Rapid whole genome amplification and sequencing of low cell numbers in a bacteraemia model

Catherine Anscombe [*1,2], Raju.V Misra[1] ,Saheer Gharbia[1]

[1] Genomics Research Unit, Public Health England, London, United Kingdom

[2] Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

* Corresponding Author: Catherine Anscombe, the Hospital for Tropical Diseases, 764 Vo Van Kiet, Dist. 5, Ho Chi Minh City, Vietnam. canscombe@oucru.org

Words count: Abstract; 154, Text; 2530

Abstract

Whilst next generation sequencing is frequently used to whole genome sequence bacteria from cultures, it is rarely achieved direct from sample, and even more rarely performed in a clinically relevant time frame. To demonstrate the potential of direct from blood sequencing a bacteraemia model was developed, using defibrinated horse blood to model whole human blood infections. Sample processing included removal of erythrocytes and lysis of white blood cells, before rapid and accurate none targeted amplification. The rapid approach to allow direct from sample sequencing, allowed greater than 92% genome coverage of pathogens of interest whilst limiting the sequencing of host genome (less than 7% of all reads). Analysis of *de novo* assembled reads allowed accurate genotypic antibiotic resistance prediction. The sample processing would be easily applicable to multiple sequencing platforms. Overall this model provides evidence that it is currently possible to rapidly produce whole genome bacterial data from low cell number sterile site infections.

Introduction

Bacterial sequencing of clinical isolates has been used in many settings, including for virulence determinants[1–4], population structure[5 6] and outbreak investigations [7–10]. Almost exclusively current applications of sequencing rely either upon culture or targeted amplification imparting a diagnostic bias towards known pathogens and bacteria with known and relatively simple growth requirements. Bacterial whole genome sequencing is most often performed at reference or research centres to study population structures and evolution, with results rarely available in a clinically relevant time. One exception to this is the application of next generation sequencing to the diagnosis of tuberculosis[11,12], where WGS was shown, in some cases, to be quicker than current routine methods for predicting resistance; however this process still relies on culturing of the bacteria prior to sequencing. Direct from sample sequencing of pathogens using an untargeted (metagenomic) approach has the potential to lower turnaround times and lower diagnostic bias. Metagenomic sequencing techniques have

already been applied to ecological studies [13–16], gut microbiomes [17–19] and investigation and identification of viruses [20–23].

In order to sequence a bacterial genome, at least 1 ng of DNA is needed, which is the equivalent of over a 200000 copies of the *E. coli* K12 genome. To directly sequence pathogens present at low levels (as few as 1 bacterial cell per ml blood) such as those found in sterile site infections an unbiased and high fidelity amplification enzyme is needed. Multiple displacement amplification (MDA) using ϕ 29 is an alternative method to PCR for the production of DNA in high enough amounts for sequencing. ϕ 29 MDA has the potential to decrease our reliance on primers and increase the lengths of DNA produced through amplification. This method has the advantage of being able to produce large lengths of DNA with lower errors than conventional PCR. ϕ 29 MDA has been applied to samples with very low starting DNA amounts from single cells (both prokaryotic and eukaryotic) and provided DNA in levels high enough to perform sequencing [24]. The high fidelity and 3'–5' proofreading activity reduces the amplification error rate to 1 in 10^6–10^7 bases compared to conventional Taq polymerase with a reported error rate of 1 in 90,000 [25]. A single binding reaction can incorporate over 70kb [26], meaning that this method is not limited by the length of the initial target like conventional PCR. It has been demonstrated that this method is suitable for use with single bacterial cells[24] and more recently malaria parasites directly from blood samples [27].Bacteraemia is a major global cause of morbidity and mortality, with a large range of aetiologies and is a particular problem in healthcare settings[28,29]. The current microbiological diagnostic process is culture based often involving specialised equipment, with time to positivity varying due to aetiology and pathogen load. Although recent advances in MALDI-TOF processing has allowed direct identification of bacteria from positive blood cultures [30,31,32], antibiotic sensitivities take a further 18 hours. The direct application of whole genome sequencing to blood samples would allow rapid pathogen diagnosis, along with simultaneous pathogen typing and genotypic resistance prediction. Furthermore, by applying unbiased pathogen detection method the diagnostic bias would be lowered. In order to allow to allow direct from sample sequencing, a two stage host cells removal was performed firstly red blood cells were removed using HetaSep®, which has previously been used to isolate nucleated cells

