1 **Oxford Nanopore MinION genome sequencer: performance characteristics,**

2 **optimised analysis workflow, phylogenetic analysis and prediction of**

3 **antimicrobial resistance in *Neisseria gonorrhoeae***

4 Daniel Golparian[1], Valentina Donà[2,†], Leonor Sánchez-Busó[3], Sunniva Foerster[1], Simon R.

5 Harris[3], Andrea Endimiani[2], Nicola Low[4,‡], Magnus Unemo[1*,‡]

6 [1]WHO Collaborating Centre for Gonorrhoea and other Sexually Transmitted Infections,

7 Department of Laboratory Medicine, Clinical Microbiology, Faculty of Medicine and Health,

8 Örebro University, Örebro, Sweden.

9 [2]Institute for Infectious Diseases, University of Bern, Bern, Switzerland.

10 [3]Pathogen Genomics, The Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton,

11 Cambridgeshire, United Kingdom.

12 [4]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland.

13

14 *Corresponding author. WHO Collaborating Centre for Gonorrhoea and other STIs,

15 Department of Laboratory Medicine, Microbiology, Örebro University Hospital, Örebro,

16 Sweden.   Tel: +46-19-6022038;   Fax: +46-19-127416;   E-mail:

17 magnus.unemo@regionorebrolan.se.

18 ‡Joint senior authors

19 [†]Present address: Institute of Veterinary Bacteriology, Vetsuisse Faculty, University of Bern,

20 Bern, Switzerland

21

22 Keywords: Nanopore, MinION, ONT, gonorrhoea, microbiology, genome sequencing,

23 antimicrobial resistance, third-generation sequencing, next-generation sequencing, bacteria

24

Abstract

Antimicrobial resistant (AMR) *Neisseria gonorrhoeae* strains are common and compromise

gonorrhoea treatment internationally. Rapid identification and characterisation of AMR

gonococcal strains could ensure appropriate and even personalised treatment, and support

identification and investigation of gonorrhoea outbreaks in nearly real-time. Whole-genome

sequencing is ideal for investigation of the emergence and dissemination of AMR

determinants that predict AMR in the gonococcal population and spread of AMR strains in

the human population. The novel, rapid and revolutionary long-read sequencer MinION is a

small hand-held device that can generate bacterial genomes within one day. However, the

accuracy of MinION reads has been suboptimal for many objectives and the MinION has not

been evaluated for gonococci. In this first MinION study for gonococci, we show that

MinION-derived sequences analysed with existing open-access, web-based sequence analysis

tools are not sufficiently accurate to identify key gonococcal AMR determinants.

Nevertheless, using an *in house*-developed CLC Genomics Workbench, we show that ONT-

derived sequences can be used for accurate prediction of decreased susceptibility or resistance

to recommended therapeutic antimicrobials. We also show that the ONT-derived sequences

can be useful for rapid phylogenomic-based molecular epidemiological investigations, and, in

hybrid assemblies with Illumina sequences, for producing contiguous assemblies and finished

reference genomes.

Introduction

Gonorrhoea is a frequently asymptomatic sexually transmitted infection caused by *Neisseria gonorrhoeae* (gonococcus). In 2012, the World Health Organization (WHO) estimated that there were 78 million new cases of gonorrhoea among adults worldwide[1]. Untreated gonorrhoea can lead to serious reproductive tract and (rarely) systemic complications[2-4]. The high estimated global incidence of gonorrhoea, increasing rates of reported prevalence in many more-resourced settings, and particularly the increasing levels of antimicrobial resistance (AMR) have grown to major public health concerns[1,5-8]. Resistance to ceftriaxone and azithromycin in *N. gonorrhoeae* threatens the currently recommended dual antimicrobial therapy, mainly ceftriaxone 250–500 mg intramuscularly plus azithromycin 1–2 g single oral dose, and no new antimicrobials for treatment of gonorrhoea are yet available[4,6-15]. Consequently, WHO included *N. gonorrhoeae* in its first list of AMR "priority pathogens" in February 2017[8,16]. Therefore, timely detection and surveillance of AMR gonococcal strains, their AMR determinants and their emergence and dissemination in the populations globally is crucial[5,6-8,11,17-19].

Next generation sequencing (NGS) is ideal for the elucidation of the molecular determinants of AMR, their dissemination throughout the gonococcal population, and the emergence and dissemination of AMR strains in the human population, nationally and internationally[20,21]. The demands for technologies that can operate at higher speed with less hands-on time have driven innovation into new sequencing approaches. Third-generation sequencing (TGS)[22] with benefits such as increased read length, reduction of sequencing time and reduction of sequencing bias introduced by PCR amplification steps[23], is being used increasingly. Most current DNA sequencing methods are based on either chemical cleavage of DNA molecules[24] or synthesis of new DNA strands. The synthesis-based methods sequence DNA strands labeled either through the primer from which it originates or on the newly

3

74 incorporated bases. This is the basis of the sequencing methods used on most currently

75 available platforms, including Illumina, Ion Torrent (Thermo Fisher Scientific), and Pacific

76 Biosciences (PacBio) sequencers. It has also been shown that individual DNA molecules can

77 be sequenced by monitoring their progress through various types of pores[25,26]. The benefits of

78 this approach include potentially very long and unbiased sequence reads, because neither

79 amplification nor chemical reactions are necessary for sequencing[27]. Oxford Nanopore

80 Technologies (ONT) (Oxford, UK) have introduced this approach with their genome

81 sequencing device MinION, a TGS platform that was commercialized in mid-2015[28,29]. The

82 MinION is a single-molecule nanopore sequencer in which nanopores are embedded in a

83 membrane placed over an electrical detection grid. As part of the library preparation protocol,

84 proteins are bound to the DNA fragment, and the DNA-protein complex is passed through the

85 pore as the motor protein helps to unwind the DNA, controlling the rate of single stranded

86 nucleotides that pass through the pore one at a time[30]. DNA molecules passing through the

87 pores create measureable alterations in the ionic current. These fluctuations ("squiggles") are

88 sequence dependent so the change in ionic current across the membrane can be used by a

89 base-calling algorithm to infer the sequence of nucleotides in each molecule[27,31]. Under ideal

90 conditions, the template DNA strand passes through the pore, followed by a hairpin adapter

91 and finally the complementary strand. In a run where both strands are sequenced, a consensus

92 sequence of the molecule can be produced; these consensus reads are termed two-directional

93 reads or "2D reads" (2D ONT reads) and generally have higher accuracy than reads from only

94 a single pass of the molecule (1D ONT reads). The MinION provides several new advantages:

95 small size (10 × 3 × 2 cm), portability, speed, low cost, and direct connection to a laptop

96 through a USB 3.0 interface; the library construction involves simplified methods; no

97 amplification step is required, and data acquisition and analysis occur in real time. The

98 MinION is promising for microbiological applications including describing the microbiome[32],

4

99   rapid diagnostics[33], transmission and surveillance[34,35], *de novo* assemblies[36-38], and microbial

100   AMR profiling[39,40]. However, the high error rate for the MinION sequencer[28,29] has limited its

101   ability to compete with existing sequencing technologies and the MinION has not been

102   previously evaluated for gonococci.

103   We evaluated the performance characteristics, ideal sequence analysis (tools and

104   workflow for taxonomy, assembly, assembly improvement ("polishing") and mapping),

105   phylogenomic analysis, and prediction of decreased susceptibility or resistance to

106   recommended therapeutic antimicrobials in *N. gonorrhoeae* using the Oxford Nanopore

107   MinION genome sequencer. We sequenced the 2016 WHO gonococcal reference

108   genomes/strains[41] and clinical gonococcal isolates, and evaluated the performance of ONT

109   assemblies and hybrid assemblies including both ONT and Illumina reads. The assemblies

110   were evaluated against the PacBio RS II-sequenced 2016 WHO gonococcal reference

111   genomes (Bioproject PRJEB14020)[41] for sequence variation and to assess whether the

112   assemblies were of sufficient quality for characterisation of AMR determinants and

113   phylogenomic-based molecular epidemiology of gonococci.

114

115   **Methods**

116   *N. gonorrhoeae* **isolates, culture, antimicrobial susceptibility testing and DNA isolation.**

117   Twenty-eight gonococcal isolates, including the previously PacBio-sequenced 2016 WHO

118   reference strains (n=14)[41] and 14 selected clinical AMR isolates, were examined. The clinical

119   isolates were collected as part of the Rapid Diagnosis of Antimicrobial Resistance in

120   Gonorrhoea (RaDAR-Go) project from patients attending clinics in Zurich and Bern,

121   Switzerland, in 2015-2016[42]. All isolates were cultured on GC agar plates (3.6% Difco GC

122   Media agar base (Becton Dickinson, Franklin Lakes, USA) with 1% IsoVitalex (bioMérieux,

123   Marcy-l'Étoile, France), 1% haemoglobin (Becton Dickinson) and 10% horse serum (SVA,

124   Uppsala, Sweden)) and incubated at 36°C, 5% CO_2-enriched atmosphere for 18-20 h. We

125   determined minimum inhibitory concentrations (MICs) of ceftriaxone, cefixime,

126   azithromycin, spectinomycin, ciprofloxacin, tetracycline and benzylpenicillin using the Etest

127   (bioMérieux), according to manufacturer's instructions. Resistance breakpoints from the

128   European Committee on Antimicrobial Susceptibility Testing (EUCAST;

129   www.eucast.org/clinical_breakpoints/) were applied. Nitrocefin test (Thermo Fisher

130   Scientific, Wilmington, DE, USA) was used to detect β-lactamase production. Genomic DNA

131   was isolated using the Wizard Genomic DNA Purification Kit (Promega Corporation,

132   Madison, WI, USA). Isolated DNA was quality controlled using fluorometric quantification

133   (Qubit; Thermo Fisher Scientific) and electrophoresis (Tapestation; Agilent, Santa Clara, CA,

134   USA).

135

136   **Oxford Nanopore Technologies library preparation and MinION sequencing.** Three μg

137   of genomic DNA was sheared to an average fragment length of 8 kb with g-TUBES (Covaris,

138   Woburn, WA, USA). Sequencing libraries were prepared according to the 2D library

139   preparation protocol with the SQK-LSK208 2D ligation kit (Oxford Nanopore Technologies),

140   including the DNA repair step with the NEBNext FFPE DNA repair module (New England

141   Biolabs, Ipswich, MA, USA). The sequencing libraries were purified using MyOne C1 beads

142   (Thermo Fisher Scientific), and 6 μl of sequencing library were loaded onto a R9.4 SpotON

143   flow cell and sequenced with the MinION Mk 1B sequencing device (Oxford Nanopore

144   Technologies) for 24 hours, including a top-up with additional 6 μl of DNA library after the

145   first 6 hours of the sequencing run. Base-calling was performed using the Metrichor cloud

146   software and all ONT reads (quality passed 1D and 2D) as well as only 2D reads were then

147   extracted using Poretools[43] for downstream analysis.

