
309	*Enterococcus faecalis*	isepamicin	359	*Enterococcus faecium*	amoxicillin+clavulanic acid
310	*Enterococcus faecalis*	kanamycin	360	*Enterococcus faecium*	ampicillin
311	*Enterococcus faecalis*	kasugamycin	361	*Enterococcus faecium*	ampicillin+clavulanic acid
312	*Enterococcus faecalis*	lincomycin	362	*Enterococcus faecium*	apramycin
313	*Enterococcus faecalis*	linezolid	363	*Enterococcus faecium*	arbekacin
314	*Enterococcus faecalis*	lividomycin	364	*Enterococcus faecium*	astromicin
315	*Enterococcus faecalis*	meropenem	365	*Enterococcus faecium*	azithromycin
316	*Enterococcus faecalis*	metronidazole	366	*Enterococcus faecium*	aztreonam
317	*Enterococcus faecalis*	minocycline	367	*Enterococcus faecium*	benzylkonium chloride
318	*Enterococcus faecalis*	mupirocin	368	*Enterococcus faecium*	bleomycin
319	*Enterococcus faecalis*	nalidixic acid	369	*Enterococcus faecium*	butiromycin
320	*Enterococcus faecalis*	neomycin	370	*Enterococcus faecium*	butirosin
321	*Enterococcus faecalis*	netilmicin	371	*Enterococcus faecium*	carbapenem
322	*Enterococcus faecalis*	norfloxacin	372	*Enterococcus faecium*	carbomycin
323	*Enterococcus faecalis*	oleandomycin	373	*Enterococcus faecium*	cefepime
324	*Enterococcus faecalis*	paromomycin	374	*Enterococcus faecium*	cefixime
325	*Enterococcus faecalis*	penicillin	375	*Enterococcus faecium*	cefotaxime
326	*Enterococcus faecalis*	piperacillin	376	*Enterococcus faecium*	cefotaxime+clavulanic acid
327	*Enterococcus faecalis*	piperacillin+clavulanic acid	377	*Enterococcus faecium*	cefoxitin
328	*Enterococcus faecalis*	piperacillin+tazobactam	378	*Enterococcus faecium*	ceftazidime
329	*Enterococcus faecalis*	pristinamycin ia	379	*Enterococcus faecium*	ceftazidime+avibactam
330	*Enterococcus faecalis*	pristinamycin iia	380	*Enterococcus faecium*	ceftiofur
331	*Enterococcus faecalis*	quinupristin	381	*Enterococcus faecium*	ceftriaxone
332	*Enterococcus faecalis*	quinupristin+dalfopristin	382	*Enterococcus faecium*	cephalosporins
333	*Enterococcus faecalis*	ribostamycin	383	*Enterococcus faecium*	cephalothin
334	*Enterococcus faecalis*	rifampicin	384	*Enterococcus faecium*	cephalotin
335	*Enterococcus faecalis*	sisomicin	385	*Enterococcus faecium*	cetylpyridinium chloride
336	*Enterococcus faecalis*	spectinomycin	386	*Enterococcus faecium*	chloramphenicol
337	*Enterococcus faecalis*	spiramycin	387	*Enterococcus faecium*	chlorhexidine
338	*Enterococcus faecalis*	streptomycin	388	*Enterococcus faecium*	ciprofloxacin
339	*Enterococcus faecalis*	sulfamethoxazole	389	*Enterococcus faecium*	clindamycin
340	*Enterococcus faecalis*	teicoplanin	390	*Enterococcus faecium*	colistin
341	*Enterococcus faecalis*	telithromycin	391	*Enterococcus faecium*	dalfopristin
342	*Enterococcus faecalis*	temocillin	392	*Enterococcus faecium*	dibekacin
343	*Enterococcus faecalis*	temperature	393	*Enterococcus faecium*	doxycycline
344	*Enterococcus faecalis*	tetracycline	394	*Enterococcus faecium*	ertapenem
345	*Enterococcus faecalis*	tiamulin	395	*Enterococcus faecium*	erythromycin
346	*Enterococcus faecalis*	ticarcillin	396	*Enterococcus faecium*	ethidium bromide
347	*Enterococcus faecalis*	ticarcillin+clavulanic acid	397	*Enterococcus faecium*	florfenicol
348	*Enterococcus faecalis*	tigecycline	398	*Enterococcus faecium*	fluoroquinolone
349	*Enterococcus faecalis*	tobramycin	399	*Enterococcus faecium*	formaldehyde
350	*Enterococcus faecalis*	trimethoprim	400	*Enterococcus faecium*	fortimicin

Supplementary Table 13. Continued.

401	*Enterococcus faecium*	fosfomycin	451	*Enterococcus faecium*	unknown beta-lactam
402	*Enterococcus faecium*	fusidic acid	452	*Enterococcus faecium*	unknown quinolone
403	*Enterococcus faecium*	gentamicin	453	*Enterococcus faecium*	vancomycin
404	*Enterococcus faecium*	hydrogen peroxide	454	*Enterococcus faecium*	virginiamycin m
405	*Enterococcus faecium*	hygromycin	455	*Enterococcus faecium*	virginiamycin s
406	*Enterococcus faecium*	imipenem	456	*Enterobacter cloacae*	amikacin
407	*Enterococcus faecium*	isepamicin	457	*Enterobacter cloacae*	amoxicillin
408	*Enterococcus faecium*	kanamycin	458	*Enterobacter cloacae*	amoxicillin+clavulanic acid
409	*Enterococcus faecium*	kasugamycin	459	*Enterobacter cloacae*	ampicillin
410	*Enterococcus faecium*	lincomycin	460	*Enterobacter cloacae*	ampicillin+clavulanic acid
411	*Enterococcus faecium*	linezolid	461	*Enterobacter cloacae*	apramycin
412	*Enterococcus faecium*	lividomycin	462	*Enterobacter cloacae*	arbekacin
413	*Enterococcus faecium*	meropenem	463	*Enterobacter cloacae*	astromicin
414	*Enterococcus faecium*	metronidazole	464	*Enterobacter cloacae*	azithromycin
415	*Enterococcus faecium*	minocycline	465	*Enterobacter cloacae*	aztreonam
416	*Enterococcus faecium*	mupirocin	466	*Enterobacter cloacae*	benzylkonium chloride
417	*Enterococcus faecium*	nalidixic acid	467	*Enterobacter cloacae*	bleomycin
418	*Enterococcus faecium*	neomycin	468	*Enterobacter cloacae*	butiromycin
419	*Enterococcus faecium*	netilmicin	469	*Enterobacter cloacae*	butirosin
420	*Enterococcus faecium*	norfloxacin	470	*Enterobacter cloacae*	carbapenem
421	*Enterococcus faecium*	oleandomycin	471	*Enterobacter cloacae*	carbomycin
422	*Enterococcus faecium*	paromomycin	472	*Enterobacter cloacae*	cefepime
423	*Enterococcus faecium*	penicillin	473	*Enterobacter cloacae*	cefixime
424	*Enterococcus faecium*	piperacillin	474	*Enterobacter cloacae*	cefotaxime
425	*Enterococcus faecium*	piperacillin+clavulanic acid	475	*Enterobacter cloacae*	cefotaxime+clavulanic acid
426	*Enterococcus faecium*	piperacillin+tazobactam	476	*Enterobacter cloacae*	cefoxitin
427	*Enterococcus faecium*	pristinamycin ia	477	*Enterobacter cloacae*	ceftazidime
428	*Enterococcus faecium*	pristinamycin iia	478	*Enterobacter cloacae*	ceftazidime+avibactam
429	*Enterococcus faecium*	quinupristin	479	*Enterobacter cloacae*	ceftiofur
430	*Enterococcus faecium*	quinupristin+dalfopristin	480	*Enterobacter cloacae*	ceftriaxone
431	*Enterococcus faecium*	ribostamycin	481	*Enterobacter cloacae*	cephalosporins
432	*Enterococcus faecium*	rifampicin	482	*Enterobacter cloacae*	cephalothin
433	*Enterococcus faecium*	sisomicin	483	*Enterobacter cloacae*	cephalotin
434	*Enterococcus faecium*	spectinomycin	484	*Enterobacter cloacae*	cetylpyridinium chloride
435	*Enterococcus faecium*	spiramycin	485	*Enterobacter cloacae*	chloramphenicol
436	*Enterococcus faecium*	streptomycin	486	*Enterobacter cloacae*	chlorhexidine
437	*Enterococcus faecium*	sulfamethoxazole	487	*Enterobacter cloacae*	ciprofloxacin
438	*Enterococcus faecium*	teicoplanin	488	*Enterobacter cloacae*	clindamycin
439	*Enterococcus faecium*	telithromycin	489	*Enterobacter cloacae*	colistin
440	*Enterococcus faecium*	temocillin	490	*Enterobacter cloacae*	dalfopristin
441	*Enterococcus faecium*	temperature	491	*Enterobacter cloacae*	dibekacin
442	*Enterococcus faecium*	tetracycline	492	*Enterobacter cloacae*	doxycycline
443	*Enterococcus faecium*	tiamulin	493	*Enterobacter cloacae*	ertapenem
444	*Enterococcus faecium*	ticarcillin	494	*Enterobacter cloacae*	erythromycin
445	*Enterococcus faecium*	ticarcillin+clavulanic acid	495	*Enterobacter cloacae*	ethidium bromide
446	*Enterococcus faecium*	tigecycline	496	*Enterobacter cloacae*	florfenicol
447	*Enterococcus faecium*	tobramycin	497	*Enterobacter cloacae*	fluoroquinolone
448	*Enterococcus faecium*	trimethoprim	498	*Enterobacter cloacae*	formaldehyde
449	*Enterococcus faecium*	tylosin	499	*Enterobacter cloacae*	fortimicin
450	*Enterococcus faecium*	unknown aminoglycoside	500	*Enterobacter cloacae*	fosfomycin

Supplementary Table 13. Continued.

501	*Enterobacter cloacae*	fusidic acid	551	*Enterobacter cloacae*	unknown quinolone
502	*Enterobacter cloacae*	gentamicin	552	*Enterobacter cloacae*	vancomycin
503	*Enterobacter cloacae*	hydrogen peroxide	553	*Enterobacter cloacae*	virginiamycin m
504	*Enterobacter cloacae*	hygromycin	554	*Enterobacter cloacae*	virginiamycin s
505	*Enterobacter cloacae*	imipenem	555	*Pseudomonas aeruginosa*	amikacin
506	*Enterobacter cloacae*	isepamicin	556	*Pseudomonas aeruginosa*	amoxicillin
507	*Enterobacter cloacae*	kanamycin	557	*Pseudomonas aeruginosa*	amoxicillin+clavulanic acid
508	*Enterobacter cloacae*	kasugamycin	558	*Pseudomonas aeruginosa*	ampicillin
509	*Enterobacter cloacae*	lincomycin	559	*Pseudomonas aeruginosa*	ampicillin+clavulanic acid
510	*Enterobacter cloacae*	linezolid	560	*Pseudomonas aeruginosa*	apramycin
511	*Enterobacter cloacae*	lividomycin	561	*Pseudomonas aeruginosa*	arbekacin
512	*Enterobacter cloacae*	meropenem	562	*Pseudomonas aeruginosa*	astromicin
513	*Enterobacter cloacae*	metronidazole	563	*Pseudomonas aeruginosa*	azithromycin
514	*Enterobacter cloacae*	minocycline	564	*Pseudomonas aeruginosa*	aztreonam
515	*Enterobacter cloacae*	mupirocin	565	*Pseudomonas aeruginosa*	benzylkonium chloride
516	*Enterobacter cloacae*	nalidixic acid	566	*Pseudomonas aeruginosa*	bleomycin
517	*Enterobacter cloacae*	neomycin	567	*Pseudomonas aeruginosa*	butiromycin
518	*Enterobacter cloacae*	netilmicin	568	*Pseudomonas aeruginosa*	butirosin
519	*Enterobacter cloacae*	norfloxacin	569	*Pseudomonas aeruginosa*	carbapenem
520	*Enterobacter cloacae*	oleandomycin	570	*Pseudomonas aeruginosa*	carbomycin
521	*Enterobacter cloacae*	paromomycin	571	*Pseudomonas aeruginosa*	cefepime
522	*Enterobacter cloacae*	penicillin	572	*Pseudomonas aeruginosa*	cefixime
523	*Enterobacter cloacae*	piperacillin	573	*Pseudomonas aeruginosa*	cefotaxime
524	*Enterobacter cloacae*	piperacillin+clavulanic acid	574	*Pseudomonas aeruginosa*	cefotaxime+clavulanic acid
525	*Enterobacter cloacae*	piperacillin+tazobactam	575	*Pseudomonas aeruginosa*	cefoxitin
526	*Enterobacter cloacae*	pristinamycin ia	576	*Pseudomonas aeruginosa*	ceftazidime
527	*Enterobacter cloacae*	pristinamycin iia	577	*Pseudomonas aeruginosa*	ceftazidime+avibactam
528	*Enterobacter cloacae*	quinupristin	578	*Pseudomonas aeruginosa*	ceftiofur
529	*Enterobacter cloacae*	quinupristin+dalfopristin	579	*Pseudomonas aeruginosa*	ceftriaxone
530	*Enterobacter cloacae*	ribostamycin	580	*Pseudomonas aeruginosa*	cephalosporins
531	*Enterobacter cloacae*	rifampicin	581	*Pseudomonas aeruginosa*	cephalothin
532	*Enterobacter cloacae*	sisomicin	582	*Pseudomonas aeruginosa*	cephalotin
533	*Enterobacter cloacae*	spectinomycin	583	*Pseudomonas aeruginosa*	cetylpyridinium chloride
534	*Enterobacter cloacae*	spiramycin	584	*Pseudomonas aeruginosa*	chloramphenicol
535	*Enterobacter cloacae*	streptomycin	585	*Pseudomonas aeruginosa*	chlorhexidine
536	*Enterobacter cloacae*	sulfamethoxazole	586	*Pseudomonas aeruginosa*	ciprofloxacin
537	*Enterobacter cloacae*	teicoplanin	587	*Pseudomonas aeruginosa*	clindamycin
538	*Enterobacter cloacae*	telithromycin	588	*Pseudomonas aeruginosa*	colistin
539	*Enterobacter cloacae*	temocillin	589	*Pseudomonas aeruginosa*	dalfopristin
540	*Enterobacter cloacae*	temperature	590	*Pseudomonas aeruginosa*	dibekacin
541	*Enterobacter cloacae*	tetracycline	591	*Pseudomonas aeruginosa*	doxycycline
542	*Enterobacter cloacae*	tiamulin	592	*Pseudomonas aeruginosa*	ertapenem
543	*Enterobacter cloacae*	ticarcillin	593	*Pseudomonas aeruginosa*	erythromycin
544	*Enterobacter cloacae*	ticarcillin+clavulanic acid	594	*Pseudomonas aeruginosa*	ethidium bromide
545	*Enterobacter cloacae*	tigecycline	595	*Pseudomonas aeruginosa*	florfenicol
546	*Enterobacter cloacae*	tobramycin	596	*Pseudomonas aeruginosa*	fluoroquinolone
547	*Enterobacter cloacae*	trimethoprim	597	*Pseudomonas aeruginosa*	formaldehyde
548	*Enterobacter cloacae*	tylosin	598	*Pseudomonas aeruginosa*	fortimicin
549	*Enterobacter cloacae*	unknown aminoglycoside	599	*Pseudomonas aeruginosa*	fosfomycin
550	*Enterobacter cloacae*	unknown beta-lactam	600	*Pseudomonas aeruginosa*	fusidic acid