in the blood, particularly granulocytes but has not previously been applied to aid in pathogen isolation. Secondly a selective white blood cell lysis was undertaken using saponin[33] in order to release intracellular pathogens and aid in host nucleic acid removal.

Methods

Model process

Clinical isolates were collected from the Royal Free Hospital Hampstead, where they had been stored at -80°C after isolation from septic patients. Phenotypic data was produced using the BD phoenix and was retrieved from final hospital reports. Bacteraemia models were set up by adding an estimated ten bacterial cells, (*S. aureus* or *E. coli*) calculated using serial dilutions to 1ml horse blood. The workflow depicted in Figure 1was then applied, with samples being cultured at each stage to assess bacterial survival, sample processing was repeated three times. 200 µl HetaSep® was added and the sample vortexed and incubated at 37°C for 10 minutes. 550 µl supernatant was removed and 200 µl 5% saponin was added to a final 2% solution, and incubated at room temperature for 5 minutes. 700µl sterile water was added for a water shock and incubated at room temperature for 30 seconds before salt restoration with the addition of 21µl 5M NaCl. The sample was centrifuged at 4000xg for 5 minutes and the supernatant discarded before addition of 2 µl turbo DNase1 and 5 µl of 10x buffer (Ambion). The sample was vortexed and incubated at 37°C for 15 minutes EDTA was added to a final concentration of 15nM. The sample was then centrifuged for 5 minutes at 4000xg and the supernatant removed and discarded. The pellet was washed in decreasing volumes of PBS, initially 200µl then 100µl followed by 20 µl with each stage being centrifuged at 6000xg for three minutes.

Bacteria were suspended in a total of 4 µl sterile PBS. Extraction was performed using alkaline method; briefly cell suspensions were added to 200 mM potassium hydroxide (Qiagen) and 50mM dithiothreitol (Qiagen) and incubated at 65°C for 10 minutes. The reaction was then neutralised using neutralisation buffer (Qiagen). The sample was then briefly vortexed and placed on ice.

Amplification was performed using ϕ29 MDA (Repli-g Single Cell Kit Qiagen). A master mix was prepared on ice in a total volume of 40 µl, with 29 µl reaction buffer, containing endonuclease resistant hexamer primers and 2 µl (40 U) of ϕ29polymerase (Qiagen, REPLI-g Single Cell Kit). The extracted DNA was then added to the master mix and the sample was then incubated at 30^0C for 2 hours, and the reaction stopped by heating to 65^0C for 3 minutes.

Sequencing

DNA was quantified using Qubit BR kit and 3 µg of ϕ29 MDA DNA was de-branched using S1 nuclease in a 90 µl reaction as follows, 3 µl 10x buffer, 3 µl 0.5M NaCl, 10 µl S1 nuclease (1U/ µl) with water to make the volume to 90µl. The digestion reaction was left at room temperature for 30 minutes and the enzyme deactivated by incubating at 70^oC with 6 µl 0.5M EDTA. The DNA was then fragmented by nebulisation at 30psi for 180 seconds. The DNA was sequenced using the 454 Junior using the manufacturer's recommended methods. A negative library was produced by using sterile PBS in the ϕ29 MDA, this was then sequenced using the same method.

