bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Targeted surveillance strategies for efficient detection of novel antibiotic**
2 **resistance variants**
3
4 Allison L. Hicks[1], Stephen M. Kissler[1], Tatum D. Mortimer[1], Kevin C. Ma[1], George
5 Taiaroa[2], Melinda Ashcroft[2], Deborah A. Williamson[2], Marc Lipsitch[1,3], Yonatan H.
6 Grad[1,4,*]
7
8 1 Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of
9 Public Health, Boston, Massachusetts, United States of America
10 2 Department of Microbiology and Immunology, The University of Melbourne at The
11 Peter Doherty Institute for Infection and Immunity, Melbourne, Australia
12 3 Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.
13 H. Chan School of Public Health, Boston, Massachusetts, United States of America,
14 4 Division of Infectious Diseases, Department of Medicine, Brigham and Women's
15 Hospital, Harvard Medical School, Boston, Massachusetts, United States of America
16
17
18 * Corresponding author: ygrad@hsph.harvard.edu
19

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20 **Abstract:**

21 Genotype-based diagnostics for antibiotic resistance represent a promising alternative to
22 empiric therapy, reducing inappropriate and ineffective antibiotic use. However, because
23 such assays infer resistance phenotypes based on the presence or absence of known
24 genetic markers, their utility will wane in response to the emergence of novel resistance.
25 Maintenance of these diagnostics will therefore require surveillance designed to ensure
26 early detection of novel resistance variants, but efficient strategies to do so remain to be
27 defined. Here, we evaluate the efficiency of targeted sampling approaches informed by
28 patient and pathogen characteristics in detecting genetic variants associated with
29 antibiotic resistance or diagnostic escape in *Neisseria gonorrhoeae*, focusing on this
30 pathogen because of its high burden of disease, the imminent threat of treatment
31 resistance, and the use and ongoing development of genotype-based diagnostics. We
32 show that incorporating patient characteristics, such as demographics, geographic
33 regions, or anatomical sites of isolate collection, into sampling approaches is not a reliable
34 strategy for increasing variant detection efficiency. In contrast, sampling approaches
35 informed by pathogen characteristics, such as genomic diversity and genomic
36 background, are significantly more efficient than random sampling in identifying genetic
37 variants associated with antibiotic resistance and diagnostic escape.

38

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

39 **Introduction:**

40 Nucleic acid-based diagnostics that enable rapid pathogen identification and
41 prediction of drug susceptibility profiles can improve clinical decision-making, reduce
42 inappropriate antibiotic use, and help address the challenge of antibiotic resistance [1-3].
43 However, the sensitivity of such diagnostics may be undermined by undetected genetic
44 variants [4-12]. Pathogen surveillance programs aimed at early detection of novel variants
45 are crucial to ensuring the clinical utility and sustainability of these diagnostics.

46 Use of traditional nucleic acid amplification tests (NAATs) for pathogen
47 identification and genotype-based diagnostics for antibiotic resistance can select for
48 genetic variants that escape detection [13]. Mutations and/or deletions at the NAAT target
49 locus that cause an amplification failure have arisen in *Neisseria gonorrhoeae, Chlamydia*
50 *trachomatis, Staphylococcus aureus,* and *Plasmodium falciparum*, resulting in false
51 negative diagnostic errors only detected when using another diagnostic platform [5-7,11].
52 Diagnostic escape associated with genotype-based diagnostics for antibiotic resistance
53 are the result of resistance-conferring variants (*e.g.*, mutations or accessory genes) not
54 accounted for in the diagnostic's panel of resistance markers [4] and require phenotypic
55 testing to be uncovered.

56 We recently presented a framework to quantify the sampling rate for early
57 detection of novel antibiotic resistance variants, defining the number of isolates that would
58 need to undergo confirmatory phenotyping from those predicted by genotype to be
59 susceptible [14]. Underlying this model are assumptions of unbiased sampling across a
60 population and independence among all isolates. However, these assumptions may not
61 hold in practice, as some subsets of the population (*e.g.*, demographics and/or

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

62 geographic regions) may be more likely to be sampled than others, and clonal
63 transmission may result in repeated sampling of closely related isolates [15-18]. The real-
64 world application of this model may also be challenging for pathogens with high case
65 incidence, such as *N. gonorrhoeae*, as the cost of phenotyping required by this model for
66 timely detection of novel resistance variants is likely to be high [14].

67 Implementing a practical surveillance system thus requires improving efficiency
68 over unbiased testing by prioritizing samples in which novel diagnostic escape variants
69 are most likely to be found. There are numerous hypotheses for how to focus sampling
70 and most quickly identify these variants. Novel variants may be more likely to emerge or
71 spread in certain anatomical niches, demographics, or geographic regions [19-22], some of
72 which may be systematically under-sampled [23] and thus may provide a basis for sampling
73 priority. Data on such characteristics may be obtained from metadata recorded during
74 clinical encounters. Alternatively, they may be inferred from pathogen genomic data.
75 Isolates or clades that are genetically divergent from the majority of isolates in a
76 population may reflect travelers, their contacts, or otherwise under-sampled lineages [24-
77 27]. Some pathogen genomic backgrounds may be more conducive to the evolution of
78 novel resistance mechanisms [28], and markers of these genomic backgrounds (*e.g.*,
79 variants associated with a range of resistance mechanisms and/or resistance to other
80 drugs) may help improve sampling efficiency. Similarly, given historical patterns of
81 antibiotic use, novel resistance may emerge on a background of existing resistance [29].
82 Thus, genetic markers of resistance to certain drugs may facilitate identification of
83 lineages more likely to have experienced selective pressures leading to emergence of
84 novel resistance variants.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

85 Here, we test the performance of sampling strategies informed by these
86 hypotheses using *N. gonorrhoeae* surveillance data. *N. gonorrhoeae* offers a useful
87 model, given the increasing drug resistance and recent focus on developing sequence-
88 based resistance diagnostics [2,30]. We present targeted sampling approaches informed by
89 patient (*i.e.,* demographics, anatomical site of isolate collection, geographical region,
90 recent travel history, or sex worker status) and pathogen (*i.e.,* phylogenetic or genomic
91 background) information. We assess the efficiency of each of these strategies to detect
92 rare (<10% prevalence) resistance variants associated with current or recent first-line
93 recommended antibiotics (*i.e.*, azithromycin [AZM] and extended spectrum
94 cephalosporins [ESCs]), as well as rare genetic variants associated with diagnostic
95 escape, across five genomic surveys with various demographic, geographic, and
96 temporal ranges. We show that phylogeny- and genomic background-aware sampling
97 approaches can increase the detection efficiency of known variants over random
98 sampling, whereas patient feature-based sampling approaches do not. Our results
99 suggest that implementation of such targeted sampling approaches into surveillance
100 programs may reduce the number of cases of novel resistance that occur before it is
101 detected, as well as the resources required to undertake surveillance, compared to
102 random sampling of a population.
103

104 **Results:**

105 **Composition of the datasets.**

106 The datasets (**Table 1**) were biased across patient demographics and/or geographic
107 regions (**Tables S1** and **S2**). Isolates from men and men who have sex with men (MSM)

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

108 were overrepresented in datasets 1 and 2 compared to overall gonorrhea incidence in
109 men and MSM in the US and Australia, respectively, during the study periods (**Table S2**,
110 *P* < 0.001 for both datasets by chi-squared test of men vs. women and MSM vs. non-
111 MSM in dataset vs. reported incidence). Dataset 4 was comprised exclusively of isolates
112 from men [31]. While it is difficult to estimate the prevalence of pharyngeal gonococcal
113 infections, as they tend to be asymptomatic [32], pharyngeal isolates represented 4% and
114 18% of isolates with reported anatomical site of collection in datasets 1 and 2,
115 respectively. This suggests either sampling bias across anatomical sites in at least one
116 of the datasets or substantial variation across the two study populations in prevalence of
117 pharyngeal gonococcal infections. Similarly, the geographic distribution of isolates in
118 dataset 3 was significantly different from the reported case incidence across countries
119 (**Table S2**, *P* < 0.001 by chi-squared test of prevalence for each of the countries in dataset
120 3 vs. the reported overall incidence for each of the countries).
121

122 **Table 1.** Summary of datasets.

Dataset	Temporal range	$N_{isolates}$	Geographic range	Metadata available	SRA study ID/Reference
1	2011-2015	896	New York, NY, US	Gender, sexual behavior, anatomical site of isolation	ERP011192 [Mortimer et al., 2020, *in preparation*]
2	2016-2017	2186	Victoria, Australia	Gender, sexual behavior, anatomical site of isolation, travel history, sex worker status	SRP185594 [33]
3	2013	1054	Europe	Country of sample collection	ERP010312 [34]
4	2015	244	Japan	Prefecture of sample collection	DRP004052 [31]
5	2014-2015	398	New Zealand	N/A	SRP111927 [35]

123

124 **Targeted sampling based on patient characteristics.**

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

125 We investigated whether sampling evenly across demographic groups (demography-
126 aware sampling), anatomical sites of isolate collection (niche-aware sampling), and
127 geographic regions (geography-aware sampling) increased detection efficiency of
128 resistance variants by ameliorating some of the demographic, niche, or geographic
129 sampling biases. We further investigated whether preferentially sampling patients with
130 recent overseas sexual encounters or recent sex work, two characteristics hypothesized
131 to be associated with the introduction and/or increased transmission of resistance [19,21,22],
132 increased the detection efficiency of resistance variants. To do so, we simulated and
133 compared the detection efficiency of three genetic resistance variants (**Table 2**) using
134 each of these targeted sampling strategies and random sampling.