Supplementary Table 13. Continued.

601	*Pseudomonas aeruginosa*	gentamicin	651	*Pseudomonas aeruginosa*	vancomycin
602	*Pseudomonas aeruginosa*	hydrogen peroxide	652	*Pseudomonas aeruginosa*	virginiamycin m
603	*Pseudomonas aeruginosa*	hygromycin	653	*Pseudomonas aeruginosa*	virginiamycin s
604	*Pseudomonas aeruginosa*	imipenem	654	*Proteus mirabilis*	amikacin
605	*Pseudomonas aeruginosa*	isepamicin	655	*Proteus mirabilis*	amoxicillin
606	*Pseudomonas aeruginosa*	kanamycin	656	*Proteus mirabilis*	amoxicillin+clavulanic acid
607	*Pseudomonas aeruginosa*	kasugamycin	657	*Proteus mirabilis*	ampicillin
608	*Pseudomonas aeruginosa*	lincomycin	658	*Proteus mirabilis*	ampicillin+clavulanic acid
609	*Pseudomonas aeruginosa*	linezolid	659	*Proteus mirabilis*	apramycin
610	*Pseudomonas aeruginosa*	lividomycin	660	*Proteus mirabilis*	arbekacin
611	*Pseudomonas aeruginosa*	meropenem	661	*Proteus mirabilis*	astromicin
612	*Pseudomonas aeruginosa*	metronidazole	662	*Proteus mirabilis*	azithromycin
613	*Pseudomonas aeruginosa*	minocycline	663	*Proteus mirabilis*	aztreonam
614	*Pseudomonas aeruginosa*	mupirocin	664	*Proteus mirabilis*	benzylkonium chloride
615	*Pseudomonas aeruginosa*	nalidixic acid	665	*Proteus mirabilis*	bleomycin
616	*Pseudomonas aeruginosa*	neomycin	666	*Proteus mirabilis*	butiromycin
617	*Pseudomonas aeruginosa*	netilmicin	667	*Proteus mirabilis*	butirosin
618	*Pseudomonas aeruginosa*	norfloxacin	668	*Proteus mirabilis*	carbapenem
619	*Pseudomonas aeruginosa*	oleandomycin	669	*Proteus mirabilis*	carbomycin
620	*Pseudomonas aeruginosa*	paromomycin	670	*Proteus mirabilis*	cefepime
621	*Pseudomonas aeruginosa*	penicillin	671	*Proteus mirabilis*	cefixime
622	*Pseudomonas aeruginosa*	piperacillin	672	*Proteus mirabilis*	cefotaxime
623	*Pseudomonas aeruginosa*	piperacillin+clavulanic acid	673	*Proteus mirabilis*	cefotaxime+clavulanic acid
624	*Pseudomonas aeruginosa*	piperacillin+tazobactam	674	*Proteus mirabilis*	cefoxitin
625	*Pseudomonas aeruginosa*	pristinamycin ia	675	*Proteus mirabilis*	ceftazidime
626	*Pseudomonas aeruginosa*	pristinamycin iia	676	*Proteus mirabilis*	ceftazidime+avibactam
627	*Pseudomonas aeruginosa*	quinupristin	677	*Proteus mirabilis*	ceftiofur
628	*Pseudomonas aeruginosa*	quinupristin+dalfopristin	678	*Proteus mirabilis*	ceftriaxone
629	*Pseudomonas aeruginosa*	ribostamycin	679	*Proteus mirabilis*	cephalosporins
630	*Pseudomonas aeruginosa*	rifampicin	680	*Proteus mirabilis*	cephalothin
631	*Pseudomonas aeruginosa*	sisomicin	681	*Proteus mirabilis*	cephalotin
632	*Pseudomonas aeruginosa*	spectinomycin	682	*Proteus mirabilis*	cetylpyridinium chloride
633	*Pseudomonas aeruginosa*	spiramycin	683	*Proteus mirabilis*	chloramphenicol
634	*Pseudomonas aeruginosa*	streptomycin	684	*Proteus mirabilis*	chlorhexidine
635	*Pseudomonas aeruginosa*	sulfamethoxazole	685	*Proteus mirabilis*	ciprofloxacin
636	*Pseudomonas aeruginosa*	teicoplanin	686	*Proteus mirabilis*	clindamycin
637	*Pseudomonas aeruginosa*	telithromycin	687	*Proteus mirabilis*	colistin
638	*Pseudomonas aeruginosa*	temocillin	688	*Proteus mirabilis*	dalfopristin
639	*Pseudomonas aeruginosa*	temperature	689	*Proteus mirabilis*	dibekacin
640	*Pseudomonas aeruginosa*	tetracycline	690	*Proteus mirabilis*	doxycycline
641	*Pseudomonas aeruginosa*	tiamulin	691	*Proteus mirabilis*	ertapenem
642	*Pseudomonas aeruginosa*	ticarcillin	692	*Proteus mirabilis*	erythromycin
643	*Pseudomonas aeruginosa*	ticarcillin+clavulanic acid	693	*Proteus mirabilis*	ethidium bromide
644	*Pseudomonas aeruginosa*	tigecycline	694	*Proteus mirabilis*	florfenicol
645	*Pseudomonas aeruginosa*	tobramycin	695	*Proteus mirabilis*	fluoroquinolone
646	*Pseudomonas aeruginosa*	trimethoprim	696	*Proteus mirabilis*	formaldehyde
647	*Pseudomonas aeruginosa*	tylosin	697	*Proteus mirabilis*	fortimicin
648	*Pseudomonas aeruginosa*	unknown aminoglycoside	698	*Proteus mirabilis*	fosfomycin
649	*Pseudomonas aeruginosa*	unknown beta-lactam	699	*Proteus mirabilis*	fusidic acid
650	*Pseudomonas aeruginosa*	unknown quinolone	700	*Proteus mirabilis*	gentamicin

Supplementary Table 13. Continued.

701	*Proteus mirabilis*	hydrogen peroxide	751	*Proteus mirabilis*	virginiamycin m
702	*Proteus mirabilis*	hygromycin	752	*Proteus mirabilis*	virginiamycin s
703	*Proteus mirabilis*	imipenem	753	*Bacteroides fragilis*	amikacin
704	*Proteus mirabilis*	isepamicin	754	*Bacteroides fragilis*	amoxicillin
705	*Proteus mirabilis*	kanamycin	755	*Bacteroides fragilis*	amoxicillin+clavulanic acid
706	*Proteus mirabilis*	kasugamycin	756	*Bacteroides fragilis*	ampicillin
707	*Proteus mirabilis*	lincomycin	757	*Bacteroides fragilis*	ampicillin+clavulanic acid
708	*Proteus mirabilis*	linezolid	758	*Bacteroides fragilis*	apramycin
709	*Proteus mirabilis*	lividomycin	759	*Bacteroides fragilis*	arbekacin
710	*Proteus mirabilis*	meropenem	760	*Bacteroides fragilis*	astromicin
711	*Proteus mirabilis*	metronidazole	761	*Bacteroides fragilis*	azithromycin
712	*Proteus mirabilis*	minocycline	762	*Bacteroides fragilis*	aztreonam
713	*Proteus mirabilis*	mupirocin	763	*Bacteroides fragilis*	benzylkonium chloride
714	*Proteus mirabilis*	nalidixic acid	764	*Bacteroides fragilis*	bleomycin
715	*Proteus mirabilis*	neomycin	765	*Bacteroides fragilis*	butiromycin
716	*Proteus mirabilis*	netilmicin	766	*Bacteroides fragilis*	butirosin
717	*Proteus mirabilis*	norfloxacin	767	*Bacteroides fragilis*	carbapenem
718	*Proteus mirabilis*	oleandomycin	768	*Bacteroides fragilis*	carbomycin
719	*Proteus mirabilis*	paromomycin	769	*Bacteroides fragilis*	cefepime
720	*Proteus mirabilis*	penicillin	770	*Bacteroides fragilis*	cefixime
721	*Proteus mirabilis*	piperacillin	771	*Bacteroides fragilis*	cefotaxime
722	*Proteus mirabilis*	piperacillin+clavulanic acid	772	*Bacteroides fragilis*	cefotaxime+clavulanic acid
723	*Proteus mirabilis*	piperacillin+tazobactam	773	*Bacteroides fragilis*	cefoxitin
724	*Proteus mirabilis*	pristinamycin ia	774	*Bacteroides fragilis*	ceftazidime
725	*Proteus mirabilis*	pristinamycin iia	775	*Bacteroides fragilis*	ceftazidime+avibactam
726	*Proteus mirabilis*	quinupristin	776	*Bacteroides fragilis*	ceftiofur
727	*Proteus mirabilis*	quinupristin+dalfopristin	777	*Bacteroides fragilis*	ceftriaxone
728	*Proteus mirabilis*	ribostamycin	778	*Bacteroides fragilis*	cephalosporins
729	*Proteus mirabilis*	rifampicin	779	*Bacteroides fragilis*	cephalothin
730	*Proteus mirabilis*	sisomicin	780	*Bacteroides fragilis*	cephalotin
731	*Proteus mirabilis*	spectinomycin	781	*Bacteroides fragilis*	cetylpyridinium chloride
732	*Proteus mirabilis*	spiramycin	782	*Bacteroides fragilis*	chloramphenicol
733	*Proteus mirabilis*	streptomycin	783	*Bacteroides fragilis*	chlorhexidine
734	*Proteus mirabilis*	sulfamethoxazole	784	*Bacteroides fragilis*	ciprofloxacin
735	*Proteus mirabilis*	teicoplanin	785	*Bacteroides fragilis*	clindamycin
736	*Proteus mirabilis*	telithromycin	786	*Bacteroides fragilis*	colistin
737	*Proteus mirabilis*	temocillin	787	*Bacteroides fragilis*	dalfopristin
738	*Proteus mirabilis*	temperature	788	*Bacteroides fragilis*	dibekacin
739	*Proteus mirabilis*	tetracycline	789	*Bacteroides fragilis*	doxycycline
740	*Proteus mirabilis*	tiamulin	790	*Bacteroides fragilis*	ertapenem
741	*Proteus mirabilis*	ticarcillin	791	*Bacteroides fragilis*	erythromycin
742	*Proteus mirabilis*	ticarcillin+clavulanic acid	792	*Bacteroides fragilis*	ethidium bromide
743	*Proteus mirabilis*	tigecycline	793	*Bacteroides fragilis*	florfenicol
744	*Proteus mirabilis*	tobramycin	794	*Bacteroides fragilis*	fluoroquinolone
745	*Proteus mirabilis*	trimethoprim	795	*Bacteroides fragilis*	formaldehyde
746	*Proteus mirabilis*	tylosin	796	*Bacteroides fragilis*	fortimicin
747	*Proteus mirabilis*	unknown aminoglycoside	797	*Bacteroides fragilis*	fosfomycin
748	*Proteus mirabilis*	unknown beta-lactam	798	*Bacteroides fragilis*	fusidic acid
749	*Proteus mirabilis*	unknown quinolone	799	*Bacteroides fragilis*	gentamicin
750	*Proteus mirabilis*	vancomycin	800	*Bacteroides fragilis*	hydrogen peroxide

Supplementary Table 13. Continued.