Data analysis

Data analysis was performed in three stages, firstly host and contaminating reads were removed before abundance trimming, secondly reads were classified using lowest common ancestor (LCA) analysis and finally classified reads were assembled and analysed. Initially reads were mapped using Newbler standard parameters against host and a local contamination library consisting of sequenced negative controls and reads identified as contamination in previous runs. Unmapped reads were written into a new fastq file using a custom python script (supplementary material) and taken forward for further analysis. The remaining reads were abundance trimmed using Khmer[34] using two passes to a maximum depth of 50, to remove over represented reads produce by the ϕ29 MDA method. Following this the reads were error trimmed using Prinseq[35], with ends being trimmed to a Q20 cut off. Blastn was then used to assign the reads against the nr database. Once completed lowest common ancestor (LCA) analysis was performed using MEGAN[36], and reads associated with the bacterial species of interest were extracted to a new fastq file. These reads were then *de novo* assembled using standard parameters in SPAdes[37] assembly outputs were assessed using QUAST (Quality Assessment

Tool for Genome Assemblies)[35]. Reference assemblies were performed using the closest reference sequence identified using the LCA analysis. Antibiotic prediction was performed using Mykrobe[11] for *S. aureus* and ResFinder[38] for *E. Coli* using the *de novo* assembled reads.

Results

Sample processing

When the full blood processing method was applied to *E. coli* and *S. aureus*, good survival rates was found for both bacteria through-out all stages with the final survival rate being 100%. Details of survivial at each stage of processing can be found in Table 1

Sequencing results S. aureus

After processing and sequencing the horse blood spiked with *S. aureus* the number of reads passing basic filter was 128500. Once known contaminants were removed and error and abundacnce trimming complete 124,145 reads remained. Using Blastn and MEGAN, 62.1% reads were identified as *S. aureus*. 6.76% of the reads were identified as the genus Equus, and 1.72% were identified as *Parascaris equorum*. When examining the *S. aureus* reads closer 4254 reads were identified to the subspecies level, (*Staphylococcus aureus subsp. aureus* HO 5096 0412), this subspecies has a complete genome available (GenBank GCA_000284535.1) and was used as a reference for reference mapping. Reference mapping produced 451 contigs and covered 92% of the reference genome. When the reads identified as *Staphylococcus* using the LCA analysis were extracted and the reads *de novo* assembled 1212 contigs were produced with an N50 of 3882. When this was compared to the reference sequence 83% of the genome was covered with 10 misassemblies. Mykrobe analysis using the *de novo* assembly gave genotypic result for 12 antibiotics. When comparing the genotypic and phenotypic (BD Phoenix™) results matched in 11 of the 12 antibiotics. The results for ciprofloxacin were inconclusive in genotypic tests, but resistant by phenotypic methods.

Sequencing results E. coli

Post sequencing 173597 reads passed the initial filter, once contaminants were removed and the reads error and abundance trimmed 170243 reads remained. Overall 73% of all reads remaining after

the analysis pipeline were identified as *E. coli.* 31959 (18%) reads had no identity. Unlike *S. aureus* it was not possible to type the *E. coli*, as 150 reads were assigned to O7:K1 and 211 reads were assigned to JJ1886. Reads which were identified as *Enterobacteriaceae, Escherichia* and *E. coli* by BLAST and LCA analysis were extracted from the fastq produced after pipeline completion. The genome of JJ1886 was available from the Integrated Microbial Genomes database (ID 2558309052), and the chromosomal sequence was used as the reference against which the extracted reads were assembled. After reference mapping assembly 93.5% of the genome was covered in 548 contigs. When the reads were *de novo* assembled, 89% of the same reference was covered in 1334 contigs. The *de novo* assembly was used to identify several resistance markers using ResFinder. The resistance markers included dfrA, conferring trimethoprim resistance, gyrA conferring resistance to fluoroquinolones. mdtK an efflux pump conferring resistance to norfloxacin and blaCMY-2 which is an AmpC, conferring resistance to beta-lactams including cephalosporins. Additionally, eight drug efflux systems were identified. These results were concordant to the phenotypic data.

Discussion

Direct diagnostics using WGS from clinical samples would, in many ways, provide the ideal diagnostic method. By providing all the information from whole genome data with the speed of direct sample testing. However, there are various obstacles, including low pathogen numbers, high host background and difficulty in interpreting genotypic data. Here, a model was prepared to demonstrate the potential for direct from sample sequencing from whole blood.

