

1 Target enrichment improves culture-independent 2 detection of *Neisseria gonorrhoeae* direct from sample 3 with Nanopore sequencing.

4 5 1. Author names

6 Teresa L. Street^{1,2*} (ORCID: 0000-0001-8491-240X), Nicholas D. Sanderson^{1,2*} (ORCID: 0000-
7 0001-6370-1961), Leanne Barker¹, James Kavanagh¹, Kevin Cole,³ The GonFast Investigators
8 Group, Martin Llewelyn³, David W. Eyre^{2,4} (ORCID: 0000-0001-5095-6367)
9

10 2. Affiliation(s)

11 ¹Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK;

12 ²National Institute for Health Research Oxford Biomedical Research Centre, John Radcliffe
13 Hospital, Oxford, UK;

14 ³Department of Microbiology and Infection, University Hospitals Sussex NHS Trust, Brighton,
15 UK;

16 ⁴Big Data Institute, Nuffield Department of Population Health, University of Oxford, UK.

17 *These authors contributed equally to this manuscript. Author order was determined by
18 mutual agreement.
19

20 2.1 Corresponding author and email address

21 Teresa L. Street

22 teresa.street@ndm.ox.ac.uk
23

24 2.2 Keywords

25 *Neisseria gonorrhoeae*, target enrichment, culture-independent, Nanopore sequencing,
26 SureSelect, AMR.
27

28 2.3 Repositories

29 European Nucleotide Archive study code PRJEB64347.

30 3. Abstract

31 Multi-drug resistant *Neisseria gonorrhoeae* infection is a significant public health risk.
32 Rapidly detecting *N. gonorrhoeae* and antimicrobial resistant (AMR) determinants by
33 metagenomic sequencing of urine is possible, although high levels of host DNA and
34 overgrowth of contaminating species hamper sequencing and limit *N. gonorrhoeae* genome
35 coverage. We performed Nanopore sequencing of nucleic acid amplification test-positive
36 urine samples and culture-positive urethra swabs with and without probe-based target

37 enrichment, using a custom SureSelect panel to investigate selectively enriching for *N.*
38 *gonorrhoeae* DNA. Probes were designed to cover the entire *N. gonorrhoeae* genome, with
39 10-fold enrichment of probes covering selected AMR determinants. Multiplexing was tested
40 in subset of samples. The proportion of sequence bases classified as *N. gonorrhoeae*
41 increased in all samples after enrichment, from a median (IQR) of 0.05% (0.01-0.1%) to 76%
42 (42-82%), giving a corresponding median improvement in fold genome coverage of 365-
43 times (112-720). Over 20-fold coverage, required for robust AMR determinant detection,
44 was achieved in 13/15 (87%) samples, compared to 2/15 (13%) without enrichment. The
45 four samples multiplexed together also achieved >20x genome coverage. Coverage of AMR
46 determinants was sufficient to predict resistance where present, and genome coverage also
47 enabled phylogenetic relationships to be reconstructed. Probe-based target enrichment can
48 improve *N. gonorrhoeae* genome coverage when sequencing DNA extracts directly from
49 urine or urethral swabs, allowing for robust detection of AMR determinants. Additionally,
50 multiplexing prior to enrichment provided enough genome coverage for AMR detection and
51 reduces the costs associated with this method.
52

53 **4. Impact statement**

54 *Neisseria gonorrhoeae* infection presents a significant public health risk, with multi-drug
55 resistance present globally. Early detection helps control the spread of antimicrobial
56 resistant strains. Genome sequencing can be used to detect infections in samples collected
57 directly from patients, without the need to grow any microorganisms in a laboratory first,
58 and this has already been demonstrated for gonorrhoea using urine samples. With enough
59 sequence information it is also possible to detect antimicrobial resistance (AMR), allowing
60 both detection of infection and information on treatment choices from the same test. This
61 study assesses a method for enriching *N. gonorrhoeae* sequence data from urine and
62 urethral swabs, and analyses the impact of enrichment on the detection of genes known to
63 cause antibiotic resistance. We show that enriching for *N. gonorrhoeae* DNA prior to
64 sequencing can improve the detection of some AMR genes, and by testing several samples
65 at the same time we can reduce the costs associated with this method.
66

67 **5. Data summary**

68 The sequence data generated in this study are deposited in the European Nucleotide Archive
69 (ENA, <https://www.ebi.ac.uk/ena/browser/>) and are publicly available under study code
70 PRJEB64347. The authors confirm all supporting data, code and protocols are provided
71 within the article or through supplementary data files.
72

73 **6. Introduction**

74 Multi-drug resistant *Neisseria gonorrhoeae* infection presents a significant global public
75 health risk(1), with resistance to ceftriaxone and/or azithromycin detected in countries
76 worldwide(2–4). Additionally, cases of *N. gonorrhoeae* infection are increasing. In 2019,
77 117,881 cases were reported in the EU/EEA, over half of which were reported in the UK with

78 116 cases/100,000 population(5). Early detection and effective treatment is vital to prevent
79 the spread of drug-resistant strains.

80

81 Metagenomic sequencing (mNGS) directly from clinical samples offers the potential to
82 reduce the time to diagnosis and can provide additional information such as detection of
83 antimicrobial resistance(6), particularly useful in settings where culture is not routine. mNGS
84 has previously been used to both identify *N. gonorrhoeae*(7) and detect antimicrobial
85 resistance (AMR) determinants(8) directly from urine samples. A limitation of mNGS is the
86 high levels of contaminating human DNA present in extracts from clinical samples which
87 limits the amount of pathogen sequence data generated(6). Target enrichment is a
88 technique that allows enrichment of genomes or genomic regions of interest prior to
89 sequencing: Biotinylated RNA oligonucleotide probes, designed to target specific sequences,
90 allow capture and enrichment of these regions of interest by hybridisation and magnetic
91 pulldown. Target enrichment has previously been demonstrated to improve the amount of
92 pathogen sequence generated from mNGS(9–11), improving genome coverage and
93 detection of AMR determinants(12).

94

95 This study assesses the ability of probe-based target enrichment to enrich for *N.*
96 *gonorrhoeae* DNA directly from urine and urethral swab samples for the detection of AMR
97 determinants using the Oxford Nanopore Technologies (ONT) sequencing platform. We also
98 tested the efficiency of target enrichment after multiplexing samples together, aiming to
99 reduce the costs associated with both probe-based enrichment and sequencing.

100

101 **7. Methods**

102 **7.1 Samples**

103 Samples were selected from those collected as part of a wider study, conducted with NHS
104 Research Ethics approval (reference 19/EM/0029). Participants were recruited at sexual
105 health clinics at Oxford University Hospitals NHS Trust, UK, and University Hospitals Sussex
106 NHS Trust, Brighton, UK. Male patients presenting with symptomatic urethritis were eligible
107 to participate and were recruited following informed consent. In Brighton, urine samples
108 were collected into universal tubes containing boric acid (Medline Scientific) for stabilization
109 during transportation to Oxford. In Oxford, samples were collected into universal tubes
110 without boric acid. Urethral swabs were placed into Sigma VCM preservation medium
111 (MWE). All samples were tested for *N. gonorrhoeae* using the BD Viper system (Becton
112 Dickinson), with confirmatory testing for *N. gonorrhoeae* undertaken using the BD Max
113 platform (Oxford) or with the FTD gonorrhea confirmation NAAT assay (Fast Track
114 Diagnostics) (Brighton). Antimicrobial susceptibility testing of cultured bacteria from urethral
115 swabs was performed manually according to EUCAST methods by disc diffusion (for nalidixic
116 acid, penicillin, azithromycin, cefuroxime and ciprofloxacin) and E-tests (for cefixime and
117 cefotaxime) in Brighton, and by E-tests (for ciprofloxacin, azithromycin and ceftriaxone) in
118 Oxford.