801	*Bacteroides fragilis*	hygromycin	851	*Bacteroides fragilis*	virginiamycin s
802	*Bacteroides fragilis*	imipenem	852	*Staphylococcus epidermidis*	amikacin
803	*Bacteroides fragilis*	isepamicin	853	*Staphylococcus epidermidis*	amoxicillin
804	*Bacteroides fragilis*	kanamycin	854	*Staphylococcus epidermidis*	amoxicillin+clavulanic acid
805	*Bacteroides fragilis*	kasugamycin	855	*Staphylococcus epidermidis*	ampicillin
806	*Bacteroides fragilis*	lincomycin	856	*Staphylococcus epidermidis*	ampicillin+clavulanic acid
807	*Bacteroides fragilis*	linezolid	857	*Staphylococcus epidermidis*	apramycin
808	*Bacteroides fragilis*	lividomycin	858	*Staphylococcus epidermidis*	arbekacin
809	*Bacteroides fragilis*	meropenem	859	*Staphylococcus epidermidis*	astromicin
810	*Bacteroides fragilis*	metronidazole	860	*Staphylococcus epidermidis*	azithromycin
811	*Bacteroides fragilis*	minocycline	861	*Staphylococcus epidermidis*	aztreonam
812	*Bacteroides fragilis*	mupirocin	862	*Staphylococcus epidermidis*	benzylkonium chloride
813	*Bacteroides fragilis*	nalidixic acid	863	*Staphylococcus epidermidis*	bleomycin
814	*Bacteroides fragilis*	neomycin	864	*Staphylococcus epidermidis*	butiromycin
815	*Bacteroides fragilis*	netilmicin	865	*Staphylococcus epidermidis*	butirosin
816	*Bacteroides fragilis*	norfloxacin	866	*Staphylococcus epidermidis*	carbapenem
817	*Bacteroides fragilis*	oleandomycin	867	*Staphylococcus epidermidis*	carbomycin
818	*Bacteroides fragilis*	paromomycin	868	*Staphylococcus epidermidis*	cefepime
819	*Bacteroides fragilis*	penicillin	869	*Staphylococcus epidermidis*	cefixime
820	*Bacteroides fragilis*	piperacillin	870	*Staphylococcus epidermidis*	cefotaxime
821	*Bacteroides fragilis*	piperacillin+clavulanic acid	871	*Staphylococcus epidermidis*	cefotaxime+clavulanic acid
822	*Bacteroides fragilis*	piperacillin+tazobactam	872	*Staphylococcus epidermidis*	cefoxitin
823	*Bacteroides fragilis*	pristinamycin ia	873	*Staphylococcus epidermidis*	ceftazidime
824	*Bacteroides fragilis*	pristinamycin iia	874	*Staphylococcus epidermidis*	ceftazidime+avibactam
825	*Bacteroides fragilis*	quinupristin	875	*Staphylococcus epidermidis*	ceftiofur
826	*Bacteroides fragilis*	quinupristin+dalfopristin	876	*Staphylococcus epidermidis*	ceftriaxone
827	*Bacteroides fragilis*	ribostamycin	877	*Staphylococcus epidermidis*	cephalosporins
828	*Bacteroides fragilis*	rifampicin	878	*Staphylococcus epidermidis*	cephalothin
829	*Bacteroides fragilis*	sisomicin	879	*Staphylococcus epidermidis*	cephalotin
830	*Bacteroides fragilis*	spectinomycin	880	*Staphylococcus epidermidis*	cetylpyridinium chloride
831	*Bacteroides fragilis*	spiramycin	881	*Staphylococcus epidermidis*	chloramphenicol
832	*Bacteroides fragilis*	streptomycin	882	*Staphylococcus epidermidis*	chlorhexidine
833	*Bacteroides fragilis*	sulfamethoxazole	883	*Staphylococcus epidermidis*	clindamycin
834	*Bacteroides fragilis*	teicoplanin	884	*Staphylococcus epidermidis*	colistin
835	*Bacteroides fragilis*	telithromycin	885	*Staphylococcus epidermidis*	dalfopristin
836	*Bacteroides fragilis*	temocillin	886	*Staphylococcus epidermidis*	dibekacin
837	*Bacteroides fragilis*	temperature	887	*Staphylococcus epidermidis*	doxycycline
838	*Bacteroides fragilis*	tetracycline	888	*Staphylococcus epidermidis*	ertapenem
839	*Bacteroides fragilis*	tiamulin	889	*Staphylococcus epidermidis*	erythromycin
840	*Bacteroides fragilis*	ticarcillin	890	*Staphylococcus epidermidis*	ethidium bromide
841	*Bacteroides fragilis*	ticarcillin+clavulanic acid	891	*Staphylococcus epidermidis*	florfenicol
842	*Bacteroides fragilis*	tigecycline	892	*Staphylococcus epidermidis*	fluoroquinolone
843	*Bacteroides fragilis*	tobramycin	893	*Staphylococcus epidermidis*	formaldehyde
844	*Bacteroides fragilis*	trimethoprim	894	*Staphylococcus epidermidis*	fortimicin
845	*Bacteroides fragilis*	tylosin	895	*Staphylococcus epidermidis*	fusidic acid
846	*Bacteroides fragilis*	unknown aminoglycoside	896	*Staphylococcus epidermidis*	gentamicin
847	*Bacteroides fragilis*	unknown beta-lactam	897	*Staphylococcus epidermidis*	hydrogen peroxide
848	*Bacteroides fragilis*	unknown quinolone	898	*Staphylococcus epidermidis*	hygromycin
849	*Bacteroides fragilis*	vancomycin	899	*Staphylococcus epidermidis*	imipenem
850	*Bacteroides fragilis*	virginiamycin m	900	*Staphylococcus epidermidis*	isepamicin

Supplementary Table 13. Continued.

901	*Staphylococcus epidermidis*	kanamycin	951	*Staphylococcus hominis*	amoxicillin+clavulanic acid
902	*Staphylococcus epidermidis*	kasugamycin	952	*Staphylococcus hominis*	ampicillin
903	*Staphylococcus epidermidis*	lincomycin	953	*Staphylococcus hominis*	ampicillin+clavulanic acid
904	*Staphylococcus epidermidis*	linezolid	954	*Staphylococcus hominis*	apramycin
905	*Staphylococcus epidermidis*	lividomycin	955	*Staphylococcus hominis*	arbekacin
906	*Staphylococcus epidermidis*	meropenem	956	*Staphylococcus hominis*	astromicin
907	*Staphylococcus epidermidis*	metronidazole	957	*Staphylococcus hominis*	azithromycin
908	*Staphylococcus epidermidis*	minocycline	958	*Staphylococcus hominis*	aztreonam
909	*Staphylococcus epidermidis*	mupirocin	959	*Staphylococcus hominis*	benzylkonium chloride
910	*Staphylococcus epidermidis*	nalidixic acid	960	*Staphylococcus hominis*	bleomycin
911	*Staphylococcus epidermidis*	neomycin	961	*Staphylococcus hominis*	butiromycin
912	*Staphylococcus epidermidis*	netilmicin	962	*Staphylococcus hominis*	butirosin
913	*Staphylococcus epidermidis*	norfloxacin	963	*Staphylococcus hominis*	carbapenem
914	*Staphylococcus epidermidis*	oleandomycin	964	*Staphylococcus hominis*	carbomycin
915	*Staphylococcus epidermidis*	paromomycin	965	*Staphylococcus hominis*	cefepime
916	*Staphylococcus epidermidis*	penicillin	966	*Staphylococcus hominis*	cefixime
917	*Staphylococcus epidermidis*	piperacillin	967	*Staphylococcus hominis*	cefotaxime
918	*Staphylococcus epidermidis*	piperacillin+clavulanic acid	968	*Staphylococcus hominis*	cefotaxime+clavulanic acid
919	*Staphylococcus epidermidis*	piperacillin+tazobactam	969	*Staphylococcus hominis*	cefoxitin
920	*Staphylococcus epidermidis*	pristinamycin ia	970	*Staphylococcus hominis*	ceftazidime
921	*Staphylococcus epidermidis*	pristinamycin iia	971	*Staphylococcus hominis*	ceftazidime+avibactam
922	*Staphylococcus epidermidis*	quinupristin	972	*Staphylococcus hominis*	ceftiofur
923	*Staphylococcus epidermidis*	quinupristin+dalfopristin	973	*Staphylococcus hominis*	ceftriaxone
924	*Staphylococcus epidermidis*	ribostamycin	974	*Staphylococcus hominis*	cephalosporins
925	*Staphylococcus epidermidis*	rifampicin	975	*Staphylococcus hominis*	cephalothin
926	*Staphylococcus epidermidis*	sisomicin	976	*Staphylococcus hominis*	cephalotin
927	*Staphylococcus epidermidis*	spectinomycin	977	*Staphylococcus hominis*	cetylpyridinium chloride
928	*Staphylococcus epidermidis*	spiramycin	978	*Staphylococcus hominis*	chloramphenicol
929	*Staphylococcus epidermidis*	streptomycin	979	*Staphylococcus hominis*	chlorhexidine
930	*Staphylococcus epidermidis*	sulfamethoxazole	980	*Staphylococcus hominis*	ciprofloxacin
931	*Staphylococcus epidermidis*	teicoplanin	981	*Staphylococcus hominis*	clindamycin
932	*Staphylococcus epidermidis*	telithromycin	982	*Staphylococcus hominis*	colistin
933	*Staphylococcus epidermidis*	temocillin	983	*Staphylococcus hominis*	dalfopristin
934	*Staphylococcus epidermidis*	temperature	984	*Staphylococcus hominis*	dibekacin
935	*Staphylococcus epidermidis*	tetracycline	985	*Staphylococcus hominis*	doxycycline
936	*Staphylococcus epidermidis*	tiamulin	986	*Staphylococcus hominis*	ertapenem
937	*Staphylococcus epidermidis*	ticarcillin	987	*Staphylococcus hominis*	erythromycin
938	*Staphylococcus epidermidis*	ticarcillin+clavulanic acid	988	*Staphylococcus hominis*	ethidium bromide
939	*Staphylococcus epidermidis*	tigecycline	989	*Staphylococcus hominis*	florfenicol
940	*Staphylococcus epidermidis*	tobramycin	990	*Staphylococcus hominis*	fluoroquinolone
941	*Staphylococcus epidermidis*	trimethoprim	991	*Staphylococcus hominis*	formaldehyde
942	*Staphylococcus epidermidis*	tylosin	992	*Staphylococcus hominis*	fortimicin
943	*Staphylococcus epidermidis*	unknown aminoglycoside	993	*Staphylococcus hominis*	fosfomycin
944	*Staphylococcus epidermidis*	unknown beta-lactam	994	*Staphylococcus hominis*	fusidic acid
945	*Staphylococcus epidermidis*	unknown quinolone	995	*Staphylococcus hominis*	gentamicin
946	*Staphylococcus epidermidis*	vancomycin	996	*Staphylococcus hominis*	hydrogen peroxide
947	*Staphylococcus epidermidis*	virginiamycin m	997	*Staphylococcus hominis*	hygromycin
948	*Staphylococcus epidermidis*	virginiamycin s	998	*Staphylococcus hominis*	imipenem
949	*Staphylococcus hominis*	amikacin	999	*Staphylococcus hominis*	isepamicin
950	*Staphylococcus hominis*	amoxicillin	1000	*Staphylococcus hominis*	kanamycin

Supplementary Table 13. Continued.