Fresh whole horse blood was used to model bacteraemia, as it was readily available. The process was developed to remove RBCs early in the process, as they represent the largest proportion of the cellular makeup of blood (up to 96%[39]), debulking the sample, and preventing release of oxidative agents. Selective lysis on WBCs allowed the release of any intracellular pathogens and exposed host nucleic acid to nucleases. *S. aureus* and *E. coli* were chosen due to different cell wall types and cell morphology; both demonstrated good survival during the developed sample processing.

The majority of sequencing reads from spiked horse blood were associated with the spiked bacteria (62.1% S. *aureus* and 73% *E. coli*), considering the horse genome is over 500 times larger than the *E. coli* genome this shows that the vast majority of the host material was removed. Overall there was good concordance of phenotypic and genotypic results showing the potential for rapid genotypic prediction of antibiotic resistance from 10 bacterial cells in 1 ml of host blood. The inconclusive ciprofloxacin results demonstrate the need for improved understanding of the mechanisms of resistance. Ciprofloxacin resistance is harder to predict as it is chromosomal mutation rather than gene acquisition. Three mechanisms of fluoroquinolone resistance have been proposed in *S. aureus,* Topoisomerase IV gene mutations, DNA gyrase gene mutations and an active efflux pump (NorA)[40]. The complexity of predicting ciprofloxacin resistance suggests that the database may be lacking in its ability to predict ciprofloxacin resistance, and so this is the most likely cause of the inconclusive result for ciprofloxacin resistance. Additionally the creators of Mykrobe (Bradley et al[11]) found a false negativity rate of 4.6% for ciprofloxacin resistance.

In addition to identifying the isolates resistance to beta-lactams the database was able to identify the blaZ gene and MecA gene. Genotypic testing will never entirely replace phenotypic susceptibility testing, due to its inability to identify novel resistance determinants and the comprehensive nature of phenotypic testing. However, in this scenario, of invasive sepsis, the gain in speed provided by not having to culture the organism to determine susceptibility could be life-saving.

Multiple antibiotic resistance factors were identified in the *E. coli*, which gave good concordance with the phenotypic data. Using the genotypic data, it was possible to rapidly identify the beta-lactamase present as BlaCMY-2. The rapid identification of the specific resistance genes in bacteria could help identify outbreaks by providing more information than a simple antibiogram. Additionally, it could help monitor novel resistance genes, or genes that are increasing in incidence. Large amounts of horizontal genome transfer amongst Gram negative bacteria has the potential to cause outbreaks of resistance bacteria through genes or plasmids[41], which would be more complex to track. Rapid identification of genes causing the resistance in isolates could help inform epidemiological and outbreak studies which could involve several species of bacteria. The output for the resistance

196 prediction was complex to interpret with several genes identified which didn't always have specific
197 drug resistance associated. This highlights a down side of generic databases, as they are often time-
198 consuming to interpret. A study of NGS sequencing data from bacteraemia isolates of *E. coli* have
199 shown resistance prediction specificity of 97%[2], if this was coupled with direct from sample
200 sequencing genotypic prediction could inform treatment more rapidly than phenotypic test.

201 Limitations of this model include differences in horse and human blood, and difference in sample
202 between spiked bacteria and bacteria causing a true infection. Additionally, factors in the blood such
203 as immune and inflammatory reactions could alter the efficiency of the preparation method. However,
204 this model shows proof of principal that direct unbiased sequencing directly from blood is possible
205 and could be used to diagnose and inform treatment of bacterial bloodstream infections.

206 The method presented here is easily adapted to allow application to other sequencing platforms.
207 Overall the method presented allows sufficient DNA for whole genome sequencing of pathogens in
208 blood to be produced within a single day.