148

6

149   **Illumina library preparation and sequencing.** The sequencing libraries for all clinical

150   isolates (n=14) were prepared using Nextera XT DNA library preparation kit (Illumina, San

151   Diego, CA, USA) and sequenced on the MiSeq Platform (Illumina), according to

152   manufacturer's instructions, resulting in an average of 967,369 reads with an average read

153   length of 257 bp after quality control and average coverage of 86.5× per base. The raw

154   sequence files for the 2016 WHO gonococcal reference strains[41] were obtained from the

155   European Nucleotide Archive (ENA; Bioproject PRJEB14020). The 2D ONT reads and

156   Illumina reads were species confirmed using the online tool One Codex

157   (www.onecodex.com). All sequenced reads are available from the European Nucleotide

158   Archive with the following accession number: PRJEB25703.

159

160   **Assembly and assessment.** To identify the ideal tool for obtaining a high quality *de novo*

161   assembly using only ONT reads, we evaluated Canu (v1.6)[44], Miniasm (vr122)[45], PBcR

162   (v8.3)[46], and SMARTdenovo (available from https://github.com/ruanjue/smartdenovo) for

163   assembly of all ONT reads (1D and 2D) and only the extracted 2D ONT reads. Canu was

164   executed with default parameters including error correction and --genomesize 2.2M. Miniasm

165   was run with default parameters. PBcR was executed with the following parameters: -length

166   500, -partitions 200 and –genomeSize=2200000. SMARTdenovo was run with default

167   parameters and –c 1 to run consensus step. Subsequently, hybridSPAdes (v3.11.1)[47] and

168   MaSuRCA (v3.2.2)[48] were used separately to produce hybrid assemblies by including paired-

169   end Illumina reads in the assembly process. hybridSPAdes was executed with the careful

170   option and –nanopore command. MaSuRCA was performed using the following parameters:

171   GRAPH_KMER_SIZE=auto, USE_LINKING_MATES=0, LIMIT_JUMP_COVERAGE=60,

172   CA_PARAMETERS=cgwErrorRate=0.25, NUM_THREADS=64, JF_SIZE=23000000,

173   DO_HOMOPOLYMER_TRIM=0.

174   The assemblies obtained from Miniasm using 2D ONT reads were error corrected twice

175   using Racon (v1.3.0)[49] and polished with the signal-level consensus software, nanopolish

176   (v0.7.1) (available from: https://github.com/jts/nanopolish). Briefly, 2D ONT reads were

177   mapped as all-vs-all read self-mapping using Minimap with the following options: -S, –w 5, -

178   L 100, -m 0. Subsequently, Miniasm was executed with default options. The 2D ONT reads

179   were mapped back to the assembly with Minimap (default parameters) and corrected with

180   Racon with default parameters, and this was repeated one time. Burrows-Wheeler Aligner

181   (v0.7.17-r1188) (BWA-MEM)[50] with the –x ont2d option was then used to map all reads back

182   to the Racon-corrected assembly to be used as the input for Nanopolish. Nanopolish was run

183   with default parameters and --min-candidate-frequency 0.1. Furthermore, Canu assemblies

184   were error corrected with Pilon (v1.22)[51], Circlator (v1.5.3)[52] and Nanopolish. All error

185   corrections for the Canu assemblies with the above-mentioned tools were performed with

186   default parameters.

187   Finally, to evaluate the accuracy of our assemblies, the ONT (corrected and non-

188   corrected) and hybrid assemblies for the WHO reference strains were compared with the

189   finished and closed 2016 WHO gonococcal reference genomes[41]. All 2D ONT reads were

190   aligned to the genome sequence of respective WHO reference strain, to be able to easily and

191   directly determine the quality and accuracy using BWA-MEM (for mapping and

192   phylogenetics) and Quast (v4.6.0) (for *de novo* assemblies). The top three assessed ONT *de*

193   *novo* assemblies and the best hybrid assemblies were subjects for downstream analysis. Gene

194   coding sequences (CDS) were annotated using Prokka (v1.12) on the chosen assemblies[53].

195

196   **Multiple-sequence alignment and phylogenetics.** All Illumina and 2D ONT reads were

197   mapped to the chromosome of FA1090 (GenBank: NC_002946.2) separately using BWA-

198   MEM, the 2D ONT reads were specified using the nanopore option (-x –ont2d). All

8

199   consensus sequences were merged into a multiple sequence alignment, single-nucleotide

200   polymorphisms (SNPs) were called, and a maximum-likelihood phylogenetic tree based on

201   SNPs was obtained using RAxML (version 8.2.8)[54]. This was also performed on Illumina

202   sequences and 2D ONT sequences separately and the phylogeny was compared in a

203   tanglegram.

204

205   **Antimicrobial resistance determinants.** To detect AMR determinants relevant for

206   gonococci as quick and straight-forwardly as possible for future routine use of MinION

207   sequencing, we evaluated several open-access and user-friendly web-based sequence analysis

208   tools. The top three assessed ONT assemblies and the best hybrid assembly were examined

209   using the Whole Genome Sequence Analysis (WGSA; www.wgsa.net)[21] and the PubMLST

210   *N. gonorrhoeae* AMR (still in development) scheme (www.pubmlst.org)[55]. The focus was on

211   AMR determinants for ceftriaxone and cefixime (*penA, mtrR, penB*), azithromycin (*23S*

212   *rDNA*, *mtrR*), ciprofloxacin (*gyrA*), tetracycline (*rpsJ, tet(M)* (plasmid-mediated resistance)),

213   and penicillin (*penA*, *ponA, mtrR*, *penB*, *bla*$_{TEM}$ (plasmid-mediated resistance)). We only

214   detected the AMR determinants on the plasmids and no additional downstream analysis was

215   performed on the plasmids.

216   Furthermore, the 2D ONT reads were examined in our *in house*-customised CLC

217   Genomics Workbench workflow to characterise the AMR determinants based on alleles in the

218   *N. gonorrhoeae* Sequence Typing for Antimicrobial Resistance (NG-STAR) database

219   (www.ngstar.canada.ca)[56]. The NG-STAR database lacked the possibility to run contig

220   assemblies at the time of this study. The workflow employed *de novo* assembly without

221   mapping and used BLAST for the identification, with the highest (%) hit being reported[57].

222   The best performing tool of assembly and subsequent detection of AMR determinants

223   was subsequently used to characterise the clinical isolates (n=14).

224

225   **Results**

226   **Overview of sequenced data.** The 14 2016 WHO reference strains[41] and 14 clinical isolates

227   were sequenced using the ONT MinION platform. In these MinION runs, we obtained

228   between 105 Mb and 922 Mb of 2D sequences with longest reads ranging from 26.0 to 58.6

229   kb and an average read length of 4.6-6.4 kb. The number of passed 2D ONT reads ranged

230   from 17,500 to 147,700 with an average of 33% of all reads. For evaluation of the ONT

231   sequences and further downstream analysis (taxonomy, hybrid assemblies and phylogeny), we

232   used the 2×100 bp Illumina HiSeq paired-end reads with >250 depth for each WHO reference

233   genome[41] and 2×300 bp Illumina MiSeq paired-end reads with 80-100× depth for each

234   clinical isolate. Taxonomy was defined for all reads using One Codex and 97-99% and 100%

235   of all reads were classified as *N. gonorrhoeae* using ONT (Fig. 1) and Illumina MiSeq,

236   respectively.

237

238   **Comparison of MinION *de novo* assemblies with the 2016 WHO gonococcal reference**

239   **genomes[41].** Assemblies were compared and selected based on the Quast results when

240   comparing to the WHO reference genomes using the following criteria: lowest number of

241   mismatches, misassemblies and number of contigs, and highest fraction of genome covered

242   (Supplementary Table). We selected and analysed ONT assemblies produced by Canu using

243   only 2D reads and polished with Nanopolish, Miniasm twice error corrected with Racon and

244   polished with Nanopolish, and SMARTdenovo. Furthermore, the MaSuRCA hybrid

245   assemblies (based on the ONT plus Illumina reads) were selected for downstream analysis

246   (Table 1). Generally, ONT assemblies based on only the 2D ONT reads had a higher accuracy

247   compared with assemblies using all reads (1D and 2D ONT reads) with fewer mismatches but

248   very modest to no change in contiguity of the assemblies and fraction of the WHO reference

249   genome covered. All error corrections were made using only 2D ONT reads and statistics

250   such as mismatches and indels improved. Nanopolish had the best performance of the tested

251   error correction tools and reduced the number of indels by up to 13 times. Assemblies

252   produced with hybridSPAdes were more affected by the lower numbers of ONT reads, i.e.