135

136 **Table 2.** Summary by dataset of the prevalence and distribution of the genetic markers
137 of resistance and resistance phenotypes tested.

Variant		Genetic			Phenotypic	
		RplD G70D	**23S rRNA C2611T (2-4 alleles)**	***penA* XXXIV**	**CRO-RS (≥0.12 μg/mL)**	**CFX-R (>0.25 μg/mL)**
Drug		AZM [36]	AZM [37]	ESCs [38]	N/A	N/A
Prevalence of variant in dataset	**1**	10.04%[a]	0.11%	5.25%	1.47%	0.11%
	2	1.14%	1.24%	1.69%	0%	0%
	3	2.47%	0.95%	15.68%[a]	1.04%	0.76%
	4	11.07%[a]	1.23%	0.41%	6.56%	8.20%
	5	0.75%	0.50%	2.26%	0.25%	0%
Phylogenetic D statistic for variant in dataset	**1**	-0.18	17.50	-0.29	N/A	N/A
	2	-0.10	0.46	-0.24	N/A	N/A
	3	0.05	0.30	-0.20	N/A	N/A
	4	-0.16	1.83	1.81	N/A	N/A
	5	0.83	1.12	-0.15	N/A	N/A

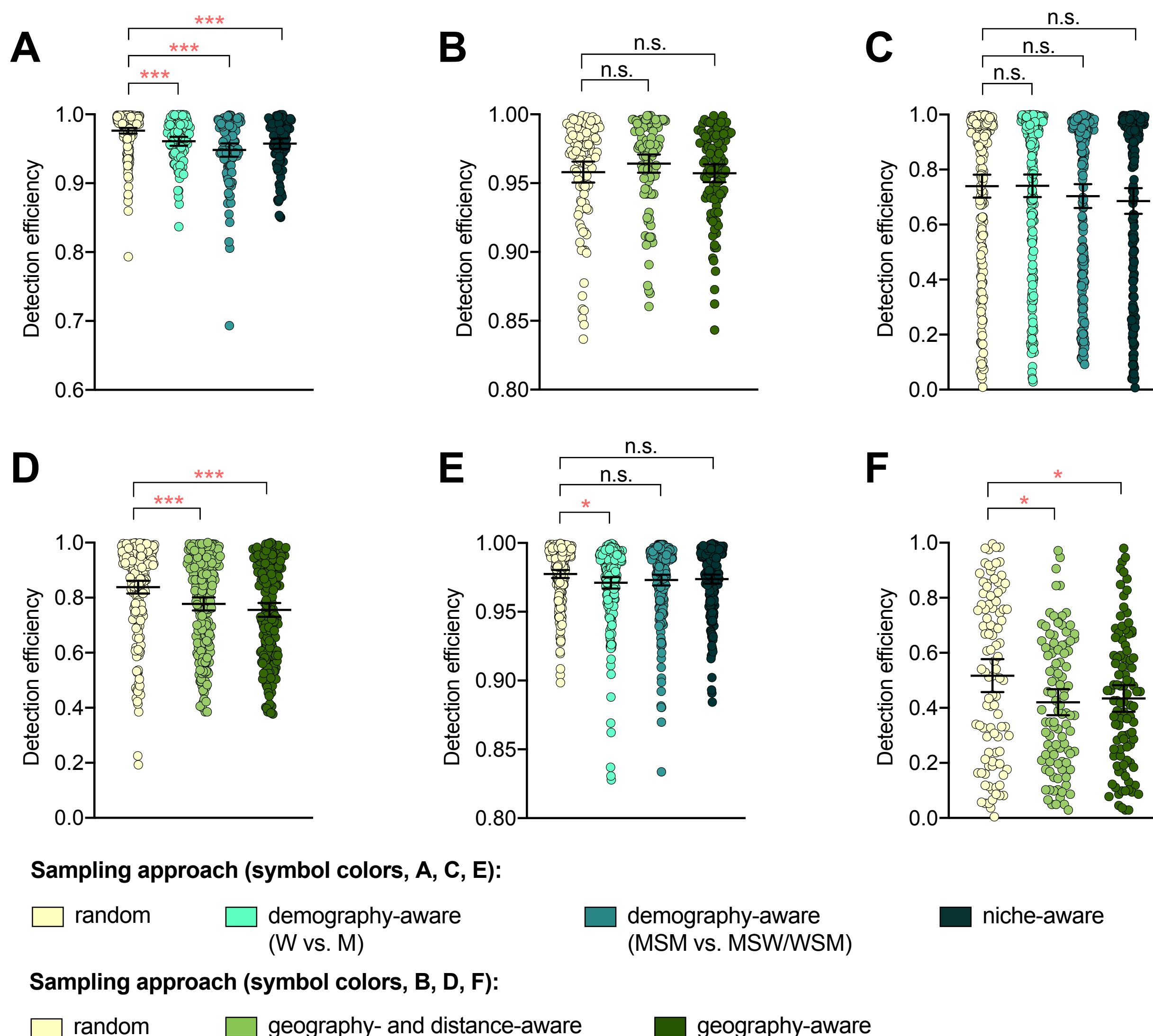
138 [a]Given the >10% prevalence of RplD G70D in datasets 1 and 4 and *penA* XXXIV in
139 dataset 3, these variants were excluded from sampling simulations.
140 AZM, azithromycin; ESC, extended-spectrum cephalosporin; CRO-RS, ceftriaxone
141 reduced susceptibility; CFX-R, cefixime resistance
142

143 The detection efficiency was not improved by demography-, niche-, geography-
144 aware sampling compared to random sampling for any of the resistance variants (**Table**

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

145 **S3, Fig. 1**). In several cases, detection efficiency significantly decreased in demography-
146 or geography-aware sampling compared to random sampling, reflecting enrichment of
147 the resistance variant in the overrepresented demographic or geographic region.
148 However, no significant association between a given resistance variant and demographic
149 group was observed across both dataset 1 and dataset 2, and no demographics or
150 geographic regions were significantly enriched for all variants (**Table S1**), suggesting that
151 preferential sampling of any of these demographics or geographic regions would not be
152 a reliable strategy for increasing novel variant detection efficiency. For example, while
153 *penA* XXXIV was significantly enriched in MSM compared to men who have sex with
154 women and women who have sex with men (MSW/WSM) patients in dataset 2 ($P < 0.003$,
155 Fisher's exact test), there was no significant difference in the proportions of MSM and
156 MSW/WSM with *penA* XXXIV in dataset 1 ($P = 0.461$, Fisher's exact test). Similarly, while
157 the AZM-R-associated RplD G70D mutation in dataset 3 was at highest prevalence in
158 patients from Malta and Greece (10% and 6.25%, respectively) and absent from patients
159 from Denmark, the AZM-R-associated 23S C2611T variant was at highest prevalence in
160 patients from Denmark (5.45%) and absent from patients from Malta or Greece.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



161
162 **Figure 1. The impact of demography-, niche-, and geography-aware sampling on**
163 **the detection efficiency of genetic resistance variants.** Dot plots showing the
164 detection efficiency (with lines indicating the mean and 95% confidence intervals from
165 100 simulations) for resistance variants RplD G70D (**A-B**), 23S rRNA C2611T (**C-D**), and
166 *penA* XXXIV (**E-F**) in datasets 1 and 2. In datasets 1 and 2, targeted sampling was
167 informed by demographic (gender and sexual behavior) and anatomical site of isolate
168 collection (niche) information (**A, C,** and **E**), and in datasets 3 and 4, targeted sampling
169 was informed by country or prefecture of sample collection (**B, D,** and **F**). Dot colors
170 indicate the sampling approach, and asterisks indicate a significant difference ($P < 0.05$
171 by Mann-Whitney U test) in detection efficiency between the demography-, niche- or
172 geography-aware approach compared to random sampling (*$P < 0.05$, **$P < 0.01$, ***$P <$
173 0.001; red asterisks indicate significantly lower detection efficiency of demography- or
174 geography-aware approaches compared to random sampling). Note that sampling
175 simulations were not performed for RplD G70D in datasets 1 and 4 or for *penA* XXXIV in
176 dataset 3 as prevalence of the variants in these datasets was >10%. n.s., not significant

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

177 at α = 0.05; M, men; W, women; MSM, men who have sex with men; MSW, men who
178 have sex with women; WSM; women who have sex with men.
179

