

1 **Genomic surveillance of multidrug-resistant organisms**
2 **based on long-read sequencing**

3
4

5 **Fabian Landman**[#], Casper Jamin[#], Angela de Haan, Sandra Witteveen, Jeroen Bos, Han G.J. van der
6 Heide, Leo M. Schouls and Antoni P.A. Hendrickx*, Dutch CPE and MRSA surveillance study groups.

7 [#]These authors contributed equally.

8

9 Centre for Infectious Disease Control, National Institute for Public Health and the Environment
10 (RIVM), Bilthoven, the Netherlands.

11

12 ***Corresponding author:** Antoni P.A. Hendrickx, antoni.hendrickx@rivm.nl, +31 (0)30 274 3196.
13 ORCID 0000-0003-4431-9323

14

15 Keywords: genomic surveillance, CPE, CPPA, CRAB, MRSA, Long-read sequencing, Nanopore, MLST,
16 wgMLST, MLVA.

17

18

19 **Background** Multidrug-resistant organisms (MDRO) pose a significant threat to public-health world-
20 wide. The ability to identify antimicrobial resistance determinants, to assess changes in molecular
21 types, and to detect transmission are essential for effective surveillance and infection prevention of
22 MDRO. Molecular characterization based on long-read sequencing has emerged as a promising
23 alternative to short-read sequencing. The aim of this study was to rapidly and accurately
24 characterize MDRO for surveillance and transmission studies based on long-read sequencing only.

25 **Methods** Genomic DNA of 356 MDRO was automatically extracted using the Maxwell-RSC48. The
26 MDRO included 106 *Klebsiella pneumoniae* isolates, 85 *Escherichia coli*, 15 *Enterobacter cloacae*
27 complex, 10 *Citrobacter freundii*, 34 *Pseudomonas aeruginosa*, 16 *Acinetobacter baumannii*, and 69
28 methicillin-resistant *Staphylococcus aureus* (MRSA), of which 24 were from an outbreak. MDRO were
29 sequenced using both short-read (Illumina NextSeq 550) and long-read (Nanopore Rapid Barcoding
30 Kit-24-V14, R10.4.1) whole-genome sequencing (WGS). Basecalling was performed for two distinct
31 models using Dorado-0.3.2 duplex mode. Long-read data was assembled using Flye, Canu, Miniasm,
32 Unicycler, Necat, Raven and Redbean assemblers. Long-read WGS data with >40x coverage was used
33 for multi-locus sequence typing (MLST), whole-genome MLST (wgMLST), *in silico* multiple locus
34 variable-number of tandem repeat analysis (iMLVA) for MRSA, and identification of resistance genes
35 (Abricate).

36 **Results** Comparison of wgMLST profiles based on long-read and short-read WGS data revealed >95%
37 of wgMLST profiles within the species-specific cluster cut-off, except for *P. aeruginosa*. The wgMLST
38 profiles obtained by long-read and short-read WGS differed only one to nine wgMLST alleles for *K.*
39 *pneumoniae*, *E. coli*, *E. cloacae* complex, *C. freundii*, *A. baumannii* complex and MRSA. For *P.*
40 *aeruginosa* differences were up to 27 wgMLST alleles between long-read and short-read wgMLST.
41 MLST sequence types and *in silico* MLVA types were concordant between long-read and short-read
42 WGS data and conventional MLVA typing. Antimicrobial resistance genes were detected in long-read
43 sequencing data with high sensitivity/specificity (92-100%/99-100%). Long-read sequencing enabled
44 analysis of an MRSA outbreak.

45 **Conclusions** We demonstrate that molecular characterization of automatically extracted DNA
46 followed by long-read sequencing is as accurate and more cost-effective compared to short-read
47 sequencing. Long-read sequencing is suitable for molecular typing and outbreak analysis as part of
48 genomic surveillance of MDRO. However, the analysis of *P. aeruginosa* requires further
49 improvement which may be obtained by other basecalling algorithms. The low implementation
50 costs, low price per isolate, and rapid library preparation for long-read sequencing of MDRO extends
51 its applicability to resource-constrained settings and low-income countries world-wide.

52 **Introduction.** Multidrug-resistant organisms (MDRO) such as methicillin-resistant *Staphylococcus*
53 *aureus* (MRSA), carbapenemase-producing *Enterobacterales* (CPE) and *Pseudomonas aeruginosa*
54 (CPPA) and carbapenem-resistant *Acinetobacter baumannii* (CRAB) pose a growing public health
55 concern worldwide. The ability to type MDRO accurately and rapidly and to identify resistance
56 determinants is essential for effective surveillance and infection control. National reference
57 laboratories and medical microbiology laboratories historically typed MDRO using various classical
58 typing methods such as phage typing [1], pulsed-field gel electrophoresis (PFGE) [2,3], amplified
59 fragment length polymorphism (AFLP) [4], Staphylococcal protein A (Spa)-typing [5,6], multiple locus
60 variable-number of tandem repeat analysis (MLVA), but also DNA sequencing-based typing methods
61 such as multi-locus sequence typing (MLST) [7–10]. These methods are being phased out by whole-
62 genome sequencing (WGS)-based methods such as core genome (cg) or whole genome (wg) MLST,
63 as these methods significantly increased typing resolution. In addition, WGS data enable
64 identification of resistance and virulence genes [11]. Current WGS-based methods typically use
65 short-read sequencing technologies with read lengths of 150 bases, such as Illumina next-generation
66 sequencing, and have been widely used for public health genomic surveillance, molecular typing,
67 and outbreak detection of MDRO [12,13]. Short-read WGS costs are high and force concessions for
68 national and local surveillance to sequence only a limited selection of MDRO [14]. Additionally,
69 short-read sequencing technologies have limitations in detecting structural variations, such as large
70 insertions, deletions, inversions, repetitive sequence elements and antimicrobial resistance (AMR)
71 plasmids, all common in bacterial genomes [13]. Reconstruction of complete bacterial genomes by
72 *de novo* assembly is unattainable through short-read sequencing as these structural variations and
73 genomic elements are larger than individual reads. Thus, chromosomes and AMR encoding plasmids
74 fail to assemble into complete assemblies. Hybrid assemblies of short-read and long-read WGS data
75 enabled reconstruction of MDRO genomes but is laborious and more expensive than short-read
76 sequencing alone. Long-read sequencing technologies, such as Nanopore sequencing, can overcome
77 these problems and have recently emerged as a promising alternative to short-read sequencing or
78 hybrid assemblies [15–18]. Furthermore, the accuracy of Nanopore long-read sequencing
79 significantly improved over time (<https://nanoporetech.com/platform/accuracy>). The major
80 objective of this study was to determine whether long-read sequencing can replace short-read
81 sequencing to enable classical and new typing methods such as MLVA, MLST and wgMLST in the
82 genomic surveillance and outbreak analysis of MDRO.

83

84 **Methods**

85 **Multidrug-resistant organisms (MDRO) from the Dutch national surveillance 2023**

86 Medical microbiology laboratories (MML) in the Netherlands are requested to submit
87 *Enterobacterales*, *P. aeruginosa*, *A. baumannii* suspected for carbapenemase production or
88 carbapenem resistance, and MRSA isolates cultured for patient care (from symptomatic infections or
89 asymptomatic carriage), to the National Institute for Public Health and the Environment (RIVM)
90 [14,19]. Between 1 May and 30 September 2023, 3,138 suspected MDRO were received by the RIVM
91 of which the majority 2,235 were MRSA isolates. Illumina short-read sequencing and Nanopore long-
92 read sequencing were performed on 356 MDRO in this study (*vide infra*). The MDRO included 106
93 *Klebsiella pneumoniae* isolates, 85 *Escherichia coli*, 15 *Enterobacter cloacae* complex, 10 *Citrobacter*
94 *freundii*, 34 *Pseudomonas aeruginosa*, 16 *Acinetobacter baumannii*, 69 methicillin-resistant
95 *Staphylococcus aureus* (MRSA) and other species (Table 1, Supplementary Table 1).

96

97 **Automated Oxford Nanopore Technologies long-read sequencing**

98 The genomic DNA of MDRO was isolated using The Maxwell® (Promega), an automated nucleic acid
99 extraction and purification platform. The Maxwell RSC Cultured Cells DNA kit (AS1620) kit was used
100 to isolate up to 48 bacterial isolates per run. Manufacturer's instructions were followed, except
101 using nuclease-free water instead of TE buffer to create the cell suspension and without RNase
102 treatment. The Oxford Nanopore protocol for rapid sequencing DNA V14 – barcoding SQK-
103 RBK114.24 was used (Oxford Nanopore Technologies [ONT], <https://nanoporetech.com/>). In brief,
104 barcoded transposome complexes were used to tagment the DNA and simultaneous attachment of a
105 pair of barcodes. Routine isolations had shown DNA concentrations within 50-300 ng/μl, thus, for
106 maintaining high throughput capacity, DNA concentrations were not measured prior to library
107 preparation. All samples were pooled and after clean-up the sequencing adapters (RAP reagent)
108 were added. Sequencing buffer and library beads were added, and the final library was loaded onto
109 a MinION flow cell (FLO-MIN114, R10.4.1)
110 ([https://community.nanoporetech.com/docs/prepare/library_prep_protocols/rapid-sequencing-](https://community.nanoporetech.com/docs/prepare/library_prep_protocols/rapid-sequencing-gdna-barcoding-sqk-rbk114/v/rbk_9176_v114_revm_27nov2022)
111 [gdna-barcoding-sqk-rbk114/v/rbk_9176_v114_revm_27nov2022](https://community.nanoporetech.com/docs/prepare/library_prep_protocols/rapid-sequencing-gdna-barcoding-sqk-rbk114/v/rbk_9176_v114_revm_27nov2022)). The 48-h sequence run was
112 started with live basecalling and demultiplexing enabled using the MinKNOW software through a
113 GridION device with 5 kHz data acquisition enabled for all samples. However, the actual basecalling
114 for analysis in this study was performed on the resulting barcoded pod5 directories using Dorado
115 0.3.2 duplex mode with the optional flags --guard-gpus, -b 448 and writing data to Fastq format. We