253   when using only 2D ONT reads. The hybrid assembly using SPAdes had less contiguity (up

254   to 5.5 times more contigs) with higher number of mismatches (up to 4.4 more mismatches per

255   100 kb) and introductions of misassemblies (Supplementary Table). In general, the 2D ONT

256   reads were used more efficiently by MaSuRCA for hybrid assembly, with overall improved

257   statistics especially for contiguity and mismatches.

258   Although the length of the assemblies based on the ONT sequences did not largely differ

259   from the WHO reference genomes, the number of CDS was vastly different. The 2016 WHO

260   gonococcal reference genomes include 2295-2450 CDS[41]. However, 3210-4712 CDS were

261   identified using Prokka in our selected ONT assemblies. In contrast, Prokka identified 2228-

262   2400 CDS in the MaSuRCA hybrid assemblies, which is in line with the published WHO

263   reference genomes (Table 1).

264   Error corrections using Pilon and Circlator, which is not designed for error correction, did

265   not substantially improve the assemblies (Supplementary Table) and the assemblies error

266   corrected with these tools were not used in the downstream analysis.

267

268   **MinION for molecular epidemiology, including rapid outbreak investigations.** The

269   Illumina reads mapped to the PacBio-produced WHO reference genomes (n=14)[41] with no

270   SNPs detected except for WHO M (one SNP) and a range of 93.3% to 96.4% of the reads

271   mapped to the respective PacBio reference genome. The ONT reads mapped with 2-58 SNPs

272   per genome sequence and 95.2% to 98.3% of the reads mapped. By executing BWA-MEM

273   with the nanopore –x ont2d option, no SNPs between the ONT reads and the respective

274   reference genomes were detected, but we found variation in the number of 2D ONT reads and

275   percentage mapped to the FA1090 reference genome (Fig. 2). Therefore, the BWA-MEM

276   nanopore option was used when mapping all 2D ONT reads to the FA1090 reference genome

277   for the phylogenetic analysis (Fig. 3). By creating the phylogeny with the ONT-produced

278   genomes (n=28) and Illumina MiSeq genomes (n=28) separately, we showed that the ONT

279   sequenced libraries produced a phylogenetic tree topology that was comparable with the one

280   using the Illumina dataset (Fig. 3b). The main difference in the tree topology was that WHO

281   G and WHO N were closer to the root when Illumina sequences were used for the phylogeny.

282   Moreover, all identical isolates separately sequenced with ONT and Illumina clustered

283   together, except for some (n=7) of the clinical isolates where the Illumina sequences were

284   more closely related. In four of these cases (11 cervix (C)/12 rectum (R) and 13 (R)/14 urethra

285   (U)), the isolates were from the same patients but from different anatomical sites. For three

286   isolates (5, 6 and 7), the Illumina sequences were also more closely related but all of the

287   separately sequenced datasets were still on the same branch showing that these isolates were

288   very similar (Fig. 3a).

289

290   **MinION for detection of AMR determinants.** To investigate the four final *de novo*

291   assemblies using Canu (polished), Miniasm (polished), SMARTdenovo, and MaSuRCA

292   (Illumina hybrid) for detection of AMR determinants in the 2016 WHO gonococcal reference

293   genomes[41], we used the WGSA (www.wgsa.net)[21] and PubMLST (www.pubmlst.org).

294 WGSA is an easy-to-use web interface where you drag-and-drop the assemblies into the

295 web browser and the characterisation is done in minutes[21]. We used this interface for all

296 chosen assemblies (n=112) and the analysis of AMR determinants took 15 minutes. We

297 observed good concordance overall with the verified AMR determinants in the reference

298 genomes (Table 2), with the exception of the 23S *rDNA* macrolide resistance determinant

299 C2611T, which was incorrectly identified in the Canu (polished) and Miniasm (polished)

300 assemblies. No ONT assembly was able to correctly characterise *penA* in WHO W and none,

301 including the MaSuRCA hybrid assembly, was able to correctly characterise *penA* in WHO K

302 (Table 2). The characterisation of plasmid-mediated AMR determinants was inaccurate in all

303 four chosen assemblies for the detection of *tet(M)*. Also, the chosen assemblies failed to

304 detect bla_{TEM} in WHO M, WHO O and WHO V and incorrectly reported presence of bla_{TEM}

305 in WHO G in the Miniasm (polished), SMARTdenovo and MaSuRCA assemblies (Table 3).

306 Out of the 2D ONT assemblies, the SMARTdenovo assembly performed the best.

307 We also used the PubMLST database, which is another easy-to-use web based tool, that

308 can mine assembly data for various purposes. The database uses assemblies as input and

309 delivers results in few minutes. The PubMLST database (www.pubmlst.org), using 2D ONT

310 assemblies, was only able to characterise the *mtrR* AMR determinant and this was only done

311 in the Canu assemblies. However, using the MaSuRCA hybrid assembly PubMLST had 100%

312 concordance with the published reference genomes for detection of *penA, mtrR*, *penB, ponA,*

313 *23S rDNA, gyrA,* and *tet(M)*. PubMLST does not include *rpsJ* in its gonococcal AMR

314 characterisation module and also failed to detect the bla_{TEM} gene in WHO M and WHO N

315 (Table 3).

316 The NG-STAR database[56], included in the WHO CC *in house*-customised CLC

317 Genomics Workbench, gave 100% concordance for the presence of all AMR determinants

318 characterised in NG-STAR[56] with the reference genomes (Table 2). All the plasmid-mediated

13

319   AMR determinants (Table 3) and the AMR determinants *penB*, *16S rDNA,* and *parC* were

320   also correctly characterised. The customised CLC Genomics Workbench was therefore used

321   to characterise the AMR determinants in the clinical isolates using only the 2D ONT reads.

322

323   ***In house*-customised CLC Genomics Workbench for characterisation of AMR**

324   **determinants in clinical gonococcal isolates.** The 14 isolates expressed decreased

325   susceptibility (MIC≥0.032 mg/L)[58] or resistance to ceftriaxone (11/14) and cefixime (11/14),

326   azithromycin (3/14), and ciprofloxacin (13/14) (Table 4). No isolate displayed spectinomycin

327   resistance or produced β-lactamase. The *in house*-customised CLC Genomics Workbench

328   including the NG-STAR database[56] correctly characterised all AMR determinants in the

329   isolates using only the 2D ONT reads and found 12, 1, and 13 of the 14 isolates carrying

330   AMR mutations in *penA* (cefixime and ceftriaxone resistance determinant), *23S rDNA*

331   (azithromycin resistance determinant), and *gyrA* (ciprofloxacin resistance determinant),

332   respectively. Twelve of the 14 isolates also harboured *parC* resistance mutations, which

333   contribute to the high MICs in isolates with high-level ciprofloxacin resistance. Specific

334   mutations in *mtrR*, resulting in an over-expression of the MtrCDE efflux pump, and *porB1b*

335   (*penB* AMR determinant), causing a decreased influx of antimicrobials through PorB, were

336   identified in 12 of the 14 isolates. No *16S rDNA* resistance mutations (spectinomycin

337   resistance determinant) or plasmid-mediated AMR determinants (high-level resistance to

338   penicillin and tetracycline) were detected. Resistance mutation in *rpsJ* (decreased

339   susceptibility/resistance to tetracycline) was found in all isolates. There was 100%

340   concordance between the AMR determinants detected using the 2D ONT reads only and

341   Illumina reads. Consequently, resistance or decreased susceptibility to extended-spectrum

342   cephalosporins, azithromycin, spectinomycin, ciprofloxacin, tetracycline, and penicillin could

343   be accurately predicted using solely 2D ONT reads (Table 4).

14

344

Discussion

346 In the present study, we show that gonococcal genomes can be *de novo* assembled with high

347 accuracy and contiguity running assemblies with MinION 2D ONT reads in combination with

348 Illumina reads using MaSuRCA. These assemblies varied from one to eight contigs and could

349 be further investigated using tools like Circlator[52] to obtain circular and finished gonococcal

350 genomes. Furthermore, it is possible to obtain accurate *de novo* assemblies for AMR

351 determinant detection by initially performing Illumina-based correction of individual ONT

352 reads and subsequently using the reads for *de novo* assembly. However, this strategy makes

353 the ONT platform dependent on additional Illumina sequencing. Nevertheless, this is another

354 approach for producing accurate finished genomes and, accordingly, an alternative to PacBio

355 sequencing of microbial reference genomes.

