1

Emergence and evolution of antimicrobial resistance genes and mutations in

Neisseria gonorrhoeae

4

5   Koji Yahara[1*], Kevin C. Ma[2], Tatum D. Mortimer[2], Ken Shimuta[1,3], Shu-ichi Nakayama[3], Aki

6   Hirabayashi[1], Masato Suzuki[1], Michio Jinnai[4], Hitomi Ohya[4], Toshiro Kuroki[4,#a], Yuko

7   Watanabe[4], Mitsuru Yasuda[5], Takashi Deguchi[6], Vegard Eldholm[7], Odile B. Harrison[8],

8   Martin C. J. Maiden[8], Yonatan H. Grad[2], Makoto Ohnishi[3,*]

9   [1] Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, Tokyo, Japan

10 [2] Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health,

11 Boston, Massachusetts, United States of America

12 [3] Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo, Japan

13 [4] Department of Microbiology, Kanagawa Prefectural Institute of Public Health, Kanagawa, Japan

14 [5] Center for Nutrition Support and Infection Control, Gifu University Hospital, Gifu, Japan

15 [6] Department of Urology, Kizawa Memorial Hospital, Gifu, Japan

16 [7] Division of Infection Control and Environmental Health, Norwegian Institute of Public Health,

17 Norway

18 [8] Department of Zoology, University of Oxford, Oxford, UK

19 [#a]Current address: Faculty of Veterinary Medicine, Okayama University of Science, 1-3

20 Ikoinooka, Imabari, Ehime, Japan, 794-8555

21

22 *Corresponding author:

23 Email: k-yahara@nih.go.jp (KY), ohnishi7@niid.go.jp (OM)

Abstract

Antimicrobial resistance in *Neisseria gonorrhoeae* is a global health concern. Strains from two internationally circulating sequence types, ST-7363 and ST-1901, have acquired resistance to treatment with third-generation cephalosporins mainly due to the emergence of mosaic *penA* alleles. These two STs were first detected in Japan; however, when and how the mosaic *penA* alleles emerged and spread to other countries remains unknown. Here, we addressed the evolution of *penA* alleles by obtaining complete genomes from three Japanese ST-1901 clinical isolates harboring mosaic *penA* allele 34 (*penA*-34) dating from 2005 and generating a phylogenetic representation of 1,075 strains sampled from 37 countries. We also sequenced the genomes of 103 Japanese ST-7363 *N. gonorrhoeae* isolates from 1996–2005 and reconstructed a phylogeny including 88 previously sequenced genomes. Based on an estimate of the time of emergence of ST-1901 harboring mosaic *penA-34* and ST-7363 harboring mosaic *penA*-10, and >300 additional genome sequences of Japanese strains representing multiple STs isolated in 1996-2015, we suggest that *penA*-34 in ST-1901 was generated from *penA*-10 via recombination with another *Neisseria* species, followed by a second recombination event with a gonococcal strain harboring wildtype *penA*-1. Following the acquisition of *penA*-10 in ST-7363, a dominant sub-lineage rapidly acquired fluoroquinolone resistance mutations at GyrA 95 and ParC 87-88, possibly due to independent mutations rather than horizontal gene transfer. Literature data suggest the emergence of these resistance determinants may reflect selection from the standard treatment regimens in Japan at that time. Our findings highlight how recombination and antibiotic use across and within *Neisseria* species intersect in driving the emergence and spread of drug-resistant gonorrhea.

Keywords: recombination, horizontal gene transfer, genomic epidemiology, antimicrobial resistance, phylogeny, surveillance, evolution, *Neisseria gonorrhoeae*

Author summary

Antimicrobial resistance is recognized as one of the greatest threats to human health, and *Neisseria gonorrhoeae* resistance is classified as one of the most urgent. The two major internationally spreading lineages resistant. to first line drugs likely originated in Japan, but when and how their genetic resistance determinants emerged remain unknown. In this study, we conducted an evolutionary analysis using clinical *N. gonorrhoeae* isolates from 37 countries, including a historical collection of Japanese isolates, to investigate the emergence of resistance in each of the two major lineages. We showed that the *penA* allele responsible for resistance to cephalosporins, the first-line treatment for gonorrhea, was possibly generated by two recombination events, one from another *Neisseria* species and one from another *N. gonorrhoeae* lineage. We also showed that mutations responsible for resistance to a previously widely used antibiotic treatment occurred twice independently in one of the two major lineages. The emergence of the genetic resistance determinants potentially reflects selection from the standard treatment regimen at that time. Our findings highlight how recombination (horizontal gene transfer) and antibiotic use across and within a bacterial species intersect in driving the emergence and spread of antimicrobial resistance genes and mutations.

Introduction

69

70   Antimicrobial resistance (AMR) is one of the greatest threats to human health, urgently

71   requiring effective global surveillance [1]. Gonorrhea is one of the most common sexually

72   transmitted bacterial infections worldwide causing substantial morbidity and economic loss [2-

73   4], with antimicrobial resistance an increasing concern. Among antimicrobial resistance

74   mechanisms, *Neisseria gonorrhoeae* resistance to third-generation cephalosporins (3GCs) and

75   fluoroquinolones has been defined as "high" priority by the World Health Organization (WHO).

76   Among 3GCs, cefixime is no longer recommended for single-dose treatment, leaving

77   ceftriaxone as the only remaining option for empirical first-line monotherapy in most countries

78   [5]. Reduced susceptibility to cefixime and ceftriaxone is mainly caused by mutations in *penA*,

79   the gene encoding the 3GC target, penicillin-binding protein 2 (PBP2). Mosaic *penA* alleles are

80   deemed 'mosaic,' as the 3′ segments of their DNA sequences have been imported from other

81   *Neisseria spp.* via homologous recombination events [6]. Many mosaic alleles have been

82   documented, reflecting both distinct recombination events and additional mutations [7], and

83   many of these mosaic alleles increase resistance to 3GCs.

84   Two internationally-disseminated sequence types (STs), defined by *Neisseria* multilocus

85   sequence typing (MLST), ST-7363 and ST-1901, have acquired mosaic *penA* alleles and can

86   be resistant to 3GCs. ST-1901, harboring mosaic *penA*-34, accounted for the majority of

87   isolates with reduced susceptibility to 3GCs in the USA and Europe from the 2000s to at least

88   the early 2010s [8, 9]. In Japan, ST-7363 accounted for the highest proportion (66%) among

89   149 isolates with reduced susceptibility to cefixime, isolated from 1998 to 2005 [10], and for

90   21% cases among 90 isolates obtained in 2015 [11]. These two STs harbor the mosaic *penA*

91   alleles 34 and 10, which have identical nucleotide and amino acid sequences, except for those

92   at the C-terminus (encoding 33 amino acids, approximately 6% of the entire sequence) [7, 11].

93   Both STs likely originated in Japan [12]. Recently, we analyzed whole genome sequence data

94   combined with antimicrobial susceptibility testing results from 204 isolates from genomic

95   surveillance in 2015 of a region where the first extensively drug resistant (XDR) *N.*

96   *gonorrhoeae* resistant to ceftriaxone was isolated. This analysis was complemented with data

97   from 67 further genomes from other time frames (from 1996 to 2015) and locations within

98   Japan [11]. We first clustered STs that were closely related at the core-genome level to ST-

99   1901 and ST-7363, resulting in ST-1901-associated and ST-7363-associated core genome

100   groups. We found distinct evolutionary pathways of mosaic *penA* acquisition for these two

101   core-genome groups: the ST-7363-associated core-genome group acquired *penA*-10 once,

102   whereas the ST-1901-associated core-genome group had multiple independent acquisitions of

103   *penA*-10 and *penA*-34. The previous study analyzed only Japanese *N. gonorrhoeae* isolates;

104   thus, when and how the mosaic *penA* alleles—particularly *penA*-34, which is now dominant in

105   USA and Europe—spread to other countries remained unknown. In addition, the majority of

106   previously analyzed isolates were from 2015, and inclusion of additional genomes from other

107   time-frames did not provide sufficient resolution to identify the time of emergence of the

108   mosaic *penA* alleles with a narrow credibility interval [11]. Although the generation of *penA*-

109   10 was explained by a single horizontal gene transfer (HGT) or recombination event, which

110   imported the 3′ *penA* segment from a commensal *Neisseria spp.*, whether generation of *penA*-

111   34 is similarly explained by a single HGT event or by multiple successive events remained

112   unknown. Furthermore, the distribution of fluoroquinolone-resistance determinants, i.e.,

113   mutations in *gyrA* and *parC* [9], was examined only in strains isolated in 2015; however, when

114   and how those mutations emerged and spread was also unclear.