116 assessed the performance of two basecalling models, dna_r10.4.1_e8.2_400bps_sup@v4.2.0 and
117 res_dna_r10.4.1_e8.2_400bps_sup@2023-09-22_bacterial-methylation. These models are referred
118 to as Dorado (<https://github.com/nanoporetech/dorado>) and Rerio
119 (<https://github.com/nanoporetech/erio>), respectively. Data was assembled using an in-house
120 Snakemake workflow. All bioinformatic tools were used with default parameters unless specified
121 otherwise. First, Chopper v0.6.0 is used to extract all Q12 reads >1000-bp [20]. Additionally, 80-bp is
122 cropped from both sides to remove possible adapters. Subsequently, FiltLong 0.2.1 was used to keep
123 the 90% best scoring reads based on length and identity (<https://github.com/rrwick/Filtlong>). To
124 assess the performance of different *de novo* assemblers long-read data was assembled using Canu
125 v2.2 [21], Flye v2.9.2 [22], Minimap v2.2.26 followed by Miniasm v0.3 [23], Minipolish v0.1.3, Necat
126 v0.0.1 [24], Raven v1.8.1 [25], Redbean wtdbg2.5 [26], and Unicycler v0.5.0 [27] long-read option.
127 Furthermore, genome assembly polishing with Medaka v1.8.0
128 (<https://github.com/nanoporetech/medaka>) was used to test if this improved *de novo* assemblies,
129 compared to unpolished assemblies. Medaka polishing was only done for Dorado basecalled reads,
130 as there was no Medaka polishing model for Rerio basecalled data available during this study.

131

132 **Automated Illumina short-read sequencing**

133 For in-house Illumina short-read sequencing of MDRO, the same DNA isolation method as described
134 above was followed. DNA libraries were prepared using Nextera DNA Flex Library Prep kit (Illumina,
135 San Diego, USA), followed by paired-end sequencing (2 x 150-bases) on the Illumina NextSeq550
136 platform (Illumina, USA), according to the manufacturer's instructions. Read quality analysis and *de*
137 *nov*o assembly was performed with the Juno-assembly v2.0.2 pipeline ([https://github.com/RIVM-](https://github.com/RIVM-bioinformatics/juno-assembly)
138 [bioinformatics/juno-assembly](https://github.com/RIVM-bioinformatics/juno-assembly)). Briefly, read quality assessment and filtering were done using FastQC
139 and FastIP [28,29]. Genomes were assembled using SPAdes [30] and curated with QUAST [31],
140 CheckM [32] and Bbtools (<https://sourceforge.net/projects/bbmap/>).

141

142 **Molecular typing from long-read and short-read sequencing data and data analyses**

143 Sixty-nine MRSA isolates, including 24 outbreak isolates were characterized by multiple-locus
144 variable number of tandem repeat analysis (MLVA) *in vitro* at BaseClear (Leiden, the Netherlands)
145 and simultaneous PCR detection of the *mecA/mecC*, and *lukF-PV/lukS-PV* genes, the latter indicative
146 of PVL production, as previously described [33,34]. In addition, *in silico* (i)MLVA on MRSA isolates
147 was done in the RIVM on long-read WGS data only. The identification of resistance genes and

148 replicons of MDRO was performed using ResFinder (version 4.1.11) and PlasmidFinder (version 2.1)
149 databases from the Center for Genomic Epidemiology, using ABRicate (v1.0.1,
150 <https://github.com/tseemann/abricate>). The resulting FASTA files were subjected to their
151 corresponding species-specific multi-locus sequence typing (MLST) and whole-genome MLST
152 (wgMLST) schemes using Ridom software in SeqSphere v9.0.1 if available [35]. The following in-
153 house wgMLST schemes were used for analyses: the *A. baumannii* wgMLST scheme was based on
154 3,473 genes (2,390 core-genome genes and 1,083 accessory-genome genes, cut-off 15 wgMLST
155 alleles), using *A. baumannii* strain ACICU (Genbank accession number NC_010611.1, September
156 2021) as a reference genome. The *K. pneumoniae* wgMLST scheme comprised 4,978 genes (3,471
157 core-genome and 1,507 accessory-genome targets, cut-off 20 wgMLST alleles) using *K. pneumoniae*
158 MGH 78,578 (NC_009648.1) as a reference genome [19]. The *E. coli* wgMLST scheme comprised
159 4,503 genes (3,199 core-genome and 1,304 accessory-genome targets, cut-off 25 wgMLST alleles)
160 using *E. coli* 536 (CP000247.1) as a reference genome [19]. The *P. stuartii* scheme comprised 3,744
161 genes (3,079 core-genome and 665 accessory-genome targets, cut-off 15 wgMLST alleles) with *P.*
162 *stuartii* CP014024.2 as reference genome (Witteveen *et al.*, 2024 submitted). The *P. aeruginosa*
163 wgMLST scheme included 6,442 genes (6,117 core-genome and 325 accessory-genome targets, cut-
164 off 15 wgMLST alleles) using *P. aeruginosa* PAO1 (NC_002516.2) as a reference genome. The *P.*
165 *mirabilis* wgMLST scheme included 3,517 genes (2,675 core-genome and 842 accessory-genome
166 targets, cut-off 15 wgMLST alleles) using *P. mirabilis* HI4320 (NZ_CP042907.1) as a reference
167 genome. The *C. freundii* wgMLST scheme included 4,495 genes (2964 core-genome and 1531 842
168 accessory-genome targets, cut-off 20 wgMLST alleles) using *C. freundii* strain HM38 (CP024672.1) as
169 a reference. For MRSA, the COL-based wgMLST scheme comprised 2,567 genes (1,861 core-genome
170 and 706 accessory-genome targets, cut-off 15 wgMLST alleles) was used [14,36]. For *E. cloacae*
171 complex a pgMLST scheme comprised of 9,829 genes from references CP001918, CP017186 and
172 CP017184 was used, cut-off 20 wgMLST alleles [37]. Both MLST and wgMLST profiles were imported
173 into BioNumerics version 8.1.1 (Applied Maths, Sint-Martens-Latem, Belgium) and used in cluster
174 analyses. Missing data were ignored in the analyses. The Illumina short-read sequencing results were
175 considered golden standard method for MLST sequence type determination, wgMLST allele calling,
176 AMR gene and replicon identification, and long-read assembly methods throughout this study. All
177 further data analyses and visualization was done using Python (v3.11.5,
178 <https://www.python.org/downloads/release/python-3115/>), Pandas (v2.0.3,
179 <https://pandas.pydata.org/docs/index.html>) and Plotly (v5.18.0,
180 <https://github.com/plotly/plotly.py>).

181

182 **Data availability**

183 Raw short-read and long-read sequencing data of 356 Dutch CPE/CPPA/CRAB/MRSA surveillance
184 isolates have been deposited in the SRA database under BioProject numbers PRJNA1076692,
185 PRJNA1076808, and PRJNA903550 (Supplementary Table 1). Relevant code was made available
186 through <https://github.com/RIVM-bioinformatics/in-silico-mlva>. The authors confirm that all
187 supporting data, code, protocols and accession numbers have been provided within the article and
188 through supplementary data files.

189 Results

190 Assessing sequencing coverage and basecalling model for wgMLST on long-read sequenced MDRO

191 From May to September 2023, the National Institute for Public Health and the Environment (RIVM)
192 received 3,138 MDRO. In this study, 356 genetically highly diverse MDRO were included; 269 CPE, of
193 which the majority were *K. pneumoniae* ($n=106$) and *E. coli* ($n=85$) and nine other species ($n=78$),
194 CPPA ($n=34$), CRAB ($n=17$) and MRSA ($n=69$) (Table 1, Supplementary Table 1). We also included 24
195 PVL-negative MRSA isolates belonging to a recurring fusidic acid-resistant impetigo-associated MRSA
196 outbreak in the Netherlands [38]. We isolated genomic DNA using the Maxwell and sequenced 356
197 MDRO using both Nanopore and Illumina WGS platforms. Sequencing coverage for long-read
198 sequencing was assessed from 356 MDRO to determine a suitable cut-off for sequencing depth. The
199 median relative number of wgMLST alleles (Nanopore versus Illumina assemblies) were 0.45, 0.93,
200 0.99 for coverages up to 0-10x, 10-20x and 20-30x, respectively (Figure 1A). The number of wgMLST
201 alleles identified remained stable at 1.00 with higher sequencing depth, indicating that the same
202 number of alleles were identified in the Nanopore assemblies compared to the Illumina assemblies.
203 This trend was seen across all species tested in this study (data not shown). The median number of
204 different wgMLST alleles compared to the Illumina assemblies was 97, 45, 17, 10, 7, 4, 5 for
205 coverages up to 0-10x, 10-20x, 20-30x, 30-40x, 40-50x, 50-60x, 60-70x, respectively (Figure 1B). The
206 median distance with Illumina assemblies was relatively stable, between 4 to 7 wgMLST alleles, with
207 coverages from 40x and higher. At a coverage 100x several outliers are observed, with differences
208 up to 300 alleles. These outliers belonged all to one single *P. aeruginosa* isolate basecalled with
209 Dorado and assembled with the different *de novo* assemblers used in this study. The 196 long-read
210 sequenced MDRO with a coverage higher than 40x were used for further analyses. To determine
211 which long-read sequencing data basecalling model performed best and if Medaka polishing (for
212 Dorado only) improved assembly when comparing long-read assemblies to the Illumina short-read
213 sequencing golden standard, we determined the number of wgMLST alleles difference for 196
214 MDRO using these methods (Figure 2). Across all species, Rerio performed best and had the lowest
215 number of faulty wgMLST allele calls. The median number of wgMLST alleles difference for Dorado
216 duplex, Dorado duplex with Medaka, and Rerio duplex was 2, 2, 2 (*A. baumannii*), 5.5, 3, 1 (*C.*
217 *freundii*), 2, 1, 1 (*E. cloacae* complex), 6, 3, 3, (*E. coli*), 23, 19.5, 9 (*K. pneumoniae*), 120, 79, 26.5 (*P.*
218 *aeruginosa*), and 4, 2, 2 (*S. aureus*), respectively. Rerio was able to improve the long-read
219 sequencing data in such a way that the number of faulty wgMLST allele calls dropped from 87-148
220 (depending on the assembler) to merely 0-5 faulty alleles for one particular *C. freundii* isolate.
221 Therefore, Rerio was used for further analyses.