356 When relying only on the ONT data, the assemblies contained high numbers of

357 mismatches (27-607 mismatches per 100 kb) and indels (641-1125 indels per 100 kb), using

358 even the best ONT assemblies obtained using SMARTdenovo, Canu and Miniasm with 2D

359 ONT reads. There are several possible reasons for the high and incorrect number of CDSs in

360 the ONT assemblies using Prokka including sequencing artefacts, homopolymers and overall

361 high error rate (SNPs, insertions and deletions) in the ONT reads causing the assemblies to

362 contain incorrect internal stop codons and false pseudogenes. Generally, the 2D ONT reads

363 generated more accurate assemblies than the 1D ONT reads and were used in the majority of

364 our downstream analysis. We aimed to obtain highly accurate *de novo* assemblies to be able

365 to rapidly identify relevant gonococcal AMR determinants to predict AMR, using only user-

366 friendly and rapid online sequence analysis tools such as WGSA (www.wgsa.net)[21] and

367 PubMLST (www.pubmlst.org). Using these online tools, the characterisation of AMR

368 determinants was generally inaccurate using ONT assemblies and, as expected, highly

15

369   accurate with hybrid assemblies. However, the *in house*-customised CLC Genomics

370   Workbench workflow, including the NG-STAR database[56], provided 100% concordance with

371   the AMR determinants of the PacBio sequenced 2016 WHO gonococcal reference genomes[41]

372   using only 2D ONT reads. We can also extract additional AMR determinants and in general

373   genes of interest that are not included in the online tools, because our AMR database is fully

374   customisable. The software workflow is based on extraction of the genes of interest in the

375   customised database from a *de novo* assembly using BLAST algorithms optimised for each

376   AMR determinant and reporting the highest hit, i.e. not only the 100% hit, in also genes with

377   premature stop codons. This is essential because the ONT *de novo* assemblies are prone to

378   errors. The main limitations of any CLC Genomics Workbench are that it is commercial (not

379   an open-source online tool) and it has fairly high system requirements (16-32 GB RAM).

380   Nevertheless, it has a simple general user interface for users who are not familiar with

381   command line bioinformatics and is available to all widely used operating systems.

382   Hopefully, open-source tools and public databases that can handle error-prone assemblies will

383   be inspired by this approach and incorporate e.g. NG-STAR data[56] rendering costly software

384   obsolete. NG-STAR can currently not use genome assemblies as input for determining the

385   AMR profiles, which limits its possibilities when performing WGS.

386   Easy and rapid genome sequencing, using platforms such as the ONT MinION, in

387   combination with algorithms that can appropriately predict AMR profiles using only genetic

388   data could be very valuable in future surveillance of gonococcal AMR and spread of AMR

389   gonococcal strains, nationally and internationally. Using the customised CLC Genomics

390   Workbench, we showed that all isolates expressing decreased susceptibility to extended-

391   spectrum cephalosporins (ESCs; cefixime and ceftriaxone) were carrying a mosaic *penA* allele

392   or a *penA* A501 substitution, which are associated with decreased susceptibility or resistance

393   to ESCs. Furthermore, one azithromycin resistant isolate (MIC=4 mg/L) contained the

16

394   C2611T mutation in four alleles of the 23S rRNA gene, conferring azithromycin resistance.

395   This macrolide-resistance mutation can be challenging to detect accurately in gonococci

396   because of the four alleles of this gene in the gonococcal genome. The number of 23S rRNA

397   gene alleles containing the C2611T mutation is associated with the level of resistance to

398   azithromycin. We included a separate analysis to detect the frequency of the mutated 23S

399   rRNA gene alleles by mapping the 2D ONT reads to a *23S rDNA* reference and analysing the

400   frequency of variation across the gene. We also showed that all ciprofloxacin resistant isolates

401   contained mutations in the *gyrA* gene (resulting in amino acid substitution S91F) and isolates

402   that also had mutations in *parC* (resulting in amino acid substitutions D86N, S87R and E91G)

403   displayed high-level resistance to ciprofloxacin. Finally, the decreased susceptibility and

404   resistance to penicillin could mainly be explained by mutations in *penA* and *ponA*, while

405   resistance to tetracycline appeared to be due to factors other than only *rpsJ* and the *tet*(*M*)

406   carrying plasmid.

407   The use 2D ONT reads for phylogenomics of gonococci, e.g. for molecular

408   epidemiological surveillance purposes, would be exceedingly valuable for outbreak

409   investigations and monitoring in nearly real time in local, national and international

410   gonococcal surveillance programmes, which aim to replace traditional, labour intensive and

411   less accurate genotyping techniques such as *N. gonorrhoeae* multi-antigen sequence typing

412   (NG-MAST) and multi-locus sequence typing (MLST) with WGS techniques[21]. Accordingly,

413   we examined the accuracy of performing phylogenomic analysis using 2D ONT reads

414   mapped to a reference genome using BWA-MEM with the nanopore option, and building a

415   phylogeny using the multiple sequence alignment. The ONT data produced a phylogenetic

416   tree topology that was comparable with the one using the Illumina dataset and similarly

417   clustered all related isolates (Fig. 3). However, the number of isolates was limited and the

418   genomic heterogeneity of the strains was high, which might have slightly biased our analysis.

419   Interestingly, the clustering of isolates was not highly dependent on the number of 2D

420   reads (Fig. 1). This suggests that the read length provides sufficient genome coverage even

421   when a relatively low number of reads are available. For example, the lower number of 2D

422   reads for WHO G (Fig 1) still provided a 99.96-100% genome fraction (Supplementary

423   Table) and, accordingly, the read length compensated for the low number of reads. For the

424   different WHO reference strains, 89.93-100% fraction of the PacBio-sequenced genomes

425   were covered by the ONT reads (Supplementary Table). Consequently, for some purposes

426   including gonococcal phylogenomic analysis, the MinION sequencer has likely collected

427   enough reads in less than one day, which opens up new possibilities in running WGS to

428   investigate gonorrhoea outbreaks in nearly real time, as well as directly from clinical samples.

429   The main benefits of using the MinION for genome sequencing included the very rapid

430   turn-around time, high accessibility by connecting the small hand-held device to a laptop, and

431   low cost of the sequencer. The main limitations included the high cost of each sequencing run

432   and that the sequencing reads remain less accurate and consistent than Illumina and PacBio

433   sequencing reads.

434   In conclusion, we show, in the first MinION study for gonococci, that ONT sequences

435   analysed with currently existing open-access, web-based sequence analysis tools are not

436   sufficiently accurate to identify key gonococcal AMR determinants. However, using an

437   appropriate analysis workflow such as an *in house*-developed CLC Genomics Workbench, we

438   show that ONT sequence data can be used for rapid and accurate identification of AMR

439   determinants to predict decreased susceptibility or resistance to recommended therapeutic

440   antimicrobials. We also show that ONT sequence data can be useful for phylogenomics of *N.*

441   *gonorrhoeae*, e.g. for molecular epidemiological investigations in nearly real time, and, using

442   ONT-Illumina hybrid assemblies, for producing contiguous assemblies and finished

443   gonococcal reference genomes.

18

444

References

1. Newman, L. *et al.* Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. *PLoS One*. **10**, e0143304; 10.1371/journal.pone.0143304 (2015).

2. Hook, E. W. III & Handsfield H. Gonococcal infections in adults. *In:* Sexually Transmitted Infections (eds. Holmes K. K. *et al*), 4th edition, The McGraw-Hill Companies Inc., USA (2008).

3. Kerle, K. K., Mascola, J. R. & Miller T. A. Disseminated gonococcal infection. *Am Fam Physician*. **45**, 209-214 (1992).

4. World Health Organization (WHO). *WHO guidelines for the treatment of Neisseria gonorrhoeae.* http://www.who.int/reproductivehealth/publications/rtis/gonorrhoea-treatment-guidelines/en/ (2016).

5. World Health Organization. *Global action plan to control the spread and impact of antimicrobial resistance in Neisseria gonorrhoeae.* http://whqlibdoc.who.int/publications/2012/9789241503501_eng.pdf?ua=1 (2012).

6. Unemo, M. & Shafer, W. M. Antimicrobial resistance in *Neisseria gonorrhoeae* in the 21st century: past, evolution, and future. *Clin Microbiol Rev*. **27**, 587–613 (2014).

7. Unemo, M. *et al.* Sexually transmitted infections: challenges ahead. *Lancet Infect Dis.* **17**, e235-e279 (2017).

8. Wi, T. *et al.* Antimicrobial resistance in *Neisseria gonorrhoeae*: Global surveillance and a call for international collaborative action. *PLoS Med.* **14**, e1002344; 10.1371/journal.pmed.1002344 (2017).

9. Terkelsen, D. *et al.* Multidrug-resistant *Neisseria gonorrhoeae* infection with ceftriaxone resistance and intermediate resistance to azithromycin, Denmark, 2017. *Euro Surveill*. **22**, 10.2807/1560-7917.ES.2017.22.42.17-00659 (2017).

10. Fifer, H. *et al.* Failure of dual antimicrobial therapy in treatment of gonorrhea. *N Engl J Med*. **374**, 2504-2506 (2016).

11. Cole, M. J. *et al.* Overall low extended-spectrum cephalosporin resistance but high azithromycin resistance in *Neisseria gonorrhoeae* in 24 European countries, 2015. *BMC Infect Dis*. **17**, 617; 10.1186/s12879-017-2707-z (2017).

12. Katz, A. R. *et al.* Cluster of *Neisseria gonorrhoeae* isolates with high-level azithromycin resistance and decreased ceftriaxone susceptibility, Hawaii, 2016. *Clin Infect Dis*. **65**, 918-923 (2017).

13. Alirol, E. *et al*. Multidrug-resistant gonorrhea: A research and development roadmap to discover new medicines. *PLoS Med*. **14**, e1002366 (2017).

14. British Government. Multi-drug resistant gonorrhoea in England: 2018. March 2018. England: British Government; 2018. Available from: https://www.gov.uk/government/publications/multi-drug-resistant-gonorrhoea-in-england-2018

15. Australian Government Department of Health. Multi-drug resistant gonorrhoea. April 2018. Australia: Department of Health; 2018. Available from: http://www.health.gov.au/internet/main/publishing.nsf/Content/mr-yr18-dept-dept004.htm

16. World Health Organization (WHO). *Global priority list of antibiotic-resistance bacteria to guide research, discovery, and development of new antibiotics*. http://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf?ua=1 (2017).