180 Isolates from patients with recent overseas sex were associated with significantly
181 longer terminal branches compared to patients that had only engaged in sex locally (**Fig.**
182 **S1**), in support of the hypothesis that international travel may be associated with the
183 importation of novel or divergent strains, or, more generally, that isolates from travelers
184 may be more likely to be associated with under-sampled lineages. Preferentially sampling
185 from patients with recent overseas sex significantly improved detection efficiency of the
186 RplD G70D mutation and the *penA* XXXIV allele, as these were at marginally higher
187 prevalence in isolates from patients with recent overseas sex compared to those from
188 patients who had only engaged in sex locally (3.03% overseas vs. 0.98% local and 2.02%
189 overseas vs. 1.67% local, respectively, *P* = 0.090 and 0.683, respectively, by Fisher's
190 exact test for both variants). In contrast, the 23S C2611T mutation was exclusively
191 present in isolates from patients who had engaged in sex locally (**Tables S1** and **S4**).
192 Similarly, while the 23S C2611T mutation was marginally enriched in isolates from
193 patients who had engaged in recent sex work compared to patients who had not (2.33%
194 in sex workers vs. 1.31% in non-sex workers, *P* = 0.327 by Fisher's exact test), and thus
195 preferentially sampling from sex workers significantly improved detection efficiency of this
196 variant compared to sampling from the full patient population, detection efficiencies for
197 the RplD G70D mutation and the *penA* XXXIV allele were not significantly improved by
198 preferentially sampling from sex workers (**Tables S1** and **S4**).

199 Together, these results suggest that while targeted sampling based on patient
200 characteristics may increase detection efficiency of some novel variants, it is difficult to

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

201 predict which groups to target for all potential novel variants.

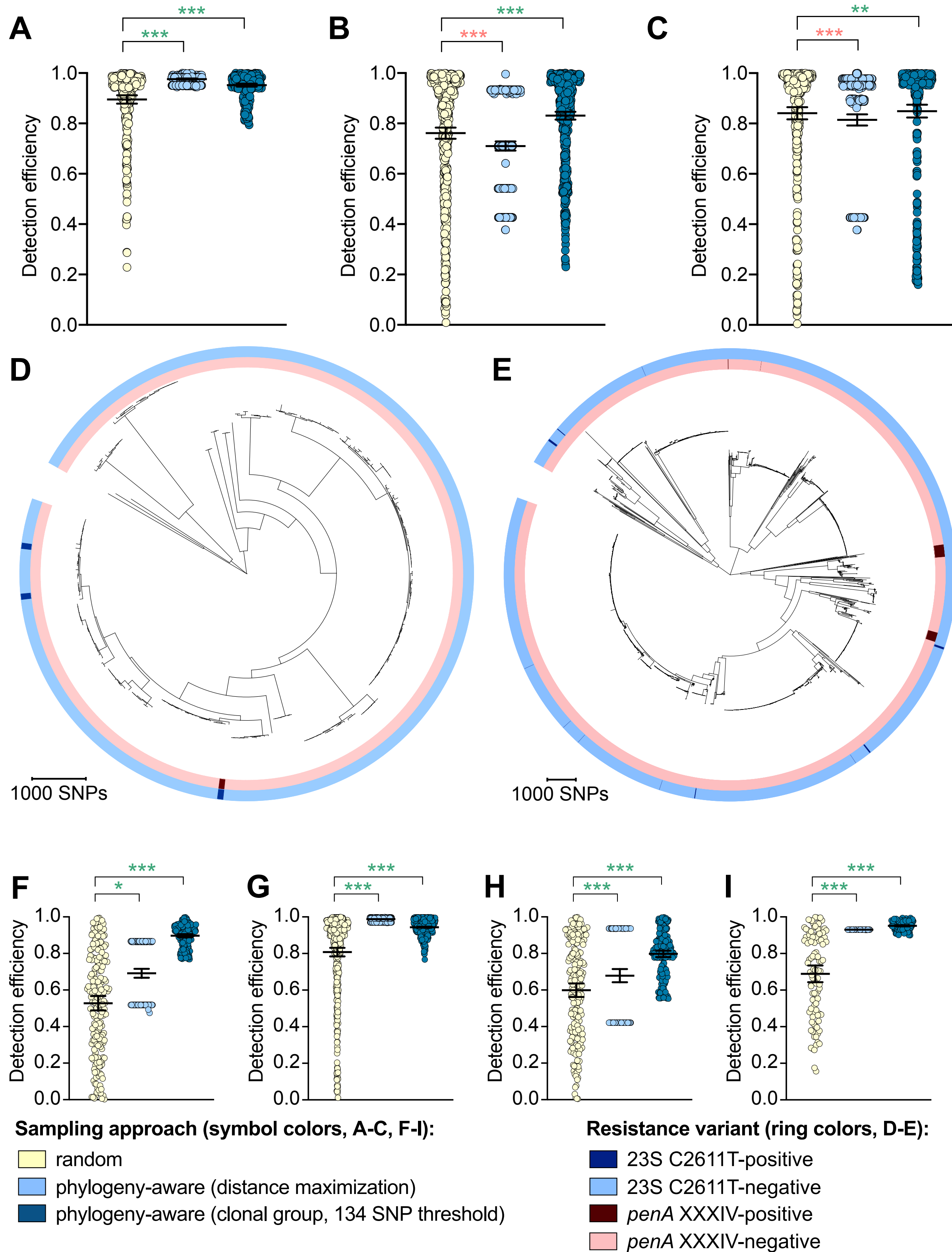
202

203 **Targeted sampling based on genetic diversity.**

204 To assess whether preferential sampling of lineages that are divergent from those that
205 have been previously sampled may increase detection efficiency of genetic resistance
206 variants over random sampling, we simulated phylogeny-aware sampling using two
207 methods: 1) maximization of the phylogenetic distance covered with each isolate sampled
208 (distance maximization) and 2) even sampling across phylogenetic lineages (clonal
209 group).

210 While the distance maximization approach increased detection efficiency
211 compared to random sampling for some variants, there were numerous instances in
212 which this approach, which led to preferential sampling of isolates associated with long
213 branches, substantially decreased detection efficiency (**Fig. 2, Table S5**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



214

215 **Figure 2. The impact of phylogeny-aware sampling on the detection efficiency of**
216 **genetic resistance and diagnostic escape variants.** Scatter dot plots showing the

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

217 detection efficiency (with lines indicating the mean and 95% confidence intervals from
218 100 simulations) for resistance variants RplD G70D (**A**), 23S rRNA C2611T (**B**), and *penA*
219 XXXIV (**C**) in datasets 1-5. Note that sampling simulations were not performed for RplD
220 G70D in datasets 1 and 4 or for *penA* XXXIV in dataset 3 as prevalence of the variants in
221 these datasets was >10%. Maximum-likelihood phylogenies produced from
222 pseudogenome alignments (with predicted regions of recombination removed) of isolates
223 from dataset 4 (**D**) and dataset 2 (**E**). Presence or absence of the 23S rRNA C2611T
224 mutation (in at least 2/4 alleles) and the mosaic *penA* XXXIV allele is indicated by colored
225 rings. Scatter dot plots showing the detection efficiency (with lines indicating the mean
226 and 95% confidence intervals from 100 simulations) for diagnostic-associated variants
227 16S rRNA C1209A (**F**), *N. meningitidis*-like *porA* (**G**), *cppB* deletion (**H**), and DR-9A
228 G168A (**I**) in all datasets in which the variant was present. Dot colors in **A-C** and **F-I**
229 indicate the sampling approach, and asterisks indicate a significant difference ($P < 0.05$
230 by Mann-Whitney U test) in detection efficiency between the phylogeny-aware approach
231 compared to random sampling (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; red asterisks indicate
232 significantly lower detection efficiency of the phylogeny-aware approach compared to
233 random sampling, and green asterisks indicate significantly higher detection efficiency of
234 the phylogeny-aware approach compared to random sampling). n.s., not significant at α
235 $= 0.05$.
236
237 The clonal group sampling approach prevents repeated sampling of very closely
238 related isolates until all unique phylogenetic clusters have been sampled. Thus, for both
239 rare variants that are clonally distributed and rare variants that are more randomly
240 dispersed throughout the phylogeny (*e.g.*, *penA* XXXIV and 23S rRNA C2611T mutations,
241 respectively, **Table 2**), this approach increases detection efficiency in cases where 1)
242 there is substantial clonality among isolates and 2) a substantial proportion of variant-
243 positive isolates do not occur in clonal lineages dominated by variant-negative isolates
244 (**Fig. 2E**). In such datasets, effectively collapsing large variant-negative lineages into a
245 single representative increases the effective prevalence of the variants and thus the
246 detection efficiency of the clonal group approach compared to random sampling. The
247 clonal group sampling approach significantly decreased detection efficiency in only one
248 instance (*i.e.*, the 23S rRNA C2611T variant in dataset 4, **Table S5**), where all isolates

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

249 with the variant appeared in large clonal lineages of predominately variant-negative
250 isolates (**Fig. 2D**).