115   In the present study, we addressed these issues by (i) obtaining complete, closed, genome

116   sequences of three Japanese ST-1901 *N. gonorrhoeae* isolates harboring *penA*-34 from 2005,

117   and (ii) reconstructing a dated phylogeny based on core-genome alignment of 1,075 genome

118   sequences from isolates sampled from 37 countries. We also sequenced the genomes of 103

119   Japanese ST-7363 *N. gonorrhoeae* isolates from 1996–2005 and reconstructed a dated

120   phylogeny of these genomes and 88 previously-sequenced Japanese ST-7363 genomes dating

121 from 1996 to 2015. We further examined how these estimated dates corresponded to

122 contemporary antibiotic use and treatment regimens. This generated a detailed narrative of the

123 emergence and evolution of the cephalosporin-resistance genes and fluoroquinolone-resistance

124 mutations, based on the dated phylogenies of these two global prevalent STs. These

125 observations improve our understanding of how resistance determinants evolve via interactions

126 among lineages and species under the selective pressure of antibiotic use.

127

Results

ST-1901-associated core-genome group and resistance to the 3GCs

130 A clonal dated phylogeny of ST-1901-associated core-genome group gonococci with branch

131 lengths corrected to account for homologous recombination was inferred using

132 ClonalFrameML, followed by BactDating (Fig 1) for 1,075 isolates sampled from 37 countries

133 dating from 1992 to 2016 (S1 Table). Identified core-genome groups corresponded to the

134 recently designated "core-genome group cluster 3", which had been identified using a locus

135 threshold of 400 or fewer locus differences [13] (S1 Table), except for 2 isolates, which

136 belonged to cgc_400 groups 18 and 221. Parameter estimates from ClonalFrameML were

137 consistent with those observed previously [11] (Supplementary text). In general, isolates

138 harboring the mosaic *penA* alleles (colored cyan in Fig 1) and others were separated into two

139 clusters, hereafter named 'sub-lineage 34' and 'susceptible sub-lineage', respectively. The sub-

140 lineage 34 accounts for the majority of isolates encoding mosaic *penA* alleles, and harbors

141 *penA*-34 and its variants (highlighted in the 1st column titled "sub-lineage 34" of Fig 1 in

142 yellow). This lineage was estimated to have emerged (red circle in the tree) between May 1990

143 and July 1999 (95% credibility interval). The sub-lineage 34 is shown as an enlargement in

144 addition to its phylogenetic neighbor located at the bottom in S1 Fig. As shown in the 2nd

145 column headed "continent" in Fig 1, most strains were isolated from North America (light pink,

146 50.6%) and Europe (pink, 42.6%) mainly through the Gonococcal Isolate Surveillance Project

147 (GISP) [8] (accounting for 80.1% of the strains isolated from North America) and the European

148 Gonococcal Antimicrobial Surveillance Programme (Euro-GASP) [14] (accounting for 64.3%

149 of the strains isolated from Europe), respectively, while very few strains were isolated from

150 Asia (blue, 3.7%), Oceania (brown, 2.3%), and South America (light blue, 0.9%)In the sub-

151 lineage 34, strains with the oldest isolation date were from 2005, among which 3 were from

152 Japan and 2 from the USA. Complete genome sequences of the three Japanese strains were

153 phylogenetically located at the base of the sub-lineage 34 tree (highlighted in the 3rd column

154 "complete genome" of Fig 1 in blue). The three Japanese isolates (GU029, GU058, and

155 GU092) were collected in the Japanese prefectures of Aichi and Gifu in February–March 2005,

156 indicating that none of them was likely closely related to an ancestor of the strains that spread

157 to other countries.

158   To provide an outgroup for sub-lineage 34, we obtained the complete genome sequence of

159 an isolate, DRR129099 (12-032 in our previous study [11]), from the Japanese prefecture of

160 Kanagawa in July 2000. This isolate harbors a non-mosaic *penA* allele and is phylogenetically

161 close to the sub-lineage 34 (highlighted in green in the 3rd column headed "complete genome"

162 of Fig 1). Alignment of the complete genomes of the four isolates and the reference isolate

163 WHO_Y (F89) included in the sub-lineage 34 (highlighted in pink in the 3rd column headed

164 "complete genome" of Fig 1), revealed conserved genomic synteny among the reference isolate

165 in the sub-lineage 34, DRR129099 harboring a non-mosaic *penA*-5 allele, and one of the

166 isolates (GU092) encoding *penA*-34 (1st, 4th, and 5th genome in S2 Fig). This indicated that the

167 overall genomic structure was maintained during evolution from the ancestor of the isolates in

168 the sub-lineage 34 that spread to other countries.

169   In the whole ST-1901-associated core-genome group, ST-1901 and its single locus variants,

170 ST7360 and ST1579, are colored in the "ST" column in Fig 1. The 91 (8.5%) other strains

171 comprised 21 STs. The top five STs (ST-9365, ST-10312, ST-8153, ST-10241, and ST-13840)

172 accounted for 62.6% of the samples (S3 Fig) and belonged to the ST-1901-associated core-

173 genome group, although the nucleotide sequences of the seven loci differed from those of ST-

-7-

174 1901. Similarly, in the sub-lineage 34, 51.7% of the strains were NG-MAST1407 (colored in

175 the "NG-MAST1407" column in Fig 1), while the other strains were classified into 274 types

176 or as undetermined according to the database of nucleotide sequences of the two highly variable

177 NG-MAST loci (*porB* and *tbpB*).

178   Most isolates harboring the 'non-mosaic' *penA*-5 were located in the susceptible sub-

179 lineage, whereas other isolates harboring *penA*-5 were present immediately outside the sub-

180 lineage 34, suggesting that *penA*-5 was ancestral and that *penA*-34 evolved from it (Fig 1).

181   The sub-lineage harboring *penA*-34 and its variants appeared to be monophyletic, whereas

182 the entire phylogeny of ST-1901-associated core-genome group included three separate mosaic

183 *penA*-10 allele acquisitions, one of which was observed in the present study to be

184 phylogenetically close to sub-lineage 34. This included the oldest isolates harboring *penA*-10

185 from 2000–2001 in Japan [15] (S1 Fig), which were not included in the previous study [11].

186 This suggests that *penA*-34 was generated via *penA*-10 from the ancestral *penA*-5 allele.

187   A small cluster of isolates encoding mosaic *penA*-35 was found in the sub-lineage 34, which

188 consisted of 27 isolates that were isolated in India, Canada, UK, and USA from 2008 to 2011,

189 and were not resistant to cefixime and ceftriaxone. A comparison of the amino acid sequences

190 of *penA*-34 and 35 indicated that the differences in the amino acid residues across the sequence

191 were likely consequences of HGT events spanning the entire *penA* locus from an unknown

192 source outside the sub-lineage 34 (S4 Fig).

193 **Dated phylogeny of ST-7363-associated core-genome group, susceptibility to the 3GCs,**

194 **and *penA* alleles**

195 A clonal dated phylogeny of ST-7363-associated core-genome group, corresponding to

196 recently designated "core-genome group cluster 8" [13] (Fig 2), showed that the isolates

197 broadly separated into two clusters, corresponding to the presence or absence of a mosaic *penA*

198 allele. In contrast to the ST-1901-associated core-genome group, we observed that the majority

199 of the ancestral non-mosaic *penA* allele was *penA*-2, the nucleotide sequence of which is

200 identical to that of *penA*-5 except for the region from nucleotides 1598–1752, the 3′ terminus

201 of the coding sequence (*penA*-2 was also referred to as "*penA*-5 variant" [11]). Consistent with

202 the results of our previous report, the majority of the mosaic *penA* was *penA*-10, with the

203 exception of isolate H041 which was resistant to ceftriaxone [7], and one isolate encoding

204 *penA*-34 (ERS311596, isolated in 2001) [15]. The phylogeny showed that *penA*-10 in ST-7363

205 was generated from the ancestral *penA*-2 (Fig 2) (95% credibility interval September

206 1991–May 1995). In agreement with a prior observation [11], we noted that *penA*-150

207 (designated in the NG-STAR database, Fig 2), susceptible to 3GCs, appeared to be a product

208 of the recombination between the mosaic *penA*-10 or 37 (H041-type) alleles and *penA*-5.