119

120 **7.2 Custom probe capture library design**

121 A SureSelectXT custom probe library was designed with Agilent, the manufacturer, to cover
122 17 closed genomes of *N. gonorrhoeae* identified by the authors to represent the diversity
123 present across the *N. gonorrhoeae* species (Table S1). Additionally, probes designed to cover
124 genes known to confer antimicrobial resistance in *N. gonorrhoeae* were represented 10-fold
125 in the final library (Table S2). The final probe library consisted of consisting of approximately
126 50,000 120mer RNA probes.

127

128 **7.3 DNA extraction, library preparation and sequencing**

129 DNA was extracted from nucleic acid amplification test (NAAT)-positive urine samples and
130 urethral swabs using the QIAamp UCP pathogen mini kit (Qiagen), as previously
131 described(7). Samples without enrichment were prepared for sequencing on version 9.4.1
132 flow cells following library preparation with the rapid PCR barcoding kit (SQK-RPB004, ONT)
133 also as previously described(7). One sample was run per flow cell on a GridION sequencer.
134 For SureSelect enrichment, samples were prepared following the Sequence Capture protocol
135 (ONT), using the ligation kit (SQK-LSK110, ONT) for sequencing with a single sample per flow
136 cell.

137

138 **7.4 Preparation for multiplexed sequencing**

139 To test the feasibility of multiplexing, 4 samples were barcoded with a modified protocol
140 using the rapid PCR barcoding kit (SQK-RPB004, ONT). Briefly, 200ul DNA extract was
141 fragmented by tagmentation in a 10 µl reaction with 2.5 µl FRM, then barcoded by PCR in a
142 double volume reaction for 17 cycles with a 5-minute elongation time at 65°C. Post-PCR,
143 barcoded DNA was processed following a modified Sequence Capture protocol, as follows.
144 DNA was end-repaired and PCR adapters ligated as per the original protocol. The initial PCR
145 was omitted, and instead equimolar amounts of each barcoded, amplified sample were
146 pooled together prior to hybridisation with the custom probes. Post-hybridisation, the
147 Sequence Capture protocol was followed as described by ONT and enriched DNA was
148 prepared for sequencing using the ligation kit, as above.

149

150 **7.5 Illumina sequencing of cultured isolates**

151 Cultures from those samples with a viable stock were prepared for sequencing on an
152 Illumina MiniSeq sequencer. *N. gonorrhoeae* DNA was extracted from cultured cells grown
153 overnight on chocolate agar at 37°C with 5% CO₂ using the QuickGene DNA tissue kit
154 (Kurabo) on the QuickGene-Mini80 system (MP Biomedicals), and prepared for sequencing
155 with the Nextera XT DNA library prep kit (Illumina).

156

157 **7.6 Bioinformatic analysis**

158 Nanopore sequences were base called and demultiplexed, where necessary, using Guppy
159 (ONT, Version 5.0 or higher) automatically on the GridION sequencer. Sequencing reads
160 classified as human were removed at source with the CRuMPIT workflow, as previously
161 described(13). Sequences were mapped to the NC_011035.1 *N. gonorrhoeae* reference
162 genome using minimap2 (v2.24-r1122) and variant called with clair2 (v2.1.1). Variants were

163 filtered with a random forest classifier, and antimicrobial resistance genes were
164 characterised from consensus genome sequences or assembled plasmid sequences, both as
165 previously described(8). The Nextflow workflow is available at
166 <https://gitlab.com/ModernisingMedicalMicrobiology/genericbugontworkflow>.

167
168 Illumina sequences from cultured isolates were mapped to the NC_011035.1 reference
169 genome with SNIPPY(14) v4.6.0 to generate SNP-only consensus sequences. Positions with
170 depth lower than 10x coverage were masked as Ns. Hence, both Nanopore and Illumina data
171 were mapped to the same reference genome and consensus sequences generated from ONT
172 or Illumina-specific variant calling pipelines. Both Illumina and ONT consensus genomes
173 were masked for repeat regions using a self-self blast, with a minimum window of 200bp and
174 90% identity score.

175
176 A phylogenetic tree was constructed using IQ-TREE v2.1.4-beta(15) and adjusted for
177 recombination events using clonalframeML v1.12(16) using the runlistcompare wrapper
178 script v0.3.8 (<https://github.com/davideyre/runListCompare>) for samples with sufficient
179 genome coverage.

180

181 **8. Results**

182 **8.1 Samples summary**

183 Twelve urine samples and four urethral swabs were selected for use in this study. Eleven
184 (11/12, 92%) urines were NAAT-positive for *N. gonorrhoeae* and one (1/12, 8%; sample
185 382UB) was NAAT-negative with a corresponding culture-positive urethral swab. All four
186 (100%) urethral swabs were culture-positive for *N. gonorrhoeae*, with corresponding urines
187 also NAAT-positive. A sample summary is provided in Figure 1 and Figure S1.

188

189 A total of 9/11 (82%) NAAT-positive urines had an additional corresponding culture-positive
190 urethral swab, with 2/11 (18%) being culture-negative. Antimicrobial susceptibility testing
191 results on cultured isolates were used for comparisons with mNGS-based predictions.
192 Results were available for 14 urethral swabs (13 tested in Brighton and 1 in Oxford; 9 with
193 NAAT-positive urines that were sequenced, 1 with a NAAT-negative urine that was
194 sequenced, and 4 where the urethral swab was sequenced). All were sensitive to both
195 cephalosporins and azithromycin (Table 1). For quinolone susceptibility, the single sample
196 from Oxford (80U) was sensitive to ciprofloxacin, and of the Brighton samples 3/13 (23%)
197 were sensitive and 10/13 (77%) resistant to nalidixic acid (Table 1). Brighton additionally
198 tested for susceptibility to penicillin (sensitive in 2/13 (15%)) and for the presence of beta-
199 lactamase (detected in 9/13, 69%).

200 **8.2 Effect of enrichment on *Neisseria gonorrhoeae* genome coverage**

201 Probe enrichment increased the proportion of *N. gonorrhoeae* sequencing reads compared
202 to no enrichment and did not affect the total sequencing output of the runs (Figure 2). A
203 comparison of total bases, bases classified as *N. gonorrhoeae* plus genome coverage breadth
204 and depth before and after enrichment is provided in Table 2. The total number of bases per
205 sample was similar between enriched (median 10.9Gb, IQR 8.1-17.3) and without
206 enrichment (median 9.4Gb, IQR 8.4-10.8). Multiplexing samples onto a single flow cell

207 reduced the number of bases per sample, as expected (median 478Mb, IQR 321-918) (Figure
208 2a). The median proportion of bases classified as *N. gonorrhoeae* increased from less than
209 0.5% of total sample bases without enrichment (median 0.05%, IQR 0.01-0.1) to more than
210 75% with enrichment (median 76%, IQR 42-82). Probe enrichment with multiplexing
211 generated a lower median proportion of *N. gonorrhoeae* bases (33%, IQR 26-45) (Figure 2b).
212 Average coverage depth of aligned reads increased from a median 5-fold (IQR 4.4-6.7)
213 without enrichment to 2405-fold after enrichment (IQR 492-5423) and 63-fold (IQR 36-153)
214 after multiplexed enrichment (Figure 2c). This resulted in all but two samples enriched with
215 SureSelect probes achieving >96.5% of their genome sequenced at a depth of 300-fold or
216 more (Table 2). Comparison of genome coverage with and without enrichment demonstrates
217 no evidence that any bias was introduced by the Sequence Capture protocol (Figure 3).