251 In cases where the clonal group sampling approach did not perform better than
252 random sampling, adjusting the threshold for clonal grouping and/or a marginal increase
253 in the prevalence of variant-positive isolates could elevate the relative performance of this
254 targeted approach. We chose 134 SNPs as an example threshold for clonal grouping, as
255 it represents the lower 95% confidence interval of the mean of SNP distances between
256 each CFX-R resistant and the closest susceptible isolate in datasets 1-5 (see Methods).
257 In the case of the 23S rRNA C2611T variant in dataset 4, the average prevalence of the
258 variant across clonal groups (*i.e.*, the total number of variant-positive isolates, counting
259 each variant-positive isolate as [1 / [1 + the total number of additional isolates that are ≤
260 134 SNPs of the isolate]], divided by the number of clonal groups) is 0.005, lower than
261 the actual prevalence of 0.012. However, if the threshold for clonal grouping was lower in
262 this instance (*e.g.*, 50 SNPs), the effective prevalence of the variants would be 0.020,
263 greater than the actual prevalence of 0.012. Similarly, using the 134 SNP threshold, if
264 one additional isolate that was > 134 SNPs from any other isolates in this dataset had the
265 23S rRNA C2611T mutation, the average prevalence of the variant across clonal groups
266 would be 0.036, greater than the actual prevalence of 0.016, and thus the clonal group
267 approach would outperform random sampling.

268 To further assess the performance of phylogeny-aware sampling in the context of
269 rare genetic variants that may have emerged in response to diagnostic pressure, we
270 simulated random and phylogeny-aware sampling to assess detection efficiency of an
271 additional set of variants. Specifically, we assessed a panel of *N. gonorrhoeae* diagnostic

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

272 escape variants: the 16S rRNA C1209A mutation, the *N. meningitidis*-like *porA*, and the
273 *cppB* deletion, all of which have been previously associated with diagnostic failure [7-10]
274 and were present in one or more of datasets 1-5 at low prevalence (**Table 3**). The G168A
275 mutation in the primer binding region of DR-9A, the target of the COBAS 4800 CT/NG
276 (Roche) diagnostic, has not previously been documented but was present in 0.1% of
277 strains from dataset 2. All of the diagnostic-associated variants assessed appeared in
278 divergent backgrounds and were thus detected more efficiently by phylogeny-aware
279 sampling compared to random sampling (**Fig. 2F-I, Table S6**). Like the results from the
280 simulations based on resistance variants, the distance maximization approach maximized
281 detection efficiency for some of the diagnostic-associated variants, but superiority of this
282 approach to random sampling was not consistent across all variants. However, the clonal
283 group approach performed significantly better than random sampling for all diagnostic-
284 associated variants across all datasets.

285

286 **Table 3.** Summary of the potential diagnostic escape variants assessed.

Variant	Diagnostic assay	Documented association with diagnostic failure	Prevalence in dataset				
			1	2	3	4	5
16S rRNA C1209A (4 alleles)	Aptima GC Combo	Yes [7]	0.11%	0.09%	0%	0%	0%
N. meningitidis-like *porA*	In-house [39,40]	Yes [8,9]	0.11%	0.05%	0%	0%	0%
cppB deletion	In-house [41,42]	Yes [10]	1.12%	0.05%	0.47%	0%	7.29%
DR-9A G168A	Roche COBAS 4800 CT/NG	No	0%	0.09%	0%	0%	0%

287
288

289 The relative performance of the clonal group sampling approach compared to
290 random sampling was generally consistent across multiple thresholds based on

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

291 pseudogenomes (*i.e.,* ≤ 134 SNPs, ≤ 422 SNPs, and fastBAPS groups); relative
292 performance of clonal group sampling using MLSTs, however, was less consistent and
293 was significantly worse than random sampling for several variants (**Fig. S2, Tables S5-**
294 **S6**). Together, these results suggest that preferentially sampling isolates that, based on
295 whole genome sequencing (WGS), are phylogenetically divergent from those that have
296 previously been sampled may increase detection efficiency of novel resistance variants.
297

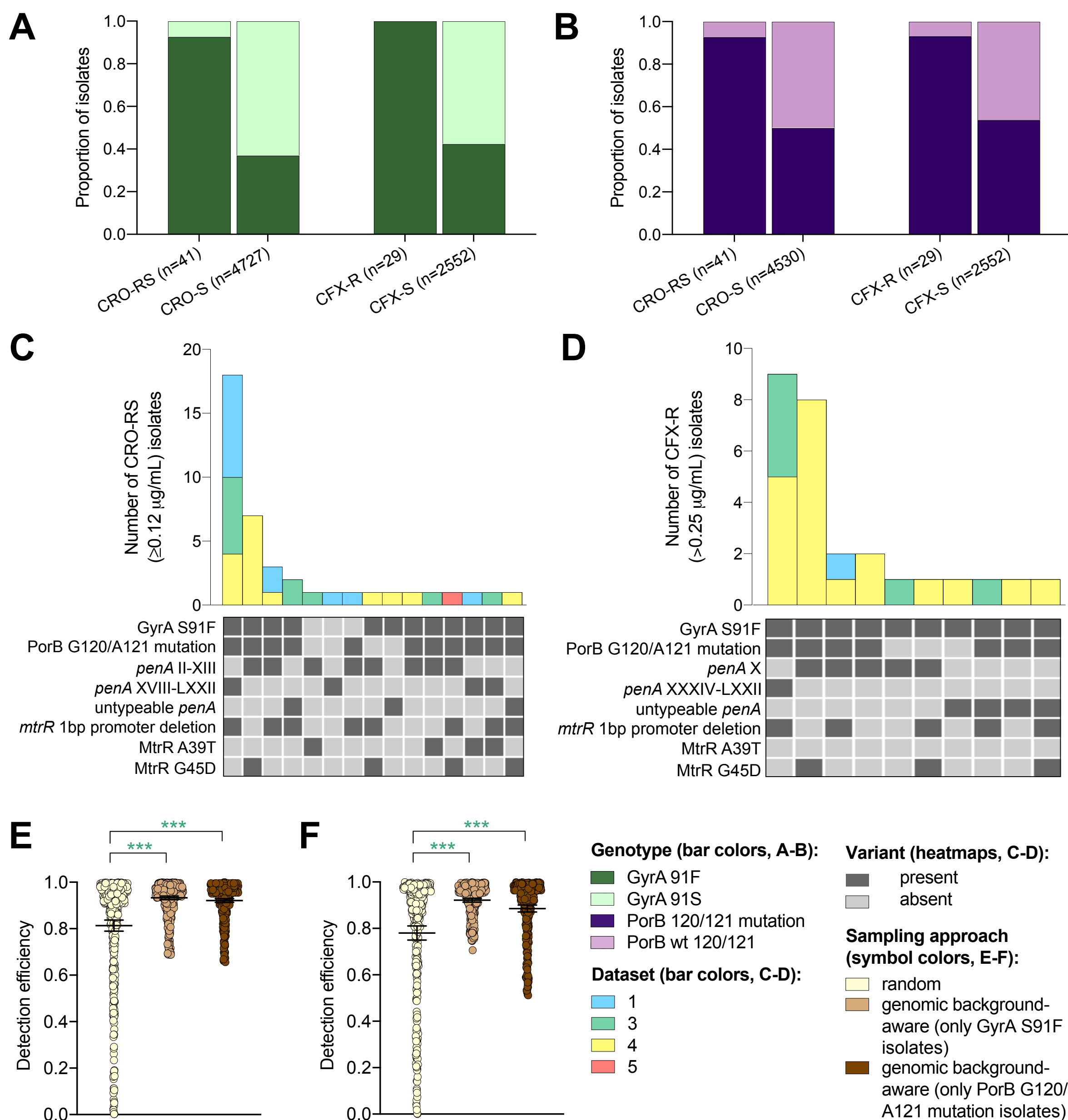
298 **Targeted sampling based on genetic markers.**

299 Multiple drug resistance is more common in pathogenic bacteria than one would expect
300 from the product of frequencies of resistance to individual drugs [43,44]. This suggests that
301 novel resistance mechanisms might be more likely to arise and spread in bacterial strains
302 that are already resistant to other drugs, a phenomenon that has been documented in *N.*
303 *gonorrhoeae* [45]. It may therefore be fruitful to look for novel resistance variants for one
304 drug in genetic backgrounds that are resistant to other drugs. It may be similarly effective
305 to sample preferentially isolates with genetic markers that have been associated with a
306 range of resistance mechanisms (*e.g.*, through epistatic interactions with other genetic
307 variants) within and/or across different antibiotics when screening for a novel resistance
308 variant. For example, as ciprofloxacin was the recommended first-line therapy for
309 uncomplicated gonorrhea through 2005 in the United Kingdom [46], 2007 in the United
310 States [47], and more recent years in other countries [48-50], we investigated whether
311 resistance to ESCs is significantly more likely to occur in the background of genotypic
312 ciprofloxacin resistance (*i.e.*, in strains with the GyrA S91F mutation). Similarly, as
313 mutations at positions 120 and/or 121 in PorB, the major outer membrane protein in