222

223 **Comparison of long-read assemblers for MLST and wgMLST on long-read sequenced MDRO**

224 To assess the best assembler for *de novo* assembly of long-read only data in the absence of Illumina
225 short-read sequencing data, Longcyclus (Unicyclus with long-reads only), Miniasm, Raven, Necat,
226 Canu, Flye and Redbean were compared based on the median (50 percentile) wgMLST allele
227 difference and the 95 percentile wgMLST alleles difference. With all species combined, the
228 difference was 1; 18 (Canu), 1; 21.55 (Flye), 2.5; 15.55 (Longcyclus), 2; 17 (Miniasm), 7; 35.55
229 (Necat), 3; 34.65 (Raven) and 6; 38.65 (Redbean) for the median and 95% percentile, respectively
230 (Figure 3, Table 2). Although the median was slightly lower for Miniasm compared to Longcyclus, the
231 95% performed better for Longcyclus, and was therefore used for subsequent analyses. For each
232 MDRO, the median and 95% percentile allele difference for Longcyclus was 1; 2.0 (*A. baumannii*), 1;
233 4 (*C. freundii*), 1; 3.5 (*E. cloacae*), 2; 17 (*E. coli*), 4; 12.6 (*K. pneumoniae*), 2; 6 (*S. aureus*), and 25.5; 43
234 (*P. aeruginosa*), respectively (Table 2). Long-read sequencing-derived MLST sequence types (ST)
235 were compared with short-read sequencing derived MLST STs among the MDRO tested. A few
236 differences were observed, for two *E. coli* isolates the Illumina MLST ST did not call the *adk* gene,
237 which was present in all other *E. coli* isolates and the long-read MLST STs all called the same locus.
238 All other differences resulted from the long-read derived MLST missing a call compared to the
239 Illumina MLST ST. For three out of four *A. baumannii* the *gdhB* gene was not called. Interruption by
240 IS elements in this element has been previously noted and could explain its absence in all of the
241 long-read MLST STs [39,40]. For Miniasm only one gene was not called in *K. pneumoniae* but
242 otherwise was in concordance with all Illumina MLST STs and thus performed best. Longcyclus also
243 missed one additional gene in an *E. coli* isolate, but all other MLST STs were identical to the Illumina
244 MLST STs (Supplementary Table 2).

245

246 **Detection of AMR genes and plasmid replicons in long-read sequencing data**

247 Next, we compared the detection of AMR genes and plasmid replicons using both long-read and
248 short-read sequencing platforms. The sensitivity and specificity of AMR gene identification was
249 assessed. The specificity of AMR gene calling for Redbean ranged from 68.2 to 100 (median 87.2)
250 and performed worst of all assemblers tested (Table 3A). Canu had the highest specificity in AMR
251 gene calling (99.7%), followed by Flye (99.6%) and Longcyclus (98.6%). Even though Canu seemed to
252 perform best, multiple copies of the same AMR gene were detected (Supplementary Figure 1),
253 suggesting *de novo* assembly artifacts. The median number of AMR genes for all species together
254 was 16 for Canu assemblies versus 11 to 13 for all other assemblers (Supplementary Figure 1). The
255 effect of this phenomenon was seen in the median genome size over the entire dataset (5.6-Mbp

256 Canu versus 5.2-Mbp for all other long read assemblers, Supplementary Figure 2). Furthermore, the
257 number of replicons was at least double for Canu assemblies versus any other assembler
258 (Supplementary Figure 3). The unweighted average sensitivity for AMR gene calling was 98.1%
259 independent of which *de novo* assembler was used and seemed to be a species-specific effect (Table
260 3). Only for Flye and Canu assemblers all carbapenemase genes were correctly detected within the
261 MDRO tested (Supplementary Table 3). The specificity of plasmid replicon detection was best for
262 Canu assembled long-read data (Table 3B), but as mentioned previously, was hampered by the
263 multi-copy assembly artifact issue (Supplementary Figure 3). Next to Canu, Flye performed excellent
264 with an unweighted average specificity of 98.5% over all species (Table 3B). Identifying the correct
265 plasmid replicon was excellent and did not seem to differ much among assemblers as the sensitivity
266 was between 98.9% (Canu) and 99.9% (Flye).

267

268 **Long-read sequencing enables analysis of an MRSA outbreak**

269 There was an impetigo-associated MRSA outbreak with MLVA-type MT4627 in the Mid-East of the
270 Netherlands in 2019 (Figure 4A + B) [38]. The number of MRSA MT4627 isolates in the South-West
271 region increased from 14 (in 2021) and 28 (in 2022) to 40 MRSA MT4627 isolates in 2023 and were
272 obtained from 40 persons (Figure 4A). Characterization of a subset of the 2023 MRSA outbreak
273 isolates using both long-read ($n=16$) and short-read ($n=16$) sequencing-based wgMLST analysis
274 revealed genetic clustering (≥ 2 isolates differing ≤ 15 wgMLST alleles) with 2019 outbreak and
275 isolates from 2020, 2021, 2022 in the minimum spanning tree (Figure 4B). Long-read sequenced
276 MRSA outbreak isolates were in close proximity of their short-read sequenced counterparts (Figure
277 4B). MLST, MLVA and *in silico* MLVA (iMLVA) analyses revealed that MLST sequence type and iMLVA
278 type could be retrieved from long-read sequencing data with $>40x$ coverage and were concordant
279 with fragment size-based *in vitro* MLVA types. The outbreak MRSA isolates were from MLST ST121,
280 clonal complex (CC) CC121, MLVA type MT4627 and iMLVA type MT4627. A subset of the MRSA
281 isolates from 2023 have diversified over time. Long-read sequencing also yielded identical AMR
282 genes *aac(6')-aph(2'')*, *mecA*, *fusC*, *blaZ*, and *dfgG*, plasmid replicons, and virulence genes, when
283 compared to short-read sequencing (Figure 4C, Supplementary Table 4). The plasmid-borne *ermC*
284 gene was lacking three long-read sequenced isolates. Long-read sequencing outperformed detection
285 of sortase B substrate genes encoding microbial surface components recognizing adhesive matrix
286 molecules (MSCRAMM) such as *clfB*, *sdrC*, *sdrD*, and *sdrE* (Figure 4C).

287

288 Discussion

289 We demonstrate that automated DNA extraction followed by Nanopore R10.4.1 long-read
290 sequencing is a reliable and cost-effective method for the molecular typing (iMLVA, MLST, wgMLST),
291 AMR gene, plasmid replicon, virulence gene identification, and outbreak detection of MDRO and can
292 therefore be applied for genomic surveillance. A sufficient sequencing depth is required, and in this
293 study a minimum of 40x sequencing depth was used. Previous studies using Illumina short-read
294 sequencing used 30x coverage as a minimum [41,42]. Older Nanopore flowcell generations had
295 higher error rates of raw Nanopore reads compared to Illumina raw reads and may explain that still
296 a higher coverage is required for R10.4.1 with current models. Long-read based wgMLST seemed to
297 be more efficient for species with a small chromosome e.g., *S. aureus* than for those with larger
298 chromosomes like *P. aeruginosa*. Unique signatures in *P. aeruginosa* data and or a lack of
299 representation in the training set for the models is one explanation for poor performance. Another
300 possibility is that even higher sequencing depth is required for bacteria with large chromosomes to
301 obtain sufficient good quality reads. To overcome this problem, one could reload the DNA library
302 prep during a long-read sequencing run. Alternatively, we recently performed a Promethion run with
303 96 MDRO which boosted coverage from 52x to 669x, with median coverage of 220x (data not
304 shown). Two different basecalling models were tested using Dorado duplex mode with and without
305 Medaka polishing and Rerio, this model has an increased performance for highly methylated DNA
306 motifs. Overall, Rerio performed best for wgMLST allele calling using in-house wgMLST schemes.
307 Research models such as Rerio are continuously updated and if proven successful, implemented into
308 the normal workflow as has been the case for this research model where the same training set was
309 used and now included in dna_r10.4.1_e8.2_400bps_sup@v4.3.0. Although these updates provide a
310 constant improvement of the ability to perform molecular characterization of bacterial isolates, it
311 imposes a problem for laboratories using Nanopore long-read sequencing under quality assurance
312 systems. The implementation of guidelines such as ISO15189 for national reference and medical
313 microbiology laboratories, or ISO23418 for food-borne pathogen sequencing requires the validation
314 of every new basecalling model before implementation and is slowing down the introduction of
315 improved methods.