218
219 Enrichment failed for two samples (265UB, 321UB). Genome coverage breadth and depth
220 were similar before and after enrichment: 59% breadth and 3.8-fold depth before and 73%
221 breadth and 5-fold depth after for sample 265UB, and 58% breadth and 4.4-fold depth
222 before and 53% breadth and 4-fold depth after for sample 321UB. Due to the high cost of
223 reagents and the need to order these in batches, we were unable to repeat these
224 experiments. A detailed summary of all sequence data generated is provided in Table S3.
225

226 **8.3 Detection of antimicrobial resistance determinants**

227 Our previous study determined that a minimum 20-fold genome coverage depth is required
228 for robust AMR prediction(8). In this study 13/15 (87%) of samples achieved this after
229 enrichment, with a minimum coverage depth of 314-fold (Table S4). Sequence data was
230 interrogated for variants within genes known to confer resistance to macrolides, quinolones,
231 cephalosporins, and for the presence of plasmid-mediated genes conferring resistance to
232 penicillin and tetracycline (Table S4). None of the samples in this study were phenotypically
233 resistant to cephalosporins or azithromycin, and no genetic determinants known to confer
234 resistance to these were detected, despite >140-fold coverage of *mtrR* and >90-fold
235 coverage of *penA*, *ponA* and *porB* in those samples with successful enrichment (Figure 4). Ten
236 (10/16, 63%) samples were reported as resistant to nalidixic acid. Mutations in *gyrA* (S91F,
237 D95G) were detected in two samples before enrichment. After enrichment, *gyrA* was
238 sequenced to a depth of at least 115-fold and mutations were detected in 9/9 (100%) of
239 those samples with singleplex enrichment: singleplex enrichment sequence data was not
240 available for one sample (347UB). Additionally, *gyrA* mutations S91F and D95G were found
241 in sample 361UB after enrichment, which did not grow in culture and therefore had no
242 susceptibility data reported by the microbiology laboratory.
243

244 For plasmid-mediated resistance, no sequence data containing *ermB*, *ermC*, *ermF* or *mef* was
245 detected, corresponding with laboratory results of susceptibility to azithromycin. Beta-
246 lactamase was detected in 9/16 (56%) samples by the laboratory, and sequencing without
247 enrichment detected the presence of *blaTEM-1* in 6 of these samples, and additionally in
248 sample 361UB which did not have susceptibility data reported After enrichment, *blaTEM-1*
249 was only detected in 2 samples using the criteria defined previously(8) , and in 3 samples
250 when considering coverage breadths lower than the predefined cutoff threshold. Both
251 samples detected above the threshold were positive by laboratory testing, and one of which
252 was an additional detection over no enrichment. Although the samples were not tested for

253 susceptibility to tetracycline, *tetM* was detected in 3 samples before enrichment and five
254 samples after enrichment (4 above and one below the cutoff threshold).

255
256 Despite 10-fold representation in final probe library, the AMR genes assessed in this study
257 did not demonstrate a higher coverage depth than the rest of the genome (Figure S2). We
258 also observed uneven enrichment across the target AMR genes, with poor improvements in
259 coverage of the 23S gene after enrichment (median (IQR) coverage depth before enrichment
260 of 1.6 (1.0-3.4) and after enrichment of 6.7 (3.6-12.5)).

261

262 **8.4 Evaluation of multiplexing prior to enrichment**

263 Four samples (80U, 303D, 347UB, 367UB) were multiplexed before hybridisation with the
264 SureSelect probes. The multiplexed run generated 8.2 gigabases (Gb) of sequence data.
265 Once bases from misclassified barcodes and unclassified bases were removed, 1.7Gb were
266 classified as bacterial and, despite enrichment, 1.4Gb classified as human. The median
267 number of bacterial bases classified as *N. gonorrhoeae* was 111 megabases (Mb) (IQR 55-
268 305), with a median *N. gonorrhoeae* single-fold genome coverage of 77% (IQR 67-85) (Table
269 3). In comparison to no enrichment, multiplexed enrichment achieved a median 12-fold (IQR
270 4.8-33) improvement in genome coverage depth, which is less than the median 365-fold
271 (IQR 112-720) improvement seen after singleplex enrichment (Table 2). Despite hybridisation
272 of equimolar amounts of barcoded DNA with the SureSelect probes there was uneven
273 coverage across the 4 samples, with >700Mb separating the highest and lowest number of
274 bases classified as *N. gonorrhoeae*.

275

276 All 4 samples achieved genome coverage depth of greater than 20-fold (Table 3), which is the
277 minimum depth required for robust detection of AMR determinants(8). Three of the four
278 were resistant to a quinolone (nalidixic acid; 303D, 347UB, 367UB) and one was sensitive
279 (ciprofloxacin, 80U). Depth was insufficient to resolve the sequence of the *gyrA* or *parC*
280 genes in any sample without enrichment, and after multiplexed enrichment adequate depth
281 was achieved to detect mutations in *gyrA* (S91F) in 2/3 resistant samples (303D and 367UB).
282 Sequencing depth was inadequate to determine resistance in sample 347U, with only 7-fold
283 coverage depth of *gyrA*. *blaTEM-1* was detected in samples 303D and 367UB without
284 enrichment, corresponding with the presence of beta-lactamase detected by the laboratory,
285 but was not detected in sample 347UB despite detection by the lab. After multiplexed
286 enrichment, *blaTEM-1* was detected below the cutoff threshold in 1/3 samples reported
287 positive by the laboratory, and in an additional sample without laboratory results (80U).

288

289 **8.5 Assessment of relatedness and comparison of Nanopore with Illumina sequence** 290 **data**

291 Comparison of Nanopore (from urine and swab sequencing) and Illumina genomes (from
292 isolate sequencing) in a phylogenetic tree showed concordance between the sequencing
293 methods, demonstrating the ability to infer relatedness: 13/15 (87%) samples had
294 singleplexed enrichment Nanopore data and 11/15 had Illumina data for analysis, giving a
295 total of 11 samples with paired Nanopore and Illumina data (Figure 5). The median (IQR)
296 [range] number of SNPs detected between Nanopore and Illumina sequence pairs generated
297 from matched urine/swab samples and cultured isolates from the same patient was 44 (38-

298 52) [24-245]. This likely represents a combination of sequencing artefact and true
299 differences between the samples.

300
301 Samples genetically further away from the reference genome used showed greater
302 differences between Illumina and Nanopore results, with samples 303D, 342UB and 358UB
303 forming sequencing technology-specific clades instead of grouping by patient, as seen in all
304 the other samples.

305
306 Two samples without enrichment (202D, 301D) generated enough genome coverage for
307 inclusion in this analysis and were almost identical to their corresponding SureSelect-
308 enriched pair (2 and 0 SNPs different, respectively). Only one multiplexed sample, 303D,
309 generated enough data to reconstruct the genome for this analysis.