492   17. World Health Organization (WHO). *Global surveillance network for gonococcal*

493   *antimicrobial   susceptibility*.

494   http://apps.who.int/medicinedocs/documents/s16348e/s16348e.pdf (1990).

495   18. Spiteri, G. *et al.* The European gonococcal antimicrobial surveillance programme

496   (Euro-GASP)–a sentinel approach in the European Union (EU)/European economic

497   area (EEA). *Sex Transm Infect*. **89**, Suppl 4:iv16-8; 10.1136/sextrans-2013-051117

498   (2013).

499   19. Kirkcaldy, R. D. *et al*. *Neisseria gonorrhoeae* antimicrobial susceptibility

500   surveillance–the gonococcal isolate surveillance project, 27 sites, United States, 2014.

501   *MMWR Surveill Summ*. **65**, 1-19 (2016).

502   20. Donà, V., Low, N., Golparian, D. & Unemo, M. Recent advances in the development

503   and use of molecular tests to predict antimicrobial resistance in *Neisseria*

504   *gonorrhoeae*. *Expert Rev Mol Diagn*. **17**, 845-859 (2017).

505   21. Harris, S. R. *et al*. Public health surveillance of multidrug-resistant clones of *Neisseria*

506   *gonorrhoeae* in Europe: a genomic survey. *Lancet Infect Dis.* 2018 May 15. pii:

507   S1473-3099(18)30225-1.

508   22. Bleidorn, C. Third generation sequencing: technology and its potential impact on

509   evolutionary biodiversity research. *Syst Biodiver*. **14**, 1-8 (2016).

510   23. Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing.

511   *Hum Mol Genet*. **19**, R227-R240 (2010).

512   24. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc Natl Acad*

513   *Sci*. **74**, 560–564 (1977).

514   25. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of

515   individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci*.

516   **93**, 13770–13773 (1996).

26. Venkatesan, B. M. & Bashir, R. Nanopore sensors for nucleic acid analysis. *Nat Nanotechnol*. **6**, 615-624 (2011).

27. Yang, Y. *et al.* Advances in nanopore sequencing technology. *J Nanosci Nanotechnol.* **13**, 4521–4538 (2013).

28. Camilla, L. C. *et al*. MinION analysis and reference consortium: Phase 1 data release and analysis. *F1000Res*. **4**, 1075; 10.12688/f1000research.7201.1 (2015).

29. Miten, J. *et al.* MinION analysis and reference consortium: Phase 2 data release and analysis of R9.0 chemistry. *F1000Res*. **6**, 760; 10.12688/f1000research.11354.1 (2017).

30. Clarke, J. *et al*. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*. **4**, 265–270 (2009).

31. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci*. **106**, 7702–7707 (2009).

32. Kerkhof, L. J., Dillon, K. P., Häggblom, M. M. & McGuinness, L. R. Profiling bacterial communities by MinION sequencing of ribosomal operons. *Microbiome*. **5**, 116; 10.1186/s40168-017-0336-9 (2017).

33. Imai, K. *et al*. A novel diagnostic method for malaria using loop-mediated isothermal amplification (LAMP) and MinION™ nanopore sequencer. *BMC Infect Dis*. **17**, 621; 10.1186/s12879-017-2718-9 (2017).

34. Faria, N. R. *et al*. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature*. **546**, 406-410 (2017).

35. Quick J. *et al*. Real-time, portable genome sequencing for Ebola surveillance. *Nature.* **530**, 228-232 (2016).

36. Deschamps S. *et al.* Characterization, correction and de novo assembly of an Oxford Nanopore genomic dataset from *Agrobacterium tumefaciens*. *Sci Rep*. **6**, 28625; 10.1038/srep28625 (2016).

37. Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat Methods*. **12**, 733–735 (2015).

38. Istace, B. *et al*. *de novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience*. **6**, 1-13 (2017).

39. van der Helm, E. *et al.* Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res*. **45**, e61; 10.1093/nar/gkw1328 (2017).

40. Judge, K., Harris, S. R., Reuter, S., Parkhill, J. & Peacock, S. J. Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes. *J Antimicrob Chemother*. **70**, 2775-2778 (2015).

41. Unemo, M. *et al*. The novel 2016 WHO *Neisseria gonorrhoeae* reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization. *J Antimicrob Chemother*. **71**, 3096-3108 (2016).

42. Donà, V. *et al*. Multiplex real-time PCR assay with high-resolution melting analysis for characterization of antimicrobial resistance in *Neisseria gonorrhoeae*. *J Clin Microbiol*. **54**, 2074-2081 (2016).

43. Loman, N. J. & Quinlan A. R. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics*. **30**, 3399-3401 (2014).

44. Koren, S., Walenz, B. P., Berlin, K., Miller J. R. & Phillippy A. M. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. **27**, 722-736 (2017).

45. Li, H. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics*. **32**, 2103-2110 (2016).

23

46. Koren, S. *et al*. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat Biotechnol*. **30**, 693-700 (2012).

47. Antipov, D., Korobeynikov, A., McLean, J. S. & Pevzner, P. A. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics*. **32**, 1009-1015 (2016).

48. Zimin, A. V. The MaSuRCA genome assembler. *Bioinformatics*. **29**, 2669-2677 (2013).

49. Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res*. **27**, 737-746 (2017).

50. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. **26**, 589-595 (2010).

51. Walker, B. J. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. **9**, e112963; 10.1371/journal.pone.0112963 (2014).

52. Hunt, M. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol*. **16**, 294; 10.1186/s13059-015-0849-0 (2015).

53. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. **30**, 2068-2069 (2014).

54. Stamatakis, R. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312-1313 (2014).

55. Jolley, K. A. & Maiden M. C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. **11**, 595; 10.1186/1471-2105-11-595 (2010).

56. Demczuk, W. *et al*. *Neisseria gonorrhoeae* Sequence Typing for Antimicrobial Resistance, a novel antimicrobial resistance multilocus typing scheme for tracking

591   global dissemination of *N. gonorrhoeae* strains. *J Clin Microbiol.* **55**, 1454-1468

592   (2017).

593   57. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J*

594   *Antimicrob Chemother.* **67**, 2640-2644 (2012).

595   58. Golparian, D., Hellmark, B., Fredlund, H. & Unemo, M. Emergence, spread and

596   characteristics of *Neisseria gonorrhoeae* isolates with in vitro decreased susceptibility

597   and resistance to extended-spectrum cephalosporins in Sweden. *Sex Transm Infect.* **86**,

598   454-460 (2010).

599

Acknowledgment

611

Authors' contributions

613 D.G., V.D., A.E., N.L. and M.U. planned and designed the study. V.D. performed the

614 MinION sequencing. D.G. performed the Illumina sequencing and all the bioinformatic

615   analysis, with support from S.H. and L.S-B. D.G. supported by M.U wrote a draft of the

616   manuscript. All authors commented on and approved the final manuscript.

617

618   Competing interests

619   The author(s) declare no competing interests.

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635 Figure 1. Readcount for the MinION (Oxford Nanopore Technologies (ONT)) dataset and

636 read taxonomy classification of 2D reads, when sequencing the 2016 WHO *Neisseria*

637 *gonorrhoeae* reference strains (n=14)[41] and 14 clinical *N. gonorrhoeae* isolates. Percentage

638 within each bar represents the proportion of reads classified as *N. gonorrhoeae*. The number

639 above each bar represents the number of reads that were classified.

640

641 Figure 2. Overview of the number of MinION (Oxford Nanopore Technologies (ONT)) 2D

642 reads and the mapability of the reads to the genome of the *Neisseria gonorrhoeae* reference

643 strain FA1090, when sequencing the 2016 WHO *N. gonorrhoeae* reference strains (n=14)[41]

644 and 14 clinical *N. gonorrhoeae* isolates.

645

646 Figure 3. (A) Phylogenetic tree of the genome sequences of the 2016 WHO *N. gonorrhoeae*

647 reference strains (n=14)[41] and clinical gonococcal isolates (n=14) sequenced with Illumina

648 MiSeq and MinION (Oxford Nanopore Technologies (ONT)). The tree uses using the genome

649 of the *N. gonorrhoeae* reference strain FA1090 as reference (shown with black bar). The

650 platform used, number of reads, average read length, and number of single nucleotide

651 polymorphisms (SNPs) are displayed as colored bars next to each node in the tree. The

652 numbers inside the SNP-bars is the pairwise distance between the Illumina and ONT

653 sequences. (B) Tanglegram to compare the phylogenetic networks based on sequencing using