316 For *de novo* assembly, Canu was unsuitable for AMR gene detection as it resulted in the erroneous
317 detection of multiple copies of the same gene. This is likely due to the inability of this algorithm to
318 properly overlap contig ends. Although Tricycler [43] is among the best long-read only *de novo*
319 assembler in the community, this tool was not used as it requires multiple manual curation steps in
320 each assembly, thereby hampering the ability for automated high-throughput bacterial assemblies.
321 The use of Miniasm or Longcycler (Unicycler supplied with long-reads only) resulted in the lowest

322 number of different wgMLST alleles compared to their Illumina short-read sequenced counterparts.
323 Besides *Pseudomonas*, all other species tested had only a few discrepant wgMLST allele calls (0 to 12
324 alleles). This was on the same level of variation in a recent multi-center study where
325 Enterobacterales, Enterococci and Staphylococci isolates were sent to participating laboratories for
326 short-read sequencing and subsequent molecular typing analysis [41,42]. For identification of AMR
327 genes and plasmid replicons Flye performed better than Longcyclus and Miniasm, even though these
328 two assemblers were best for wgMLST allele calling. Overall, genotyping was excellent and on par
329 with other studies investigating the inter-laboratory reproducibility of Illumina short-read
330 sequencing-based genotyping, where they found >99% performance [41,44]. Notably, long-read
331 sequencing-based *de novo* assembly methods can better discriminate multi-copy AMR genes in
332 mobile genetic elements such as plasmids, as these regions are impossible to assemble with short-
333 read sequencing. Therefore, the evaluation of the specificity for AMR genes and replicon
334 identification may not be the best metric to evaluate the performance of long-read sequencing
335 methods when using Illumina short-read sequencing data as golden standard. Additionally, short-
336 read assembly is unable to truly identify multi-copy AMR genes. However, no other reference is
337 available. Finally, it should be noted, genotype to phenotype is still difficult to infer, and large
338 discrepancies have been observed among methods used in a multi-centre study investigating this
339 challenge [45]. Furthermore, Nanopore long-read sequencing was superior over Illumina short-read
340 sequencing of genes encoding MSCRAMMs. MRSCRAMMs are known to harbor multiple repetitive
341 domains and are implicated in binding to collagen, fibrinogen, cytokeratin components of the
342 extracellular matrix [46].

343 In conclusion, for laboratories wanting to implement Nanopore long-read sequencing of MDRO we
344 therefore recommend using a minimum 40x coverage, Rerio basecalling and Miniasm or Longcyclus
345 as *de novo* assembler for molecular typing and outbreak detection for genomic surveillance. For the
346 best performing AMR gene and plasmid replicon detection we recommend using Flye instead of
347 Canu, Miniasm or Longcyclus, as Canu generated assembly artifacts and Miniasm and Longcyclus did
348 not perform as good as Flye on sensitivity and specificity. Future studies are needed to optimize the
349 performance of Nanopore long-read sequencing for *P. aeruginosa*. The use of long-read sequencing
350 can provide additional valuable insights into virulence determinants, resistance plasmids and
351 resistance gene copy number of MDRO. This may help to inform and guide effective control
352 measures for MDRO which were previously not possible using short-read sequencing. Importantly,
353 the relatively low purchase and implementation costs of long-read sequencing, coupled with the low
354 price per isolate and rapid library preparation not only enables genomic surveillance and outbreak

355 analysis but extends its applicability to resource-constrained settings and low-income countries
356 world-wide.

357

358 **Competing interests**

359 The authors do not have any financial or non-financial competing interests that may undermine the
360 objectivity, integrity and value of this study. Illumina and Oxford Nanopore Technologies were not
361 involved in the design, execution and analyses, nor the interpretation of data or conclusions from
362 this study.

363

364 **Financial support**

365 This research was funded by the Dutch Ministry of Health, Welfare and Sport (V/150302/22/BR).

366

367 **Ethical statement**

368 Ethical approval was not required for the present study, since it is based on genomic surveillance
369 data only. Samples from which the isolates were cultured, were all collected as part of routine health
370 care.

371

372 **Acknowledgements**

373 We thank all the members of the Dutch CPE and MRSA surveillance study Groups and the Dutch
374 medical microbiology laboratories for submitting MDRO isolates to the RIVM for the national
375 CPE/MRSA surveillance program. We thank Dr. Romy D. Zwittink, Dr. Rob Mariman and Dr. Daan W.
376 Notermans for critical reading of the manuscript. Members of the Dutch CPE and MRSA Surveillance
377 Study Groups:

- 378 • A.L.E. van Arkel, ADRZ medisch centrum, Department of Medical Microbiology, Goes
- 379 • M.A. Leversteijn-van Hall, Alrijne Hospital, Department of Medical Microbiology, Leiden
- 380 • W. van den Bijllaardt, Amphia Hospital, Microvida Laboratory for Microbiology, Breda
- 381 • R. van Mansfeld, Amsterdam UMC - location AMC, Department of Medical Microbiology and
382 Infection Prevention, Amsterdam
- 383 • K. van Dijk, Amsterdam UMC - location Vumc, Department of Medical Microbiology and
384 Infection Control, Amsterdam
- 385 • B. Zwart, Atalmedial, Department of Medical Microbiology, Amsterdam
- 386 • B.M.W. Diederens, Bravis Hospital/ZorgSaam Hospital Zeeuws-Vlaanderen, Department of
387 Medical Microbiology, Roosendaal/Terneuzen

- 388 • H. Berkhout, Canisius Wilhelmina Hospital, Department of Medical Microbiology and
389 Infectious Diseases, Nijmegen
- 390 • D.W. Notermans, Centre for Infectious Disease Control, National Institute for Public Health
391 and the Environment, Bilthoven
- 392 • A. Ott, Certe, Department of Medical Microbiology Groningen & Drenthe, Groningen
- 393 • K. Waar, Certe, Department of Medical Microbiology Friesland & Noordoostpolder,
394 Leeuwarden
- 395 • W. Ang, Comicro, Department of Medical Microbiology, Hoorn
- 396 • J. da Silva, Deventer Hospital, Department of Medical Microbiology, Deventer
- 397 • A.L.M. Vlek, Diaconessenhuis Utrecht, Department of Medical Microbiology and
398 Immunology, Utrecht
- 399 • A.G.M. Buiting, Elisabeth-TweeSteden (ETZ) Hospital, Department of Medical Microbiology
400 and Immunology, Tilburg
- 401 • L.G.M. Bode, Erasmus University Medical Center, Department of Medical Microbiology and
402 Infectious Diseases, Rotterdam
- 403 • Jansz, Eurofins PAMM, Department of Medical Microbiology, Veldhoven
- 404 • S. Paltansing, Franciscus Gasthuis & Vlietland, Department of Medical Microbiology and
405 Infection Control, Rotterdam
- 406 • A.J. van Griethuysen, Gelderse Vallei Hospital, Department of Medical Microbiology, Ede
- 407 • J.R. Lo Ten Foe, Gelre Hospital, Department of Medical Microbiology and Infection Control,
408 Apeldoorn
- 409 • M.J.C.A. van Trijp, Groene Hart Ziekenhuis, Department of Medical Microbiology and
410 Infection Prevention, Gouda
- 411 • M. Wong, Haga Hospital, Department of Medical Microbiology, 's-Gravenhage
- 412 • A.E. Muller, HMC Westeinde Hospital, Department of Medical Microbiology, 's-Gravenhage
- 413 • M.P.M. van der Linden, IJsselland hospital, Department of Medical Microbiology, Capelle a/d
414 IJssel
- 415 • M. van Rijn, Ikazia Hospital, Department of Medical Microbiology, Rotterdam
- 416 • S.B. Debast, Isala Hospital, Laboratory of Medical Microbiology and Infectious Diseases,
417 Zwolle
- 418 • E. Kolwijck, Jeroen Bosch Hospital, Department of Medical Microbiology and Infection
419 Control, 's-Hertogenbosch
- 420 • N. Al Naiemi, LabMicTA, Regional Laboratory of Microbiology Twente Achterhoek, Hengelo
- 421 • T. Schulin, Laurentius Hospital, Department of Medical Microbiology, Roermond
- 422 • S. Dinant, Maasstad Hospital, Department of Medical Microbiology, Rotterdam
- 423 • S.P. van Mens, Maastricht University Medical Centre, Department of Medical Microbiology,
424 Infectious Diseases & Infection Prevention, Maastricht
- 425 • D.C. Melles, Meander Medical Center, Department of Medical Microbiology, Amersfoort
- 426 • J.W.T. Cohen Stuart, Noordwest Ziekenhuisgroep, Department of Medical Microbiology,
427 Alkmaar
- 428 • P. Gruteke, OLVG Lab BV, Department of Medical Microbiology, Amsterdam
- 429 • A. van Dam, Public Health Service, Public Health Laboratory, Amsterdam
- 430 • Maat, Radboud University Medical Center, Department of Medical Microbiology, Nijmegen
- 431 • Maraha, Regional Laboratory for Microbiology, Department of Medical Microbiology,
432 Dordrecht
- 433 • J.C. Sinnige, Regional Laboratory of Public Health, Department of Medical Microbiology,
434 Haarlem
- 435 • E. van der Vorm, Reinier de Graaf Groep, Department of Medical Microbiology, Delft
- 436 • M.P.A. van Meer, Rijnstate Hospital, Laboratory for Medical Microbiology and Immunology,
437 Velp

- 438 • M. de Graaf , Saltro Diagnostic Centre, Department of Medical Microbiology, Utrecht
- 439 • E. de Jong, Slingeland Hospital, Department of Medical Microbiology, Doetinchem
- 440 • S.J. Vainio , St Antonius Hospital, Department of Medical Microbiology and Immunology,
- 441 Nieuwegein
- 442 • E. Heikens, St Jansdal Hospital, Department of Medical Microbiology, Harderwijk
- 443 • M. den Reijer, Star-shl diagnostic centre, Department of Medical Microbiology , Rotterdam
- 444 • J.W. Dorigo-Zetsma, TergooiMC, Central Bacteriology and Serology Laboratory, Hilversum
- 445 • A. Troelstra, University Medical Center Utrecht, Department of Medical Microbiology,
- 446 Utrecht
- 447 • E. Bathoorn, University of Groningen, Department of Medical Microbiology, Groningen
- 448 • J. de Vries, VieCuri Medical Center, Department of Medical Microbiology, Venlo
- 449 • D.W. van Dam, Zuyderland Medical Centre, Department of Medical Microbiology and
- 450 Infection Control, Sittard-Geleen
- 451 • E.I.G.B. de Brauwier, Zuyderland Medical Centre, Department of Medical Microbiology and
- 452 Infection Control, Heerlen
- 453 • R. Steingrover, St. Maarten Laboratory Services, Department of Medical Microbiology, Cay
- 454 Hill (St. Maarten)
- 455 • Analytical Diagnostic Center N.V. Curaçao, Department of Medical Microbiology, Willemstad
- 456 (Curaçao)

457 **Figure legends**

458 **Table 1.** Selected MDRO from the Dutch national surveillance May-September 2023.