310

311 9. Discussion

312 This study demonstrates the utility of target enrichment for the culture-free detection of
313 *Neisseria gonorrhoeae* directly from urine and urethral swabs by sequencing. We also show,
314 for the first time with Nanopore sequencing, that target enrichment can be achieved after
315 multiplexing of samples prior to enrichment.

316

317 Where enrichment was successful it was possible to achieve >96.5% genome coverage
318 breadth at a depth of 300-fold or greater, representing a minimum 40-fold improvement in
319 coverage depth. Multiplexing with enrichment still provided >20-fold coverage depth in all
320 samples tested. The depth and breadth of *N. gonorrhoeae* genome coverage achieved after
321 enrichment enabled detection of *gyrA* mutations known to confer resistance to quinolones
322 in all samples with enriched sequence data reported as resistant to nalidixic acid, and
323 additionally detected mutations in a sample with no susceptibility data due to failed growth
324 on culture. Targeted enrichment allowed sequencing at high depth of other chromosomal
325 genes known to confer antimicrobial resistance, e.g., the *mtrR* gene where a mutation at
326 amino acid position 45 confers both multi-drug and macrolide resistance. This offers
327 confidence that resistance should be detected if it were present.

328

329 mNGS holds promise for use in routine clinical diagnostics(17) but is hampered by high levels
330 of host DNA, which dominates when nucleic acid is extracted directly from a sample without
331 initial culture, and subsequently overwhelms the sequence data such that information on
332 pathogens is often very limited (for example(18)). Target enrichment enables selection of the
333 genome of interest over contaminating host DNA, and is useful where the target pathogen is
334 known in advance of the diagnostic test. Prevalence of multi-drug resistant *N. gonorrhoeae*
335 and the need for a fast diagnostic that can detect both the pathogen and any corresponding
336 antimicrobial resistance makes targeted sequencing an appealing prospect.

337

338 Susceptibility testing relies on culture, but *N. gonorrhoeae* can sometimes be difficult to
339 grow in the laboratory. This study demonstrates the additional value of sequencing with
340 enrichment, detecting *gyrA* mutations with a mean coverage depth of 116-fold in a sample
341 that did not grow in culture.

342

343 Detection of plasmid-mediated resistance genes was less successful, with enrichment
344 missing detection of beta-lactamase known to be present by laboratory testing in the
345 majority of samples, and despite detection of *bla*TEM-1 in the sequence data of six samples
346 before enrichment. Inefficient amplification of plasmid DNA pre-hybridisation, or the short
347 SureSelect probes not effectively pulling out the plasmid-bound gene during hybridisation,
348 could be possible explanations for the unsuccessful detection of *bla*TEM-1 but further work
349 is required to determine the exact reason.

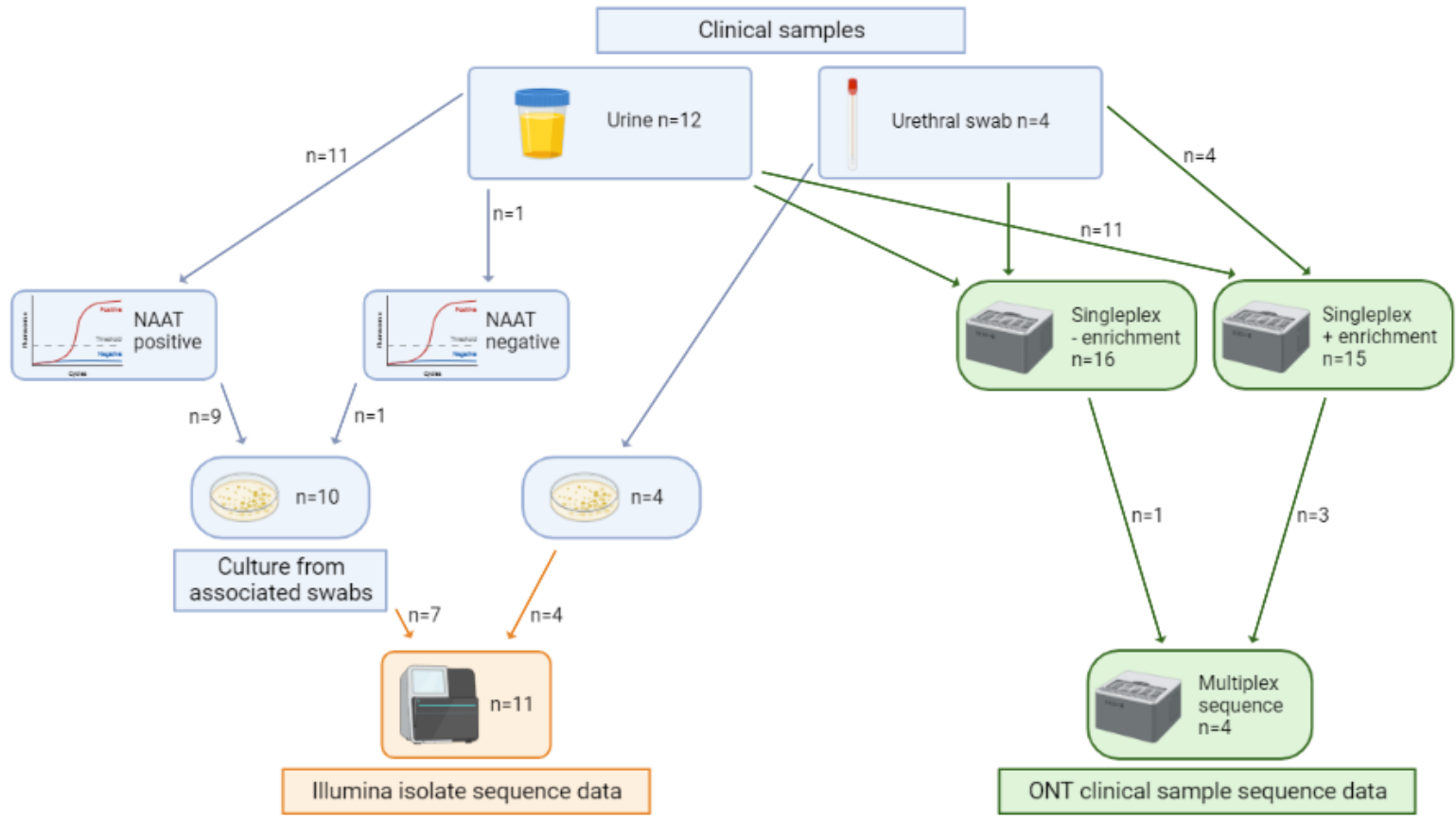
350
351 Two samples failed enrichment. Analysis of their post-enrichment sequence data identified
352 predominantly laboratory kit-based contaminants in addition to low numbers of reads
353 mapping to *N. gonorrhoeae*. We propose that this is due to the very low amount of *N.*
354 *gonorrhoeae* DNA in the original urine extract for these two samples, as has been described
355 by others studying low bacterial load samples(19,20).

356
357 Multiplexing allowed a median 12-fold improvement in genome coverage when four
358 samples were barcoded and pooled prior to hybridisation and detection of relevant AMR
359 determinants. Demultiplexing samples after sequencing leads to loss of data as some reads
360 remain unclassified due to a barcoding score lower than the default requirement in Guppy.
361 The larger than expected number of unclassified reads generated from the multiplexed
362 samples could be a consequence of the combination of barcoding and ligation library
363 preparation used to prepare these samples, but further work would be required to
364 determine this. The number of unclassified reads could potentially be reduced by changing
365 the minimum barcoding score requirements, although this carries the risk of increased false-
366 positive calls.