1001	*Staphylococcus hominis*	kasugamycin	1051	*Staphylococcus haemolyticus*	ampicillin
1002	*Staphylococcus hominis*	lincomycin	1052	*Staphylococcus haemolyticus*	ampicillin+clavulanic acid
1003	*Staphylococcus hominis*	linezolid	1053	*Staphylococcus haemolyticus*	apramycin
1004	*Staphylococcus hominis*	lividomycin	1054	*Staphylococcus haemolyticus*	arbekacin
1005	*Staphylococcus hominis*	meropenem	1055	*Staphylococcus haemolyticus*	astromicin
1006	*Staphylococcus hominis*	metronidazole	1056	*Staphylococcus haemolyticus*	azithromycin
1007	*Staphylococcus hominis*	minocycline	1057	*Staphylococcus haemolyticus*	aztreonam
1008	*Staphylococcus hominis*	mupirocin	1058	*Staphylococcus haemolyticus*	benzylkonium chloride
1009	*Staphylococcus hominis*	nalidixic acid	1059	*Staphylococcus haemolyticus*	bleomycin
1010	*Staphylococcus hominis*	neomycin	1060	*Staphylococcus haemolyticus*	butiromycin
1011	*Staphylococcus hominis*	netilmicin	1061	*Staphylococcus haemolyticus*	butirosin
1012	*Staphylococcus hominis*	norfloxacin	1062	*Staphylococcus haemolyticus*	carbapenem
1013	*Staphylococcus hominis*	oleandomycin	1063	*Staphylococcus haemolyticus*	carbomycin
1014	*Staphylococcus hominis*	paromomycin	1064	*Staphylococcus haemolyticus*	cefepime
1015	*Staphylococcus hominis*	penicillin	1065	*Staphylococcus haemolyticus*	cefixime
1016	*Staphylococcus hominis*	piperacillin	1066	*Staphylococcus haemolyticus*	cefotaxime
1017	*Staphylococcus hominis*	piperacillin+clavulanic acid	1067	*Staphylococcus haemolyticus*	cefotaxime+clavulanic acid
1018	*Staphylococcus hominis*	piperacillin+tazobactam	1068	*Staphylococcus haemolyticus*	cefoxitin
1019	*Staphylococcus hominis*	pristinamycin ia	1069	*Staphylococcus haemolyticus*	ceftazidime
1020	*Staphylococcus hominis*	pristinamycin iia	1070	*Staphylococcus haemolyticus*	ceftazidime+avibactam
1021	*Staphylococcus hominis*	quinupristin	1071	*Staphylococcus haemolyticus*	ceftiofur
1022	*Staphylococcus hominis*	quinupristin+dalfopristin	1072	*Staphylococcus haemolyticus*	ceftriaxone
1023	*Staphylococcus hominis*	ribostamycin	1073	*Staphylococcus haemolyticus*	cephalosporins
1024	*Staphylococcus hominis*	rifampicin	1074	*Staphylococcus haemolyticus*	cephalothin
1025	*Staphylococcus hominis*	sisomicin	1075	*Staphylococcus haemolyticus*	cephalotin
1026	*Staphylococcus hominis*	spectinomycin	1076	*Staphylococcus haemolyticus*	cetylpyridinium chloride
1027	*Staphylococcus hominis*	spiramycin	1077	*Staphylococcus haemolyticus*	chloramphenicol
1028	*Staphylococcus hominis*	streptomycin	1078	*Staphylococcus haemolyticus*	chlorhexidine
1029	*Staphylococcus hominis*	sulfamethoxazole	1079	*Staphylococcus haemolyticus*	ciprofloxacin
1030	*Staphylococcus hominis*	teicoplanin	1080	*Staphylococcus haemolyticus*	clindamycin
1031	*Staphylococcus hominis*	telithromycin	1081	*Staphylococcus haemolyticus*	colistin
1032	*Staphylococcus hominis*	temocillin	1082	*Staphylococcus haemolyticus*	dalfopristin
1033	*Staphylococcus hominis*	temperature	1083	*Staphylococcus haemolyticus*	dibekacin
1034	*Staphylococcus hominis*	tetracycline	1084	*Staphylococcus haemolyticus*	doxycycline
1035	*Staphylococcus hominis*	tiamulin	1085	*Staphylococcus haemolyticus*	ertapenem
1036	*Staphylococcus hominis*	ticarcillin	1086	*Staphylococcus haemolyticus*	erythromycin
1037	*Staphylococcus hominis*	ticarcillin+clavulanic acid	1087	*Staphylococcus haemolyticus*	ethidium bromide
1038	*Staphylococcus hominis*	tigecycline	1088	*Staphylococcus haemolyticus*	florfenicol
1039	*Staphylococcus hominis*	tobramycin	1089	*Staphylococcus haemolyticus*	fluoroquinolone
1040	*Staphylococcus hominis*	trimethoprim	1090	*Staphylococcus haemolyticus*	formaldehyde
1041	*Staphylococcus hominis*	tylosin	1091	*Staphylococcus haemolyticus*	fortimicin
1042	*Staphylococcus hominis*	unknown aminoglycoside	1092	*Staphylococcus haemolyticus*	fosfomycin
1043	*Staphylococcus hominis*	unknown beta-lactam	1093	*Staphylococcus haemolyticus*	fusidic acid
1044	*Staphylococcus hominis*	unknown quinolone	1094	*Staphylococcus haemolyticus*	gentamicin
1045	*Staphylococcus hominis*	vancomycin	1095	*Staphylococcus haemolyticus*	hydrogen peroxide
1046	*Staphylococcus hominis*	virginiamycin m	1096	*Staphylococcus haemolyticus*	hygromycin
1047	*Staphylococcus hominis*	virginiamycin s	1097	*Staphylococcus haemolyticus*	imipenem
1048	*Staphylococcus haemolyticus*	amikacin	1098	*Staphylococcus haemolyticus*	isepamicin
1049	*Staphylococcus haemolyticus*	amoxicillin	1099	*Staphylococcus haemolyticus*	kanamycin
1050	*Staphylococcus haemolyticus*	amoxicillin+clavulanic acid	1100	*Staphylococcus haemolyticus*	kasugamycin

Supplementary Table 13. Continued.

1101	*Staphylococcus haemolyticus*	lincomycin	1151	*Staphylococcus capitis*	ampicillin+clavulanic acid
1102	*Staphylococcus haemolyticus*	linezolid	1152	*Staphylococcus capitis*	apramycin
1103	*Staphylococcus haemolyticus*	lividomycin	1153	*Staphylococcus capitis*	arbekacin
1104	*Staphylococcus haemolyticus*	meropenem	1154	*Staphylococcus capitis*	astromicin
1105	*Staphylococcus haemolyticus*	metronidazole	1155	*Staphylococcus capitis*	azithromycin
1106	*Staphylococcus haemolyticus*	minocycline	1156	*Staphylococcus capitis*	aztreonam
1107	*Staphylococcus haemolyticus*	mupirocin	1157	*Staphylococcus capitis*	benzylkonium chloride
1108	*Staphylococcus haemolyticus*	nalidixic acid	1158	*Staphylococcus capitis*	bleomycin
1109	*Staphylococcus haemolyticus*	neomycin	1159	*Staphylococcus capitis*	butiromycin
1110	*Staphylococcus haemolyticus*	netilmicin	1160	*Staphylococcus capitis*	butirosin
1111	*Staphylococcus haemolyticus*	norfloxacin	1161	*Staphylococcus capitis*	carbapenem
1112	*Staphylococcus haemolyticus*	oleandomycin	1162	*Staphylococcus capitis*	carbomycin
1113	*Staphylococcus haemolyticus*	paromomycin	1163	*Staphylococcus capitis*	cefepime
1114	*Staphylococcus haemolyticus*	penicillin	1164	*Staphylococcus capitis*	cefixime
1115	*Staphylococcus haemolyticus*	piperacillin	1165	*Staphylococcus capitis*	cefotaxime
1116	*Staphylococcus haemolyticus*	piperacillin+clavulanic acid	1166	*Staphylococcus capitis*	cefotaxime+clavulanic acid
1117	*Staphylococcus haemolyticus*	piperacillin+tazobactam	1167	*Staphylococcus capitis*	cefoxitin
1118	*Staphylococcus haemolyticus*	pristinamycin ia	1168	*Staphylococcus capitis*	ceftazidime
1119	*Staphylococcus haemolyticus*	pristinamycin iia	1169	*Staphylococcus capitis*	ceftazidime+avibactam
1120	*Staphylococcus haemolyticus*	quinupristin	1170	*Staphylococcus capitis*	ceftiofur
1121	*Staphylococcus haemolyticus*	quinupristin+dalfopristin	1171	*Staphylococcus capitis*	ceftriaxone
1122	*Staphylococcus haemolyticus*	ribostamycin	1172	*Staphylococcus capitis*	cephalosporins
1123	*Staphylococcus haemolyticus*	rifampicin	1173	*Staphylococcus capitis*	cephalothin
1124	*Staphylococcus haemolyticus*	sisomicin	1174	*Staphylococcus capitis*	cephalotin
1125	*Staphylococcus haemolyticus*	spectinomycin	1175	*Staphylococcus capitis*	cetylpyridinium chloride
1126	*Staphylococcus haemolyticus*	spiramycin	1176	*Staphylococcus capitis*	chloramphenicol
1127	*Staphylococcus haemolyticus*	streptomycin	1177	*Staphylococcus capitis*	chlorhexidine
1128	*Staphylococcus haemolyticus*	sulfamethoxazole	1178	*Staphylococcus capitis*	ciprofloxacin
1129	*Staphylococcus haemolyticus*	teicoplanin	1179	*Staphylococcus capitis*	clindamycin
1130	*Staphylococcus haemolyticus*	telithromycin	1180	*Staphylococcus capitis*	colistin
1131	*Staphylococcus haemolyticus*	temocillin	1181	*Staphylococcus capitis*	dalfopristin
1132	*Staphylococcus haemolyticus*	temperature	1182	*Staphylococcus capitis*	dibekacin
1133	*Staphylococcus haemolyticus*	tetracycline	1183	*Staphylococcus capitis*	doxycycline
1134	*Staphylococcus haemolyticus*	tiamulin	1184	*Staphylococcus capitis*	ertapenem
1135	*Staphylococcus haemolyticus*	ticarcillin	1185	*Staphylococcus capitis*	erythromycin
1136	*Staphylococcus haemolyticus*	ticarcillin+clavulanic acid	1186	*Staphylococcus capitis*	ethidium bromide
1137	*Staphylococcus haemolyticus*	tigecycline	1187	*Staphylococcus capitis*	florfenicol
1138	*Staphylococcus haemolyticus*	tobramycin	1188	*Staphylococcus capitis*	fluoroquinolone
1139	*Staphylococcus haemolyticus*	trimethoprim	1189	*Staphylococcus capitis*	formaldehyde
1140	*Staphylococcus haemolyticus*	tylosin	1190	*Staphylococcus capitis*	fortimicin
1141	*Staphylococcus haemolyticus*	unknown aminoglycoside	1191	*Staphylococcus capitis*	fosfomycin
1142	*Staphylococcus haemolyticus*	unknown beta-lactam	1192	*Staphylococcus capitis*	fusidic acid
1143	*Staphylococcus haemolyticus*	unknown quinolone	1193	*Staphylococcus capitis*	gentamicin
1144	*Staphylococcus haemolyticus*	vancomycin	1194	*Staphylococcus capitis*	hydrogen peroxide
1145	*Staphylococcus haemolyticus*	virginiamycin m	1195	*Staphylococcus capitis*	hygromycin
1146	*Staphylococcus haemolyticus*	virginiamycin s	1196	*Staphylococcus capitis*	imipenem
1147	*Staphylococcus capitis*	amikacin	1197	*Staphylococcus capitis*	isepamicin
1148	*Staphylococcus capitis*	amoxicillin	1198	*Staphylococcus capitis*	kanamycin
1149	*Staphylococcus capitis*	amoxicillin+clavulanic acid	1199	*Staphylococcus capitis*	kasugamycin
1150	*Staphylococcus capitis*	ampicillin	1200	*Staphylococcus capitis*	lincomycin

Supplementary Table 13. Continued.

