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] Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, United Kingdom

* 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

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

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

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

70 **Methods**

71 ***Model process***

72 Clinical isolates were collected from the Royal Free Hospital Hampstead, where they had been
73 stored at -80°C after isolation from septic patients. Phenotypic data was produced using the BD
74 phoenix and was retrieved from final hospital reports. Bacteraemia models were set up by adding an
75 estimated ten bacterial cells, (*S. aureus* or *E. coli*) calculated using serial dilutions to 1ml horse blood.
76 The workflow depicted in Figure 1was then applied, with samples being cultured at each stage to
77 assess bacterial survival, sample processing was repeated three times. 200 µl HetaSep® was added
78 and the sample vortexed and incubated at 37°C for 10 minutes. 550 µl supernatant was removed and
79 200 µl 5% saponin was added to a final 2% solution, and incubated at room temperature for 5
80 minutes. 700µl sterile water was added for a water shock and incubated at room temperature for 30
81 seconds before salt restoration with the addition of 21µl 5M NaCl. The sample was centrifuged at
82 4000xg for 5 minutes and the supernatant discarded before addition of 2 µl turbo DNase1 and 5 µl of
83 10x buffer (Ambion). The sample was vortexed and incubated at 37°C for 15 minutes EDTA was
84 added to a final concentration of 15nM. The sample was then centrifuged for 5 minutes at 4000xg and
85 the supernatant removed and discarded. The pellet was washed in decreasing volumes of PBS,
86 initially 200µl then 100µl followed by 20 µl with each stage being centrifuged at 6000xg for three
87 minutes.

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

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

97 ***Sequencing***

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

105 ***Data analysis***

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

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

122 **Results**

123 ***Sample processing***

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

127 ***Sequencing results S. aureus***

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

142 ***Sequencing results E. coli***

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

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

158 **Discussion**

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

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

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

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

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

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

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

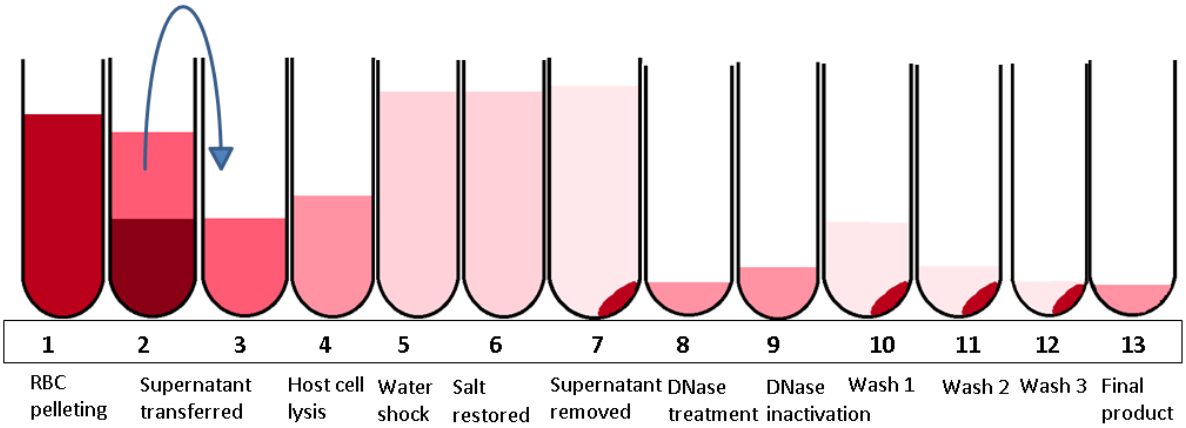
207 The method presented here is easily adapted to allow application to other sequencing platforms.
208 Overall the method presented allows sufficient DNA for whole genome sequencing of pathogens in
209 blood to be produced within a single day.

210

211

212

213 ***Tables and Figures***



214

215 **Figure 1** Illustration of the process for bacterial isolation from whole blood using HetaSep and

216 selective lysis with Saponin. Numbers refer to sampling points where bacterial recovery was

217 investigated.

218

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

219 **Table 1** the average number of CFU isolated at each processing stage for bacterial isolation from

220 whole blood using HetaSep and selective lysis with Saponin after workflow improvements. Numbers

221 refer to sampling points where bacterial recovery was investigated as illustrated in **Figure 1**

222

223

	Phenotypic-(MIC)	Genotypic	Comment 224
	E. coli		
Ampicillin	R (>8)	R	blaCMY-2 225
Ceftriaxone	R (>4)	R	blaCMY-2
Ciprofloxacin	R (>1)	R	gyrA 226
Colistin	S (<=0.5)	S	227
Gentamicin	S (<=0.25)	S	
Levofloxacin	R (>2)	R	gyrA 228
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*

229

230 References

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

233 2. Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella
234 pneumoniae isolates using whole genomic sequence data. *J. Antimicrob. Chemother.* **68,**
235 2234–44 (2013).

236 3. Leopold, S. R., Goering, R. V., Witten, A., Harmsen, D. & Mellmann, A. Bacterial whole-
237 genome sequencing revisited: Portable, scalable, and standardized analysis for typing and
238 detection of virulence and antibiotic resistance genes. *J. Clin. Microbiol.* **52,** 2365–2370
239 (2014).

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

242 5. Kennedy, A. D. *et al.* Epidemic community-associated methicillin-resistant Staphylococcus
243 aureus: recent clonal expansion and diversification. *Proc. Natl. Acad. Sci.* **105,** 1327–1332
244 (2008).

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

247 7. Eyre, D. W. *et al.* A pilot study of rapid benchtop sequencing of Staphylococcus aureus and
248 Clostridium difficile for outbreak detection and surveillance. *BMJ Open* **2,** e001124–e001124
249 (2012).

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

252 9. Snitkin, E. S. *et al.* Tracking a Hospital Outbreak of Carbapenem-Resistant Klebsiella
253 pneumoniae with Whole-Genome Sequencing. *Sci. Transl. Med.* **4,** 148ra116-148ra116
254 (2012).

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

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

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

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

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

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

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

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

271 18. Lamendella, R., Domingo, J. W. S., Ghosh, S., Martinson, J. & Oerther, D. B. Comparative
272 fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiol.* **11,**
273 103 (2011).

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

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

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

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

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

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

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

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

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

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

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

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

298 doi:10.1371/journal.pone.0127962

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

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

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

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

310 34. Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B. & Brom, T. H. A Reference-Free
311 Algorithm for Computational Normalization of Shotgun Sequencing Data. *arXiv* **1203.4802,**
312 1–18 (2012).

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

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

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

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

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

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

324 41. Tofteland, S., Naseer, U., Lislevand, J. H., Sundsfjord, A. & Samuelsen, ??rjan. A Long-Term
325 Low-Frequency Hospital Outbreak of KPC-Producing Klebsiella pneumoniae Involving
326 Intergenus Plasmid Diffusion and a Persisting Environmental Reservoir. *PLoS One* **8,** 1–8
327 (2013).

328 42. Wilson, J. *et al.* Trends among pathogens reported as causing bacteraemia in England, 2004-
329 2008. *Clin. Microbiol. Infect.* **17,** 451–458 (2011).

330

331