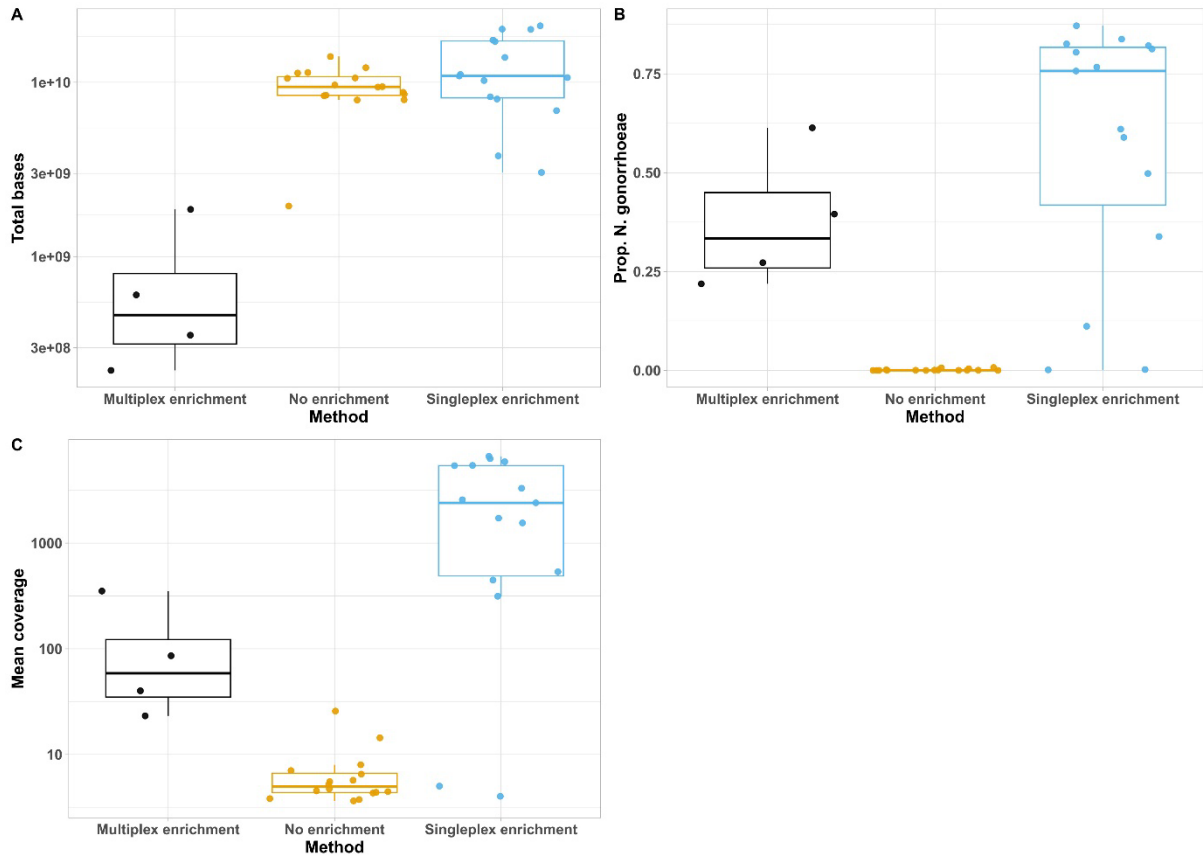
367
368 Other limitations of this study include the large starting amount of DNA required for this
369 hybridisation method. Often, DNA extracts directly from clinical samples fail to generate the
370 required 3.5µg required for the ONT Sequence Capture protocol. Extracts from urine and
371 urethral swabs do, in general, achieve the required amount of DNA. However, samples from
372 sterile sites without much host DNA contamination, for example CSF samples, or from
373 samples with a limited volume or size, would not be good candidates for probe-based
374 enrichment with Nanopore sequencing. The method is performed over two days, with a 16-
375 hour overnight hybridisation, which makes it a lengthy assay. Although there could
376 potentially be some optimisation of the protocol, it currently cannot compete with the time
377 it takes to perform a NAAT or other molecular assay, but is still likely faster than culture.
378 There are also limitations arising from artefacts introduced by ONT sequencing, as seen for
379 some samples in Figure 5. However, this study was performed with version 9.4.1 flowcells,
380 and the more recent generation of flowcells, versions 10 onwards, may address this(21). In
381 our data we observed uneven enrichment across target AMR genes, with the 23S gene in
382 particular under-represented in enriched sequence data. Future work could optimise probes
383 in this region of the genome to increase coverage of this important gene.

384
385 The nucleic acid amplification tests are a reliable method for detecting *N. gonorrhoeae* but
386 cannot give the whole picture of AMR, an important consideration given the increasing
387 prevalence of multi-drug resistant gonorrhoea. This study demonstrates sequencing with
388 target enrichment can enable high-depth chromosomal antimicrobial resistance gene

389 characterisation and relatedness detection that was not previously possible for these sample
390 types with direct metagenomic sequencing. Additionally, multiplexing prior to enrichment
391 allows higher sample throughput and reduction in the costs associated with both probe-
392 based enrichment and sequencing.



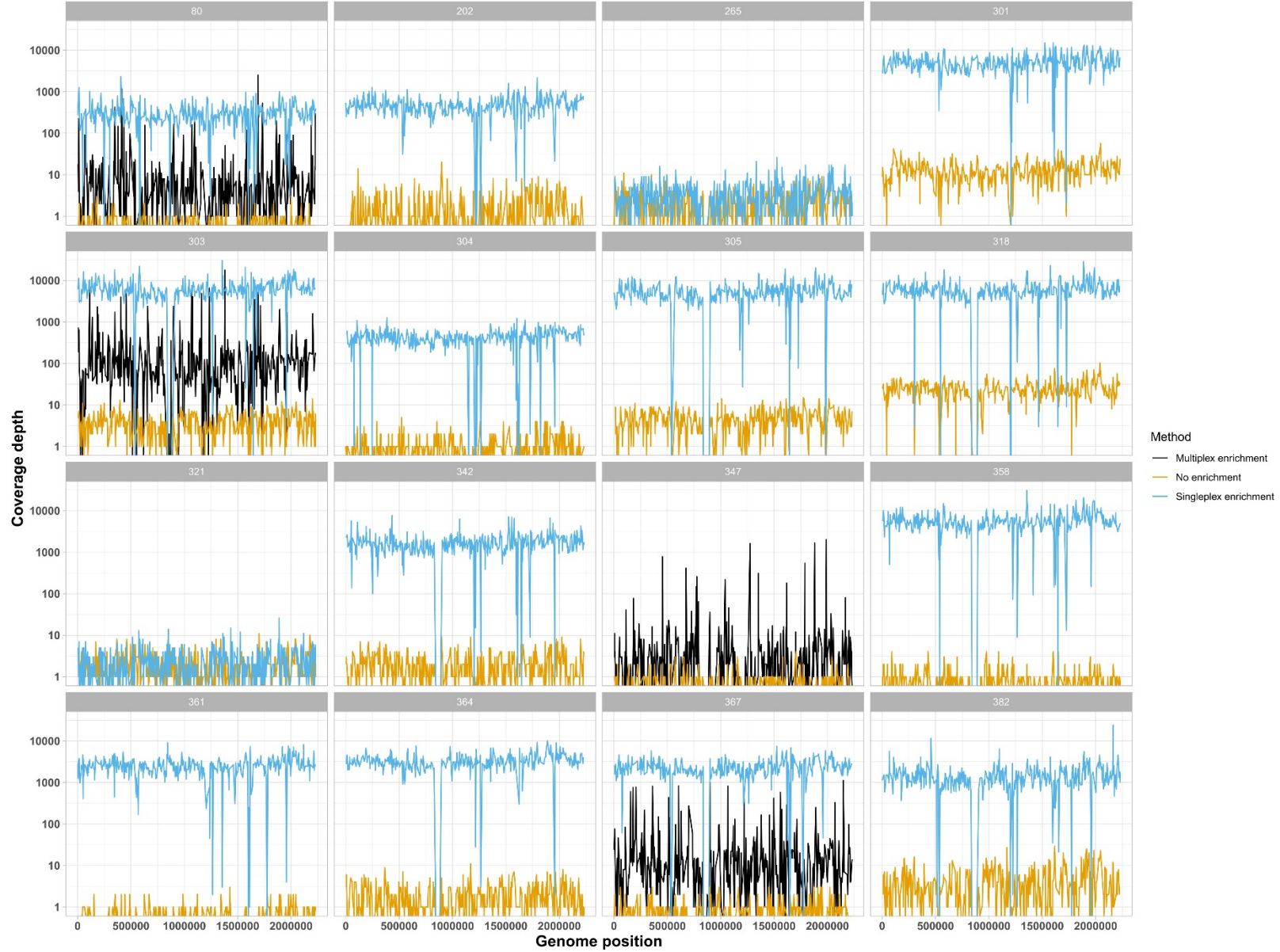
394
395 Figure 1. Sample summary. Figure created with BioRender