459 **Table 2.** Median allele and 95 percentile wgMLST allele difference for each combination of species
460 and *de novo* assembler.

461 **Table 3.** A) Percentage sensitivity (lower panel) and specificity (upper panel) of AMR gene
462 identification for each species and assembler tested in this study. B) Percentage sensitivity (lower
463 panel) and specificity (upper panel) of plasmid replicon detection for each species and assembler
464 were tested in this study. Only Rerio as basecalling and long-read sequenced isolates with a coverage
465 higher than 40x coverage were used.

466 **Figure 1.** A) Nanopore long-read sequencing coverage versus relative number of wgMLST alleles
467 identified in each assembly versus the Illumina genome assembly. B) Nanopore sequencing coverage
468 versus wgMLST allele distance to Illumina assemblies.

469 **Figure 2.** Nanopore long-read sequencing wgMLST alleles difference compared to Illumina
470 assemblies, for Dorado super accurate, Rerio, and Medaka polishing method.

471 **Figure 3.** Nanopore long-read sequencing wgMLST alleles difference relative to Illumina short-read
472 sequences per bacterial species and long-read sequence data assembler.

473 **Figure 4. Long-read sequencing-based analyses of an impetigo-associated MRSA outbreak in the**
474 **Netherlands 2023.** A) Geographic localization of persons with an MT4627 MRSA in the Netherlands.
475 The initial outbreak in 2019 has been described previously [38]. After this outbreak, this type mainly
476 occurred in the province of Zuid-Holland, The Netherlands, towards the end of 2023. B) Minimum
477 spanning tree of MT4627 MRSA analyzed by wgMLST. In total 16 isolates from 2023 were long-read
478 sequenced to 40x coverage and included in the analysis (green, Table 1). Of these 16 isolates, the
479 short-read counterpart was included in the figure (red). A wgMLST cluster cut-off of 15 was used.
480 Halo's indicate isolates varying ≤ 15 wgMLST alleles. C) Comparison of short-read and long-read
481 sequenced resistomes, plasmid replicons, and genes encoding microbial surface components
482 recognizing adhesive matrix molecules (MSCRAMM) of the outbreak-associated cluster isolates.

483 **Supplementary Figure 1** Number of AMR genes called per assembler and species

484 **Supplementary Figure 2** Genome size for each species and assembler.

485 **Supplementary Figure 3.** Number of plasmid replicons called for each assembler and species.

486

487 References

- 488 1. Williams RE, Rippon JE, Dowsett LM. Bacteriophage typing of strains of *Staphylococcus aureus*
489 from various sources. *Lancet*. 1953;1:510–4.
- 490 2. Tenover FC, Arbeit RD, Goering RV. How to select and interpret molecular strain typing methods
491 for epidemiological studies of bacterial infections: a review for healthcare epidemiologists.
492 Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. *Infect*
493 *Control Hosp Epidemiol*. 1997;18:426–39.
- 494 3. Ichiyama S, Ohta M, Shimokata K, Kato N, Takeuchi J. Genomic DNA fingerprinting by pulsed-field
495 gel electrophoresis as an epidemiological marker for study of nosocomial infections caused by
496 methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol*. 1991;29:2690–5.
- 497 4. Savelkoul PHM, Aarts HJM, de Haas J, Dijkshoorn L, Duim B, Otsen M, et al. Amplified-Fragment
498 Length Polymorphism Analysis: the State of an Art. *Journal of Clinical Microbiology*. 1999;37:3083–
499 91.
- 500 5. Frénay HM, Theelen JP, Schouls LM, Vandenbroucke-Grauls CM, Verhoef J, van Leeuwen WJ, et al.
501 Discrimination of epidemic and nonepidemic methicillin-resistant *Staphylococcus aureus* strains on
502 the basis of protein A gene polymorphism. *J Clin Microbiol*. 1994;32:846–7.
- 503 6. Harmsen D, Claus H, Witte W, Rothgänger J, Claus H, Turnwald D, et al. Typing of methicillin-
504 resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa*
505 repeat determination and database management. *J Clin Microbiol*. 2003;41:5442–8.
- 506 7. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. Multilocus sequence typing for
507 characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*.
508 *J Clin Microbiol*. 2000;38:1008–15.
- 509 8. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a
510 portable approach to the identification of clones within populations of pathogenic microorganisms.
511 *Proc Natl Acad Sci U S A*. 1998;95:3140–5.
- 512 9. Maiden MCJ. Multilocus sequence typing of bacteria. *Annu Rev Microbiol*. 2006;60:561–88.
- 513 10. Urwin R, Maiden MCJ. Multi-locus sequence typing: a tool for global epidemiology. *Trends*
514 *Microbiol*. 2003;11:479–87.
- 515 11. van Alphen LB, von Wintersdorff CJH, Savelkoul PHM. Epidemiological Typing Using WGS. In:
516 Moran-Gilad J, Yagel Y, editors. *Application and Integration of Omics-powered Diagnostics in Clinical*
517 *and Public Health Microbiology* [Internet]. Cham: Springer International Publishing; 2021 [cited 2024
518 Jan 31]. p. 69–87. Available from: https://doi.org/10.1007/978-3-030-62155-1_5
- 519 12. Aanensen DM, Carlos CC, Donado-Godoy P, Okeke IN, Ravikumar KL, NIHR Global Health
520 Research Unit on Genomic Surveillance of Antimicrobial Resistance. Implementing Whole-Genome
521 Sequencing for Ongoing Surveillance of Antimicrobial Resistance: Exemplifying Insights Into
522 *Klebsiella pneumoniae*. *Clin Infect Dis*. 2021;73:S255–7.
- 523 13. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol*.
524 2015;13:787–94.

- 525 14. Schouls LM, Witteveen S, van Santen-Verheuvél M, de Haan A, Landman F, van der Heide H, et al.
526 Molecular characterization of MRSA collected during national surveillance between 2008 and 2019
527 in the Netherlands. *Commun Med (Lond)*. 2023;3:123.
- 528 15. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. Single-nucleotide discrimination in
529 immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci U S A*.
530 2009;106:7702–7.
- 531 16. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly.
532 *Genomics, Proteomics & Bioinformatics*. 2016;14:265–79.
- 533 17. Foster-Nyarko E, Cottingham H, Wick RR, Judd LM, Lam MMC, Wyres KL, et al. Nanopore-only
534 assemblies for genomic surveillance of the global priority drug-resistant pathogen, *Klebsiella*
535 *pneumoniae*. *Microb Genom*. 2023;9:mgen000936.
- 536 18. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, et al. MinION nanopore
537 sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat*
538 *Biotechnol*. 2015;33:296–300.
- 539 19. Hendrickx APA, Landman F, de Haan A, Witteveen S, van Santen-Verheuvél MG, Schouls LM, et
540 al. *bla* OXA-48-like genome architecture among carbapenemase-producing *Escherichia coli* and
541 *Klebsiella pneumoniae* in the Netherlands. *Microb Genom*. 2021;7.
- 542 20. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing
543 data. *Bioinformatics*. 2023;39:btad311.
- 544 21. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate
545 long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27:722–
546 36.
- 547 22. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat
548 graphs. *Nat Biotechnol*. 2019;37:540–6.
- 549 23. Li H. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences.
550 *Bioinformatics*. 2016;32:2103–10.
- 551 24. Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, et al. Efficient assembly of nanopore reads via
552 highly accurate and intact error correction. *Nat Commun*. 2021;12:60.
- 553 25. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci*.
554 2021;1:332–6.
- 555 26. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods*. 2020;17:155–8.
- 556 27. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from
557 short and long sequencing reads. *PLoS Comput Biol*. 2017;13:e1005595.
- 558 28. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control.
559 *F1000Res*. 2018;7:1338.
- 560 29. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*.
561 2018;34:i884–90.