209 **Proposed origin of mosaic *penA*-34 and 10 alleles**

210 To further analyze the origins of mosaic *penA*-34 and *penA*-10, we compared their nucleotide

211 sequences to those of the susceptible alleles in the ST-1901 and ST-7363-associated core-

212 genome groups. Nucleotide sequence alignment of *penA*-5, 10, and 34 and their downstream

213 sequences in ST-1901 is shown schematically in Fig 3, and at the nucleotide level in S5 Fig.

214 The 1st to 293rd nucleotides (region (1) in Fig 3A) were identical, followed by the recombined

215 mosaic region from 294th nucleotide (region (2), indicated by the red line in *penA*-10 in Fig

216 3A, and between red arrows in S5 Fig), in which sequence identity between *penA*-5 and *penA*-

217 10 with their downstream sequences was 89.0%, whereas sequence identity between *penA*-10

218 and *penA*-34 with their downstream sequences was 98.2%. The coding sequences (CDS) of

219 *penA*-10 and 34 were identical, with the exception of 105 bp at the 3′ end, indicated by the left

220 part of the orange line at the end of *penA*-34 in Fig 3A. In the following region (3), the

221 downstream sequences of *penA*-5 and *penA*-10 were identical, whereas those of *penA*-10 and

222 34 harbored four individual polymorphisms and 99.6% sequence identity. In the following

223 region (4), the nucleotide sequences of *penA*-5, 10, and 34 exhibited ≥ 99.9% sequence identity.

224 Based on the results of nucleotide sequence comparison, the origin of *penA*-10 was most

225 parsimoniously explained by an approximately 2.4 kb recombination event from the start of

226 mosaic region (2) in a *penA*-5 background. The source is unknown, as BLASTn search against

227 the NCBI nr database did not yield any hit in other *Neisseria* species with high (> 95%)

228 sequence identity.

229 PenA 34 and 10 differed in the C-terminus, and the difference continued in their

230 downstream nucleotide sequences in the regions (2) and (3). To investigate the possibility that

231 *penA*-34 emerged via HGT in the background of *penA*-10, we aligned the nucleotide sequence

232 corresponding to the orange line in regions (2) and (3) (approximately 2.5 kb, between orange

233 arrows in S5 Fig) to the genome sequences of 140 strains isolated in 1996–1997 in the Japanese

234 prefecture of Kanagawa (S3 Table, most of them newly sequenced in the present study) and

235 the 204 Japanese strains isolated in 2015 [11], representing multiple STs. A maximum

236 likelihood tree constructed from this alignment (S7 Fig) revealed that ST-1901 isolates

237 encoding *penA*-34 and its variant were clustered with ST-1594 strains encoding *penA*-1. In

238 particular, the nucleotide sequence was identical between *penA*-34 and *penA*-1, with the

239 exception of a single polymorphism whereas the *penA*-34 variant harbored a polymorphism at

240 the C-terminus. The next most similar sequence was found in a strain (09-021) harboring *penA*-

241 2 (located next to the ST-1594 strains encoding *penA*-1 in S7 Fig) and containing five

242 polymorphisms compared to the sequence of *penA*-1. These results suggested that the origin of

243 *penA*-34 can be explained by two recombination events: first, a recombination event from a

244 commensal *Neisseria* into a *penA*-5 background, resulting in *penA*-10 (red line in *penA*-10 in

245 Fig 3A); second, a recombination event from *penA*-1 into the *penA*-10 background resulting in

246 *penA*-34 (orange line in *penA*-34 in Fig 3A and orange rectangle in Fig 4).

247 Although no study has reported the prevalence of the *penA*-1 allele in Japan, analysis of

248 genome sequence data of the 140 isolates in 1996–1997 and those of 204 isolates in 2015

249 collected via surveillance of symptomatic cases, showed that the *penA*-1 allele was encoded

250 only in ST-1594 that accounted for 3.6% and 7.8% cases (the 7th and 5th most frequent ST)

251 among the collected strains in 1996–1997 and 2015, respectively.

252 Similar to ST-1901, *penA*-2 and 10 of ST-7363 exhibited identical nucleotide sequence

-10-

253 from nucleotide 1 to 293, followed by the mosaic region from nucleotide 294 to 390 bp

254 downstream of the end of *penA* (region (2)', indicated by the red line in Fig 3B and between

255 red arrows in S6 Fig). This is most parsimoniously explained by a recombination event,

256 although the source of this DNA segment is unknown (Fig 4). Comparison of the length of the

257 recombined regions between ST-1901 and ST-7363 indicated that *penA*-10 in ST-7363 was

258 unlikely to be the origin of *penA*-10 in ST-1901, as the recombined fragment was shorter than

259 that of ST-1901 (Fig 4).

260 **Analysis of susceptibility to fluoroquinolones and examination of mutations in *gyrA* and**

261 ***parC* in the dated phylogenies**

262 For fluoroquinolone resistance, the dated phylogenies also revealed distinct evolutionary

263 paths for the ST-1901- and ST-7363-associated core-genome groups. Almost all isolates in

264 both core-genome groups were resistant to fluoroquinolones (minimal inhibitory

265 concentration (MIC) > 0.06 µg/mL, using the EUCAST breakpoint) and harbored the GyrA

266 91F substitution. In the ST-1901-associated core-genome group, the substitutions at GyrA

267 D95G and ParC S87R were shared among most isolates, including the clade encoding the

268 non-mosaic *penA* (Fig 1). The substitution at ParC S88P was found only in a small cluster of

269 11 strains (bottom of Fig 1). In contrast, in the ST-7363-associated core-genome group, a

270 dominant sub-lineage contained amino acid substitutions at GyrA D95GN, ParC S87R, and

271 ParC S88P, which arose within a short time (highlighted as a green horizontal line in Fig 2;

272 95% credibility interval August 1996–March 1998). These substitutions arose after the

273 acquisition of *penA*-10, which is inferred to have occurred on a more ancestral node: the 95%

274 credibility interval for the dates of these two events do not overlap. The output of

275 ClonalFrameML used for constructing the clonal phylogeny did not include a signature of the

276 recombination importing the GyrA and ParC substitutions, suggesting that they arose via

277 independent mutations.

278

Discussion

An extensive analysis of *N. gonorrhoeae* ST-1901 and ST-7363-associated core-genome groups, including the 1,075 isolates sampled from 37 countries for ST-1901 and 191 Japanese isolates sampled for ST-7363, has improved the resolution of the time-resolved phylogenies of these important gonococci. Newly sequenced, historical, Japanese isolates included three ST-1901 gonococci harboring the mosaic *penA*-34 (NG-MAST1407), which accounted for the majority of isolates with reduced susceptibility to 3GC in the USA and Europe. In the ST-1901-associated core-genome group, the three earliest isolates were obtained in Japan from February–March, 2005, while two originated in the USA (GCGS0944, GCGS0920) in 2005 (S1 Fig). The time-resolved tree suggested that the three Japanese isolates were phylogenetically closest to the unsampled ancestor, consistent with the hypothesis that this sub-lineage likely originated in Japan [12].

The time-resolved trees of ST-1901 and ST-7363-associated core-genome groups enabled us to estimate the timing of the emergence of three genetic determinants of antimicrobial resistance: (i) mosaic *penA*-34 in ST-1901 (May 1990–July 1999); (ii) mosaic *penA*-10 in ST-1901 (April 1986–August 1995) and in ST-7363 (September 1991–May 1995); and (iii) GyrA D95GN and ParC S87R-S88P in ST-7363 (August 1996–March 1998).

How do these estimated dates correlate to antibiotic use? Although there were no national data on antibiotic usage at that time in Japan, a 1999 survey on the use of antimicrobial drugs in 14 hospitals and 4 clinics mostly in a prefecture in Japan showed that 50% cases (9 out of 18) exclusively used fluoroquinolones for the treatment of urethritis (Dr. Mitsuru Yasuda, unpublished data). Similarly, another survey of clinics in Fukuoka in Japan in the late 1990s reported that "Fluoroquinolones were most frequently used (46%) as first-line treatment for gonorrhea, followed by penicillin with or without a β-lactamase inhibitor (28%), cephems (16%), and others (10%). Of the various fluoroquinolones available in Japan, a 7-day course of levofloxacin (200 mg twice a day or 100 mg three times a day) was most frequently used in the treatment of gonococcal infections (M. Tanaka, unpublished data)" [16]. The regimen used

-12-

307   in Japan in the late 1990s exceeded the single ciprofloxacin dose recommended in the USA,

308   and may have contributed to selection of the *gyrA* and *parC* mutations, resulting in increased

309   fluoroquinolone resistance.