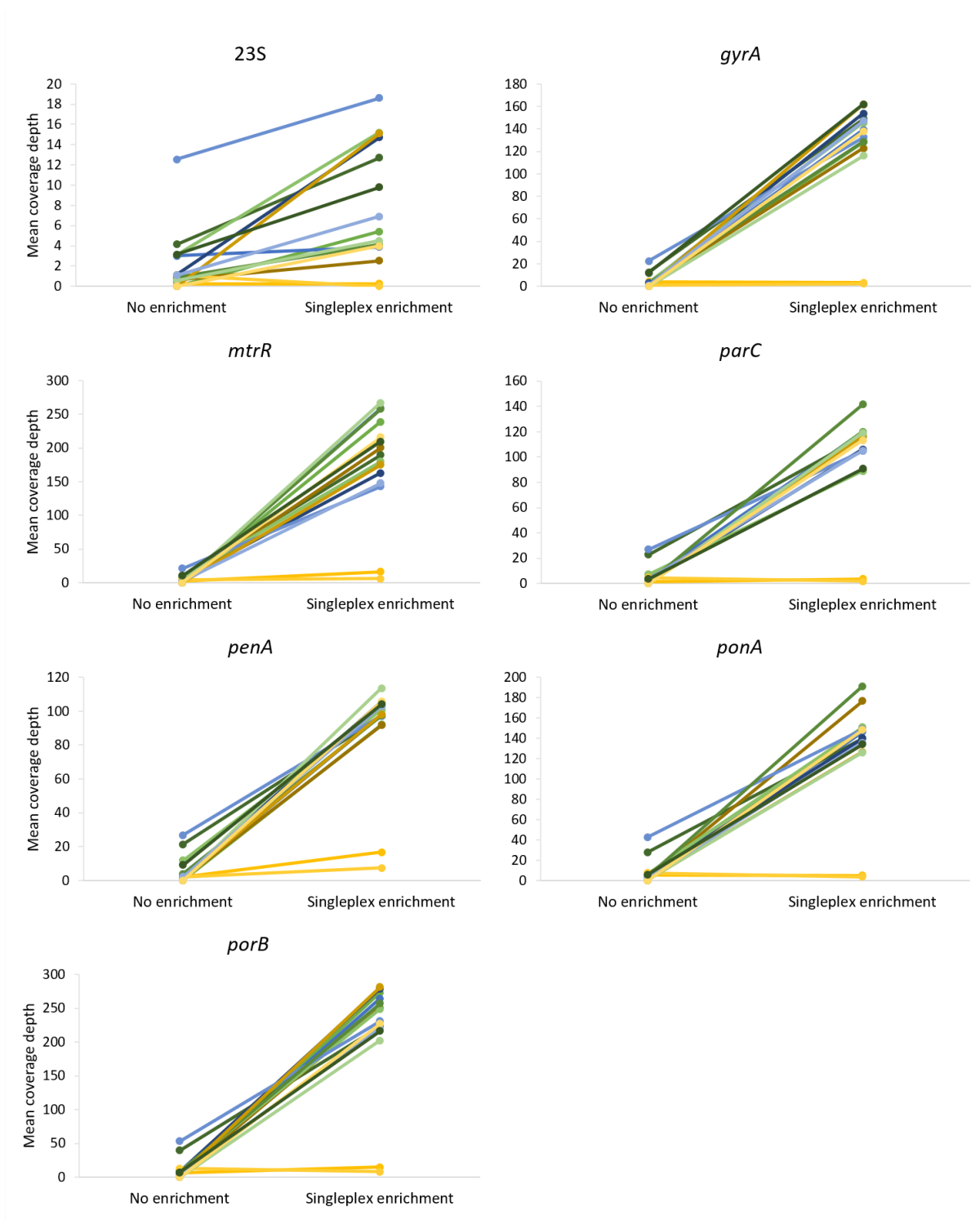
396
397 **Figure 2. Comparison of sequence data before and after enrichment. Total number of bases per**
398 **sequencing run (A); Percentage of bases classified as *N. gonorrhoeae* (B); Average coverage depth**
399 **of *N. gonorrhoeae* genome per sequencing run (C) for samples sequenced without enrichment**
400 **(orange), with singleplex enrichment (blue) or with enrichment in a multiplex of 4 samples (black).**