1201	*Staphylococcus capitis*	linezolid
1202	*Staphylococcus capitis*	lividomycin
1203	*Staphylococcus capitis*	meropenem
1204	*Staphylococcus capitis*	metronidazole
1205	*Staphylococcus capitis*	minocycline
1206	*Staphylococcus capitis*	mupirocin
1207	*Staphylococcus capitis*	nalidixic acid
1208	*Staphylococcus capitis*	neomycin
1209	*Staphylococcus capitis*	netilmicin
1210	*Staphylococcus capitis*	norfloxacin
1211	*Staphylococcus capitis*	oleandomycin
1212	*Staphylococcus capitis*	paromomycin
1213	*Staphylococcus capitis*	penicillin
1214	*Staphylococcus capitis*	piperacillin
1215	*Staphylococcus capitis*	piperacillin+clavulanic acid
1216	*Staphylococcus capitis*	piperacillin+tazobactam
1217	*Staphylococcus capitis*	pristinamycin ia
1218	*Staphylococcus capitis*	pristinamycin iia
1219	*Staphylococcus capitis*	quinupristin
1220	*Staphylococcus capitis*	quinupristin+dalfopristin
1221	*Staphylococcus capitis*	ribostamycin
1222	*Staphylococcus capitis*	rifampicin
1223	*Staphylococcus capitis*	sisomicin
1224	*Staphylococcus capitis*	spectinomycin
1225	*Staphylococcus capitis*	spiramycin
1226	*Staphylococcus capitis*	streptomycin
1227	*Staphylococcus capitis*	sulfamethoxazole
1228	*Staphylococcus capitis*	teicoplanin
1229	*Staphylococcus capitis*	telithromycin
1230	*Staphylococcus capitis*	temocillin
1231	*Staphylococcus capitis*	temperature
1232	*Staphylococcus capitis*	tetracycline
1233	*Staphylococcus capitis*	tiamulin
1234	*Staphylococcus capitis*	ticarcillin
1235	*Staphylococcus capitis*	ticarcillin+clavulanic acid
1236	*Staphylococcus capitis*	tigecycline
1237	*Staphylococcus capitis*	tobramycin
1238	*Staphylococcus capitis*	trimethoprim
1239	*Staphylococcus capitis*	tylosin
1240	*Staphylococcus capitis*	unknown aminoglycoside
1241	*Staphylococcus capitis*	unknown beta-lactam
1242	*Staphylococcus capitis*	unknown quinolone
1243	*Staphylococcus capitis*	vancomycin
1244	*Staphylococcus capitis*	virginiamycin m
1245	*Staphylococcus capitis*	virginiamycin s

"Our future lies not in our enthusiasm to

innovate but rather in our courage to regulate."

-Raymond Moodley

8 CONCLUSION AND FUTURE WORK

We stand at a remarkable juncture in the field of metagenomic sequencing. Driven by the advent of cost-effective sequencing techniques and the enhancement of computational processing, alongside other scientific advancements, this field has seen significant leaps forward. The promise that metagenomic sequencing holds for microbiology is transformative - it holds the potential to identify species agnostically and detect antimicrobial resistance (AMR) rapidly. In this thesis, I have developed a pipeline for direct-from-blood culture sequencing to predict species and antimicrobial sensitivity that can be employed in real-time, offering an alternative to traditional culture-based and current molecular based diagnostics. By enhancing our diagnostic tools and techniques, I've aimed to reveal a more comprehensive picture of the composition of bloodstream infections, offering the potential for targeted therapy in a more clinically meaningful way. Although not

discussed in this thesis, metagenomics has the additional benefit of tracking virulent or antibiotic-resistant clones, and identifying clusters of infection, which is critical for managing infectious disease outbreaks at the hospital, community, and public health level.

In Chapter 3, I conducted a comprehensive systematic review and meta-analysis examining metagenomic sequencing as a pathogen-agnostic tool for diagnosing clinical infections, on numerous clinical samples such as blood, cerebrospinal fluid, and orthopaedic fluids among others. The review emphasizes that despite the impressive progress made in metagenomic sequencing, the diagnostic field is still evaluating the level of accuracy necessary, given its unique advantages and disadvantages. The metagenomic approach can identify pathogens overlooked by culture techniques, however, accurately predicting AMR from metagenomic data poses significant difficulties. I found the categorical AMR agreement rate with phenotyping to be at 88%, while the major and very major error rates were at 5% and 24%, respectively, exceeding the FDA's current regulatory thresholds. This indicates that a better understanding and cataloguing of the mechanisms underlying AMR is required, and the algorithms to detect these within metagenomic data need optimization before accurate resistance prediction is possible. Despite these challenges, the potential for the evolution and advancement of clinical metagenomics is convincing. This systematic analysis has significantly contributed to the development of my own research methodology by pinpointing the robust and deficient aspects of current studies. Remarkably, it became evident that no investigations had employed a species-agnostic technique for processing clinical blood culture samples to predict species and resistance. Such an approach is instrumental in achieving an accurate performance estimate for clinical

applicability. Additionally, only a few studies (n=4) incorporated over 200 clinical samples, with none of these focused on blood or blood cultures from the literature search. An assessment of the quality controls employed in clinical metagenomic studies revealed a general deficiency, which informed my decision to incorporate negative, positive, and internal controls into my research design. The salient issue of contamination, seen in numerous studies, necessitated the adoption of meticulous laboratory procedures from the inception of my study. While many past studies were conducted using Illumina sequencing, my review found that newer studies employing Oxford Nanopore technology were viable and offered quicker results using real-time sequencing technology.

A significant impediment in the field of direct-from-sample sequencing lies in determining the optimal nucleic acid extraction methodologies from an array of existing extraction kits and techniques. In Chapter 4, to establish an optimized DNA extraction pipeline specifically designed for direct application to blood culture, I conducted a comparative assessment of prevalent DNA extraction methods and human DNA depletion techniques. In this experiment, the BiOstic Bacteraemia kit excelled in the extraction of total DNA, extracting 3x and 6x the amount compared to the QIAamp UCP Pathogen Mini kit and MolYsisTM Complete5 kits, respectively. This superior DNA yield could potentially enhance the sensitivity of bacterial detection, particularly in samples characterized by a low bacterial DNA concentration. Nevertheless, one must also consider the potential risk of escalated contamination due to higher DNA concentrations, necessitating a balance to optimize DNA concentration and minimize contamination. None of the human DNA depletion methods examined showed a significant impact on the extracted bacterial copy number. However, the saponin method demonstrated a noteworthy

range of 7-142x fold depletion in human copy numbers compared to other methodologies. The BiOstic kit achieved a nearly 137-fold higher bacterial-to-human DNA ratio compared to the Molysis kit, which was similar to the extraction ratio of the QIAamp kit. Negative controls for both water and culture-media showed insignificant differences concerning human and bacterial copy numbers. This finding suggests an absence of bacterial DNA in negative culture bottles, a positive sign for future metagenomic studies possibly sampling directly from these bottles. Regarding contamination, the Molysis extraction kit emerged as the primary offender, seemingly due to its complex protocol, with increased contamination observed across both human and bacterial controls. Contrastingly, the BiOstic kit displayed narrower ranges for copy numbers in negative controls, indicating reduced contamination concerns. From the findings of this experiment, the BiOstic Bacteraemia kit is recommended due to its superior DNA extraction capabilities, coupled with relatively lower observed contamination. The saponin method is suggested as the preferred choice for human DNA depletion, although its necessity could be negated when dealing with positive blood cultures containing high bacterial concentrations.

I then conducted a pilot study to further test if saponin pre-treatment vs. no pre-treatment improves bacterial sequencing yield. Significantly, saponin pre-treatment led to a substantial reduction in the number of human bases sequenced, but the number of bacterial bases sequenced remained consistent when compared to the no pre-treatment group. Furthermore, saponin pre-treatment, due to increased DNA manipulation, resulted in bacterial read lengths averaging 108 bases lower and more widely distributed than in the no pre-treatment group, implying that no pre-treatment leads to longer and more consistent DNA read lengths. While barcode

contamination was observed, it was minimal. From this, I concluded that saponin pre-treatment does not appear essential for achieving high bacterial sequencing depth using Oxford Nanopore when multiplexing 2-6 samples directly extracted from positive blood culture. Furthermore, from this sequencing pilot study, I discovered that Kraken2 produced erroneous bioinformatic classifications at times, often species dependent, which resulted in less accurate quantitative results.

Addressing erroneous bioinformatic contamination is a significant and important challenge for improving diagnostics using metagenomics, as these errors might imply co-infection and result in unnecessary antibiotic treatment. Understanding the degrees of such misclassification is essential for the development of effective diagnostic workflows. In Chapter 5, I was able to define and quantify the rates of bioinformatic species misclassification by using simulated reference genomes and real-world single isolate sequencing data. This comprehensive approach allows for the quantification of bioinformatic contamination and aids in making informed choices about optimal bioinformatic tools to maximize classification accuracy. I found that the Bracken classifier, when coupled with a "standard" (or larger) database, resulted in the most accurate performance across all species. This pairing provided the most accurate species abundance estimates for both short and long-read data. Importantly, I also showed that applying a set of specific filtering thresholds to longer error-prone reads from the Oxford Nanopore can perform equally well as less error prone, but shorter reads from Illumina sequencing to identify species. I also generated the first known catalogue of misclassified species, which serves as a comprehensive reference guide for understanding the rate of misclassification that can occur when using different sequencing platforms, classification tools, and databases. With the catalogue as a benchmark, I proposed

a method to apply the misclassification thresholds to metagenomic samples. This method aids in discerning the true microbial diversity in a sample and helps differentiate between genuine polymicrobial infection and mere bioinformatic contamination. Further into the research, I discovered that misclassification tends to occur more when species share higher average nucleotide identity. The benchmarking metrics derived from this research should play a pivotal role in the selection and implementation of metagenomic sequencing tools. My hope is that this will significantly contribute to the accuracy and optimization of metagenomic workflows, with substantial impacts on both research and clinical applications.

After recognising the limitations inherent in previous metagenomic studies, optimising DNA extraction methodologies, ensuring control over bioinformatic misclassification, and successfully executing a pilot study, I was prepared to implement a comprehensive direct-from-blood culture metagenomic workflow. In Chapter 6, I demonstrate a workflow aimed to deliver direct-from-blood culture speciation and AMR results, allowing for a detailed comparison of routine results and sample processing times against conventional microbiology workflows. After accounting for plausible co-infections, the study showed a 100% sensitivity (229/229) and specificity (205/205) rate in culture-positive samples, and 100% specificity (61/61) in culture-negative samples. It identified 13 co-infections, amounting to 6% of the processed samples, which were missed by standard laboratory procedures, showing the broad spectrum of pathogens that metagenomic sequencing could reveal. For instance, it detected *Bacteroides fragilis* in a patient with a liver abscess, and *Enterococcus faecium* as a co-infection, which were missed by conventional diagnostics, but in the latter example appeared from culture taken the following day from the same patient. Furthermore, 5/6 samples were

identified by sequencing where the routine lab had completely failed to identify the species of the bacteria grown. For example, sequencing found *Capnocytophaga canimorsus* in a patient that had evidence of a bite mark, in keeping with transmission from an animal source. Overall, this workflow demonstrated improved performance over traditional methods, offering species results within approximately 3.5 hours, which is more than half of the typical laboratory turnaround time. The stringent quality control measures included the use of negative, positive, and internal controls and meticulous laboratory work in a controlled environment further enhance the credibility and reproducibility of these findings. This study found minimal barcode contamination, with only 1 incorrectly assigned read observed for every 12,739 correctly assigned reads. The findings of this study pave the way for significant improvements in diagnosing and managing bloodstream infections, setting a benchmark for sensitivity, specificity, and the ability to reveal co-infections. It's expected that the workflow I've developed here will serve as a guide for future investigations in the field of metagenomics, providing a solid structure for quality control and accurate results.

In Chapter 7, I sought to refine the prediction of antimicrobial resistance, a critical component in effective treatment strategies for bloodstream infections. This involved benchmarking various existing tools designed to predict AMR. Through rigorous comparative analysis, ResFinder, particularly when utilized with raw reads from fastq input files, emerged as the top performer. Sensitivity and specificity for antimicrobial resistance prediction using ResFinder at the two-hour interval for *Staphylococcus aureus* was 100% (12/13) and 99% (138/139), and for *Escherichia coli* was 91% (53/58 drug-organism combinations) and 94% (505/539) respectively. Across the top 10 clinically relevant species (including *Coagulase-*

negative staphylococci), the overall AMR sensitivity and specificity was 91% and 90% respectively. Similar performance could be obtained from raw reads after only two hours of sequencing. Results were obtainable within 5hrs30min from positive blood culture, with the process of extraction to sequencing taking a median of 3hrs30min, followed by a complete AMR prediction within the next 2 hours, delivering results 22 hours sooner than traditional lab methods for the comparable samples. Furthermore, I identified a potential avenue for improving AMR prediction: the integration of multiple databases to expand the resistance catalogue. I noted that certain resistance-conferring genes, which were not present in the ResFinder database, were present in other databases. Incorporating these additional genes could enhance the predictive capabilities of the tool. Despite the promising results, it was clear that more research is needed. Across all the classifiers used, only *Staphylococcus aureus, Escherichia coli, Enterococcus sp., Klebsiella pneumonia, Enterobacter cloacae,* and *Coagulase-negative staphylococci* met FDA recommended standards, indicating a need for further understanding and characterization of resistance mechanisms. Poor sensitivity for detecting phenotypic resistance was seen in *Proteus mirabilis* and *Enterobacter cloacae*, while poor specificity was seen in *Enterobacter cloacae, Klebsiella pneumoniae* and *Pseudomonas aeruginosa*. These results emphasize the need for larger scale investigations that evaluate both phenotypic and genotypic aspects of these bacteria. Such studies, potentially leveraging Genome-Wide Association Studies, would contribute to a more nuanced understanding of the correlation between the genetic makeup of these organisms and their resistance traits, ultimately improving our predictive capabilities in the field of antimicrobial resistance.