310   A survey in Fukuoka, Japan, also showed that in the 1990s, fluoroquinolones were most

311   frequently used (46%) as the first-line of treatment for gonorrhea, followed by penicillin with

312   or without a β-lactamase inhibitor (28%), and cephems (16%) [16]. In 1999, the reported MIC_{90}

313   for cefixime for clinical isolates of *N. gonorrhoeae* was 0.06 to 0.125 μg/mL, which indicated

314   the presence of isolates with cefixime MIC = 0.125 μg/mL. This prompted the conclusion that

315   a single dose of 400 mg cefixime used in the USA would not eradicate ≤ 95% of the clinical

316   strains of *N. gonorrhoeae* [17]. The first Japanese STI treatment guidelines in 1999 described

317   a regimen of six 100 mg doses of cefixime at a 12-h interval. Between 1999 and 2001, a

318   regimen of 200 mg cefixime at 6 h interval was proposed, although it eradicated only 88.2%

319   and 54.5% *N. gonorrhoeae* isolates with MIC_{90} and MIC values of 0.125 μg/mL, respectively

320   [17]. These regimens and the observed circulating resistance might have contributed to the

321   selection and dissemination of gonococcal strains with reduced susceptibility to cephalosporins.

322   Although information regarding the alleles responsible for the cefixime MIC of 0.125 μg/mL

323   at that time is lacking, the rising MICs and shifting patterns of antibiotic use reflect the dynamic

324   interplay between the evolution of cefixime resistance and shifting treatment strategies.

325   Regarding the GyrA D95GN and ParC S88P amino acid substitutions in ST-7363,

326   responsible for fluoroquinolone resistance, it is interesting to note that a sub-lineage (Fig 2)

327   harboring two substitutions (GyrA D95GN and ParC S87R), but not at ParC 88, has not been

328   sampled since 2005 (except for the GU294 strain isolated in 2015 as indicated by the long

329   branch at the bottom of Fig 2). The apparent loss of this sub-lineage suggests the importance

330   of the substitution at ParC 88 or another mutation that became dominant in ST-7363.

331   Alternatively, the stochasticity of transmission or some other factor (perhaps susceptibility to

332   local treatments) might have led to the loss of this sub-lineage.

333   We propose two potential recombination events for the generation of *penA*-34 in ST-1901:

334   recombination with another *Neisseria* species that generated *penA*-10, followed by another

335   recombination event with a strain harboring *penA*-1 that converted *penA*-10 to *penA*-34. The

336   lengths of the recombined sequence were approximately 2.3 kb and 2.5 kb, respectively, which

337   is consistent with a recent estimate (2.5 kb) of the mean of the geometrically distributed DNA

338   tract lengths transferred between donors and recipients in *N. gonorrhoeae* [18]. Similarly,

339   recombination with a donor susceptible to 3GC was recently reported in *N. gonorrhoeae* in

340   Japan [11] and the USA [8, 19], although it led to loss of the resistance phenotype. It should be

341   noted that the inference of recombination is based on sampling, and we cannot rule out the

342   possibility that recombination with a currently unsampled strain might have generated *penA*-

343   34. Nevertheless, we can speculate that a strain harboring *penA*-1 is currently the most likely

344   donor of the recombination that generated *penA*-34 from *penA*-10.

345   The reason behind the dominance of ST-1901 harboring *penA*-34, but not ST-7363

346   harboring the same allele, remains unsolved. In the ST-7363 group, *penA*-10 was dominant,

347   and there was only one isolate with *penA*-34, identified in 2001. Data indicating significant

348   difference in fitness between *penA*-10 and 34 in vitro are lacking. Possibly, the differences in

349   fitness, selection, and dissemination between *penA*-10 and 34 depended on the genetic

350   background of those sequence types and environmental factors at the time they were selected.

351   Unfortunately, data regarding the potential environmental factors such as national or

352   regional information of antimicrobial use at the time of their emergence and dissemination are

353   not accessible. Further studies are therefore warranted to prospectively collect such data and

354   isolates, and conduct integrative analyses of genome sequences, antimicrobial susceptibility,

355   and the environmental factors in order to monitor, understand, and control emergence and

356   dissemination of new resistance determinants of public health importance. A recent study that

357   analyzed the genome and antimicrobial susceptibility data of *N. gonorrhoeae* isolates from

358   1928 to 2013 in Denmark [20], reported that no isolate was interpreted to be ciprofloxacin-

359   resistant (MIC > 0.06 μg/mL) until the 1980s. The percentage of isolates resistant to

360   ciprofloxacin increased from 0 to 14.3% in 1990s. Five isolates from the 1950s–1970s

361   contained a GyrA S91T amino acid substitution, although all of them were susceptible to

362   ciprofloxacin. Compared to the recent study, the isolates analyzed in our study were all

363   collected after 1990: the oldest strain of the ST-1901-associated core-genome group was

364   collected in 1992, and that of ST-7363 in 1996. Further studies are warranted to conduct

365   genome sequencing of historical Japanese strains collected before 1990s to identify isolates

366   phenotypically susceptible to fluoroquinolones and explore when and how the individual amino

367   acid substitutions of GyrA S91F, D95G, and ParC S87R in ST-1901 and that of GyrA S91F in

368   ST-7363 may have occurred.

369   Our previous study revealed that the different evolutionary pathways of the two major core-

370   genome groups regarding the mosaic *penA* alleles responsible for resistance to 3GC [11]. The

371   present study increased our understanding by elucidating the following three major points: 1)

372   *penA*-34 in ST-1901 was likely generated from *penA*-10 via two recombination events; 2) the

373   *penA*-10 allele in the ST-1901-associated core-genome group emerged in at least three distinct

374   Japanese sub-lineages independently, one of which was phylogenetically adjacent to the sub-

375   lineage harboring *penA*-34; and, 3) the *penA*-10 allele in ST-7363, possibly generated via

376   recombination with another *Neisseria* species, was unlikely to be a source of HGT that

377   generated *penA*-10 in ST-1901-associated core-genome group, as the recombined fragment was

378   shorter than that of ST-1901-associated core-genome group. The single acquisition of *penA*-10

379   in ST-7363 was revealed in both the previous and present studies, although our understanding

380   regarding its relationship with *penA*-10 and *penA*-34 in the ST-1901-associated core-genome

381   group in terms of order of their generation and possibility of recombination between the two

382   core-genome groups were improved in the present study. Furthermore, the present study

383   demonstrated another interesting difference in fluoroquinolone resistance between the two

384   core-genome groups since 1990s; ST-7363 was originally susceptible to 3GC and harbored

385   amino acid substitution only in GyrA 91 in the 1990s; the simultaneous amino acid

386   substitutions at GyrA 95 and ParC 87 occurred in the sub-lineage in the short time after it

387   acquired the mosaic *penA*-10, whereas in the ST-1901-associated core-genome group, all the

388 substitutions in GyrA 91, 95, and ParC 87 were already observed in 1990s in most strains

389 (99.2%, 1066/1075), including the sub-lineage mostly susceptible to 3GC (at the bottom in Fig

390 1).

391   In summary, after combining the previously published dataset from 37 countries with the

392 new genome sequence data and the antimicrobial susceptibility data of historical gonococcal

393 isolates from Japan, we described the possible pathways of emergence of cephalosporin-

394 resistant genes and fluoroquinolone-resistant mutations of two globally circulating *N.*

395 *gonorrhoeae* core-genome groups. We further discussed the dynamic interplay between the

396 evolution of antibiotic resistance and treatment regimens during time period of the emergence

397 of genetic determinants of antimicrobial resistance. Such elucidation of evolutionary pathways

398 will be useful for understanding and controlling the current and future evolution and spread of

399 the pathogen, and resistance determinants driven by recombination and selective pressure of

400 antibiotic use.