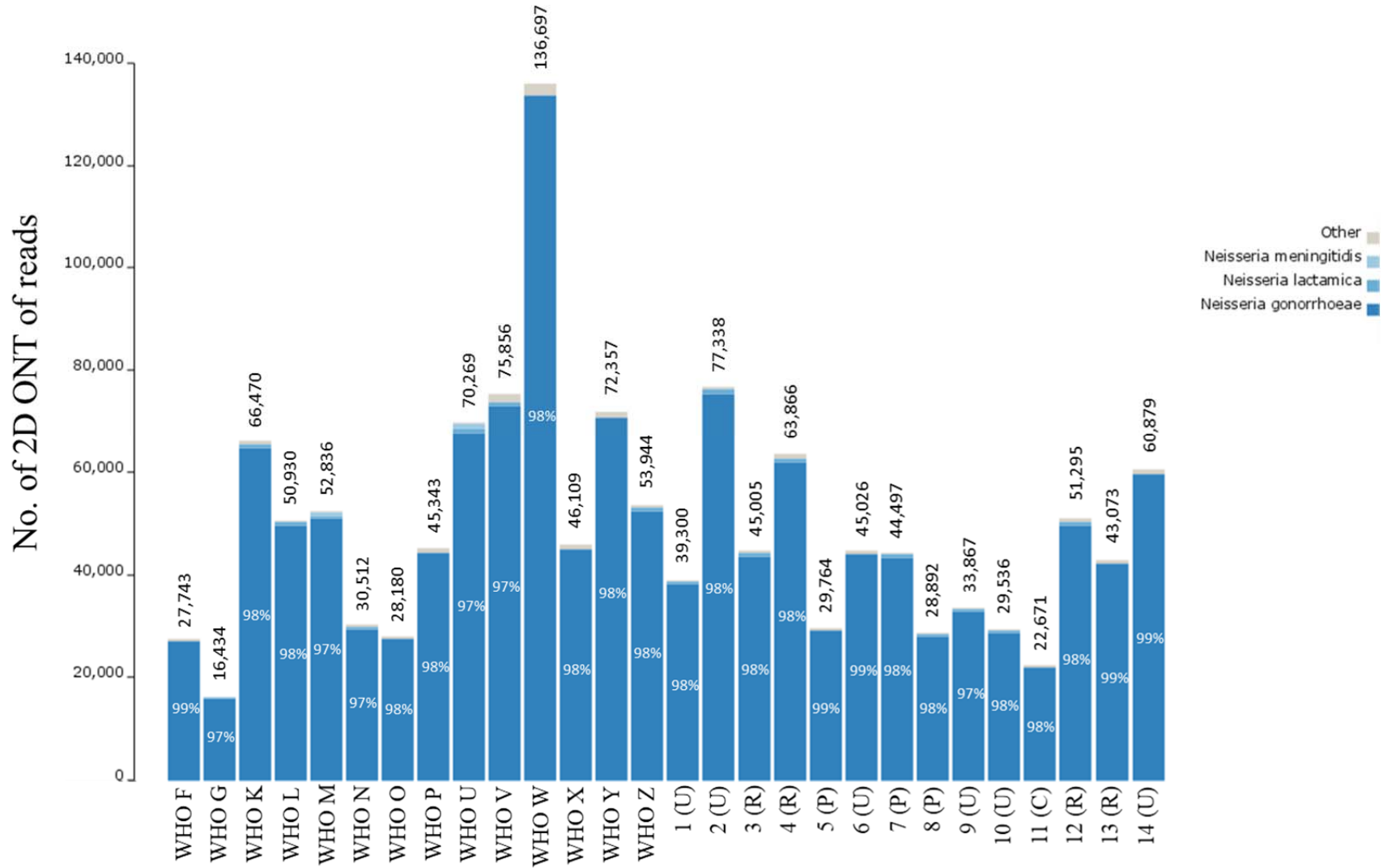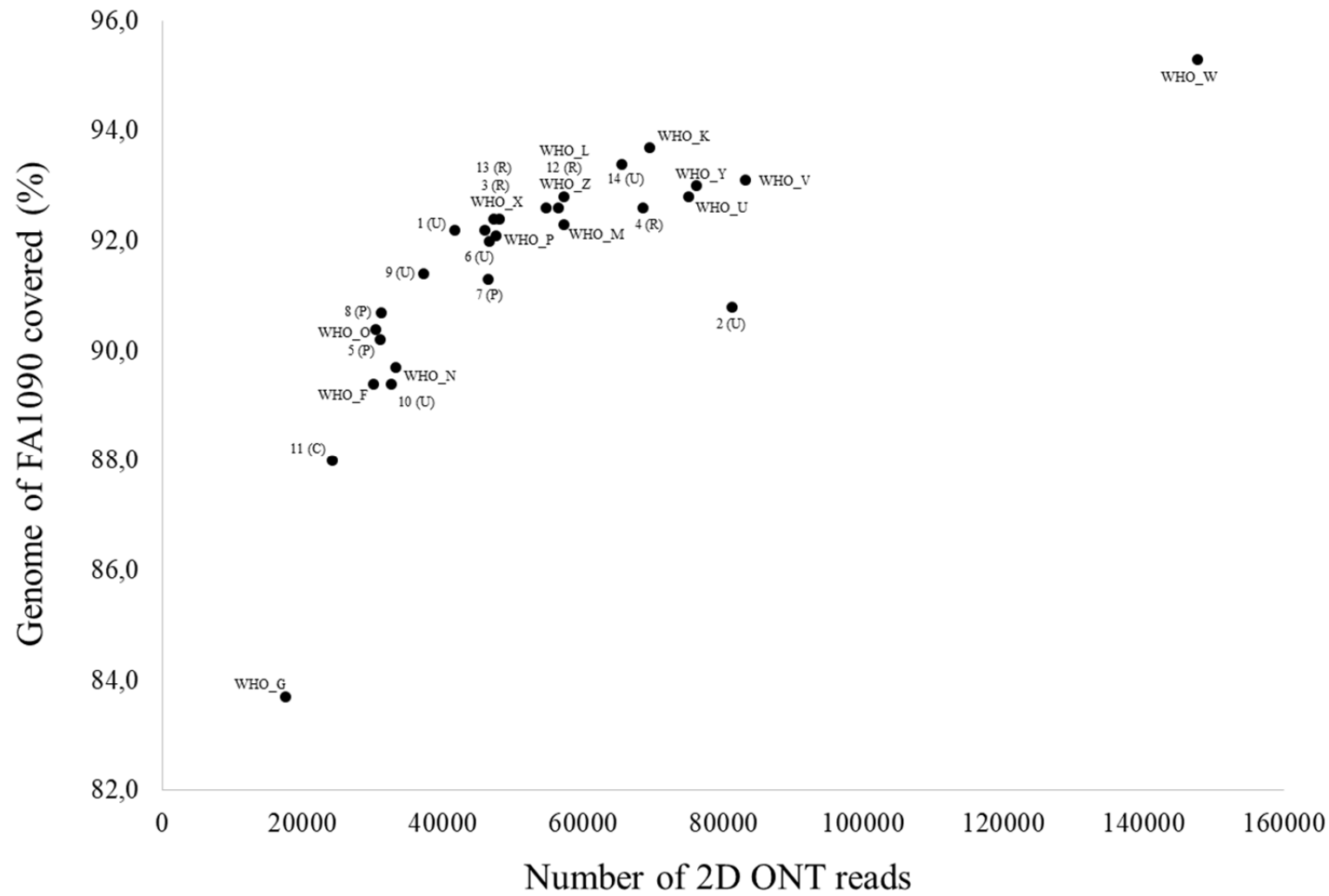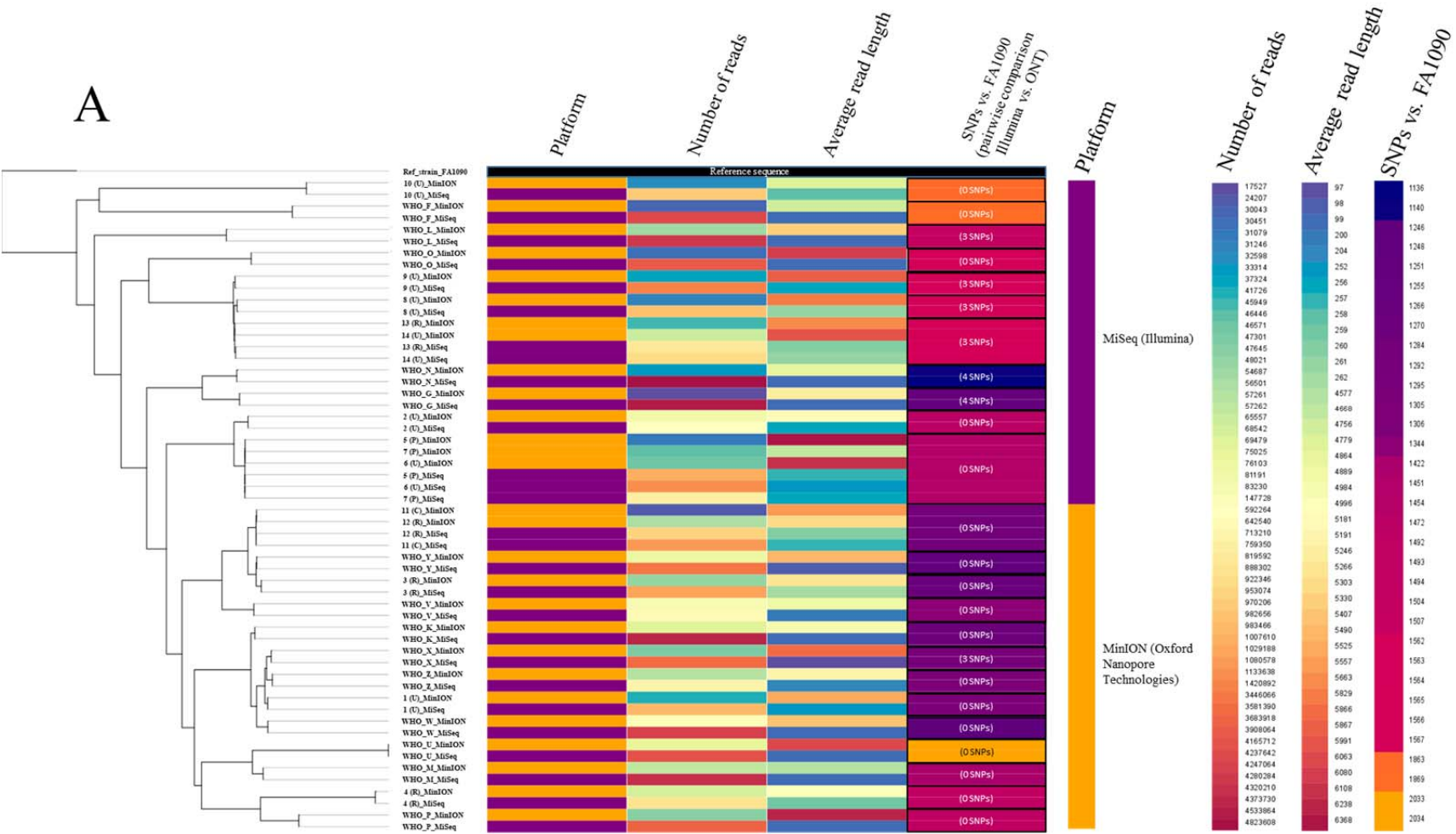654 Illumina technology (left hand side) and Oxford Nanopore technology (right hand side).
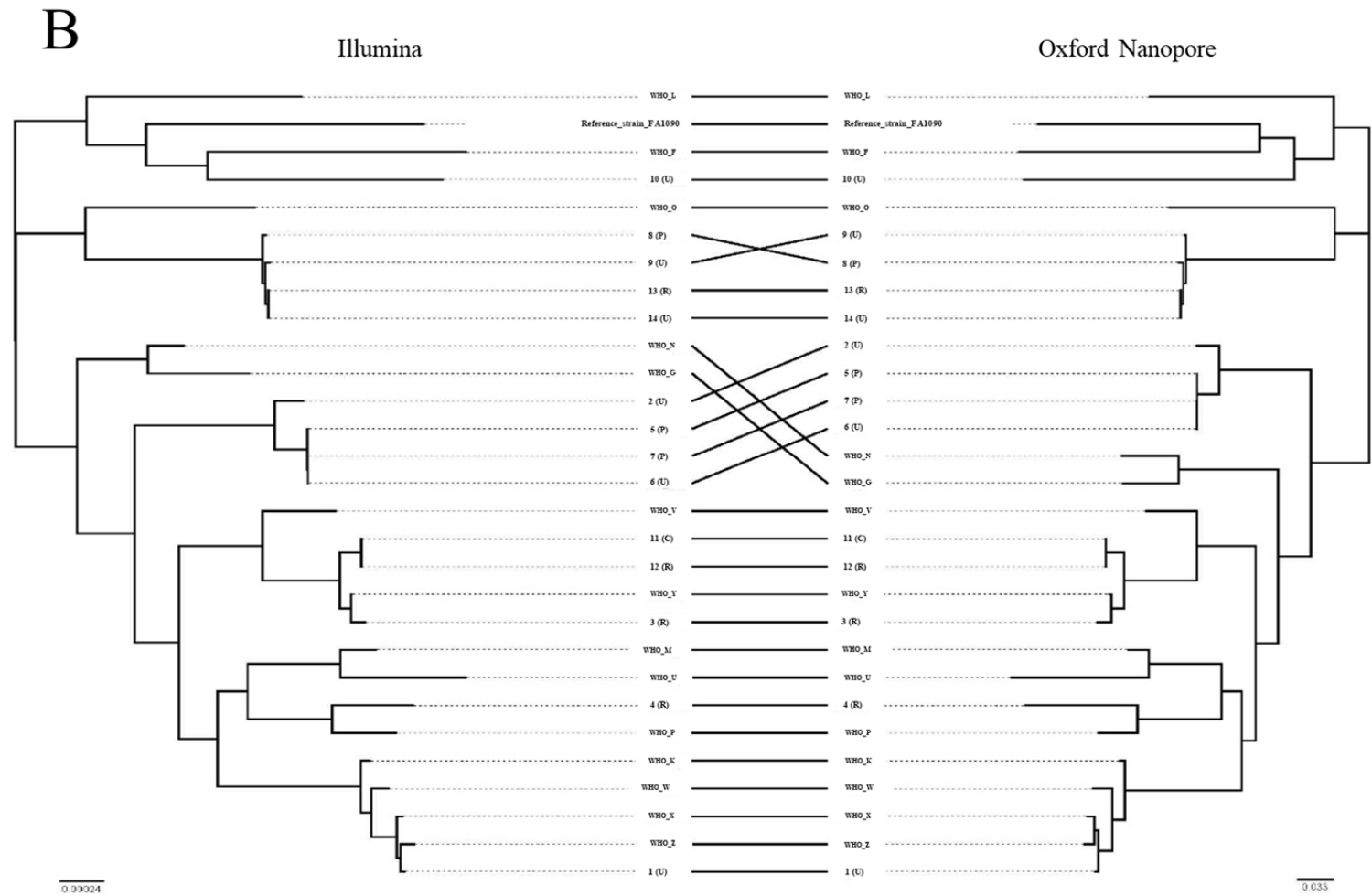
655