In comparison to routine culture methods, metagenomic sequencing provides quicker results, and more comprehensive results than both culture methods and multiplex molecular tests like the BioFire® Blood Culture Identification 2 (BCID2) panel, by identifying an unprecedented range of pathogens and offering valuable data on antimicrobial resistance in a timely manner. Whole blood metagenomic sequencing compared to cell-free DNA, captures both intracellular and extracellular microbes, yielding a more complete diagnostic picture. Despite the potential of metagenomic sequencing for diagnosing infectious diseases, several barriers need to be addressed before it can be widely implemented in clinical settings. Firstly, efficient bacterial amplification and methods for depleting human DNA are required to enable direct sequencing from blood, circumventing the need for blood culture. Secondly, the sequencing workflow needs to be automated to boost efficiency and consistency, ensuring it is more suitable for routine clinical applications. Lastly, the significant costs and labour associated with metagenomic sequencing necessitate rigorous cost-effectiveness studies to substantiate their justification compared to the potential diagnostic benefits in terms of speed and accuracy. In the meantime, owing to its current high costs and labour-intensive processes, metagenomic sequencing might initially be adopted as a reference laboratory test, offering valuable insights for complicated or unresolved cases. As cost-effectiveness studies progress, they will provide critical data to assess whether the routine use of metagenomics in clinical diagnostics can be economically justified. Overcoming these challenges will unlock the full potential of metagenomic sequencing in the diagnosis and management of infectious diseases.

The results from the workflow developed in this thesis confirms the potential of metagenomic sequencing from blood culture samples as an effective and efficient

tool for identifying species and AMR. These advantages point towards more precise treatment strategies, potentially reducing the overuse of broad-spectrum antibiotics and helping to control the spread of antimicrobial resistance. The future impact of this work is far-reaching. It paves the way for improved diagnosis and treatment of infectious diseases, moving us closer to a model of personalized medicine. With the results from this study, we can see a future where metagenomic sequencing is an integral part of routine clinical practice. As bioinformatics tools improve and our understanding of antimicrobial resistance grows, we can expect to see these techniques used more widely and effectively. In conclusion, the work done in this thesis is a significant step forward in the field of infectious disease diagnostics for bloodstream infections. It has shown the practicality and effectiveness of metagenomic sequencing and pointed towards a future where this technique becomes a standard diagnostic tool, helping us better understand and treat infectious diseases.

9 REFERENCES

1. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem Biol* **5**, R245–R249 (1998).

2. Chiu, C. Y. & Miller, S. A. Clinical metagenomics. *Nat Rev Genet* **20**, 341–355 (2019).

3. Anson, L. W. *et al.* DNA extraction from primary liquid blood cultures for bloodstream infection diagnosis using whole genome sequencing. *J Med Microbiol* **67**, 347–357 (2018).

4. Wilson, M. *et al.* Chronic Meningitis Investigated via Metagenomic Next-Generation Sequencing. *JAMA Neurol* **75**, 947–955 (2018).

5. Doyle, R. M. *et al.* Direct whole-genome sequencing of sputum accurately identifies drug-resistant mycobacterium tuberculosis faster than MGIT culture sequencing. *J Clin Microbiol* **56**, e00666-18 (2018).

6. Public Health England. *UK Standards for Microbiology Investigations: Investigation of blood cultures (for organisms other than Mycobacterium species)*. (2019).

7. Schieffer, K. M. *et al.* Multicenter evaluation of the Sepsityper™ extraction kit and MALDI-TOF MS for direct identification of positive blood culture isolates using the BD BACTEC™ FX and VersaTREK® diagnostic blood culture systems. *J Appl Microbiol* **116**, 934–941 (2014).

8. Weinstein, M. P. *et al.* Controlled evaluation of BacT/Alert standard aerobic and FAN aerobic blood culture bottles for detection of bacteremia and fungemia. *J Clin Microbiol* **33**, 978–981 (1995).

9. Petti, C. A. *et al.* Utility of extended blood culture incubation for isolation of Haemophilus, Actinobacillus, Cardiobacterium, Eikenella, and Kingella organisms: a retrospective multicenter evaluation. *J Clin Microbiol* **44**, 257–259 (2006).

10. World Health Organisation. *Thirteenth general programme of work, 2019–2023: promote health, keep the world safe, serve the vulnerable*. https://apps.who.int/iris/handle/10665/324775 (2019).

11. Eliopoulos, G. M., Cosgrove, S. E. & Carmeli, Y. The impact of antimicrobial resistance on health and economic outcomes. *Clin. Infect. Dis.* **36**, 1433–1437 (2003).

12. Ventola, C. L. The antibiotic resistance crisis: part 1: causes and threats. *Pharm. Ther.* **40**, 277 (2015).

13. Centers for Disease Control and Prevention. *Antibiotic/antimicrobial resistance (AR/AMR): biggest threats and data*. *CDC http://www. cdc. gov/drugresistance/threat-report-2013/(updated 26 Nov 2018)* (2019).

14. Antoñanzas, F. & Goossens, H. The economics of antibiotic resistance: a claim for personalised treatments. *Eur J Health Econ* **20**, 483–485 (2019).

15. van Belkum, A. *et al.* Developmental roadmap for antimicrobial susceptibility testing systems. *Nat Rev Microbiol* **17**, 51–62 (2019).

16. Centers for Disease Control. *World Health Day: media fact sheet: antimicrobial resistance: no action today, no cure tomorrow*. (2013).

17. Reller, L. B., Weinstein, M., Jorgensen, J. H. & Ferraro, M. J. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin. Infect. Dis.* **49**, 1749–1755 (2009).

18. Stefaniuk, E., Baraniak, A., Gniadkowski, M. & Hryniewicz, W. Evaluation of the BD Phoenix automated identification and susceptibility testing system in clinical microbiology laboratory practice. *Eur. J. Clin. Microbiol. Infect. Dis.* **22**, 479–485 (2003).

19. Shetty, N., Hill, G. & Ridgway, G. L. The Vitek analyser for routine bacterial identification and susceptibility testing: protocols, problems, and pitfalls. *J Clin Pathol* **51**, 316–323 (1998).

20. Idelevich, E. A. *et al.* Evaluation of an automated system for reading and interpreting disk diffusion antimicrobial susceptibility testing of fastidious bacteria. *PLoS One* **11**, e0159183 (2016).

21. Lutgring, J. D. *et al.* Evaluation of the Accelerate Pheno system: results from two academic medical centers. *J Clin Microbiol* **56**, e01672-17 (2018).

22. Choi, J. *et al.* Rapid drug susceptibility test of Mycobacterium tuberculosis using microscopic time-lapse imaging in an agarose matrix. *Appl Microbiol Biotechnol* **100**, 2355–2365 (2016).

23. Freitas, A. R. & Werner, G. Antibiotic susceptibility testing for therapy and antimicrobial resistance surveillance: genotype beats phenotype? *Future Microbiol* **17**, 1093–1097 (2022).

24. Kohler, P. *et al.* Prevalence of vancomycin-variable Enterococcus faecium (VVE) among vanA-positive sterile site isolates and patient factors associated with VVE bacteremia. *PLoS One* **13**, e0193926 (2018).

25. Bender, J. K. *et al.* Update on prevalence and mechanisms of resistance to linezolid, tigecycline and daptomycin in enterococci in Europe: Towards a common nomenclature. *Drug Resist. Updat.* **40**, 25–39 (2018).

26. Kaase, M. *et al.* Multicentre investigation of carbapenemase-producing Escherichia coli and Klebsiella pneumoniae in German hospitals. *Int. J. Med. Microbiol.* **306**, 415–420 (2016).

27. Emeraud, C. *et al.* Screening of OXA-244 producers, a difficult-to-detect and emerging OXA-48 variant? *J. Antimicrob. Chemother.* **75**, 2120–2123 (2020).

28. Smelikova, E., Tkadlec, J. & Krutova, M. How to: screening for mcr-mediated resistance to colistin. *Clin. Microbiol. Infect.* (2021).

29. Bortolaia, V. *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**, 3491–3500 (2020).

30. Cantera, J. L. *et al.* Assessment of eight nucleic acid amplification technologies for potential use to detect infectious agents in low-resource settings. *PLoS One* **14**, e0215756 (2019).

31. Fluit, A. C., Visser, M. R. & Schmitz, F.-J. Molecular detection of antimicrobial resistance. *Clin Microbiol Rev* **14**, 836–871 (2001).

32. Miller, M. B. & Tang, Y.-W. Basic concepts of microarrays and potential applications in clinical microbiology. *Clin Microbiol Rev* **22**, 611–633 (2009).

33. Li, Y., Fan, P., Zhou, S. & Zhang, L. Loop-mediated isothermal amplification (LAMP): A novel rapid detection platform for pathogens. *Microb Pathog* **107**, 54–61 (2017).

34. Lapierre, S. G. & Drancourt, M. rpoB Targeted Loop-Mediated Isothermal Amplification (LAMP) Assay for Consensus Detection of Mycobacteria Associated With Pulmonary Infections. *Front Med (Lausanne)* **5**, 332 (2018).

35. Chen, X. *et al.* The rapid and visual detection of methicillin-susceptible and methicillin-resistant Staphylococcus aureus using multiplex loop-mediated isothermal amplification linked to a nanoparticle-based lateral flow biosensor. *Antimicrob Resist Infect Control* **9**, 111 (2020).

36. Zhong, L. L. *et al.* Multiplex loop-mediated isothermal amplification (multi-LAMP) assay for rapid detection of mcr-1 to mcr-5 in colistin-resistant bacteria. *Infect Drug Resist* **12**, 1877 (2019).

37. Ramanan, P., Bryson, A. L., Binnicker, M. J., Pritt, B. S. & Patel, R. Syndromic panel-based testing in clinical microbiology. *Clin Microbiol Rev* **31**, (2018).

38. Senok, A. *et al.* Clinical Impact of the BIOFIRE Blood Culture Identification 2 Panel in Adult Patients with Bloodstream Infection: A Multicentre Observational Study in the United Arab Emirates. *Diagnostics 2023, Vol. 13, Page 2433* **13**, 2433 (2023).

39. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (1979)* **270**, 467–470 (1995).

40. Frye, J. G. *et al.* Development of a DNA microarray to detect antimicrobial resistance genes identified in the National Center for Biotechnology Information database. *Microb. Drug Resist.* **16**, 9–19 (2010).

41. Huang, W.-L., Hsu, Z.-J., Chang, T. C. & Jou, R. Rapid and accurate detection of rifampin and isoniazid-resistant Mycobacterium tuberculosis using an oligonucleotide array. *Clin. Microbiol. Infect.* **20**, O542–O549 (2014).

42. Gryadunov, D. *et al.* Evaluation of hybridisation on oligonucleotide microarrays for analysis of drug-resistant Mycobacterium tuberculosis. *Clin. Microbiol. Infect.* **11**, 531–539 (2005).

43. Freije, C. A. & Sabeti, P. C. Detect and destroy: CRISPR-based technologies for the response against viruses. *Cell Host Microbe* **29**, 689–703 (2021).

44. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333–351 (2016).

45. Kozarewa, I. *et al.* Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* **6**, 291–295 (2009).

46. Head, S. R. *et al.* Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques* **56**, 61–77 (2014).

47. Gu, W., Miller, S. & Chiu, C. Y. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu. Rev. Pathol.: Mech. Dis.* **14**, 319–338 (2019).

48. Fredericks, D. N. & Relman, D. A. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* **9**, 18–33 (1996).

49. Wilson, M. R. *et al.* Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N. Engl. J. Med.* **380**, 2327–2340 (2019).

50. Sardi, S. I. *et al.* Coinfections of zika and chikungunya viruses in bahia, Brazil, identified by metagenomic next-generation sequencing. *J Clin Microbiol* **54**, 2348–2353 (2016).

51. Rossoff, J. *et al.* Noninvasive Diagnosis of Infection Using Plasma Next-Generation Sequencing: A Single-Center Experience. *Open Forum Infect Dis* **6**, (2019).

52. Hong, D. K. *et al.* Liquid biopsy for infectious diseases: sequencing of cell-free plasma to detect pathogen DNA in patients with invasive fungal disease. *Diagn Microbiol Infect Dis* **92**, 210–213 (2018).