209

210

211

Tables and Figures

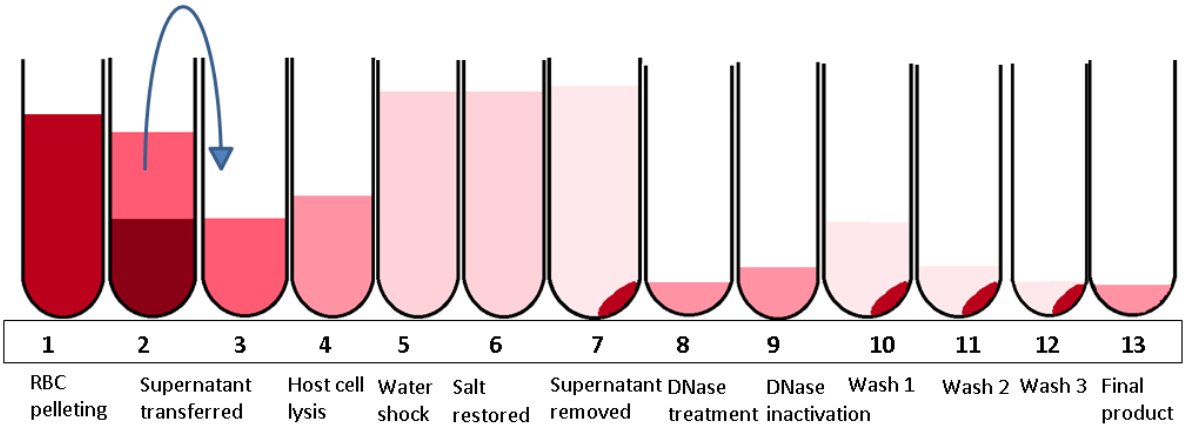

Figure 1 Illustration of the process for bacterial isolation from whole blood using HetaSep and selective lysis with Saponin. Numbers refer to sampling points where bacterial recovery was investigated.

Stage		*S. aureus*	*E. coli*
1	Spiked cells	10	11
2	RBC separation-bottom	0	0
3	RBC separation-top	10	9
4	Addition of Saponin	12	9
5	Water shock	12	12
6	Salt restoration	10	10
7	Cell pelleting-top Cell pelleting-bottom	0 9	0 9
8	DNAse treatment	11	12
9	EDTA and PBS	10	12
10	Wash 1-top Wash 1-bottom	0 9	0 11
11	Wash 2-top Wash 2–bottom	0 10	0 10
12	Wash 3-top Wash 3–bottom	0 8	0 9
13	Final pellet	11	9

Table 1 the average number of CFU isolated at each processing stage for bacterial isolation from whole blood using HetaSep and selective lysis with Saponin after workflow improvements. Numbers refer to sampling points where bacterial recovery was investigated as illustrated in **Figure 1**

222

	Phenotypic-(MIC)	Genotypic	Comment 223
	E. coli		
Ampicillin	R (>8)	R	blaCMY-2 224
Ceftriaxone	R (>4)	R	blaCMY-2
Ciprofloxacin	R (>1)	R	gyrA 225
Colistin	S (<=0.5)	S	226
Gentamicin	S (<=0.25)	S	
Levofloxacin	R (>2)	R	gyrA 227
Meropenem	S (<=0.25)	S	
Nalidixic acid	(>16)	R	ermB
Nitrofurantoin	S (<=16)	S	
Trimethoprim	R (>4)	R	dfrA
	S. aureus		
Ciprofloxacin	R (>2)	Inconclusive	
Clindamycin	S (<=0.25)	S	
Erythromycin	S (<=0.25)	S	
Fusidic acid	S (<=1)	S	
Gentamicin	S (<1)	S	
Methicillin	R (oxacillin 2)	R	MecA
Mupirocin	S (<=1)	S	
Penicillin G	R (> 0.25)	R	blaZ
Rifampicin	S (<=0.25)	S	
Tetracycline	S (<=0.5)	S	
Trimethoprim	S (<=1)	S	
Vancomycin	S (1)	S	

Table 2 phenotypic and genotypic result for antibiograms of *E. coli* and *S. aureus*

228

229 References

230 1. Gordon, N. C. *et al.* Prediction of staphylococcus aureus antimicrobial resistance by whole-
231 genome sequencing. *J. Clin. Microbiol.* **52,** 1182–1191 (2014).