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

314 gonococci, have been associated with resistance to a range of drugs from multiple
315 classes [51], we investigated whether resistance to ESCs is significantly more likely to occur
316 in strains with PorB 120 and/or 121 mutations. Isolates with CRO-RS and CFX-R were
317 significantly more likely to have the GyrA S91F mutation and the PorB G120 and/or A121
318 mutations than the wild-type GyrA S91 or wild-type PorB G120/A121 ($P < 0.001$, Fisher's
319 exact test, **Fig. 3A-B**). Further, both GyrA S91F and PorB G120 and/or A121 mutations
320 occurred across a range of ESC resistance locus haplotypes (**Fig. 3C-D**). For all datasets
321 with CRO-RS or CFX-R isolates, detection efficiency of both variants was significantly
322 increased by only sampling isolates with the GyrA S91F mutation or the PorB G120 and/or
323 A121 mutations (**Fig. 3E-F, Table S7**). Together, these results suggest that preferential
324 sampling of isolates with certain genetic markers, including markers of resistance to
325 previous first-line antibiotics, may increase the detection efficiency of novel resistance
326 variants.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



327

328 **Figure 3. The impact of genomic background-aware sampling on the detection**
329 **efficiency of phenotypic resistance variants.** Bar charts showing the proportions of
330 ceftriaxone reduced susceptibility (CRO-RS) isolates, ceftriaxone susceptible (CRO-S)
331 isolates, cefixime resistant (CFX-R) isolates, and cefixime susceptible (CFX-S) isolates
332 with GyrA S91F and GyrA S91 wild-type alleles (**A**) and with PorB G120 and/or A121
333 mutations and PorB G120 and A121 wild-type alleles (**B**) across datasets 1-5. Bar charts
334 showing the number of (**C**) CRO-RS and (**D**) CFX-R isolates with each haplotype, along
335 with heatmaps showing the presence or absence of the GyrA S19F mutation, the PorB
336 G120 and/or A121 mutations, and other alleles at loci previously associated with
337 extended spectrum cephalosporin resistance. Bar colors in (**C**) and (**D**) indicate the
338 dataset from which the isolates were derived. Scatter dot plots showing the detection
339 efficiency (with lines indicating the mean and 95% confidence intervals from 100

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

340 simulations) for CRO-RS (**E**) and CFX-R (**F**) in all datasets in which the variant was
341 present. Dot colors in **E-F** indicate the sampling approach, and asterisks indicate a
342 significant difference (P < 0.05 by Mann-Whitney U test) in detection efficiency between
343 the phylogeny-aware approach compared to random sampling (*P < 0.05, **P < 0.01,
344 ***P < 0.001; green asterisks indicate significantly higher detection efficiency of the
345 genomic background-aware approach compared to random sampling).
346

347 **Discussion**

348 With sequencing becoming more integral to routine pathogen surveillance and
349 diagnostics, it is important to ensure that models mapping genotypic information to
350 expected pathogen phenotype and/or clinical outcome are comprehensive and current [52].
351 In the case of genotype-based diagnostics, sustained phenotypic surveillance is crucial
352 for identifying resistance variants that have recently emerged and/or increased in
353 prevalence from previously undetected levels. While effective incorporation of patient
354 metadata into surveillance strategies may be challenging, availability and incorporation
355 of information on pathogen characteristics (*e.g.*, pathogen genomic data) into surveillance
356 programs may ultimately decrease the cost of surveillance to maintain the sensitivity of
357 these diagnostic tools.

358 Collection of patient metadata, including demographic and geographic information,
359 is crucial to understanding the epidemiology of drug resistance. However, it may be
360 difficult to obtain data on the relevant patient features, and the predictive power of such
361 features may rapidly decay because of patient mobility and interactions [53]. While
362 availability of patient metadata varied across the datasets assessed, our results suggest
363 that while incorporation of patient metadata into sampling strategies may increase
364 detection efficiency for some novel resistance variants, it may be difficult to generalize for
365 all potential novel resistance variants. It is possible that targeted sampling based on

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

366 patient characteristics may be more reliable in the context of pathogens, antibiotic, and/or
367 patient characteristics not assessed here.

368 Incorporation of WGS into routine pathogen surveillance by public health agencies
369 [54,55] may facilitate use of genomic information in phenotypic sampling strategies,
370 particularly with emerging metagenomic approaches that do not require bacterial culture
371 [56]. Our results show that phylogeny-aware sampling, particularly the clonal group
372 approach, which reduces the amount of repeated sampling of closely related isolates,
373 significantly improved detection efficiency over random sampling for multiple resistance
374 and diagnostic-associated variants. Further, identification of and preferential sampling of
375 isolates with genetic markers that are consistently predictive of resistance across a range
376 of mechanisms, including those associated with resistance to other drugs, may
377 supplement phylogeny-aware sampling to further optimize detection efficiency of novel
378 variants. However, the utility of sampling based on genetic markers of other resistance
379 mechanisms will likely vary substantially across different drugs and be influenced by
380 future treatment guidelines.

381 While the clonal group sampling approach increased detection efficiency for the
382 resistance and diagnostic escape variants assessed here, it may be difficult to determine
383 the most effective and reliable metric or threshold for clonal grouping, especially as this
384 is likely to vary across different clinical populations, antibiotics, and bacterial species.
385 Detection efficiency was generally consistent across the two SNP thresholds and
386 fastBAPS groupings based on WGS. However, performance of the clonal group approach
387 using MLSTs was inconsistent and, in some instances, worse than random sampling,
388 likely due to the shortcomings of MLST compared to WGS-based approaches in

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

389 distinguishing between AMR variant-positive clades and more distantly-related variant-
390 negative clades in species such as *N. gonorrhoeae* [34]. This suggests that this approach
391 is sensitive to similarity thresholds and that a low SNP threshold based on WGS
392 assemblies may be the most appropriate approach, particularly in a population where
393 there is expected to be substantial clonality among isolates and thus, even with a low
394 threshold, detection efficiency will be improved by the clonal group approach. More
395 broadly, surveillance incorporating WGS rather than MLST loci alone may further promote
396 NAAT sustainability by enabling screening for variants with previously undetected
397 mutations in target loci, such as the *N. gonorrhoeae* DR-9A G168A variants, that may be
398 associated with diagnostic escape.

399 We have assessed these targeted sampling approaches in detection of multiple
400 resistance variants across a range of populations, but these represent only a fraction of
401 resistance mechanisms in a single species. These findings may extend to other antibiotics
402 and bacterial species. For example, given the high degree of clonality among *M.*
403 *tuberculosis* isolates and the significant variation in prevalence of drug resistance and
404 resistance-conferring genotypes across clonal groups [57,58], the clonal group sampling
405 approach may similarly improve detection efficiency of novel resistance variants in *M.*
406 *tuberculosis*. For species in which drug resistance is primarily acquired through gene
407 acquisition, it is unclear if phylogeny-aware sampling based on the core genome will
408 improve detection efficiency of novel variants. K-mer distances [59,60] may provide a more
409 practical alternative generalizable to more resistance mechanisms associated with gene
410 acquisition. Further, the requirement of confirmatory phenotyping to identify novel
411 resistance may not extend to pathogens that are expected to be associated with reliably-

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

412 identifiable treatment failures, as for these pathogens, identification of treatment failure
413 likely represents the most efficient method of novel resistance variant detection [61].
414 However, for other pathogens, such as *N. gonorrhoeae* [62], treatment failures may go
415 undetected for reasons including partial abatement of symptoms or long treatment
416 regimens. Ultimately, as genotype-based diagnostics for antibiotic resistance become
417 available for more species, it will be important to assess the efficiencies of these
418 approaches across pathogens with different clinical, epidemiological, and evolutionary
419 paradigms.

420 Since we lack the datasets to assess targeted sampling of variants from the time
421 they first emerged in a population, any associations we observed between the variants
422 and patient or pathogen features do not necessarily reflect those around the time of
423 emergence. Thus, more longitudinal epidemiological and genomic studies, particularly
424 after the implementation of genotype-based diagnostics, are necessary to better
425 characterize patterns of novel resistance emergence and inform targeted surveillance
426 approaches.

427 The phylogeny-aware sampling approaches presented here are based on the
428 assumption that genomic data will be available for the pool of potential isolates from
429 incident cases that may undergo confirmatory phenotyping. However, using information
430 on isolate features to increase surveillance efficiency may be feasible even in the absence
431 of mass prospective sequencing. For example, under the general assumption that novel
432 resistance variants are more likely to appear in underrepresented lineages, phylogeny-
433 aware surveillance could be paired with a diagnostic approach such as genomic neighbor
434 typing [56], where any isolates with either susceptible or low confidence calls that appear

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

435 to be divergent from the genomes in the reference database would be prioritized for
436 confirmatory phenotyping. Similarly, a diagnostic that predicts AMR phenotypes through
437 a combination of transcriptomic and genomic typing [63] may facilitate targeted surveillance
438 by identifying isolates with ambiguous predictions (*e.g.*, isolates with transcriptional
439 signatures of resistance that lack known genomic markers of resistance) that could be
440 prioritized for confirmatory phenotyping.