53. Parize, P. *et al.* Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. *Clin. Microbiol. Infect.* **23**, 574.e1-574.e6 (2017).

54. Greninger, A. L. *et al.* Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med* **7**, 99 (2015).

55. Horiba, K. *et al.* Comprehensive detection of pathogens in immunocompromised children with bloodstream infections by next-generation sequencing. *Sci Rep* **8**, 3784 (2018).

56. Ruppé, E. *et al.* Clinical metagenomics of bone and joint infections: a proof of concept study. *Sci Rep* **7**, 7718 (2017).

57. Street, T. L. *et al.* Clinical metagenomic sequencing for species identification and antimicrobial resistance prediction in orthopedic device infection. *J Clin Microbiol* **60**, e02156-21 (2022).

58. Miller, S. *et al.* Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res* **29**, 831–842 (2019).

59. van Belkum, A. & Rochas, O. Laboratory-based and point-of-care testing for MSSA/MRSA detection in the age of whole genome sequencing. *Front Microbiol* **9**, 1437 (2018).

60. Witney, A. A. *et al.* Clinical use of whole genome sequencing for Mycobacterium tuberculosis. *BMC Med* **14**, 1–7 (2016).

61. Gordon, N. C. *et al.* Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol* **52**, 1182–1191 (2014).

62. Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. *J. Antimicrob. Chemother.* **68**, 2234–2244 (2013).

63. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinform.* **19**, 1–8 (2018).

64. Govender, K. N., Street, T. L., Sanderson, N. D. & Eyre, D. W. Metagenomic sequencing as a pathogen-agnostic clinical diagnostic tool for infectious diseases: a systematic review and meta-analysis of diagnostic test accuracy studies. *J Clin Microbiol* **59**, e02916-20 (2021).

65. Food and Drug Administration. *Class II special controls guidance document: antimicrobial susceptibility test (AST) systems; guidance for industry and FDA.* (2003).

66. CRyPTIC Consortium and the 100, 000 Genomes Project. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N. Engl. J. Med.* **379**, 1403–1415 (2018).

67. Arango-Argoty, G. *et al.* DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 1–15 (2018).

68. Gygli, S. M., Borrell, S., Trauner, A. & Gagneux, S. Antimicrobial resistance in Mycobacterium tuberculosis: Mechanistic and evolutionary perspectives. *FEMS Microbiol Rev* **41**, 354–373 (2017).

69. George, S. *et al.* Resolving plasmid structures in enterobacteriaceae using the MinION nanopore sequencer: Assessment of MinION and MinION/illumina hybrid data assembly approaches. *Microb Genom* **3**, (2017).

70. Ghurye, J. S., Cepeda-Espinoza, V. & Pop, M. Metagenomic assembly: Overview, challenges and applications. *Yale J. Biol. Med.* **89**, 353–362 (2016).

71. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824–834 (2017).

72. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 1–13 (2019).

73. Lindgreen, S., Adair, K. L. & Gardner, P. P. An evaluation of the accuracy and speed of metagenome analysis tools. *Sci Rep* **6**, 19233 (2016).

74. Bremges, A. & McHardy, A. C. Critical Assessment of Metagenome Interpretation Enters the Second Round. *mSystems* **3**, (2018).

75. Kashinskaya, E. N., Andree, K. B., Simonov, E. P. & Solovyev, M. M. DNA extraction protocols may influence biodiversity detected in the intestinal microbiome: a case study from wild Prussian carp, Carassius gibelio. *FEMS Microbiol Ecol* **93**, (2017).

76. Hasan, M. 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* **54**, 919–927 (2016).

77. Charalampous, T. *et al.* Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat Biotechnol* **37**, 783–792 (2019).

78. Gu, W. *et al.* Depletion of Abundant Sequences by Hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol* **17**, 41 (2016).

79. Corless, C. E. *et al.* Contamination and sensitivity issues with a real-time universal 16s rRNA PCR. *J Clin Microbiol* **38**, 1747–1752 (2000).

80. Direito, S. O. L., Zaura, E., Little, M., Ehrenfreund, P. & Röling, W. F. M. Systematic evaluation of bias in microbial community profiles induced by whole genome amplification. *Environ Microbiol* **16**, 643–657 (2014).

81. Robinson, E. R., Walker, T. M. & Pallen, M. J. Genomics and outbreak investigation: From sequence to consequence. *Genome Med* **5**, 36 (2013).

82. Hasman, H. *et al.* Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J Clin Microbiol* **52**, 139–146 (2014).

83. Sahl, J. W. *et al.* Phylogenetically typing bacterial strains from partial SNP genotypes observed from direct sequencing of clinical specimen metagenomic data. *Genome Med* **7**, 52 (2015).

84. Gilchrist, C. A., Turner, S. D., Riley, M. F., Petri, W. A. & Hewlett, E. L. Whole-genome sequencing in outbreak analysis. *Clin Microbiol Rev* **28**, 541–563 (2015).

85. Harbarth, S. *et al.* Inappropriate initial antimicrobial therapy and its effect on survival in a clinical trial of immunomodulating therapy for severe sepsis. *Am. J. Med.* **115**, 529–535 (2003).

86. Harbarth, S. *et al.* Epidemiology and prognostic determinants of bloodstream infections in surgical intensive care. *Arch. Surg.* **137**, 1353–1359 (2002).

87. Garrouste-Orgeas, M. *et al.* Excess Risk of Death from Intensive Care Unit--Acquired Nosocomial Bloodstream Infections: A Reappraisal. *Clin. Infect. Dis.* **42**, 1118–1126 (2006).

88. Votintseva, A. A. *et al.* Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. *J Clin Microbiol* **55**, 1285–1298 (2017).

89. Hoshijima, M. *et al.* Roles of heterotypic CCN2/CTGF-CCN3/NOV and homotypic CCN2-CCN2 interactions in expression of the differentiated phenotype of chondrocytes. *FEBS Journal* **279**, 3584–3597 (2012).

90. Perl, B. *et al.* Cost-Effectiveness of Blood Cultures for Adult Patients with Cellulitis. *Clin. Infect. Dis.* **29**, 1483–1488 (1999).

91. Udar, N. *et al.* Development and Analytical Validation of a DNA Dual-Strand Approach for the US Food and Drug Administration–Approved Next-Generation

Sequencing–Based Praxis Extended RAS Panel for Metastatic Colorectal Cancer Samples. *J. Mol. Diagn.* **22**, 159–178 (2020).

92. Linsenmeyer, K. *et al.* Culture if spikes? Indications and yield of blood cultures in hospitalized medical patients. *J Hosp Med* **11**, 336–340 (2016).

93. Burd, E. M. Validation of laboratory-developed molecular assays for infectious diseases. *Clin Microbiol Rev* **23**, 550–576 (2010).

94. Food and Drug Administration. *Infectious disease next generation sequencing based diagnostic devices: microbial identification and detection of antimicrobial resistance and virulence markers: draft guidance for industry and FDA*. (2016).

95. Blauwkamp, T. A. *et al.* Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol* **4**, 663–674 (2019).

96. Thoendel, M. J. *et al.* Identification of prosthetic joint infection pathogens using a shotgun metagenomics approach. *Clin. Infect. Dis.* **67**, 1333–1338 (2018).

97. Roca, I. *et al.* The global threat of antimicrobial resistance: Science for intervention. *New Microbes New Infect* **6**, 22–29 (2015).

98. O' Neil, J. *Review on Antibiotic resisitance. Antimicrobial Resistance : Tackling a crisis for the health and wealth of nations*. *Health and Wealth Nations* vol. 1 https://amr-review.org/sites/default/files/AMR Review Paper - Tackling a crisis for the health and wealth of nations_1.pdf (2014).

99. Anscombe, C., Misra, R. & Gharbia, S. *Rapid whole genome amplification and sequencing of low cell numbers in a bacteraemia model*. *bioRxiv* https://www.biorxiv.org/content/early/2018/04/26/153965%0Ahttp://biorxiv.org/content/early/2017/06/23/153965.abstract (2017) doi:10.1101/153965 (preprint).

100. Gwinn, M., Maccannell, D. & Armstrong, G. L. Next-Generation Sequencing of Infectious Pathogens. *JAMA - J. Am. Med. Assoc.* **321**, 893–894 (2019).

101. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* **6**, e1000097 (2009).

102. Whiting, P. F. *et al.* Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* **155**, 529–536 (2011).

103. Schwarzer, G., Carpenter, J. R. & Rücker, G. *Meta-analysis with R*. (Springer, 2015).

104. Somasekar, S. *et al.* Viral Surveillance in Serum Samples From Patients With Acute Liver Failure By Metagenomic Next-Generation Sequencing. *Clin. Infect. Dis.* **65**, 1477–1485 (2017).

105. Farnaes, L. *et al.* Community-acquired pneumonia in children: cell-free plasma sequencing for diagnosis and management. *Diagn Microbiol Infect Dis* **94**, 188–191 (2019).

106. Graf, E. H. *et al.* Unbiased Detection of Respiratory Viruses by Use of RNA Sequencing-Based Metagenomics: a Systematic Comparison to a Commercial PCR Panel. *J Clin Microbiol* **54**, 1000 LP – 1007 (2016).

107. Schlaberg, R. *et al.* Viral Pathogen Detection by Metagenomics and Pan-Viral Group Polymerase Chain Reaction in Children With Pneumonia Lacking Identifiable Etiology. *J Infect Dis* **215**, 1407–1415 (2017).

108. Street, T. L. *et al.* Molecular diagnosis of orthopedic-device-related infection directly from sonication fluid by metagenomic sequencing. *J Clin Microbiol* **55**, 2334–2347 (2017).

109. Ivy, M. I. *et al.* Direct detection and identification of prosthetic joint infection pathogens in synovial fluid by metagenomic shotgun sequencing. *J Clin Microbiol* **56**, e00402-18 (2018).

110. Xing, X. W. *et al.* Apparent performance of metagenomic next-generation sequencing in the diagnosis of cryptococcal meningitis: A descriptive study. *J Med Microbiol* **68**, 1204–1210 (2019).

111. Hong, N. T. T. *et al.* Performance of Metagenomic Next-Generation Sequencing for the Diagnosis of Viral Meningoencephalitis in a Resource-Limited Setting. *Open Forum Infect Dis* **7**, ofaa046–ofaa046 (2020).

112. Doan, T. *et al.* Metagenomic DNA Sequencing for the Diagnosis of Intraocular Infections. *Ophthalmology* **124**, 1247–1248 (2017).

113. Cheng, J. *et al.* Detection of pathogens from resected heart valves of patients with infective endocarditis by next-generation sequencing. *Int. J. Infect. Dis.* **83**, 148–153 (2019).

114. Simner, P. J., Miller, S. & Carroll, K. C. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin. Infect. Dis.* **66**, 778–788 (2018).

115. Weinstein, M. P. & Doern, G. V. A critical appraisal of the role of the clinical microbiology laboratory in the diagnosis of bloodstream infections. *J Clin Microbiol* **49**, S26 LP-S29 (2011).

116. Dolley, C. S. The Technology of Bacteria Investigation. *J Am Med Assoc* **10**, 278 (1885).

117. Amrane, S. & Lagier, J.-C. Metagenomic and clinical microbiology. *Hum Microb J* **9**, 1–6 (2018).

118. Bharucha, T. *et al.* STROBE-metagenomics: a STROBE extension statement to guide the reporting of metagenomics studies. *Lancet Infect Dis* **20**, e251-60 (2020).

119. Bossuyt, P. M. *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Clin Chem* **61**, 1446–1452 (2015).

120. Brooks, J. P. *et al.* The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol* **15**, 1–14 (2015).

121. Bag, S. *et al.* An improved method for high quality metagenomics DNA extraction from human and environmental samples. *Sci Rep* **6**, 26775 (2016).

122. Meadow, J. F. *et al.* Humans differ in their personal microbial cloud. *PeerJ* **3**, e1258 (2015).

123. Salzberg, S. L. *et al.* Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system. *Neurol.: Neuroimmunol. NeuroInflammation* **3**, (2016).

124. Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K. & Hugenholtz, P. A bioinformatician's guide to metagenomics. *Microbiol. Mol. Biol. Rev.* **72**, 557–578 (2008).

125. Nadkarni, M. A., Martin, F. E., Jacques, N. A. & Hunter, N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology (N Y)* **148**, 257–266 (2002).

126. Herrera, L. J. *et al.* Quantitative Analysis of Circulating Plasma DNA as a Tumor Marker in Thoracic Malignancies. *Clin. Chem.* **51**, (2005).

127. Rang, F. J., Kloosterman, W. P. & de Ridder, J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biol* **19**, 90 (2018).

128. Xu, Y. *et al.* Detection of viral pathogens with multiplex nanopore MinION sequencing: Be careful with cross-Talk. *Front Microbiol* **9**, (2018).