232 2. Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella
233 pneumoniae isolates using whole genomic sequence data. *J. Antimicrob. Chemother.* **68,**
234 2234–44 (2013).

235 3. Leopold, S. R., Goering, R. V., Witten, A., Harmsen, D. & Mellmann, A. Bacterial whole-
236 genome sequencing revisited: Portable, scalable, and standardized analysis for typing and
237 detection of virulence and antibiotic resistance genes. *J. Clin. Microbiol.* **52,** 2365–2370
238 (2014).

239 4. Ashton, P. M. *et al.* Insight into Shiga toxin genes encoded by Escherichia coli O157 from
240 whole genome sequencing. *PeerJ* **3,** e739 (2015).

241 5. Kennedy, A. D. *et al.* Epidemic community-associated methicillin-resistant Staphylococcus
242 aureus: recent clonal expansion and diversification. *Proc. Natl. Acad. Sci.* **105,** 1327–1332
243 (2008).

244 6. Harris, S. R. *et al.* Evolution of MRSA During Hospital Transmission and Intercontinental
245 Spread. *Science* **327,** 469–474 (2010).

246 7. Eyre, D. W. *et al.* A pilot study of rapid benchtop sequencing of Staphylococcus aureus and
247 Clostridium difficile for outbreak detection and surveillance. *BMJ Open* **2,** e001124–e001124
248 (2012).

249 8. Reuter, S. *et al.* A pilot study of rapid whole-genome sequencing for the investigation of a
250 Legionella outbreak. *BMJ Open* **3,** e002175 (2013).

9. Snitkin, E. S. *et al.* Tracking a Hospital Outbreak of Carbapenem-Resistant Klebsiella pneumoniae with Whole-Genome Sequencing. *Sci. Transl. Med.* **4,** 148ra116-148ra116 (2012).

10. Köser, C. U. *et al.* Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N.Engl.J.Med.* **366,** 2267–2275 (2012).

11. Bradley, P. *et al.* Rapid antibiotic resistance predictions from genome sequence data for S. aureus and M. tuberculosis. *bioRxiv* **6,** 18564 (2015).

12. Pankhurst, L. J. *et al.* Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. *Lancet Respir. Med.* **4,** 49–58 (2016).

13. Mason, O. U. *et al.* Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *ISME J.* **8,** 1464–75 (2014).

14. Vieira-silva, S. & Rocha, E. P. C. The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet.* **6,** (2010).

15. Sharon, I. *et al.* Comparative metagenomics of microbial traits within oceanic viral communities. *ISME J.* **5,** 1178–1190 (2011).

16. Gilbert, J. a. & Dupont, C. L. Microbial Metagenomics: Beyond the Genome. *Ann. Rev. Mar. Sci.* **3,** 347–371 (2011).

17. Erickson, A. R. *et al.* Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. *PLoS One* **7,** (2012).

18. Lamendella, R., Domingo, J. W. S., Ghosh, S., Martinson, J. & Oerther, D. B. Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiol.* **11,** 103 (2011).

19. Walker, A. W., Duncan, S. H., Louis, P. & Flint, H. J. Phylogeny, culturing, and

metagenomics of the human gut microbiota. *Trends Microbiol.* **22,** 267–274 (2014).

20. Kristensen, D. M., Mushegian, A. R., Dolja, V. V. & Koonin, E. V. New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* **18,** 11–19 (2010).

21. Mokili, J. L., Rohwer, F. & Dutilh, B. E. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* **2,** 63–77 (2012).

22. Do, L. A. H. *et al.* Respiratory Syncytial Virus and Other Viral Infections among Children under Two Years Old in Southern Vietnam 2009-2010: Clinical Characteristics and Disease Severity. *PLoS One* **11,** e0160606 (2016).