401

Methods

403 **Isolates and DNA sequencing**

404 In total, 1,075 genome sequences from *N. gonorrhoeae* isolates belonging to a ST-1901-

405 associated core-genome group were included (S1 Table). This also comprised three mosaic

406 *penA*-34-harboring ST-1901 isolates from Japan (2 isolates from Aichi and 1 isolate from Gifu

407 prefectures) isolated in February–March 2005 and newly sequenced in this study using

408 MinION and MiniSeq. These three strains were isolated via surveillance by Gifu University

409 that obtained 51 ST-1901 strains (95 strains across various STs) in 2005. In addition, a ST-

410 1901 strain (12-032) isolated in 2000 in Kanagawa, Japan [11] harboring *penA*-5 and

411 susceptible to cephalosporins was re-sequenced using the MinION platform to examine

412 genomic synteny between it and the strains harboring *penA*-34. Five other ST-1901 isolates

413 dating from 2001 to 2004 in Kanagawa, Japan, were sequenced using the Illumina MiSeq

414 platform. These newly sequenced genomes were combined with those of 1,066 publicly

415 available genomes from 14 studies covering 37 countries [8, 11, 14, 15, 19, 21-28]. The

416 publicly available genomes were selected based on a tree constructed in our previous study

417 [29] such that they formed a group at the core-genome level.

418 For MinION sequencing, a MagAttract HMW DNA Kit (Qiagen, Hilden, Germany) was

419 used for isolation of high molecular weight genomic DNA of each isolate, and the Rapid

420 Sequencing Kit (SQK-RAD004) and R9.4 flowcells were used. For MiSeq sequencing, the

421 genomic DNA of each isolate was extracted using a MagNA Pure LC DNA isolation kit on a

422 MagNA pure LC instrument (Roche Diagnostics GmbH, Mannheim, Germany), which was

423 used for Nextera XT library construction and genome sequencing (300 bp paired-end) using an

424 Illumina MiSeq Reagent Kit v3 (600-cycle).

425 For ST-7363, 103 strains that were isolated from 1996 and 2005 in Kanagawa were

426 sequenced using the Illumina MiSeq platform. These data were combined with 88 publicly

427 available genomes [11, 15] (S2 Table).

Antimicrobial susceptibility testing

429 The MICs of 3GCs (ceftriaxone and cefixime) and fluoroquinolones (ciprofloxacin and

430 levofloxacin) of the newly sequenced historical strains isolated in Kanagawa were determined

431 using agar dilution method [30] at the Kanagawa Prefectural Institute of Public Health and Gifu

432 University (S1 and S2 Tables). MICs measurements were repeated for the following strains for

433 which the genotypes and phenotypes were initially discordant, and MICs that better matched

434 the genotypes were subsequently used: GCGS0938 (ceftriaxone) and GCGS0627 (ceftriaxone)

435 in ST-1901 using the Etest method (bioMérieux), and GU250 (cefixime, ceftriaxone), GU431

436 (cefixime, ceftriaxone), and GU478 (cefixime, ceftriaxone) in ST-7363 using the agar dilution

437 method. To define susceptible/resistant phenotypes, the following MIC cut-offs were used

438 according to the European Union Committee on Antimicrobial Susceptibility Testing

439 (EUCAST; www.eucast.org/clinical_breakpoints): susceptibility ≤ 0.125 µg/mL and resistance

440 > 0.125 µg/mL for ceftriaxone and cefixime; susceptibility ≤ 0.03 µg/mL and resistance ≥ 0.06

441 µg/mL for ciprofloxacin. For strains on which MICs of fluoroquinolones were measured only

-17-

442   for levofloxacin by the Gifu University (with names starting by "GU_"), the MIC cut-off of

443   ciprofloxacin was used, as that of levofloxacin is not defined in EUCAST and a previous study

444   has shown that MIC_{50}, MIC_{90}, and MIC range of ciprofloxacin and levofloxacin were similar

445   for a set of 87 isolates [31].

446   **Genome assembly**

447   The Illumina read data (300 bp paired-end reads) of each isolate sequenced using MiSeq were

448   used for de novo assembly using SPAdes [32], or Shovill [33]; SPAdes was used at its core,

449   but speed was improved. The assembled contigs were checked to ensure that the coverage was

450   more than 10× and that total genome size was approximately equal to that of the type strain *N.*

451   *gonorrhoeae* FA1090 (2.0–2.3M bp). The MinION read data of each of the four isolates of ST-

452   1901 were used for de novo assembly using Canu version 1.7 [34]. Errors were corrected using

453   three runs of Pilon for each contig assembled from the MinION reads [35] based on mapping

454   of Illumina reads to the contigs using bowtie2 [36] with "very-sensitive" option, followed by

455   automated circularization of the contigs using Circlator [37] and adjustment of the 1st position

456   of the genome sequence to the start of *dnaA* using an in-house script. The number of contigs

457   and N50 of each isolate are summarized in S1 and S2 Tables. All isolates used in this study

458   can be found on the pubMLST.org/neisseria database where MLST, NG MAST and NG STAR

459   STs were determined along with isolate core genome groups as described previously [13].

460   MLST typing of 158 out of the 1,075 strains did not result in ST-1901, but were included in

461   our dataset, as they formed a large cluster with other ST-1901 strains as a core-genome group.

462   Each genome was annotated using Prokka [38].

463   **Construction of the clonal time-resolved phylogeny of ST-1901-associated and ST-7363-**

464   **associated core-genome groups**

465   As in previous studies [11, 39], pairwise genome alignment between the reference genome and

466   one of the other strains was performed using progressiveMauve [40] for SNP calling and

467   evolutionary analyses in ST-1901 and ST-7363, which enabled the construction of positional

468   homology alignments even for genomes with variable gene content and rearrangement. The

469   complete genome sequences of the "WHO_Y" (F89) and "WHO_X" (H041) strains [41] were

470   used as ST-1901 and ST-7363 references, respectively. The alignments were then combined

471   into a multiple whole genome alignment, in which each position corresponded to that of the

472   reference genome. This approach was validated [11] by comparing with another whole genome

473   alignment using a pipeline [42] that similarly conducts pairwise genome alignment between a

474   reference genome and one of the other genomes using MUMmer [43]. A maximum likelihood

475   tree was constructed using PhyML [44] from the genome alignment containing the 10,844 core

476   SNPs in ST-7363. A maximum likelihood tree was generated from the genome alignment

477   containing the 28,306 core SNPs in ST-1901 using RAxML [45], as PhyML failed to complete

478   the tree generating process. The default setting was used for RAxML, whereas for PhyML, we

479   used the -m GTR -c 4 -a e parameters that indicate the GTR + G4 model of DNA substitution

480   with estimation of the shape parameter of the gamma distribution by maximizing the likelihood.

481   Using this as a starting tree, a clonal phylogeny was constructed with corrected branch lengths

482   to account for homologous recombination using the standard model of ClonalFrameML [46].

483   The alignment sites that were present in at least 70% isolates were used to focus on core and

484   fairly conserved sites with missing frequency ≤ 30%, as validated in our previous study [11].

485   TempEst [47] was used to root the phylogeny by minimizing the sum of the squared residuals

486   from the regression of root-to-tips on sampling year. BactDating [48] was used, in which the

487   significance of the temporal signal in both ST-7363 and ST-1901-associated core-genome

488   groups was confirmed (the date randomization test, $P < 10^{-4}$). BactDating was then run for 10

489   million Markov chain Monte Caro (MCMC) iterations. After completion, the effective sample

490   size (ESS) of all parameters were confirmed as exceeding 400 and trace plots were confirmed

491   as demonstrating well-mixed chains. A time-resolved phylogeny and 95% confidence interval

492   for each branch in the tree was exported.

493   **Analysis of genetic AMR determinants**

494   For 3GC resistance determinants, the *penA* alleles were extracted based on a BLASTn search

495   of the nucleotide sequence of mosaic *penA* alleles-10 and-34 and that of non-mosaic *penA*-5

496 (https://figshare.com/articles/penA_nucleotide_sequence_alignment_penA_5_X_XXXIV_an

497 d_1_/12203858) against each of the other genomes. When the mosaic *penA* alleles extracted

498 based on the BLASTn search were truncated, sequence reads were mapped to the nucleotide

499 sequence of *penA* alleles-10 and-34. The nucleotide sequence identity between the alleles was

500 checked, and the allele numbering defined in NG-STAR database (https://ngstar.canada.ca)

501 [49] was assigned to each *penA* allele ["penA_allele (NG-STAR)" column in S1 and S2 Tables].

502 Fig 1 shows the grouping of "*penA*-34 variants" for alleles that were 100% aligned to *penA*-34

503 and showed sequence identity of < 100% and > 99.5%, which was higher than that observed

504 with *penA*-10. The "*penA*-10 variants" were grouped similarly.