- 562 30. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new
563 genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.*
564 2012;19:455–77.
- 565 31. QUASt: quality assessment tool for genome assemblies - PubMed [Internet]. [cited 2024 Jan 31].
566 Available from: <https://pubmed.ncbi.nlm.nih.gov/23422339/>
- 567 32. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of
568 microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.*
569 2015;25:1043–55.
- 570 33. Schouls LM, Spalburg EC, van Luit M, Huijsdens XW, Pluister GN, van Santen-Verheueve MG, et al.
571 Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with
572 pulsed-field gel electrophoresis and spa-typing. *PLoS One.* 2009;4:e5082.
- 573 34. Bosch T, Pluister GN, van Luit M, Landman F, van Santen-Verheueve M, Schot C, et al. Multiple-
574 locus variable number tandem repeat analysis is superior to spa typing and sufficient to characterize
575 MRSA for surveillance purposes. *Future Microbiol.* 2015;10:1155–62.
- 576 35. Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, et al. Updating benchtop
577 sequencing performance comparison. *Nat Biotechnol.* 2013;31:294–6.
- 578 36. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. Bacterial whole-genome sequencing
579 revisited: portable, scalable, and standardized analysis for typing and detection of virulence and
580 antibiotic resistance genes. *J Clin Microbiol.* 2014;52:2365–70.
- 581 37. Hendrickx APA, Debast S, Pérez-Vázquez M, Schoffelen AF, Notermans DW, Landman F, et al. A
582 genetic cluster of MDR *Enterobacter cloacae* complex ST78 harbouring a plasmid containing *bla* VIM-
583 1 and *mcr-9* in the Netherlands. *JAC Antimicrob Resist.* 2021;3:dlab046.
- 584 38. Vendrik KEW, Kuijper EJ, Dimmendaal M, Silvis W, Denie-Verhaegh E, de Boer A, et al. An unusual
585 outbreak in the Netherlands: community-onset impetigo caused by a meticillin-resistant
586 *Staphylococcus aureus* with additional resistance to fusidic acid, June 2018 to January 2020. *Euro*
587 *Surveill.* 2022;27:2200245.
- 588 39. Repizo GD, Espariz M, Seravalle JL, Díaz Miloslavich JJ, Steimbrüch BA, Shuman HA, et al.
589 *Acinetobacter baumannii* NCIMB8209: a Rare Environmental Strain Displaying Extensive Insertion
590 Sequence-Mediated Genome Remodeling Resulting in the Loss of Exposed Cell Structures and
591 Defensive Mechanisms. *mSphere.* 2020;5:e00404-20.
- 592 40. Tomaschek F, Higgins PG, Stefanik D, Wisplinghoff H, Seifert H. Head-to-Head Comparison of Two
593 Multi-Locus Sequence Typing (MLST) Schemes for Characterization of *Acinetobacter baumannii*
594 Outbreak and Sporadic Isolates. *PLoS One.* 2016;11:e0153014.
- 595 41. Jamin C, De Koster S, van Koevinge S, De Coninck D, Mensaert K, De Bruyne K, et al.
596 Harmonization of whole-genome sequencing for outbreak surveillance of Enterobacteriaceae and
597 Enterococci. *Microb Genom.* 2021;7:000567.
- 598 42. Mellmann A, Andersen PS, Bletz S, Friedrich AW, Kohl TA, Lilje B, et al. High Interlaboratory
599 Reproducibility and Accuracy of Next-Generation-Sequencing-Based Bacterial Genotyping in a Ring
600 Trial. *J Clin Microbiol.* 2017;55:908–13.

601 43. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, et al. Trycycler: consensus long-read
602 assemblies for bacterial genomes. *Genome Biol.* 2021;22:266.

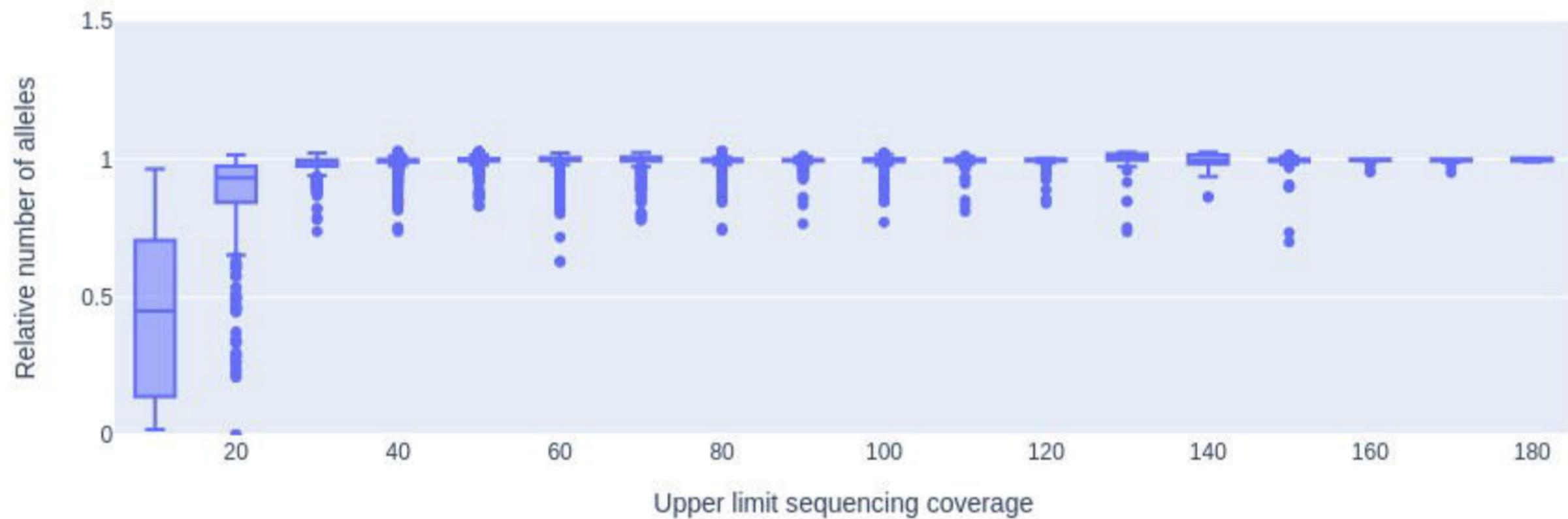
603 44. Coolen JPM, Jamin C, Savelkoul PHM, Rossen JWA, Wertheim HFL, Matamoros SP, et al. Centre-
604 specific bacterial pathogen typing affects infection-control decision making. *Microb Genom.*
605 2021;7:000612.

606 45. Doyle RM, O'Sullivan DM, Aller SD, Bruchmann S, Clark T, Coello Pelegrin A, et al. Discordant
607 bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of
608 bacterial isolates: an inter-laboratory study. *Microb Genom.* 2020;6:e000335.

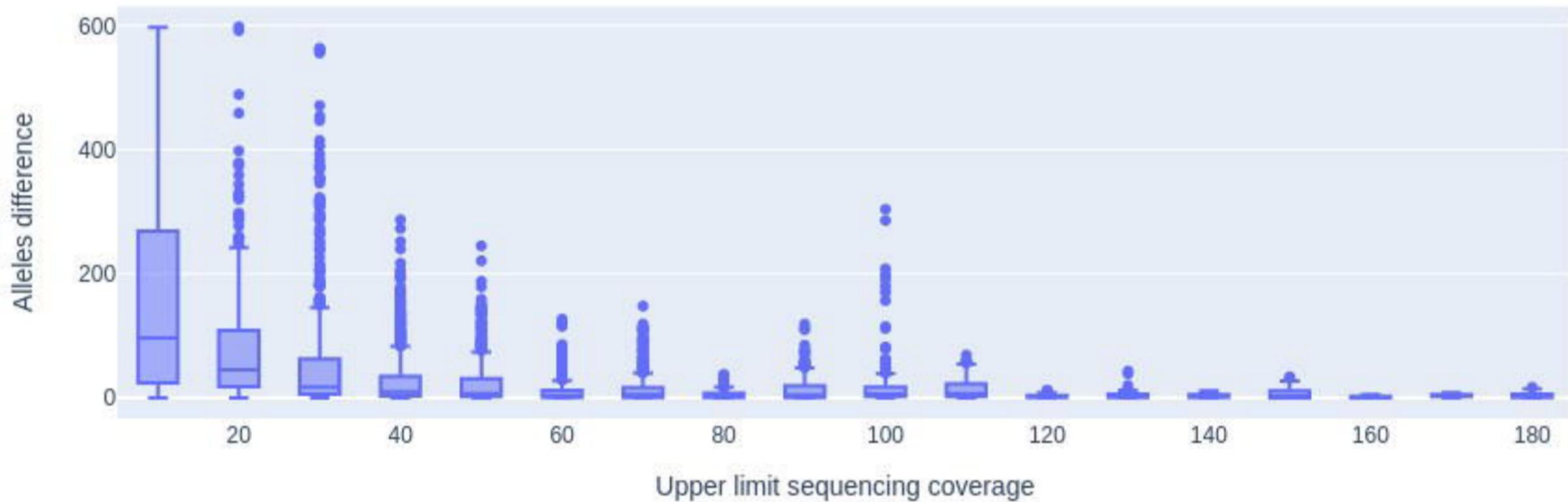
609 46. Foster TJ. The MSCRAMM Family of Cell-Wall-Anchored Surface Proteins of Gram-Positive Cocci.
610 *Trends Microbiol.* 2019;27:927–41.

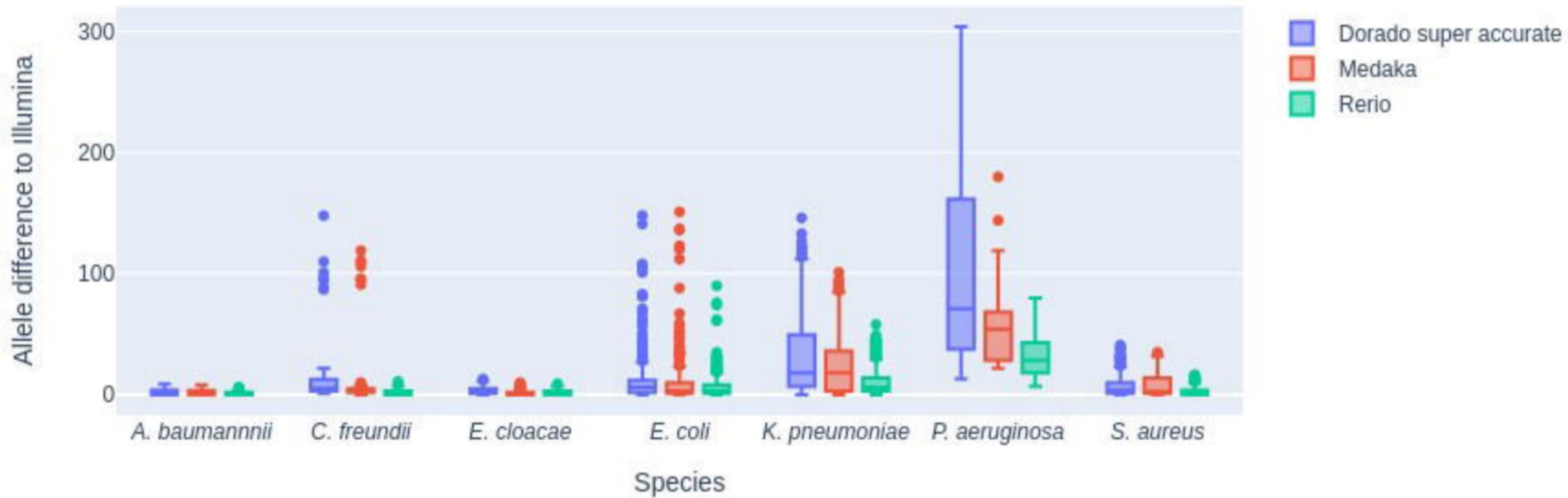
611

Sequencing coverage versus relative number of alleles found compared to Illumina assembly