23. Dacheux, L. *et al.* A preliminary study of viral metagenomics of french bat species in contact with humans: Identification of new mammalian viruses. *PLoS One* **9,** (2014).

24. Lasken, R. S. Single-cell genomic sequencing using Multiple Displacement Amplification. *Curr. Opin. Microbiol.* **10,** 510–516 (2007).

25. Hutchison, C. a, Smith, H. O., Pfannkoch, C. & Venter, J. C. Cell-free cloning using phi29 DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 17332–17336 (2005).

26. Blanco, L. *et al.* Highly Efficient DNA Synthesis by the Phage 429 DNA Polymerase. *J. Biol. Chem.* **264,** 8935–8940 (1989).

27. Cowell, A. N. *et al.* Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of Plasmodium vivax from Unprocessed Clinical Samples. *MBio* **8,** e02257-16 (2017).

28. Pien, B. C. *et al.* The Clinical and Prognostic Importance of Positive Blood Cultures in Adults. *Am. J. Med.* **123,** 819–828 (2010).

29. Prasad, N., Murdoch, D. R., Reyburn, H. & Crump, J. A. Etiology of Severe Febrile Illness in Low- and Middle-Income Countries : A Systematic Review. *PLoS One* 1–25 (2015).

297 doi:10.1371/journal.pone.0127962

298 30. Schubert, S. *et al.* Novel, improved sample preparation for rapid, direct identification from
299 positive blood cultures using matrix-assisted laser desorption/ionization time-of-flight
300 (MALDI-TOF) mass spectrometry. *J. Mol. Diagn.* **13,** 701–6 (2011).

301 31. Prod'Hom, G., Bizzini, A., Durussel, C., Bille, J. & Greub, G. Matrix-assisted laser desorption
302 ionization-time of flight mass spectrometry for direct bacterial identification From positive
303 blood culture pellets. *J. Clin. Microbiol.* **48,** 1481–1483 (2010).

304 32. Stevenson, L. G., Drake, S. K. & Murray, P. R. Rapid identification of bacteria in positive
305 blood culture broths by matrix-assisted laser desorption ionization-time of flight mass
306 spectrometry. *J. Clin. Microbiol.* **48,** 444–447 (2010).

307 33. Zelenin, S. *et al.* Microfluidic-based isolation of bacteria from whole blood for sepsis
308 diagnostics. *Biotechnol. Lett.* **37,** 825–830 (2015).

309 34. Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B. & Brom, T. H. A Reference-Free
310 Algorithm for Computational Normalization of Shotgun Sequencing Data. *arXiv* **1203.4802,**
311 1–18 (2012).

312 35. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets.
313 *Bioinformatics* **27,** 863–864 (2011).

314 36. Huson, D., Auch, A., Qi, J. & Schuster, S. MEGAN analysis of metagenome data. *Gennome*
315 *Res.* **17,** 377–386 (2007).

316 37. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to
317 Single-Cell Sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).

318 38. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob.*
319 *Chemother.* **67,** 2640–2644 (2012).

320 39. Alberts B, Johnson A, L. J. *Molecular Biology of the Cell. 4th edition.* (2002).

321 40. Tanaka, M., Wang, T., Onodera, Y., Uchida, Y. & Sato, K. Mechanism of quinolone resistance
322 in Staphylococcus aureus. *J. Infect. Chemother.* **6,** 131–139 (2000).

323 41. Tofteland, S., Naseer, U., Lislevand, J. H., Sundsfjord, A. & Samuelsen, ??rjan. A Long-Term
324 Low-Frequency Hospital Outbreak of KPC-Producing Klebsiella pneumoniae Involving
325 Intergenus Plasmid Diffusion and a Persisting Environmental Reservoir. *PLoS One* **8,** 1–8
326 (2013).

327 42. Wilson, J. *et al.* Trends among pathogens reported as causing bacteraemia in England, 2004-
328 2008. *Clin. Microbiol. Infect.* **17,** 451–458 (2011).

329

330