401

402

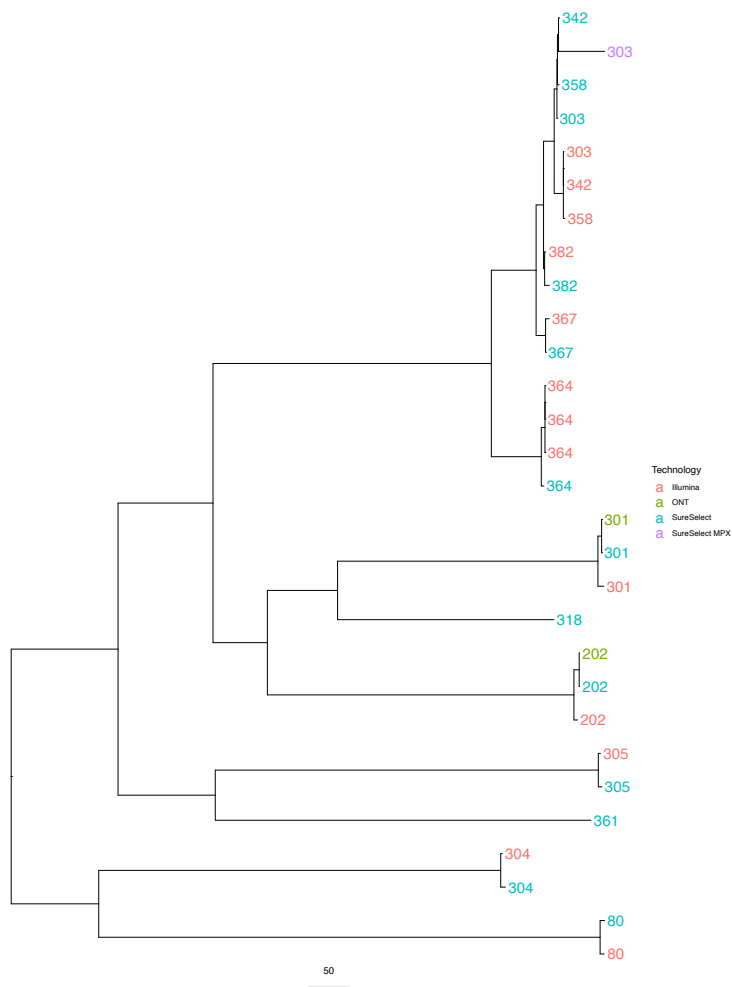


404 **Figure 3. Genome coverage depth for samples sequenced without enrichment (orange), with singleplex enrichment (blue) or with enrichment in a**
405 **multiplex of 4 samples (black). Sample 347 has sequence data without enrichment and with multiplexed enrichment only. Short regions of low coverage**
406 **seen in blue represent areas of divergence between the sample and the reference genome used for mapping and/or the genomes included in the probe**
407 **set. The similar troughs seen with the orange and blue lines suggest that divergence from the reference genome is the most likely explanation, rather**
408 **than a gap in the probe set.**



409
410
411
412

Figure 4. Comparison of mean coverage depth for 7 antimicrobial resistance determinants before and after target enrichment. Different colours are used for each sample pair with and without enrichment.



413
414 **Figure 5. Maximum likelihood tree of genomes generated from consensus sequences, comparing**
415 **the same samples sequenced with Illumina (red), Nanopore with no enrichment (green), Nanopore**
416 **with singleplex enrichment (blue) and Nanopore with multiplexed enrichment (purple). Sample**
417 **364 has three Illumina sequences generated from two rectal swab cultures (not included in**
418 **enrichment and singleplex sequencing) and one urethral swab culture.**

419 **11. Tables**

Sample	Sample type for culture	Sample type for sequencing	Ciprofloxacin/ Nalidixic acid	Cephalosporins	Azithromycin	Penicillin	B-lactamase*
80U	Urethral swab	Urine	S	S	S	no test	no test
202D	Urethral swab	Urethral swab	S	S	S	S	N
265UB	Urethral swab	Urine	R	S	S	R	P
301D	Urethral swab	Urethral swab	R	S	S	I	N
303D	Urethral swab	Urethral swab	R	S	S	R	P
304D	Urethral swab	Urethral swab	S	S	S	I	N
305UB	Urethral swab	Urine	R	S	S	R	P
318UB	No growth on culture	Urine					
321UB	Urethral swab	Urine	S	S	S	S	N
342UB	Urethral swab	Urine	R	S	S	R	P
347UB	Urethral swab	Urine	R	S	S	R	P
358UB	Urethral swab	Urine	R	S	S	R	P
361UB	No growth on culture	Urine					
364UB	Urethral swab	Urine	R	S	S	R	P
367UB	Urethral swab	Urine	R	S	S	R	P
382UB	Urethral swab	Urine	R	S	S	R	P

420
421 **Table 1. Antimicrobial susceptibility results. S, sensitive; I, intermediate; R, resistant; *for B-lactamase,**
422 **N denotes susceptible; P denotes possible resistance.**

Sample name	Without enrichment				With singleplex enrichment				Coverage depth fold-enrichment, singleplex vs no enrichment
	Ngon bases	Total bases	Genome coverage breadth (%)	Average coverage depth	Ngon bases	Total bases	Genome coverage breadth (%)	Average coverage depth	
80U	749179	8434029003	3.46	7.98	1280208433	3782522608	97.44	314.12	39.3
202D	4184726	9656053514	35.73	4.45	1550689332	13895666698	98.53	536.27	120.6
265UB	5949401	11386449183	58.55	3.82	17922316	8014502433	73.45	5.01	1.31
301D	35361341	7933536503	98.83	14.36	14947146476	17150934403	99.80	5413.55	377.0
303D	9692745	9439958602	82.96	4.73	17664552591	21087141427	97.08	6637.14	1403.8
304D	1937846	7906566477	17.88	4.32	1790026178	3038176523	97.49	447.24	103.5
305UB	12659589	9391830588	86.68	5.53	14039629756	17449531128	96.75	5432.74	983.0
318UB	62722932	8544860509	96.36	25.72	16557094125	20163378663	97.03	6295.03	244.8
321UB	7745538	12127667362	57.53	4.38	9600291	6880830294	53.10	4.02	0.92
342UB	4929172	14056225894	51.44	3.75	6489125469	10636014129	96.70	1724.94	460.4
358UB	1399857	11292620134	10.46	5.22	15428552493	20118811661	97.09	5905.35	1131.6
361UB	698401	8371283598	3.78	7.05	7748292558	10234243279	99.04	2576.51	365.4
364UB	4923268	8765729938	51.80	3.64	8974767733	10869934865	96.84	3304.73	908.2
367UB	1474533	10579314587	12.82	4.52	6713863610	8261193715	96.63	2405.44	532.2
382UB	12826754	1949351698	71.43	6.52	5518490118	11085690959	96.73	1552.06	238.2

423

424

425

426

Table 2. Summary of bases classified as *N. gonorrhoeae* (Ngon), total bases, single-fold genome coverage breadth, and average genome coverage depth for all samples before and after singleplex enrichment.

Sample	Without enrichment			With singleplex enrichment			With multiplex enrichment			Coverage depth fold-enrichment, multiplex vs no enrichment
	Total Ngon bases	Genome coverage breadth (%)	Average coverage depth	Total Ngon bases	Genome coverage breadth (%)	Average coverage depth	Total Ngon bases	Genome coverage breadth (%)	Average coverage depth	
80U	749179	3.46	7.98	1280208433	97.44	314.12	96384278	71.91	40.05	5.02
303D	9692745	82.96	4.73	17664552591	97.08	6637.14	1144125623	94.61	352.47	74.5
347UB	1066606	7.23	5.69	No singleplex data available			47875172	52.70	48688685	4.07
367UB	1474533	12.82	4.52	6713863610	96.63	2405.44	238397019	82.44	85.96	19.0

427

428

429

Table 3. Comparison of no enrichment, singleplex enrichment and multiplex enrichment. Total bases classified as *N. gonorrhoeae* (Ngon), single-fold genome coverage breadth, and genome coverage depth reported for the 4 samples sequenced as a multiplex.