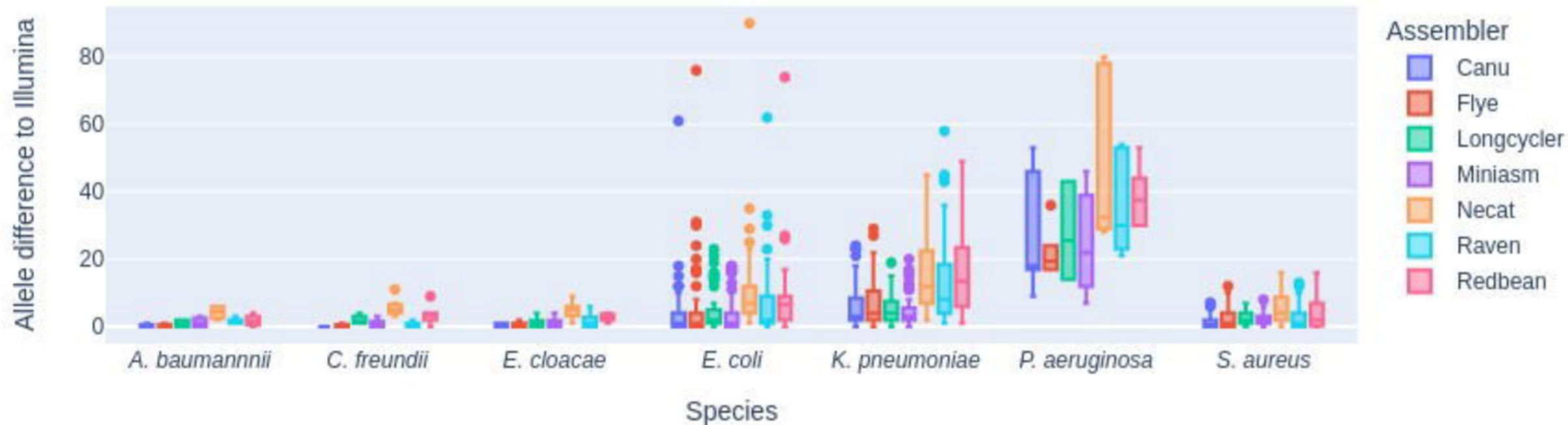


Sequencing coverage versus allele distance to Illumina assemblies

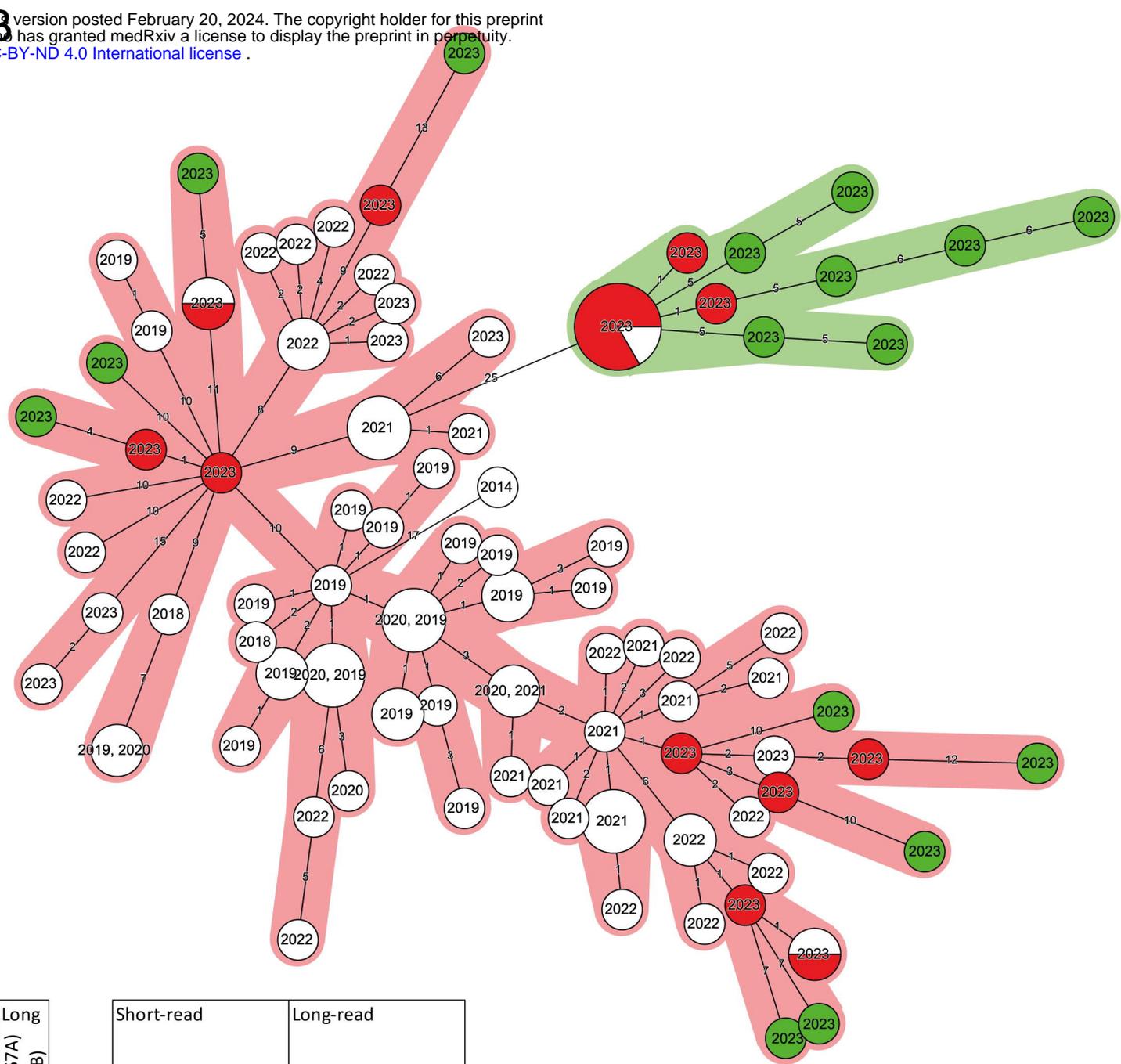
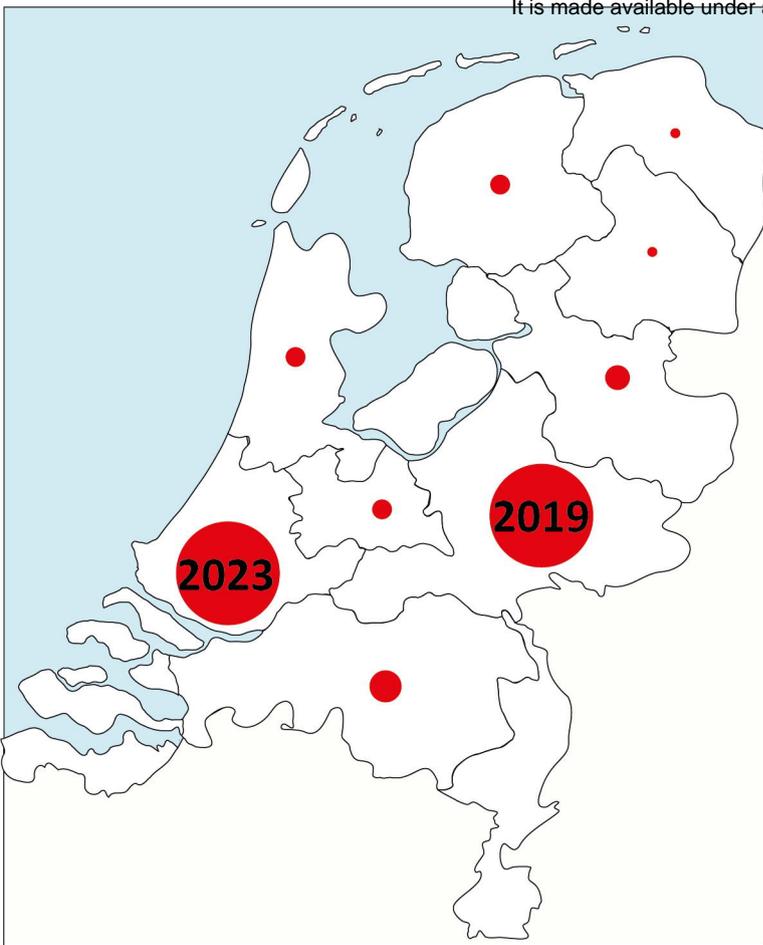