441 Advances in diagnostics, extensive sequencing of clinical isolates, and large
442 collections of clinical and pathogen data together provide new opportunities for integrating
443 data streams and optimizing surveillance efforts. As marker-based point-of-care AMR
444 diagnostics are developed and implemented, optimization of surveillance systems will
445 require assessments like those modeled here of species-, drug-, and population-specific
446 factors that may affect the emergence and distribution of diagnostic escape resistance
447 variants, as well as how the diagnostic itself may complement surveillance efforts.

448

449 **Methods:**

450 **Dataset preparation and phylogenetic reconstruction**

451 See **Table 1** for details of the *N. gonorrhoeae* datasets and **Tables 2** and **3** for the
452 variants assessed. Raw sequencing data were downloaded from the NCBI Sequence
453 Read Archive. Genomes were assembled using SPAdes v3.13 [64] with default
454 parameters and the careful option to minimize the number of mismatches. Assembly
455 quality was assessed using QUAST v4.3 [65], and contigs <500 bp in length and/or with
456 <10x average coverage were removed. Isolate reference-based pseudogenomes were
457 constructed by mapping raw reads to the NCCP11945 reference genome (RefSeq

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

458 accession number NC_011035.1) using BWA-MEM v7.12 [66], the Picard toolkit v2.8
459 (http://broadinstitute.github.io/picard) to identify duplicate reads, and Pilon v1.22 [67] to
460 determine the base call for each site, with a minimum depth of 10 and a minimum base
461 quality of 20.

462 Loci in **Tables 2** and **3** were extracted from the genome assemblies using blastn
463 [68] followed by MUSCLE alignment using default parameters [69] to assess the presence
464 or absence of the resistance variants. Presence or absence of mutations in the multi-
465 copy 16S and 23S rRNA genes and the repetitive DR-9A and DR-9B regions [70] was
466 assessed using BWA-MEM, the Picard toolkit, and Pilon, as above, to map raw reads
467 to a single 16S rRNA allele, a single 23S rRNA allele, a single DR-9A region, and a
468 single DR-9B region from the NCCP11945 reference isolate and determine the
469 mapping quality-weighted percentage of each nucleotide at the site of interest. Isolate
470 metadata and resistance variant profiles are given in **Table S1.**

471 Gubbins v2.3.4 [71] was used with default parameters to identify and mask
472 recombinant regions from the pseudogenomes and build maximum likelihood
473 phylogenies from the non-recombinant pseudogenome alignments for each dataset
474 through RAxML v8.2.12 [72]. Pairwise phylogenetic distances were calculated after removal
475 of predicted recombinant regions using the ape package in R. Phylogenetic distributions
476 of genetic resistance variants were assessed by estimating the phylogenetic D statistic [73]
477 using the caper package in R. Bayesian analysis of population structure was performed
478 on the pseudogenome alignments for each dataset using fastBAPS [74]. Multilocus
479 sequence types (MLSTs) were assigned using the PubMLST database
480 (https://pubmlst.org/neisseria/).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

481

482 **Sampling approaches**

483 For each sampling approach/dataset/variant combination, 100 simulations were carried
484 out with isolate sampling continuing until variant detection. We defined 'detection
485 efficiency' as 1 minus the fraction of isolates sampled prior to variant detection (excluding
486 any samples for which the presence or absence of the variant could not be determined).
487 Because the purpose of this study was to compare the rare variant detection efficiency
488 between random sampling and targeted sampling approaches, we did not evaluate RplD
489 G70D in datasets 1 and 4 or for the *penA* XXXIV allele in dataset 3, as the prevalence of
490 these variants in these datasets was > 10%.

491 In demography-aware sampling (datasets 1 and 2), the first isolate was selected
492 at random, and each successive isolate was randomly selected from alternating
493 demographic groups (men vs. women and men who have sex with men [MSM] vs. men
494 who have sex with women [MSW] or women who have sex with men [WSM]). For
495 anatomical site (niche)-aware sampling (datasets 1 and 2), the first isolate was selected
496 at random, and each successive isolate was randomly selected from alternating
497 anatomical sites of isolate collection (*i.e.*, cervix, urethra, rectum, and pharynx). For
498 geography-aware sampling (datasets 3 and 4), the first isolate was selected at random,
499 and each successive isolate was randomly selected from alternating geographic regions
500 (countries or prefectures). For geography- and distance-aware sampling (datasets 3 and
501 4), the first isolate was selected at random, and each successive isolate was selected
502 randomly from the region (country or prefecture) with the largest product of geographic
503 distances from previously sampled regions, only re-sampling from a given region after all

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

504 regions had been sampled in that round. For travel history- and sex work-aware sampling
505 (dataset 2), isolates were selected at random either limiting the pool to isolates from
506 patients who had recently engaged in overseas sex or sex work, respectively [33].

507 For phylogeny-aware sampling (datasets 1-5), the first isolate was selected at
508 random, and each successive isolate was either selected to maximize the product of
509 phylogenetic distances from each of the previously sampled isolates ("distance
510 maximization") or selected randomly with the exception of ensuring even sampling across
511 phylogenetic groups ("clonal group"; *i.e.*, isolates $\leq N$ SNPs from a previously sampled
512 isolate that were excluded from future sampling until all "clonal groups" had been
513 sampled). SNP cutoffs tested for the clonal group approach included 1) 134 SNPs, the
514 lower 95% confidence interval of the mean SNP distance across datasets 1-5 between
515 each isolate with phenotypic cefixime resistance (CFX-R), azithromycin resistance (AZM-
516 R), and/or ceftriaxone reduced susceptibility (CRO-RS, >0.25 μg/mL, >1 μg/mL, and
517 ≥0.12 μg/mL, respectively) and the closest susceptible isolate, and 2) 422 SNPs, the
518 lower 95% confidence interval of the mean SNP distance across datasets 1-5 between
519 each isolate with the RplD G70D mutation, the 23S rRNA C2611T mutation, and/or the
520 *penA* XXXIV allele and the closest isolate without the resistance variant. The clonal group
521 sampling approach was further tested by alternating sampling across fastBAPS and
522 MLST groups.

523 For genomic background-aware sampling, isolates were selected at random either
524 limiting the pool to isolates with genotypic ciprofloxacin resistance (*i.e.*, the GyrA S91F
525 mutation) or to isolates with a mutation at PorB G120 and/or PorB A121, which have been
526 associated with a range of resistance pathways in multiple classes of antibiotics [51].

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

527 Genomic background-aware sampling was assessed in detection of CRO-RS (datasets
528 1 and 3-5; dataset 2 had no CRO-RS isolates) and CFX-R (datasets 1 and 3-4; datasets
529 2 and 5 had no CFX-RS isolates).

530

531 **References:**

532

533 1 McAdams, D., Waldetoft, K. W., Tedijanto, C., Lipsitch, M. & Brown, S. P.
534 Resistance diagnostics as a public health tool to combat antibiotic resistance: A
535 model-based evaluation. *PLoS Biol* **17**, e3000250,
536 doi:https://doi.org/10.1371/journal.pbio.3000250 (2019).
537 2 Fingerhuth, S. M., Low, N., Bonhoeffer, S. & Althaus, C. L. Detection of antibiotic
538 resistance is essential for gonorrhoea point-of-care testing: a mathematical
539 modelling study. *BMC Med* **15**, 142, doi:10.1186/s12916-017-0881-x (2017).
540 3 Tuite, A. R. *et al.* Impact of Rapid Susceptibility Testing and Antibiotic Selection
541 Strategy on the Emergence and Spread of Antibiotic Resistance in Gonorrhea. *J*
542 *Infect Dis* **216**, 1141-1149, doi:10.1093/infdis/jix450 (2017).
543 4 Andre, E. *et al.* Novel rapid PCR for the detection of Ile491Phe rpoB mutation of
544 Mycobacterium tuberculosis, a rifampicin-resistance-conferring mutation
545 undetected by commercial assays. *Clin Microbiol Infect* **23**, 267 e265-267 e267,
546 doi:10.1016/j.cmi.2016.12.009 (2017).
547 5 Berhane, A. *et al.* Major Threat to Malaria Control Programs by Plasmodium
548 falciparum Lacking Histidine-Rich Protein 2, Eritrea. *Emerg Infect Dis* **24**, 462-
549 470, doi:10.3201/eid2403.171723 (2018).
550 6 Herrmann, B. *et al.* Emergence and spread of Chlamydia trachomatis variant,
551 Sweden. *Emerg Infect Dis* **14**, 1462-1465, doi:10.3201/eid1409.080153 (2008).
552 7 Guglielmino, C. J. D., Appleton, S., Vohra, R. & Jennison, A. V. Identification of
553 an unusual 16S rRNA mutation in Neisseria gonorrhoeae. *J Clin Microbiol*,
554 doi:10.1128/JCM.01337-19 (2019).
555 8 Whiley, D. M. *et al.* False-negative results using Neisseria gonorrhoeae porA
556 pseudogene PCR - a clinical gonococcal isolate with an N. meningitidis porA
557 sequence, Australia, March 2011. *Euro Surveill* **16** (2011).
558 9 Golparian, D., Johansson, E. & Unemo, M. Clinical Neisseria gonorrhoeae isolate
559 with a N. meningitidis porA gene and no prolyliminopeptidase activity, Sweden,
560 2011: danger of false-negative genetic and culture diagnostic results. *Euro*
561 *Surveill* **17** (2012).
562 10 Bruisten, S. M. *et al.* Multicenter validation of the cppB gene as a PCR target for
563 detection of Neisseria gonorrhoeae. *J Clin Microbiol* **42**, 4332-4334,
564 doi:10.1128/JCM.42.9.4332-4334.2004 (2004).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