656

657   **Fig 1.**

658

659   **Fig 2.**

A

B

Illumina   Oxford Nanopore

661

Fig 3.

663 **Table 1.** Statistics for the three chosen MinION (Oxford Nanopore Technologies)-only *de novo* assemblies and the MinION-Illumina hybrid

664 assembly, when sequencing the 2016 WHO *Neisseria gonorrhoeae* reference strains (n=14)[41].

		WHO F	WHO G	WHO K	WHO L	WHO M	WHO N	WHO O	WHO P	WHO U	WHO V	WHO W	WHO X	WHO Y	WHO Z
Canu[44] + Nanopolish	Genome fraction (%)	98.96	99.99	100.00	99.97	99.59	100.00	99.99	99.96	99.81	100.00	89.93	100.00	100.00	99.48
	Contigs	1	5	3	3	2	3	3	1	1	1	1	1	1	1
	Largest contig (bp)	2266147	1611568	1415719	2166845	2178610	2182677	2178602	2170535	2244376	2231681	1997317	2190700	2234645	2222403
	Mismatches / 100 kb	301.10	205.07	194.76	278.69	226.92	146.03	258.56	266.96	186.77	113.13	277.04	190.09	268.69	196.64
	Indels / 100 kb	309.39	412.92	182.73	261.07	207.51	208.12	255.61	257.39	158.88	92.15	239.01	191.75	320.37	265.95
	Annotated CDS	3893	4172	3632	3783	3717	3898	3770	3561	3726	3639	3210	3543	3863	3850
Miniasm[45] + Racon[49] + Nanopolish	Genome fraction (%)	99.48	99.96	100.00	100.00	100.00	95.67	99.99	99.89	100.00	99.99	99.97	100.00	99.99	98.22
	Contigs	7	6	6	9	7	6	4	7	6	3	7	1	1	71
	Largest contig (bp)	761198	744021	1060644	1288971	938255	2063596	2168124	1621005	1775729	2220894	1171322	2171999	2229595	1847314
	Mismatches / 100 kb	249.03	262.03	161.07	291.29	107.47	310.14	290.36	198.44	229.38	265.48	119.6	256	340.97	237.81
	Indels / 100 kb	275.87	616.64	394.5	448.49	120.41	497.18	393.58	495.94	276.87	327.2	140.89	255.68	358.28	238.45
	Annotated CDS	3997	4485	4321	4276	3789	4240	3999	4418	4146	3848	3967	3515	3969	4373
SMARTdenovo	Genome fraction (%)	99.02	100.00	100.00	100.00	100.00	99.98	100.00	100.00	99.62	100.00	100.00	100.00	100.00	100.00
	Contigs	2	2	2	2	3	3	3	1	15	1	2	1	1	1
	Largest contig (bp)	1749604	2152447	1769445	2167238	2182509	2162509	2160811	2167730	1026600	2213205	2213608	2163199	2210497	2211396
	Mismatches / 100 kb	219.03	111.24	97.38	124.36	179.05	202.09	173.35	114.91	224.47	26.65	226.22	175.3	246.44	237.6
	Indels / 100 kb	680.4	822.11	670.05	690.99	654.08	686.08	662.68	669.96	657.5	641.03	659.5	639.12	817.77	735.46
	Annotated CDS	4602	4470	4388	4523	4512	4614	4526	4372	4712	4378	4301	4342	4553	4461
MaSuRCA[48]	Genome fraction (%)	99.46	99.99	99.91	99.99	99.97	99.99	99.99	99.99	100.00	99.99	99.99	99.99	99.99	99.95
	Contigs	1	3	3	2	5	3	5	3	2	7	8	6	4	1
	Largest contig (bp)	2279985	2176816	2128117	2190405	2189696	2188414	2197012	2204260	2258223	2236154	2238072	2171685	2248815	2255505
	Mismatches / 100 kb	1.89	0.74	2.54	3.97	1.38	2.12	2.31	2.48	2.55	0.14	0.95	2.9	0.45	3.68
	Indels / 100 kb	2.06	1.02	1.71	1.38	1.01	1.66	1.71	1.89	1.61	1.35	2.43	1.98	0.99	1.88
	Annotated CDS	2304	2229	2228	2261	2307	2264	2306	2241	2312	2362	2400	2230	2333	2259

665
666 CDS, coding sequence

667   **Table 2.** Characterisation of chromosomal antimicrobial resistance (AMR) determinants in the 2016 WHO *Neisseria gonorrhoeae* reference

668   strains (n=14)[41] using Whole Genome Sequence Analysis (WGSA; www.wgsa.net)[21] and three MinION-only assemblies, MinION-Illumina

669   hybrid assembly and an *in house*-customised CLC Genomics Workbench, including NG-STAR[56] and a customised database for AMR

670   determinant detection. The discrepant results from the published reference genomes are in bold. The correct AMR determinants are in italic.