430 **12. Author statements**

431 **12.1 Author contributions**

432 TLS, NDS, ML and DWE designed the study. NDS conducted the bioinformatics analysis. TLS,
433 LB and JK performed laboratory experiments. KC organised clinical samples, performed
434 culturing and collated metadata for samples collected in Brighton. TLS, NDS and DWE
435 prepared the manuscript. The GonFast Investigators Group recruited patients and collected
436 samples. All authors read and contributed to the manuscript.
437

438 **12.2 Conflicts of interest**

439 The authors declare no conflicts of interest.
440

441 **12.3 Funding information**

442 This study was funded by the Centers for Disease Control and Prevention, USA.
443 This study was also supported by the National Institute for Health Research (NIHR) Oxford
444 Biomedical Research Centre (BRC). The funders had no role in study design, data collection
445 and interpretation, or the decision to submit the work for publication. The views expressed
446 are those of the authors and not necessarily those of the NHS, NIHR or Department of
447 Health. DWE is a Robertson Foundation Big Data Institute Fellow.
448

449 **12.4 Ethical approval**

450 This study was conducted with NHS Research Ethics committee approval (reference
451 19/EM/0029).
452

453 **12.5 Acknowledgements**

454 We thank the microbiology laboratory staff of Oxford University Hospitals NHS Foundation
455 Trust and the University Hospitals Sussex NHS Foundation Trust for providing assistance with
456 sample collection.

457 The GonFast Investigators Group includes Joanna Rees and Emily Lord (Oxfordshire Sexual
458 Health Service, Oxford University Hospitals NHS Foundation Trust, Oxford); and Suneeta
459 Soni, Celia Richardson, Joanne Jessop, and Tanya Adams (Brighton and Hove sexual health
460 and contraception service, Royal Sussex County Hospital, Brighton).
461

462 **13. References**

- 463 1. Unemo M, Shafer WM. Antimicrobial Resistance in *Neisseria gonorrhoeae* in the 21st
464 Century: Past, Evolution, and Future. *Clin Microbiol Rev.* 2014 Jul;27(3):587–613.
- 465 2. Eyre DW, Town K, Street T, Barker L, Sanderson N, Cole MJ, et al. Detection in the
466 United Kingdom of the *Neisseria gonorrhoeae* {FC428} clone, with ceftriaxone resistance
467 and intermediate resistance to azithromycin, October to December 2018. *Euro Surveill.*
468 2019 Mar;24(10).

- 469 3. Golparian D, Vestberg N, Södersten W, Jacobsson S, Ohnishi M, Fang H, et al.
470 Multidrug-resistant *Neisseria gonorrhoeae* isolate SE690: mosaic penA-60.001 gene
471 causing ceftriaxone resistance internationally has spread to the more antimicrobial-
472 susceptible genomic lineage, Sweden, September 2022. *Eurosurveillance* [Internet]. 2023
473 Mar 9 [cited 2023 Apr 27];28(10). Available from:
474 <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2023.28.10.2300125>
- 475 4. Sawatzky P, Demczuk W, Lefebvre B, Allen V, Diggle M, Hoang L, et al. Increasing
476 Azithromycin Resistance in *Neisseria gonorrhoeae* Due to NG-MAST 12302 Clonal Spread
477 in Canada, 2015 to 2018. *Antimicrob Agents Chemother*. 2022 Mar 15;66(3):e01688-21.
- 478 5. European Centre for Disease Prevention and Control. Gonorrhoea. In: ECDC. Annual
479 Epidemiological Report for 2019. Stockholm: ECDC; 2023 [Internet]. [cited 2023 Apr 27].
480 Available from: <https://www.ecdc.europa.eu/en/publications-data/gonorrhoea-annual-epidemiological-report-2019>
481
- 482 6. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet*. 2019 Jun;20(6):341–55.
- 483 7. Street TL, Barker L, Sanderson ND, Kavanagh J, Hoosdally S, Cole K, et al. Optimizing
484 DNA Extraction Methods for Nanopore Sequencing of *Neisseria gonorrhoeae* Directly
485 from Urine Samples. *J Clin Microbiol* [Internet]. 2020 Feb 24 [cited 2020 Jul 3];58(3).
486 Available from: <https://jcm.asm.org/content/58/3/e01822-19>
- 487 8. Sanderson ND, Swann J, Barker L, Kavanagh J, Hoosdally S, Crook D, et al. High
488 precision *Neisseria gonorrhoeae* variant and antimicrobial resistance calling from
489 metagenomic Nanopore sequencing. *Genome Res* [Internet]. 2020 Sep 1 [cited 2020 Sep
490 14]; Available from: <http://genome.cshlp.org/content/early/2020/09/01/gr.262865.120>
- 491 9. Cai W, Nunziata S, Rascoe J, Stulberg MJ. SureSelect targeted enrichment, a new cost
492 effective method for the whole genome sequencing of *Candidatus Liberibacter asiaticus*.
493 *Sci Rep*. 2019 Dec 12;9(1):18962.
- 494 10. Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZM, et al. Rapid
495 Whole-Genome Sequencing of *Mycobacterium tuberculosis* Isolates Directly from Clinical
496 Samples. Land GA, editor. *J Clin Microbiol*. 2015 Jul;53(7):2230–7.
- 497 11. Bowden KE, Joseph SJ, Cartee JC, Ziklo N, Danavall D, Raphael BH, et al. Whole-
498 Genome Enrichment and Sequencing of *Chlamydia trachomatis* Directly from Patient
499 Clinical Vaginal and Rectal Swabs. Rasmussen AL, editor. *mSphere*. 2021 Apr
500 28;6(2):e01302-20.
- 501 12. Doyle RM, Burgess C, Williams R, Gorton R, Booth H, Brown J, et al. Direct Whole-
502 Genome Sequencing of Sputum Accurately Identifies Drug-Resistant *Mycobacterium*
503 *tuberculosis* Faster than MGIT Culture Sequencing. Mellmann A, editor. *J Clin Microbiol*.
504 2018 Aug;56(8):e00666-18.

- 505 13. Sanderson ND, Street TL, Foster D, Swann J, Atkins BL, Brent AJ, et al. Real-time
506 analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices.
507 BMC Genomics. 2018 Sep;19(1):714.
- 508 14. Seemann, Torsten. Snippy: Rapid haploid variant calling and core genome alignment
509 [Internet]. 2015. Available from: <https://github.com/tseemann/snippy>
- 510 15. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
511 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015
512 Jan;32(1):268–74.
- 513 16. Didelot X, Wilson DJ. ClonalFrameML: Efficient Inference of Recombination in Whole
514 Bacterial Genomes. Prlic A, editor. PLOS Comput Biol. 2015 Feb 12;11(2):e1004041.
- 515 17. Miller S, Chiu C. The Role of Metagenomics and Next-Generation Sequencing in
516 Infectious Disease Diagnosis. Clin Chem. 2021 Dec 30;68(1):115–24.
- 517 18. Street TL, Sanderson ND, Atkins BL, Brent AJ, Cole K, Foster D, et al. Molecular
518 Diagnosis of OrthopedicDevice-Related Infection Directly from Sonication Fluid by
519 Metagenomic Sequencing. J Clin Microbiol. 2017;(August):JCM.00462-17.
- 520 19. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and
521 laboratory contamination can critically impact sequence-based microbiome analyses.
522 BMC Biol. 2014 Nov 12;12(1):87.
- 523 20. Paniagua Voirol LR, Valsamakis G, Yu M, Johnston PR, Hilker M. How the ‘kitome’
524 influences the characterization of bacterial communities in lepidopteran samples with low
525 bacterial biomass. J Appl Microbiol. 2021 Jun 1;130(6):1780–93.
- 526 21. Ni Y, Liu X, Simeneh ZM, Yang M, Li R. Benchmarking of Nanopore R10.4 and R9.4.1
527 flow cells in single-cell whole-genome amplification and whole-genome shotgun
528 sequencing. Comput Struct Biotechnol J. 2023;21:2352–64.

529