505 For fluoroquinolone resistance determinants, the nonsynonymous substitutions in *gyrA*

506 and *parC* described in a recent review [9] and also examined in our previous study [11] were

507 analyzed. In *N. gonorrhoeae*, quinolone resistance has been attributed to nonsynonymous

508 substitutions in specific regions of *gyrA* and *parC*, namely, amino acid positions 91 and 95 in

509 GyrA, position 87 in ParC, and less frequently positions 86 and 88 in ParC. The presence or

510 absence of the substitutions was detected using PointFinder [50]. The presence or absence of

511 these genetic AMR determinants, as well as the MICs of the antimicrobial drugs for each

512 strain, were illustrated as heat maps using Phandango [51].

513 **Inference of recombination events that generated mosaic *penA*-10 and 34 alleles**

514 To investigate the recombination events that generated the mosaic *penA* alleles, a nucleotide

515 sequence alignment of *penA* and a 5 kb region downstream of its 3′ end was prepared based on

516 a BLASTn search against the genome sequences of ST-1901- and ST-7363-associated core-

517 genome groups. The nucleotide sequences of *penA*-34 and its downstream region were used as

518 query sequences in the BLASTn search. The nucleotide sequences obtained from the BLASTn

519 were aligned using MAFFT v7.245 [52] and manually examined using Jalview [53] to analyze

520 nucleotide sequence identity and detect recombined fragments for *penA*-10 and 34 in ST-1901

521 and *penA*-10 in ST-7363. The BLASTn search was also performed against a custom database

522 of genome sequences of Japanese *N. gonorrhoeae* strains: 140 strains isolated in 1996–1997

523 in the Japanese prefecture of Kanagawa (S3 Table) and 204 strains isolated via genomic

524 surveillance in 2015 in the Japanese prefectures of Kyoto and Osaka [11], both of which were

525 not confined to ST-1901 or ST-7363, but included various STs. The nucleotide sequences

526 obtained using BLASTn were aligned using MAFFT v7.245 [52] and manually examined using

527 Jalview [53], using which an alignment of the potentially recombined fragment was extracted.

528 A maximum likelihood tree of the fragment was constructed using PhyML [44], which was

529 mid-point rooted and visualized using FigTree version 1.4.3. To investigate the potential

530 sources of the recombination, a BLASTn search was performed for each recombined fragment

531 in the NCBI nt database or the custom database of genome sequences of Japanese *N.*

532 *gonorrhoeae* strains.

533

Acknowledgments

537
538
References

541 1.   Sugden R, Kelly R, Davies S. Combatting antimicrobial resistance globally. Nat

542 Microbiol. 2016;1(10):16187. doi: 10.1038/nmicrobiol.2016.187. PubMed PMID: 27670123.

543 2.   Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, et al.

544 Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted

545 Infections in 2012 Based on Systematic Review and Global Reporting. PLoS One.

546 2015;10(12):e0143304. doi: 10.1371/journal.pone.0143304. PubMed PMID: 26646541;

547 PubMed Central PMCID: PMCPMC4672879.

548 3.   WHO. Global incidence and prevalence of selected curable sexually transmitted

549 infections 2008 [cited 2017 May 12]. Available from:

550 http://www.who.int/reproductivehealth/publications/rtis/2008_STI_estimates.pdf.

551 4. WHO. Global action plan to control the spread and impact of antimicrobial

552 resistance in Neisseria gonorrhoeae 2012 [cited 2017 May 12]. Available from:

553 http://whqlibdoc.who.int/publications/2012/9789241503501_eng.pdf.

554 5. Whittles LK, White PJ, Didelot X. Estimating the fitness cost and benefit of

555 cefixime resistance in *Neisseria gonorrhoeae* to inform prescription policy: A modelling

556 study. PLoS Med. 2017;14(10):e1002416. doi: 10.1371/journal.pmed.1002416. PubMed

557 PMID: 29088226; PubMed Central PMCID: PMCPMC5663337.

558 6. Bowler LD, Zhang QY, Riou JY, Spratt BG. Interspecies recombination between the

559 penA genes of Neisseria meningitidis and commensal Neisseria species during the emergence

560 of penicillin resistance in N. meningitidis: natural events and laboratory simulation. J

561 Bacteriol. 1994;176(2):333-7. doi: 10.1128/jb.176.2.333-337.1994. PubMed PMID: 8288526;

562 PubMed Central PMCID: PMCPMC205054.

563 7. Ohnishi M, Golparian D, Shimuta K, Saika T, Hoshina S, Iwasaku K, et al. Is

564 *Neisseria gonorrhoeae* initiating a future era of untreatable gonorrhea?: detailed

565 characterization of the first strain with high-level resistance to ceftriaxone. Antimicrob

566 Agents Chemother. 2011;55(7):3538-45. doi: 10.1128/AAC.00325-11. PubMed PMID:

567 21576437; PubMed Central PMCID: PMCPMC3122416.

568 8. Grad YH, Harris SR, Kirkcaldy RD, Green AG, Marks DS, Bentley SD, et al.

569 Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins,

570 Macrolides, and Fluoroquinolones in the United States, 2000-2013. J Infect Dis.

571 2016;214(10):1579-87. doi: 10.1093/infdis/jiw420. PubMed PMID: 27638945; PubMed

572 Central PMCID: PMCPMC5091375.

573 9. Harrison OB, Clemence M, Dillard JP, Tang CM, Trees D, Grad YH, et al. Genomic

574 analyses of *Neisseria gonorrhoeae* reveal an association of the gonococcal genetic island

575 with antimicrobial resistance. J Infect. 2016;73(6):578-87. doi: 10.1016/j.jinf.2016.08.010.

576 PubMed PMID: 27575582; PubMed Central PMCID: PMCPMC5127880.

577 10. Shimuta K, Watanabe Y, Nakayama S, Morita-Ishihara T, Kuroki T, Unemo M, et

578   al. Emergence and evolution of internationally disseminated cephalosporin-resistant

579   *Neisseria gonorrhoeae* clones from 1995 to 2005 in Japan. BMC Infect Dis. 2015;15:378.

580   doi: 10.1186/s12879-015-1110-x. PubMed PMID: 26381611; PubMed Central PMCID:

581   PMCPMC4574456.

582   11.   Yahara K, Nakayama SI, Shimuta K, Lee KI, Morita M, Kawahata T, et al. Genomic

583   surveillance of Neisseria gonorrhoeae to investigate the distribution and evolution of

584   antimicrobial-resistance determinants and lineages. Microb Genom. 2018;4(8). doi:

585   10.1099/mgen.0.000205. PubMed PMID: 30063202; PubMed Central PMCID:

586   PMCPMC6159555.

587   12.   Unemo M, Nicholas RA. Emergence of multidrug-resistant, extensively drug-

588   resistant and untreatable gonorrhea. Future Microbiol. 2012;7(12):1401-22. doi:

589   10.2217/fmb.12.117. PubMed PMID: 23231489; PubMed Central PMCID:

590   PMCPMC3629839.

591   13.   Harrison OB, Cehovin A, Skett J, Jolley KA, Massari P, Genco CA, et al. Neisseria

592   gonorrhoeae Population Genomics: Use of the Gonococcal Core Genome to Improve

593   Surveillance of Antimicrobial Resistance. J Infect Dis. 2020. doi: 10.1093/infdis/jiaa002.

594   PubMed PMID: 32163580.

595   14.   Harris SR, Cole MJ, Spiteri G, Sanchez-Buso L, Golparian D, Jacobsson S, et al.

596   Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a

597   genomic survey. Lancet Infect Dis. 2018;18(7):758-68. doi: 10.1016/S1473-3099(18)30225-

598   1. PubMed PMID: 29776807; PubMed Central PMCID: PMCPMC6010626.

599   15.   Sanchez-Buso L, Golparian D, Corander J, Grad YH, Ohnishi M, Flemming R, et al.