565 11 Lee, G. H., Pang, S. & Coombs, G. W. Misidentification of Staphylococcus
566 aureus by the Cepheid Xpert MRSA/SA BC Assay Due to Deletions in the spa
567 Gene. *Journal of clinical microbiology* **56**, doi:10.1128/JCM.00530-18 (2018).
568 12 Marks, M. *et al.* Diagnostics for Yaws Eradication: Insights From Direct Next-
569 Generation Sequencing of Cutaneous Strains of Treponema pallidum. *Clin Infect*
570 *Dis* **66**, 818-824, doi:10.1093/cid/cix892 (2018).
571 13 Smid, J. H., Althaus, C. L., Low, N., Unemo, M. & Herrmann, B. Rise and fall of
572 the new variant of Chlamydia trachomatis in Sweden: mathematical modelling
573 study. *Sex Transm Infect*, doi:10.1136/sextrans-2019-054057 (2019).
574 14 Hicks, A. L., Kissler, S. M., Lipsitch, M. & Grad, Y. H. Surveillance to maintain the
575 sensitivity of genotype-based antibiotic resistance diagnostics. *PLoS Biol* **17**,
576 e3000547, doi:10.1371/journal.pbio.3000547 (2019).
577 15 Rempel, O. R. & Laupland, K. B. Surveillance for antimicrobial resistant
578 organisms: potential sources and magnitude of bias. *Epidemiol Infect* **137**, 1665-
579 1673, doi:10.1017/S0950268809990100 (2009).
580 16 Unemo, M. *et al.* World Health Organization Global Gonococcal Antimicrobial
581 Surveillance Program (WHO GASP): review of new data and evidence to inform
582 international collaborative actions and research efforts. *Sex Health*,
583 doi:10.1071/SH19023 (2019).
584 17 Hutinel, M. *et al.* Population-level surveillance of antibiotic resistance in
585 Escherichia coli through sewage analysis. *Euro Surveill* **24**, doi:10.2807/1560-
586 7917.ES.2019.24.37.1800497 (2019).
587 18 Van Goethem, N. *et al.* Status and potential of bacterial genomics for public
588 health practice: a scoping review. *Implement Sci* **14**, 79, doi:10.1186/s13012-
589 019-0930-2 (2019).
590 19 Lewis, D. A. The role of core groups in the emergence and dissemination of
591 antimicrobial-resistant N gonorrhoeae. *Sex Transm Infect* **89 Suppl 4**, iv47-51,
592 doi:10.1136/sextrans-2013-051020 (2013).
593 20 Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S. & Laxminarayan, R.
594 Anthropological and socioeconomic factors contributing to global antimicrobial
595 resistance: a univariate and multivariable analysis. *Lancet Planet Health* **2**, e398-
596 e405, doi:10.1016/S2542-5196(18)30186-4 (2018).
597 21 Frost, I., Van Boeckel, T. P., Pires, J., Craig, J. & Laxminarayan, R. Global
598 Geographic Trends in Antimicrobial Resistance: The Role of International Travel.
599 *J Travel Med*, doi:10.1093/jtm/taz036 (2019).
600 22 Hernando Rovirola, C. *et al.* Antimicrobial resistance in Neisseria gonorrhoeae
601 isolates from foreign-born population in the European Gonococcal Antimicrobial
602 Surveillance Programme. *Sex Transm Infect*, doi:10.1136/sextrans-2018-053912
603 (2020).
604 23 Kirkcaldy, R. D., Weston, E., Segurado, A. C. & Hughes, G. Epidemiology of
605 gonorrhoea: a global perspective. *Sex Health*, doi:10.1071/SH19061 (2019).
606 24 Perrin, L., Kaiser, L. & Yerly, S. Travel and the spread of HIV-1 genetic variants.
607 *Lancet Infect Dis* **3**, 22-27 (2003).
608 25 Pham Thanh, D. *et al.* A novel ciprofloxacin-resistant subclade of H58
609 Salmonella Typhi is associated with fluoroquinolone treatment failure. *Elife* **5**,
610 e14003, doi:10.7554/eLife.14003 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

611 26 Kingsley, R. A. *et al.* Epidemic multiple drug resistant Salmonella Typhimurium
612 causing invasive disease in sub-Saharan Africa have a distinct genotype.
613 *Genome Res* **19**, 2279-2287, doi:10.1101/gr.091017.109 (2009).
614 27 Mac Aogain, M., Rogers, T. R. & Crowley, B. Identification of emergent bla CMY-
615 2 -carrying Proteus mirabilis lineages by whole-genome sequencing. *New*
616 *Microbes New Infect* **9**, 58-62, doi:10.1016/j.nmni.2015.11.012 (2016).
617 28 Borrell, S. & Gagneux, S. Strain diversity, epistasis and the evolution of drug
618 resistance in Mycobacterium tuberculosis. *Clin Microbiol Infect* **17**, 815-820,
619 doi:10.1111/j.1469-0691.2011.03556.x (2011).
620 29 Gould, I. M. & MacKenzie, F. M. Antibiotic exposure as a risk factor for
621 emergence of resistance: the influence of concentration. *J Appl Microbiol* **92**
622 **Suppl**, 78S-84S (2002).
623 30 Hook, E. W., 3rd & Kirkcaldy, R. D. A Brief History of Evolving Diagnostics and
624 Therapy for Gonorrhea: Lessons Learned. *Clin Infect Dis* **67**, 1294-1299,
625 doi:10.1093/cid/ciy271 (2018).
626 31 Yahara, K. *et al.* Genomic surveillance of Neisseria gonorrhoeae to investigate
627 the distribution and evolution of antimicrobial-resistance determinants and
628 lineages. *Microb Genom* **4**, doi:10.1099/mgen.0.000205 (2018).
629 32 Wiesner, P. J., Tronca, E., Bonin, P., Pedersen, A. H. & Holmes, K. K. Clinical
630 spectrum of pharyngeal gonococcal infection. *N Engl J Med* **288**, 181-185,
631 doi:10.1056/NEJM197301252880404 (1973).
632 33 Williamson, D. *et al.* Bridging of Neisseria Gonorrhoeae Across Diverse Sexual
633 Networks in the HIV Pre-Exposure Prophylaxis (PrEP) Era: A Clinical and
634 Molecular Epidemiological Study. *Nature Communications* **10**, 3988 (2019).
635 34 Harris, S. R. *et al.* Public health surveillance of multidrug-resistant clones of
636 Neisseria gonorrhoeae in Europe: a genomic survey. *Lancet Infect Dis* **18**, 758-
637 768, doi:10.1016/S1473-3099(18)30225-1 (2018).
638 35 Lee, R. S. *et al.* Genomic epidemiology and antimicrobial resistance of Neisseria
639 gonorrhoeae in New Zealand. *J Antimicrob Chemother* **73**, 353-364,
640 doi:10.1093/jac/dkx405 (2018).
641 36 Grad, Y. H. *et al.* Genomic Epidemiology of Gonococcal Resistance to Extended-
642 Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United
643 States, 2000-2013. *J Infect Dis* **214**, 1579-1587, doi:10.1093/infdis/jiw420 (2016).
644 37 Ng, L. K., Martin, I., Liu, G. & Bryden, L. Mutation in 23S rRNA associated with
645 macrolide resistance in Neisseria gonorrhoeae. *Antimicrob Agents Chemother*
646 **46**, 3020-3025 (2002).
647 38 Grad, Y. H. *et al.* Genomic epidemiology of Neisseria gonorrhoeae with reduced
648 susceptibility to cefixime in the USA: a retrospective observational study. *Lancet*
649 *Infect Dis* **14**, 220-226, doi:10.1016/S1473-3099(13)70693-5 (2014).
650 39 Whiley, D. M. *et al.* A new confirmatory Neisseria gonorrhoeae real-time PCR
651 assay targeting the porA pseudogene. *Eur J Clin Microbiol Infect Dis* **23**, 705-
652 710, doi:10.1007/s10096-004-1170-0 (2004).
653 40 Whiley, D. M. *et al.* A real-time PCR assay for the detection of Neisseria
654 gonorrhoeae in genital and extragenital specimens. *Diagn Microbiol Infect Dis* **52**,
655 1-5, doi:10.1016/j.diagmicrobio.2004.12.011 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