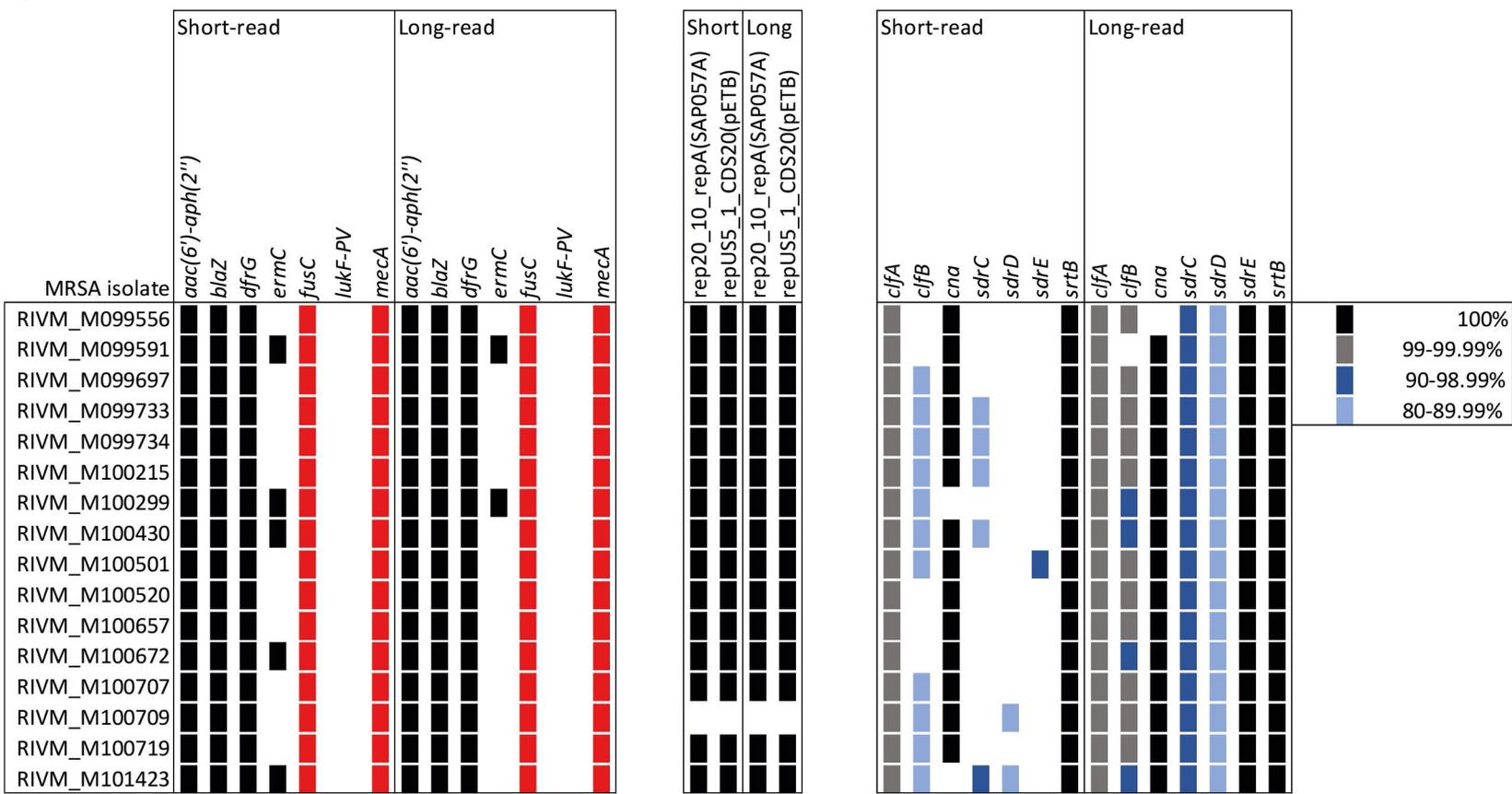
Allele differences for each species per assembler



A



C



MDRO	All coverages	>20x coverage	>30x coverage	>40x coverage
<i>Acinetobacter baumannii</i>	16	11	7	4
<i>Acinetobacter nosocomialis</i>	1	1	1	1
<i>Acinetobacter ursingii</i>	1			
<i>Citrobacter farmeri</i>	1	1	1	1
<i>Citrobacter freundii</i>	10	9	9	9
<i>Enterobacter cloacae complex</i>	15	14	11	11
<i>Escherichia coli</i>	85	77	69	54
<i>Klebsiella oxytoca</i>	6	5	3	2
<i>Klebsiella pneumoniae</i>	106	84	70	48
<i>Proteus mirabilis</i>	4	4	4	2
<i>Providencia stuartii</i>	3	2	2	2
<i>Pseudomonas aeruginosa</i>	34	23	12	7
<i>Pseudomonas nitroreducens</i>	1			
<i>Raoultella ornithinolytica</i>	4	2	1	
<i>Staphylococcus aureus</i>	45	45	44	39
<i>Staphylococcus aureus</i> (outbreak)	24	23	19	16
Total	356	301	253	196

	Canu	Flye	Longcycler	Miniasm	Necat	Raven	Redbean
<i>A. baumannii</i>	0.0 --- 0.8	0.0 --- 0.8	1.0 --- 2.0	1.0 --- 2.8	4.5 --- 6.0	1.0 --- 2.7	1.5 --- 3.7
<i>C. freundii</i>	0.0 --- 0.0	0.0 --- 1.0	1.0 --- 4.0	0.5 --- 2.6	5.0 --- 9.6	1.0 --- 1.6	4.0 --- 7.2
<i>E. cloacae</i>	0.0 --- 1.0	0.0 --- 1.5	1.0 --- 3.5	1.0 --- 3.5	5.0 --- 8.0	1.0 --- 5.0	3.0 --- 4.0
<i>E. coli</i>	1.0 --- 13.0	1.0 --- 26.1	2.0 --- 17.0	1.0 --- 17.0	7.0 --- 26.4	2.0 --- 25.4	6.5 --- 20.1
<i>K. pneumoniae</i>	3.0 --- 22.3	4.0 --- 21.6	4.0 --- 12.6	3.0 --- 15.6	12.0 --- 40.6	8.0 --- 43.6	13.5 --- 43.3
<i>P. aeruginosa</i>	18.0 --- 51.2	19.5 --- 33.0	25.5 --- 43.0	22.0 --- 44.2	32.5 --- 79.5	30.0 --- 53.8	37.5 --- 50.8
<i>S. aureus</i>	0.0 --- 6.1	1.0 --- 10.0	2.0 --- 6.0	2.0 --- 6.0	4.0 --- 14.2	1.0 --- 10.2	2.0 --- 14.0

A

Specificity and sensitivity of AMR gene calling

	Canu	Flye	Longcyclcr	Miniasm	Necat	Raven	Redbean
<i>C. farmeri</i>	100	100	100	100	100	100	100
<i>C. freundii</i>	99.2	99.2	97.7	93.2	85	94	87.2
<i>E. cloacae</i>	100	100	100	96.9	91.9	98.8	96.3
<i>E. coli</i>	99.8	99.8	97.1	99.6	98	91.2	81.8
<i>K. oxytoca</i>	100	100	100	100	100	100	75
<i>K. pneumoniae</i>	100	99.6	98.9	98.6	99.3	94.6	84.4
<i>P. aeruginosa</i>	100	100	100	100	100	100	97.8
<i>P. mirabilis</i>	100	100	98.1	98.1	100	98.1	71.2
<i>P. stuartii</i>	100	100	97.1	100	97.1	97.1	91.4
<i>S. aureus</i>	97.3	96.7	96	95.3	95.7	95.7	68.2
<i>C. farmeri</i>	100	100	100	100	100	100	100
<i>C. freundii</i>	100	100	100	100	100	100	100
<i>E. cloacae</i>	95.3	95.3	95.3	95.1	94.9	95.2	95.1
<i>E. coli</i>	99	99.2	99.2	99.2	99.2	99.1	99.3
<i>K. oxytoca</i>	100	100	100	100	100	100	100
<i>K. pneumoniae</i>	98	98	98	97.9	98	97.9	98.7
<i>P. aeruginosa</i>	97.9	97.9	97.9	97.9	97.9	97.9	100
<i>P. mirabilis</i>	92.9	92.9	92.7	92.7	92.9	92.7	90.2
<i>P. stuartii</i>	100	100	100	100	100	100	100
<i>S. aureus</i>	98.6	99	99.3	99	99.3	99.3	99

B

Specificity and sensitivity of plasmid replicon calling

	Canu	Flye	Longcyclcr	Miniasm	Necat	Raven	Redbean
<i>C. farmeri</i>	100	100	100	100	100	100	100
<i>C. freundii</i>	100	100	62.2	91.9	62.2	56.8	56.8
<i>E. cloacae</i>	97.9	100	81.2	89.6	83.3	87.5	72.9
<i>E. coli</i>	99.1	92.2	77.2	90	80.8	75.8	57.1
<i>K. oxytoca</i>	100	100	66.7	100	66.7	100	100
<i>K. pneumoniae</i>	100	96.7	82.7	93.5	86	79	63.6
<i>P. aeruginosa</i>	100	100	100	100	100	100	100
<i>P. mirabilis</i>	100	100	66.7	66.7	100	66.7	33.3
<i>P. stuartii</i>	100	100	33.3	66.7	33.3	33.3	33.3
<i>S. aureus</i>	100	95.9	83.5	87.6	89.7	88.7	42.3
<i>C. farmeri</i>	100	100	100	100	100	100	100
<i>C. freundii</i>	100	100	100	100	100	100	100
<i>E. cloacae</i>	97.9	100	100	100	100	100	100
<i>E. coli</i>	99.1	100	99.4	99.5	99.4	99.4	99.2
<i>K. oxytoca</i>	100	100	100	100	100	100	100
<i>K. pneumoniae</i>	100	99	100	100	100	100	100
<i>P. aeruginosa</i>	100	100	100	100	100	100	100
<i>P. mirabilis</i>	100	100	100	100	100	100	100
<i>P. stuartii</i>	100	100	100	100	100	100	100
<i>S. aureus</i>	92.4	100	96.4	98.8	100	97.7	97.6