600   The impact of antimicrobials on gonococcal evolution. Nat Microbiol. 2019;4(11):1941-50.

601   doi: 10.1038/s41564-019-0501-y. PubMed PMID: 31358980; PubMed Central PMCID:

602   PMCPMC6817357.

603   16.   Tanaka M, Nakayama H, Haraoka M, Saika T. Antimicrobial resistance of Neisseria

604   gonorrhoeae and high prevalence of ciprofloxacin-resistant isolates in Japan, 1993 to 1998. J

605   Clin Microbiol. 2000;38(2):521-5. PubMed PMID: 10655338; PubMed Central PMCID:

606   PMCPMC86137.

607   17.   Deguchi T, Yasuda M, Yokoi S, Ishida K, Ito M, Ishihara S, et al. Treatment of

608   uncomplicated gonococcal urethritis by double-dosing of 200 mg cefixime at a 6-h interval. J

609   Infect Chemother. 2003;9(1):35-9. doi: 10.1007/s10156-002-0204-8. PubMed PMID:

610   12673405.

611   18.   Arnold BJ, Gutmann MU, Grad YH, Sheppard SK, Corander J, Lipsitch M, et al.

612   Weak Epistasis May Drive Adaptation in Recombining Bacteria. Genetics.

613   2018;208(3):1247-60. doi: 10.1534/genetics.117.300662. PubMed PMID: 29330348;

614   PubMed Central PMCID: PMCPMC5844334.

615   19.   Grad YH, Kirkcaldy RD, Trees D, Dordel J, Harris SR, Goldstein E, et al. Genomic

616   epidemiology of *Neisseria gonorrhoeae* with reduced susceptibility to cefixime in the USA: a

617   retrospective observational study. Lancet Infect Dis. 2014;14(3):220-6. doi: 10.1016/S1473-

618   3099(13)70693-5. PubMed PMID: 24462211; PubMed Central PMCID: PMCPMC4030102.

619   20.   Golparian D, Harris SR, Sanchez-Buso L, Hoffmann S, Shafer WM, Bentley SD, et

620   al. Genomic evolution of Neisseria gonorrhoeae since the preantibiotic era (1928-2013):

621   antimicrobial use/misuse selects for resistance and drives evolution. BMC Genomics.

622   2020;21(1):116. doi: 10.1186/s12864-020-6511-6. PubMed PMID: 32013864; PubMed

623   Central PMCID: PMCPMC6998845.

624   21.   Costa-Lourenco A, Abrams AJ, Dos Santos KTB, Argentino ICV, Coelho-Souza T,

625   Canine MCA, et al. Phylogeny and antimicrobial resistance in Neisseria gonorrhoeae isolates

626   from Rio de Janeiro, Brazil. Infect Genet Evol. 2018;58:157-63. doi:

627   10.1016/j.meegid.2017.12.003. PubMed PMID: 29225148; PubMed Central PMCID:

628   PMCPMC6932625.

629   22.   Demczuk W, Lynch T, Martin I, Van Domselaar G, Graham M, Bharat A, et al.

630   Whole-genome phylogenomic heterogeneity of Neisseria gonorrhoeae isolates with decreased

631   cephalosporin susceptibility collected in Canada between 1989 and 2013. J Clin Microbiol.

632   2015;53(1):191-200. doi: 10.1128/JCM.02589-14. PubMed PMID: 25378573; PubMed

633   Central PMCID: PMCPMC4290921.

634   23.   Demczuk W, Martin I, Peterson S, Bharat A, Van Domselaar G, Graham M, et al.

635   Genomic Epidemiology and Molecular Resistance Mechanisms of Azithromycin-Resistant

636   Neisseria gonorrhoeae in Canada from 1997 to 2014. J Clin Microbiol. 2016;54(5):1304-13.

637   doi: 10.1128/JCM.03195-15. PubMed PMID: 26935729; PubMed Central PMCID:

638   PMCPMC4844716.

639   24.   Eyre DW, De Silva D, Cole K, Peters J, Cole MJ, Grad YH, et al. WGS to predict

640   antibiotic MICs for Neisseria gonorrhoeae. J Antimicrob Chemother. 2017;72(7):1937-47.

641   doi: 10.1093/jac/dkx067. PubMed PMID: 28333355; PubMed Central PMCID:

642   PMCPMC5890716.

643   25.   Ezewudo MN, Joseph SJ, Castillo-Ramirez S, Dean D, Del Rio C, Didelot X, et al.

644   Population structure of Neisseria gonorhoeae based on whole genome data and its

645   relationship with antibiotic resistance. PeerJ. 2015;3:e806. doi: 10.7717/peerj.806. PubMed

646   PMID: 25780762; PubMed Central PMCID: PMCPMC4358642.

647   26.   Kwong JC, Chow EPF, Stevens K, Stinear TP, Seemann T, Fairley CK, et al.

648   Whole-genome sequencing reveals transmission of gonococcal antibiotic resistance among

649   men who have sex with men: an observational study. Sex Transm Infect. 2018;94(2):151-7.

650   doi: 10.1136/sextrans-2017-053287. PubMed PMID: 29247013; PubMed Central PMCID:

651   PMCPMC5870456.

652   27.   Lee RS, Seemann T, Heffernan H, Kwong JC, Goncalves da Silva A, Carter GP, et

653   al. Genomic epidemiology and antimicrobial resistance of Neisseria gonorrhoeae in New

654   Zealand. J Antimicrob Chemother. 2018;73(2):353-64. doi: 10.1093/jac/dkx405. PubMed

655   PMID: 29182725; PubMed Central PMCID: PMCPMC5890773.

656   28.   Ryan L, Golparian D, Fennelly N, Rose L, Walsh P, Lawlor B, et al. Antimicrobial

657   resistance and molecular epidemiology using whole-genome sequencing of Neisseria

658   gonorrhoeae in Ireland, 2014-2016: focus on extended-spectrum cephalosporins and

659   azithromycin. Eur J Clin Microbiol Infect Dis. 2018;37(9):1661-72. doi: 10.1007/s10096-

660   018-3296-5. PubMed PMID: 29882175.

661   29.   Ma KC, Mortimer TD, Hicks AL, Wheeler NE, Sanchez-Buso L, Golparian D, et al.

662   Adaptation to the cervical environment is associated with increased antibiotic susceptibility

663   in Neisseria gonorrhoeae. Nat Commun. 2020;11(1):4126. doi: 10.1038/s41467-020-17980-1.

664   PubMed PMID: 32807804.

665   30.   Wiegand I, Hilpert K, Hancock RE. Agar and broth dilution methods to determine

666   the minimal inhibitory concentration (MIC) of antimicrobial substances. Nat Protoc.

667   2008;3(2):163-75. doi: 10.1038/nprot.2007.521. PubMed PMID: 18274517.

668   31.   Shigemura K, Okada H, Shirakawa T, Tanaka K, Arakawa S, Kinoshita S, et al.

669   Susceptibilities of Neisseria gonorrhoeae to fluoroquinolones and other antimicrobial agents

670   in Hyogo and Osaka, Japan. Sex Transm Infect. 2004;80(2):105-7. doi:

671   10.1136/sti.2003.006908. PubMed PMID: 15054169; PubMed Central PMCID:

672   PMCPMC1744816.

673   32.   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.

674   SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J

675   Comput Biol. 2012;19(5):455-77. doi: 10.1089/cmb.2012.0021. PubMed PMID: 22506599;

676   PubMed Central PMCID: PMCPMC3342519.

677   33.   Seemann T. Shovill: faster SPAdes assembly of Illumina reads 2018 [cited 2020

678   June 21]. Available from: https://github.com/tseemann/shovill.

679   34.   Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu:

680   scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation.

681   Genome Res. 2017;27(5):722-36. doi: 10.1101/gr.215087.116. PubMed PMID: 28298431;

682   PubMed Central PMCID: PMCPMC5411767.

683   35.   Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an

684   integrated tool for comprehensive microbial variant detection and genome assembly

685   improvement. PLoS One. 2014;9(11):e112963. doi: 10.1371/journal.pone.0112963. PubMed

686   PMID: 25409509; PubMed Central PMCID: PMCPMC4237348.

687   36.   Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.

688   2012;9(4):357-9. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PubMed Central

689   PMCID: PMCPMC3322381.

690   37.   Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated

691   circularization of genome assemblies using long sequencing reads. Genome Biol.

692   2015;16:294. doi: 10.1186/s13059-015-0849-0. PubMed PMID: 26714481; PubMed Central

693   PMCID: PMCPMC4699355.

694   38.   Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.

695   2014;30(14):2068-9. doi: 10.1093/bioinformatics/btu153. PubMed PMID: 24642063.

696   39.   Zhang G, Leclercq SO, Tian J, Wang C, Yahara K, Ai G, et al. A new subclass of

697   intrinsic aminoglycoside nucleotidyltransferases, ANT(3")-II, is horizontally transferred

698   among Acinetobacter spp. by homologous recombination. PLoS Genet.