656 41 Diemert, D. J., Libman, M. D. & Lebel, P. Confirmation by 16S rRNA PCR of the
657 COBAS AMPLICOR CT/NG test for diagnosis of Neisseria gonorrhoeae infection
658 in a low-prevalence population. *J Clin Microbiol* **40**, 4056-4059,
659 doi:10.1128/jcm.40.11.4056-4059.2002 (2002).
660 42 Van Dyck, E., Smet, H., Van Damme, L. & Laga, M. Evaluation of the Roche
661 Neisseria gonorrhoeae 16S rRNA PCR for confirmation of AMPLICOR PCR-
662 positive samples and comparison of its diagnostic performance according to
663 storage conditions and preparation of endocervical specimens. *J Clin Microbiol*
664 **39**, 2280-2282, doi:10.1128/JCM.39.6.2280-2282.2001 (2001).
665 43 Chang, H. H. *et al.* Origin and proliferation of multiple-drug resistance in bacterial
666 pathogens. *Microbiol Mol Biol Rev* **79**, 101-116, doi:10.1128/MMBR.00039-14
667 (2015).
668 44 Lehtinen, S., Blanquart, F., Lipsitch, M., Fraser, C. & with the Maela
669 Pneumococcal, C. On the evolutionary ecology of multidrug resistance in
670 bacteria. *PLoS Pathog* **15**, e1007763, doi:10.1371/journal.ppat.1007763 (2019).
671 45 Goldstein, E. *et al.* Factors related to increasing prevalence of resistance to
672 ciprofloxacin and other antimicrobial drugs in Neisseria gonorrhoeae, United
673 States. *Emerg Infect Dis* **18**, 1290-1297, doi:10.3201/eid1808.111202 (2012).
674 46 Whittles, L. K., White, P. J., Paul, J. & Didelot, X. Epidemiological Trends of
675 Antibiotic Resistant Gonorrhoea in the United Kingdom. *Antibiotics (Basel)* **7**,
676 doi:10.3390/antibiotics7030060 (2018).
677 47 Centers for Disease, C. & Prevention. Update to CDC's sexually transmitted
678 diseases treatment guidelines, 2006: fluoroquinolones no longer recommended
679 for treatment of gonococcal infections. *MMWR Morb Mortal Wkly Rep* **56**, 332-
680 336 (2007).
681 48 Hemarajata, P., Yang, S., Soge, O. O., Humphries, R. M. & Klausner, J. D.
682 Performance and Verification of a Real-Time PCR Assay Targeting the gyrA
683 Gene for Prediction of Ciprofloxacin Resistance in Neisseria gonorrhoeae. *J Clin*
684 *Microbiol* **54**, 805-808, doi:10.1128/JCM.03032-15 (2016).
685 49 Unemo, M. & Dillon, J. A. Mitigating the emergence and spread of multidrug- and
686 extensively drug-resistant gonorrhea: is there sufficient support in resource-poor
687 settings in Africa? *Sex Transm Dis* **41**, 238-239,
688 doi:10.1097/OLQ.0000000000000117 (2014).
689 50 Bazzo, M. L. *et al.* First nationwide antimicrobial susceptibility surveillance for
690 Neisseria gonorrhoeae in Brazil, 2015-16. *J Antimicrob Chemother* **73**, 1854-
691 1861, doi:10.1093/jac/dky090 (2018).
692 51 Mortimer, T. D. & Grad, Y. H. Applications of genomics to slow the spread of
693 multidrug-resistant Neisseria gonorrhoeae. *Ann N Y Acad Sci* **1435**, 93-109,
694 doi:10.1111/nyas.13871 (2019).
695 52 Dona, V., Low, N., Golparian, D. & Unemo, M. Recent advances in the
696 development and use of molecular tests to predict antimicrobial resistance in
697 Neisseria gonorrhoeae. *Expert Rev Mol Diagn* **17**, 845-859,
698 doi:10.1080/14737159.2017.1360137 (2017).
699 53 Goldstein, E., Pitzer, V. E., O'Hagan, J. J. & Lipsitch, M. Temporally Varying
700 Relative Risks for Infectious Diseases: Implications for Infectious Disease

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

701 Control. *Epidemiology* **28**, 136-144, doi:10.1097/EDE.0000000000000571
702 (2017).
703 54 European Centre for Disease Prevention and Control. ECDC strategic framework
704 for the integration of molecular and genomic typing into European surveillance
705 and multi-country outbreak investigations. (2019).
706 55 Brown, E., Dessai, U., McGarry, S. & Gerner-Smidt, P. Use of Whole-Genome
707 Sequencing for Food Safety and Public Health in the United States. *Foodborne*
708 *Pathog Dis* **16**, 441–450, doi:10.1089/fpd.2019.2662 (2019).
709 56 Břinda, K. *et al.* Rapid heuristic inference of antibiotic resistance and
710 susceptibility by genomic neighbor typing. *bioRxiv*,
711 doi:http://dx.doi.org/10.1101/403204 (2019).
712 57 Merker, M. *et al.* Evolutionary history and global spread of the Mycobacterium
713 tuberculosis Beijing lineage. *Nat Genet* **47**, 242-249, doi:10.1038/ng.3195 (2015).
714 58 Casali, N. *et al.* Evolution and transmission of drug-resistant tuberculosis in a
715 Russian population. *Nat Genet* **46**, 279-286, doi:10.1038/ng.2878 (2014).
716 59 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation
717 using MinHash. *Genome Biol* **17**, 132, doi:10.1186/s13059-016-0997-x (2016).
718 60 Lees, J. A. *et al.* Fast and flexible bacterial genomic epidemiology with
719 PopPUNK. *Genome Res* **29**, 304-316, doi:10.1101/gr.241455.118 (2019).
720 61 Berenger, B. M. *et al.* Genetic Characterization and Enhanced Surveillance of
721 Ceftriaxone-Resistant Neisseria gonorrhoeae Strain, Alberta, Canada, 2018.
722 *Emerg Infect Dis* **25**, 1660-1667, doi:10.3201/eid2509.190407 (2019).
723 62 Eyre, D. W. *et al.* Detection in the United Kingdom of the Neisseria gonorrhoeae
724 FC428 clone, with ceftriaxone resistance and intermediate resistance to
725 azithromycin, October to December 2018. *Euro Surveill* **24**, doi:10.2807/1560-
726 7917.ES.2019.24.10.1900147 (2019).
727 63 Bhattacharyya, R. P. *et al.* Simultaneous detection of genotype and phenotype
728 enables rapid and accurate antibiotic susceptibility determination. *Nat Med* **25**,
729 1858-1864, doi:10.1038/s41591-019-0650-9 (2019).
730 64 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its
731 applications to single-cell sequencing. *J Comput Biol* **19**, 455-477,
732 doi:10.1089/cmb.2012.0021 (2012).
733 65 Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment
734 tool for genome assemblies. *Bioinformatics* **29**, 1072-1075,
735 doi:10.1093/bioinformatics/btt086 (2013).
736 66 Li, H. Aligning sequence reads, clone sequences and assembly contigs with
737 BWA-MEM. *arXiv e-prints*, doi:arXiv:1303.3997 (2013).
738 67 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant
739 detection and genome assembly improvement. *PLoS One* **9**, e112963,
740 doi:10.1371/journal.pone.0112963 (2014).
741 68 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local
742 alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-
743 2836(05)80360-2 (1990).
744 69 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
745 throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.946533; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

746 70 Kawa, D., Lu, S.-D. & Dailey, P. Oligonucleotides, methods and kits for detecting
747 Neisseria Gonorrhoeae. WO patent EP1697541B1 (2013).
748 71 Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of
749 recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids*
750 *Res* **43**, e15, doi:10.1093/nar/gku1196 (2015).
751 72 Sommer, M. O. A., Munck, C., Toft-Kehler, R. V. & Andersson, D. I. Prediction of
752 antibiotic resistance: time for a new preclinical paradigm? *Nat Rev Microbiol* **15**,
753 689-696, doi:10.1038/nrmicro.2017.75 (2017).
754 73 Fritz, S. A. & Purvis, A. Selectivity in mammalian extinction risk and threat types:
755 a new measure of phylogenetic signal strength in binary traits. *Conserv Biol* **24**,
756 1042-1051, doi:10.1111/j.1523-1739.2010.01455.x (2010).
757 74 Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. Fast
758 hierarchical Bayesian analysis of population structure. *Nucleic Acids Res* **47**,
759 5539-5549, doi:10.1093/nar/gkz361 (2019).
760