129. Yang, C., Chu, J., Warren, R., GigaScience, I. B.- & 2017, undefined. NanoSim: nanopore sequence read simulator based on statistical characterization. *Gigascience* **6**, gix010 (2017).

130. Paim, T., Reiter, K., Oliveira, C., Control, P. D.-J. I. & 2013, undefined. MALDI-TOF MS performance to identify gram-positive cocci clinical isolates in Porto Alegre/RS, Brazil. *Brazil. J Infect Control* **2**, 112–6 (2013).

131. Oxford Nanopore Technologies. https://nanoporetech.com/products/kits (2020).

132. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**, 1–12 (2014).

133. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci* **3**, e104 (2017).

134. Walker, M. A. *et al.* GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics* **34**, 4287–4289 (2018).

135. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* **7**, 1–9 (2016).

136. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59–60 (2015).

137. Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* **12**, 902–903 (2015).

138. Milanese, A. *et al.* Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun* **10**, 1–11 (2019).

139. Ye, S. H., Siddle, K. J., Park, D. J. & Sabeti, P. C. Benchmarking metagenomics tools for taxonomic classification. *Cell* **178**, 779–794 (2019).

140. Mavromatis, K. *et al.* Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nat Methods* **4**, 495–500 (2007).

141. Meyer, F. *et al.* Assessing taxonomic metagenome profilers with OPAL. *Genome Biol* **20**, 1–10 (2019).

142. Sczyrba, A. *et al.* Critical assessment of metagenome interpretation—a benchmark of metagenomics software. *Nat Methods* **14**, 1063–1071 (2017).

143. McIntyre, A. B. R. *et al.* Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. *Genome Biol* **18**, 1–19 (2017).

144. McArdle, A. J. & Kaforou, M. Sensitivity of shotgun metagenomics to host DNA: abundance estimates depend on bioinformatic tools and contamination is the main issue. *Access Microbiol* **2**, (2020).

145. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–D745 (2016).

146. Karst, S. M. *et al.* High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat Methods* **18**, 165–169 (2021).

147. Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read simulator. *Bioinformatics* **28**, 593–594 (2012).

148. L, S. T. *et al.* Clinical Metagenomic Sequencing for Species Identification and Antimicrobial Resistance Prediction in Orthopedic Device Infection. *J Clin Microbiol* **60**, e02156-21 (2022).

149. Břinda, K., Yang, C., Chu, J., Linthorst, J. & Franus, W. karel-brinda/NanoSim-H: NanoSim-H 1.1.0.4. https://zenodo.org/record/1341250 (2018) doi:10.5281/ZENODO.1341250.

150. Salzberg, S. L. & Wood, D. E. Releasing the Kraken. *Front. bioinform.* **1**, 75 (2021).

151. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: Rapid and sensitive classification of metagenomic sequences. *Genome Res* **26**, 1721–1729 (2016).

152. Lipworth, S. *et al.* Ten-year longitudinal molecular epidemiology study of Escherichia coli and Klebsiella species bloodstream infections in Oxfordshire, UK. *Genome Med* **13**, 144 (2021).

153. Young, B. C. *et al.* Antimicrobial resistance determinants are associated with Staphylococcus aureus bacteraemia and adaptation to the healthcare environment: a bacterial genome-wide association study. *Microb Genom* **7**, 700 (2021).

154. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 1–8 (2018).

155. Koenker, R. *et al. Package 'quantreg.' Cran R-project. org* https://cran. rproject. org/web/packages/quantreg/quantreg. pdf. (2018).

156. Zheng, W. *et al.* StreptoBase: An Oral Streptococcus mitis Group Genomic Resource and Analysis Platform. *PLoS One* **11**, e0151908–e0151908 (2016).

157. Ehling-Schulz, M., Lereclus, D. & Koehler, T. M. The Bacillus cereus Group: Bacillus Species with Pathogenic Potential. *Microbiol Spectr* **7**, 10.1128/microbiolspec.GPP3-0032–2018 (2019).

158. C., M. P. *et al.* Specific Distribution within the Enterobacter cloacae Complex of Strains Isolated from Infected Orthopedic Implants. *J Clin Microbiol* **47**, 2489–2495 (2009).

159. Liang, Q., Bible, P. W., Liu, Y., Zou, B. & Wei, L. DeepMicrobes: taxonomic classification for metagenomics with deep learning. *NAR Genom Bioinform* **2**, lqaa009 (2020).

160. Tedersoo, L., Albertsen, M., Anslan, S. & Callahan, B. Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Appl Environ Microbiol* **87**, e00626-21 (2021).

161. Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**, 1–16 (2020).

162. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* **20**, 129 (2019).

163. Pearman, W. S., Freed, N. E. & Silander, O. K. Testing the advantages and disadvantages of short-and long-read eukaryotic metagenomics using simulated reads. *BMC Bioinform.* **21**, 1–15 (2020).

164. Fleischmann, C. *et al.* Assessment of global incidence and mortality of hospital-treated sepsis current estimates and limitations. *Am J Respir Crit Care Med* **193**, 259–272 (2016).

165. Ammerlaan, H. S. M. *et al.* Secular trends in nosocomial bloodstream infections: Antibiotic-resistant bacteria increase the total burden of infection. *Clin. Infect. Dis.* **56**, 798–805 (2013).

166. Kumar, A. *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* **34**, 1589–1596 (2006).

167. d'Humières, C. *et al.* The potential role of clinical metagenomics in infectious diseases: Therapeutic perspectives. *Drugs* **81**, 1453–1466 (2021).

168. Boolchandani, M., D'Souza, A. W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20**, 356–370 (2019).

169. Kafetzopoulou, L. E. *et al.* Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science (1979)* **363**, 74–77 (2019).

170. Loman, N. J. *et al.* A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of shiga-toxigenic Escherichia coli O104:H4. *J Am Med Assoc* **309**, 1502–1510 (2013).

171. Harris, P. N. A. *et al. Rapid nanopore metagenomic sequencing and predictive susceptibility testing of positive blood cultures from intensive care patients with sepsis*. *medRxiv* (2023).

172. Govender, K. N. & Eyre, D. W. Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications. *Microb Genom* **8**, 000886 (2022).

173. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).

174. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19126–19131 (2009).

175. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R news* **2**, 18–22 (2002).

176. Normand, A.-C. *et al.* Decision criteria for MALDI-TOF MS-based identification of filamentous fungi using commercial and in-house reference databases. *BMC Microbiol* **17**, 1–17 (2017).

177. Rosa, N. M., Penati, M., Fusar-Poli, S., Addis, M. F. & Tola, S. Species identification by MALDI-TOF MS and gap PCR–RFLP of non-aureus Staphylococcus, Mammaliicoccus, and Streptococcus spp. associated with sheep and goat mastitis. *Vet Res* **53**, 84 (2022).

178. Spanu, T. *et al.* Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to rpoB gene sequencing for species identification of bloodstream infection staphylococcal isolates. *Clin. Microbiol. Infect.* **17**, 44–49 (2011).

179. Jaen-Luchoro, D. *et al.* Corynebacterium sanguinis sp. nov., a clinical and environmental associated corynebacterium. *Syst Appl Microbiol* **43**, 126039 (2020).

180. Mathers, A. J. *et al.* Klebsiella quasipneumoniae provides a window into carbapenemase gene transfer, plasmid rearrangements, and patient interactions with the hospital environment. *Antimicrob Agents Chemother* **63**, e02513-18 (2019).

181. Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **20**, 593–621 (2007).

182. Lin, P.-C., Chang, C.-L., Chung, Y.-H., Chang, C.-C. & Chu, F.-Y. Revisiting factors associated with blood culture positivity: Critical factors after the introduction of automated continuous monitoring blood culture systems. *Medicine* **101**, (2022).

183. Liu, V. X. *et al.* The timing of early antibiotics and hospital mortality in sepsis. *Am J Respir Crit Care Med* **196**, 856–863 (2017).

184. Chen, H. *et al.* Clinical evaluation of cell-free and cellular metagenomic next-generation sequencing of infected body fluids. *J Adv Res* (2023) doi:https://doi.org/10.1016/j.jare.2023.02.018.

185. Chabriere, E., Bassène, H., Drancourt, M. & Sokhna, C. MALDI-TOF MS and point of care are disruptive diagnostic tools in Africa. *New Microbes New Infect* **26**, S83–S88 (2018).

186. Tran, A., Alby, K., Kerr, A., Jones, M. & Gilligan, P. H. Cost savings realized by implementation of routine microbiological identification by matrix-assisted laser desorption ionization–time of flight mass spectrometry. *J Clin Microbiol* **53**, 2473–2479 (2015).

187. Greninger, A. L. The challenge of diagnostic metagenomics. *Expert Rev Mol Diagn* **18**, 605–615 (2018).

188. Marchukov, D., Li, J., Juillerat, P., Misselwitz, B. & Yilmaz, B. Benchmarking microbial DNA enrichment protocols from human intestinal biopsies. *Front Genet* **14**, 1184473 (2023).

189. Goto, M. & Al-Hasan, M. N. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. *Clin. Microbiol. Infect.* **19**, 501–509 (2013).

190. Ikuta, K. S. *et al.* Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **400**, 2221–2248 (2022).

191. Afshari, A., Schrenzel, J., Ieven, M. & Harbarth, S. Bench-to-bedside review: Rapid molecular diagnostics for bloodstream infection-a new frontier? *Crit Care* **16**, 1–12 (2012).

192. Endimiani, A., Tamborini, A., Luzzaro, F., Lombardi, G. & Toniolo, A. A two-year analysis of risk factors and outcome in patients with bloodstream infection. *Jpn J Infect Dis* **56**, 1–7 (2003).

193. Köck, R. *et al.* Implementation of short incubation MALDI-TOF MS identification from positive blood cultures in routine diagnostics and effects on empiric antimicrobial therapy. *Antimicrob Resist Infect Control* **6**, 1–7 (2017).

194. Buehler, S. S. *et al.* Effectiveness of practices to increase timeliness of providing targeted therapy for inpatients with bloodstream infections: A laboratory medicine best practices systematic review and meta-analysis. *Clin Microbiol Rev* **29**, 59–103 (2015).

195. Courvalin, P. Genotypic approach to the study of bacterial resistance to antibiotics. *Antimicrob Agents Chemother* **35**, 1019–1023 (1991).

196. Lebovitz, E. E. & Burbelo, P. D. Commercial multiplex technologies for the microbiological diagnosis of sepsis. *Mol Diagn Ther* **17**, 221–231 (2013).

197. Alcock, B. P. *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* **48**, D517–D525 (2020).

198. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* **11**, 12728 (2021).

199. Lu, J. *et al.* Metagenome analysis using the Kraken software suite. *Nat Protoc* **17**, 1–25 (2022).

200. Gweon, H. S. *et al.* The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. *Environ Microbiome* **14**, 7 (2019).

201. Neuert, S. *et al.* Prediction of phenotypic antimicrobial resistance profiles from whole genome sequences of non-typhoidal Salmonella enterica. *Front Microbiol* **9**, 592 (2018).

202. Day, M. R. *et al.* Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of Salmonella enterica serovars Typhi and Paratyphi. *J. Antimicrob. Chemother.* **73**, 365–372 (2018).

203. Karlowsky, J. A. *et al.* Prevalence and antimicrobial susceptibilities of bacteria isolated from blood cultures of hospitalized patients in the United States in 2002. *Ann Clin Microbiol Antimicrob* **3**, 1–8 (2004).

204. Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D. & Aarestrup, F. M. ResFinder–an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom* **8**, (2022).

205. Food and Drug Administration. *Guidance for industry and FDA. Class II special controls guidance document: antimicrobial susceptibility test (AST) systems*. *Center for Devices and Radiological Health, Food and Drug Administration, US Department of Health and Human Services, Silver Spring, MD* (2009).

206. Su, M., Satola, S. W. & Read, T. D. Genome-based prediction of bacterial antibiotic resistance. *J Clin Microbiol* **57**, 10–1128 (2019).

207. Gordon, N. C. *et al.* Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol* **52**, 1182–1191 (2014).

208. Bradley, P. *et al.* Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat Commun* **6**, 10063 (2015).

209. Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. *J. Antimicrob. Chemother.* **68**, 2234–2244 (2013).

210. Von Wintersdorff, C. J. H. *et al.* Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front Microbiol* **7**, 173 (2016).

211. Burnham, C.-A. D., Leeds, J., Nordmann, P., O'Grady, J. & Patel, J. Diagnosing antimicrobial resistance. *Nat Rev Microbiol* **15**, 697–703 (2017).

212. European Committee for Antimicrobial Susceptibility Testing (EUCAST) of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID). *Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. Clin. Microbiol. Infect.* vol. 9 https://doi.org/10.1046/j.1469-0691.2003.00790.x (2003).

“Wine is just bottled music.”

-Lise du Buisson