	Assembler	Gene	WHO F	WHO G	WHO K	WHO L	WHO M	WHO N	WHO O	WHO P	WHO U	WHO V	WHO W	WHO X	WHO Y	WHO Z
wgsa.net	Canu[44] + Nanopolish	penA	WT	WT	**WT**	A501V	WT	WT	WT	WT	WT	WT	**WT**	Mosaic	Mosaic, A501P	Mosaic
		mtrR (A-deletion/G45D)	No	**No**	Yes	Yes	Yes	No	**No**	No	No	Yes	**No**	Yes	Yes	No
		penB (G101, A102)	WT, WT	WT, WT	K, D	K, D	K, D	WT, WT	K, D	WT, D	WT, WT	K, D	K, D	K, D	K, N	K, D
		23S rDNA	**C2611T**	WT	WT	WT	WT	WT	WT	**C2611T**	C2611T	A2059G	WT	**C2611T**	**C2611T**	**C2611T**
		gyrA (S91F)	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
		ponA (L421P)	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
		rpsJ (V57M)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Miniasm[45] + Racon[49] + Nanopolish	penA	WT	WT	**WT**	A501V	WT	WT	WT	WT	WT	WT	**WT**	Mosaic	A501P	Mosaic
		mtrR (A-deletion/G45D)	No	**No**	Yes	Yes	Yes	No	**No**	No	No	Yes	Yes	Yes	Yes	No
		penB (G101, A102)	WT, WT	WT, WT	K, D	K, D	K, D	WT, WT	K, D	WT, D	WT, WT	K, D	K, D	K, D	K, N	K, D
		23S rDNA	**C2611T**	WT	WT	WT	**C2611T**	WT	WT	WT	C2611T	A2059G	WT	WT	WT	WT
		gyrA (S91F)	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
		ponA (L421P)	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
		rpsJ (V57M)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	SMARTdenovo	penA	WT	WT	**WT**	A501V	WT	WT	WT	WT	WT	WT	**WT**	Mosaic	A501P	Mosaic
		mtrR (A-deletion/G45D)	No	**No**	Yes	Yes	Yes	No	**No**	No	No	No	**No**	**No**	**No**	No
		penB (G101, A102)	WT, WT	WT, WT	K, D	K, D	K, D	WT, WT	K, D	WT, D	WT, WT	K, D	K, D	K, D	K, N	K, D
		23S rDNA	WT	WT	WT	WT	WT	WT	WT	WT	C2611T	A2059G	WT	WT	WT	WT
		gyrA (S91F)	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
		ponA (L421P)	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
		rpsJ (V57M)	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	MaSuRCA[48]	penA	WT	WT	**WT**	A501V	WT	WT	WT	WT	WT	WT	**WT**	Mosaic	A501P	Mosaic
		mtrR (A-deletion/G45D)	No	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	No
		penB (G101, A102)	WT, WT	WT, WT	K, D	K, D	K, D	WT, WT	K, D	WT, D	WT, WT	K, D	K, D	K, D	K, N	K, D
		23S rDNA	WT	WT	WT	WT	WT	WT	WT	WT	C2611T	A2059G	WT	WT	WT	WT
		gyrA (S91F)	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
		ponA (L421P)	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
		rpsJ (V57M)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	In house-customised CLC Genomics Workbench	penA	WT	WT	Mosaic	A501V	WT	WT	WT	WT	WT	WT	Mosaic	Mosaic	Mosaic, A501P	Mosaic
		mtrR (A-deletion/G45D)	No	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	No
		penB (G101, A102)	WT, WT	WT, WT	K, D	K, D	K, D	WT, WT	K, D	WT, D	WT, WT	K, D	K, D	K, D	K, N	K, D
		23S rDNA	WT	WT	WT	WT	WT	WT	WT	WT	C2611T	A2059G	WT	WT	WT	WT
		gyrA (S91F)	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
		ponA (L421P)	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
		rpsJ (V57M)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

671

672

673 **Table 3.** Detection of plasmid-mediated antimicrobial resistance determinants in the 2016 WHO *Neisseria gonorrhoeae* reference strains

674 (n=14)[41] using Whole Genome Sequence Analysis (WGSA; www.wgsa.net)[21] and an *in house*-customised CLC Genomics Workbench. The

675 discrepant results from the published references are in bold.

			WHO F	WHO G	WHO K	WHO L	WHO M	WHO N	WHO O	WHO P	WHO U	WHO V	WHO W	WHO X	WHO Y	WHO Z
www.wgsa.net	Canu [44] + Nanopolish	bla$_{TEM}$	No	No	No	No	**No**	Yes	**No**	No	No	**No**	No	No	No	No
		tet(M)	No	**No**	No	No	No	**No**	No	No	No	No	No	No	No	No
	Miniasm [45] + Racon [49] + Nanopolish	bla$_{TEM}$	No	**Yes**	No	No	**No**	Yes	**No**	No	No	**No**	No	No	No	No
		tet(M)	No	**No**	No	No	No	**No**	No	No	No	No	No	No	No	No
	SMARTdenovo	bla$_{TEM}$	No	**Yes**	No	No	**No**	Yes	**No**	No	No	**No**	No	No	No	No
		tet(M)	No	**No**	No	No	No	**No**	No	No	No	No	No	No	No	No
	MaSuRCA [48]	bla$_{TEM}$	No	**Yes**	No	No	**No**	Yes	**No**	No	No	**No**	No	No	No	No
		tet(M)	No	**No**	No	No	No	**No**	No	No	No	No	No	No	No	No
	In house-customised CLC Genomics Workbench	bla$_{TEM}$	No	No	No	No	Yes	Yes	Yes	No	No	Yes	No	No	No	No
		tet(M)	No	Yes	No	No	No	Yes	No	No	No	No	No	No	No	No
	Published reference [41]	bla$_{TEM}$	No	No	No	No	Yes	Yes	Yes	No	No	Yes	No	No	No	No

		tet(M)	No	Yes	No	No	No	Yes	No	No	No	No	No	No	No	No

676   WT, Wild type.

677

678

679

680

681

682

683

684

685

Table 4. Minimum inhibitory concentrations (MIC, mg/L) of seven antimicrobials in 14 clinical *Neisseria gonorrhoeae* isolates and associated antimicrobial resistance determinants characterised using an *in house*-customised CLC Genomics Workbench with MinION (Oxford Nanopore Technology) reads.

		1 (U)	2 (U)	3 (R)	4 (R)	5 (P)	6 (U)	7 (P)	8 (P)	9 (U)	10 (U)	11 (C)	12 (R)	13 (R)	14 (U)
MIC interpretation (S/I/R)[a]	Ceftriaxone (MIC, mg/L)	S (0.032)	S (0.064)	S (0.016)	S (0.002)	S (0.064)	S (0.064)	S (0.064)	S (0.064)	S (0.032)	S (0.004)	S (0.032)	S (0.032)	S (0.064)	S (0.064)
	Cefixime (MIC, mg/L)	R (0.25)	S (0.064)	S (0.064)	S (<0.016)	S (0.064)	S (0.032)	S (0.032)	S (0.064)	S (0.016)	S (<0.016)	S (0.125)	S (0.125)	S (0.064)	S (0.064)
	Azithromycin (MIC, mg/L)	S (0.125)	S (0.25)	I (0.5)	I (0.5)	S (0.25)	S (0.5)	S (0.25)	S (0.25)	S (0.5)	R (4)	R (1)	R (1)	I (0.5)	I (0.5)
	Ciprofloxacin (MIC, mg/L)	R (>32)	R (4)	R (>32)	R (0.125)	R (16)	R (32)	R (16)	R (>32)	R (>32)	S (0.004)	R (>32)	R (>32)	R (>32)	R (>32)
	Spectinomycin (MIC, mg/L)	S (16)	S (8)	S (8)	S (8)	S (16)	S (16)	S (16)	S (8)	S (16)	S (16)	S (16)	S (16)	S (16)	S (16)
	Penicillin (MIC, mg/L)	I (1)	R (2)	R (2)	I (1)	R (4)	R (4)	R (2)	R (2)	R (2)	I (0.5)	R (2)	R (2)	R (2)	R (2)
	Tetracycline (MIC, mg/L)	I (1)	I (1)	R (2)	S (0.25)	R (2)	R (2)	I (1)	S (0.5)	S (0.25)	S (0.25)	I (1)	I (1)	S (0.5)	S (0.5)
Antimicrobial resistance determinants extracted using an *in house*-customised CLC Genomics Workbench	*penA* (Mosaic/A501V/T)	Mosaic	A501V	Mosaic	WT	A501V	A501V	A501V	A501T	A501T	WT	Mosaic	Mosaic	A501T	A501T
	mtrR (A-deletion/G45D)	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
	penB variant	Yes	Yes	Yes	WT	Yes	Yes	Yes	Yes	Yes	WT	Yes	Yes	Yes	Yes
	23S rDNA (A2059, C2611)	WT	WT	WT	WT	WT	WT	WT	WT	WT	C2611T	WT	WT	WT	WT
	16S rDNA (C1194T)	No	No	No	No	No	No	No	No	No	No	No	No	No	No
	gyrA (S91F)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
	parC (S86, S87, E91)	S87R	D86N	S87R	WT	D86N	D86N	D86N	E91G	E91G	WT	S87R	S87R	E91G	E91G
	ponA (L421P)	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
	penA (D345a)	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes
	rpsJ (V57M)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	bla$_{TEM}$ gene (plasmid)	No	No	No	No	No	No	No	No	No	No	No	No	No	No
	tet(M) (plasmid)	No	No	No	No	No	No	No	No	No	No	No	No	No	No

[a] Breakpoints according to the European Committee on Antimicrobial Susceptibility testing (www.eucast.org) were used.

37