699   2017;13(2):e1006602. doi: 10.1371/journal.pgen.1006602. PubMed PMID: 28152054.

700   40.   Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with

701   gene gain, loss and rearrangement. PloS one. 2010;5(6):e11147. doi:

702   10.1371/journal.pone.0011147. PubMed PMID: 20593022; PubMed Central PMCID:

703   PMC2892488.

704   41.   Unemo M, Golparian D, Sanchez-Buso L, Grad Y, Jacobsson S, Ohnishi M, et al.

705   The novel 2016 WHO *Neisseria gonorrhoeae* reference strains for global quality assurance of

706   laboratory investigations: phenotypic, genetic and reference genome characterization. J

707   Antimicrob Chemother. 2016;71(11):3096-108. doi: 10.1093/jac/dkw288. PubMed PMID:

708   27432602; PubMed Central PMCID: PMCPMC5079299.

709   42.   Didelot X, Pang B, Zhou Z, McCann A, Ni P, Li D, et al. The role of China in the

710   global spread of the current cholera pandemic. PLoS Genet. 2015;11(3):e1005072. doi:

711   10.1371/journal.pgen.1005072. PubMed PMID: 25768799; PubMed Central PMCID:

712   PMCPMC4358972.

713   43.   Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al.

714   Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):R12. doi:

715   10.1186/gb-2004-5-2-r12. PubMed PMID: 14759262; PubMed Central PMCID:

716   PMCPMC395750.

717   44.   Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New

718   algorithms and methods to estimate maximum-likelihood phylogenies: assessing the

719   performance of PhyML 3.0. Systematic biology. 2010;59(3):307-21. doi:

720   10.1093/sysbio/syq010. PubMed PMID: 20525638.

721   45.   Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis

722   of large phylogenies. Bioinformatics. 2014;30(9):1312-3. doi:

723   10.1093/bioinformatics/btu033. PubMed PMID: 24451623; PubMed Central PMCID:

724   PMCPMC3998144.

725   46.   Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in

726   whole bacterial genomes. PLoS Comput Biol. 2015;11(2):e1004041. doi:

727   10.1371/journal.pcbi.1004041. PubMed PMID: 25675341; PubMed Central PMCID:

728   PMCPMC4326465.

729   47.   Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of

730   heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution,.

731   2016;2(1):vew007.

732   48.   Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of

733   ancestral dates on bacterial phylogenetic trees. Nucleic Acids Res. 2018;46(22):e134. doi:

734   10.1093/nar/gky783. PubMed PMID: 30184106; PubMed Central PMCID:

735   PMCPMC6294524.

736   49.   Demczuk W, Sidhu S, Unemo M, Whiley DM, Allen VG, Dillon JR, et al. Neisseria

737   gonorrhoeae Sequence Typing for Antimicrobial Resistance, a Novel Antimicrobial

738   Resistance Multilocus Typing Scheme for Tracking Global Dissemination of N. gonorrhoeae

739   Strains. J Clin Microbiol. 2017;55(5):1454-68. doi: 10.1128/JCM.00100-17. PubMed PMID:

-28-

740   28228492; PubMed Central PMCID: PMCPMC5405263.

741   50.   Zankari E, Allesoe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM.

742   PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated

743   with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother.

744   2017;72(10):2764-8. doi: 10.1093/jac/dkx217. PubMed PMID: 29091202; PubMed Central

745   PMCID: PMCPMC5890747.

746   51.   Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR.

747   Phandango: an interactive viewer for bacterial population genomics. bioRxiv. 2017.

748   52.   Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:

749   improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. doi:

750   10.1093/molbev/mst010. PubMed PMID: 23329690; PubMed Central PMCID:

751   PMCPMC3603318.

752   53.   Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a

753   multiple sequence alignment editor and analysis workbench. Bioinformatics.

754   2009;25(9):1189-91. doi: 10.1093/bioinformatics/btp033. PubMed PMID: 19151095;

755   PubMed Central PMCID: PMCPMC2672624.

756   54.   Shimuta K, Unemo M, Nakayama S, Morita-Ishihara T, Dorin M, Kawahata T, et al.

757   Antimicrobial resistance and molecular typing of *Neisseria gonorrhoeae* isolates in Kyoto

758   and Osaka, Japan, 2010 to 2012: intensified surveillance after identification of the first strain

759   (H041) with high-level ceftriaxone resistance. Antimicrob Agents Chemother.

760   2013;57(11):5225-32. doi: 10.1128/AAC.01295-13. PubMed PMID: 23939890; PubMed

761   Central PMCID: PMCPMC3811299.

762

763
764

765

Figure captions

767

Fig 1. Whole genome sequence dated phylogeny, resistance patterns of the antimicrobials, and genetic polymorphisms in the ST-1901-associated core-genome group.

Left: A clonal dated phylogeny with corrected branch lengths to account for homologous recombination. In the heat map, the 1st column titled "sub-lineage 34", the sub-lineage harboring *penA*-34 is colored. The 2nd column titled "continent" shows Asia (deep blue), South America (light blue), Europe (pink), North America (light pink), or Oceania (brown). The 3rd column titled "complete genome" shows the three Japanese strains harboring *penA*-34 from 2005 in blue and the ancestral Japanese strain (at the bottom) harboring *penA*-5 in black, whereas the WHO_Y (F89) strain is in pink. In the 4th column titled "ST", ST-1901 and its single locus-variants are colored using different colors as shown in our previous study [11]. In the 5th column titled "NG-MAST1407", strains of NG-MAST1407 are colored. In the 6th and 7th column, susceptible/resistant (S/R) categories according to the EUCAST minimal inhibitory concentration (MIC) breakpoint (susceptibility ≤ 0.125 µg/mL and resistance > 0.125 µg/mL) of 3GCs (cefixime CFM and ceftriaxone CRO) are shown. The columns were colored grey when the MIC values were missing. In the 8th column, the presence (light yellow-green) or absence of any non-mosaic *penA* allele is shown. In the 9th–12th columns, the presence (light yellow green) or absence of a specific non-mosaic *penA* allele is shown. The 13th column shows the presence (cyan) or absence any mosaic *penA* allele. The 14th–18th columns show presence (cyan) or absence a specific mosaic *penA* allele (specifically, 34 and its variants [54], 10 and its variants [54], and 35 [22]). The 19th column shows the susceptible/resistant (S/R) categories of fluoroquinolones (mostly ciprofloxacin, and much less frequently, levofloxacin) according to the EUCAST breakpoint. The 20th–21st columns show the presence (yellow-green) or absence of nonsynonymous amino acid changes compared to the wild type in GyrA. The 22nd–24th columns show the presence (green) or absence of nonsynonymous amino acid

792   changes compared to the wild type in ParC. In the clonal dated phylogeny at the left, a red

793   circle indicates emergence of the sub-lineage 34, whereas two purple lines indicate 95%

794   confidence intervals examined in the main text (emergence time of the sub-lineage 34, and that

795   of one of the three sub-lineages harboring *penA*-10).

796

797   **Fig 2. Whole genome sequence dated phylogeny, resistance patterns of the antimicrobials,**

798   **and genetic polymorphisms in the ST-7363-associated core-genome group.**

799   The columns in the heat map are almost the same as those in Fig 1, although the first five

800   columns in Fig 1 have been omitted here. The 1st and 2nd columns show the susceptible/resistant

801   (S/R) categories according to the EUCAST breakpoint of cefixime (CFM) and ceftriaxone

802   (CRO). The 3rd column shows the presence (light yellow-green) or absence of any non-mosaic

803   *penA* allele. The 4th–6th columns show the presence (light yellow-green) or absence of a specific

804   non-mosaic *penA* allele. The 7th column shows the presence (cyan) or absence of any mosaic

805   *penA* allele. The 8th–9th columns show the presence (cyan) or absence a specific mosaic *penA*

806   allele (specifically, 10, and 34 or 37 (H041-type)[7]). The 10th column shows the

807   susceptible/resistant (S/R) categories of fluoroquinolones (mostly ciprofloxacin, and much less

808   frequently, levofloxacin) according to the EUCAST breakpoint. The 11th–12th columns show

809   the presence (yellow green) or absence of nonsynonymous amino acid changes compared to

810   the wild type in GyrA. The 13th–14th columns show the presence (green) or absence of

811   nonsynonymous amino acid changes compared to the wild type in ParC. In the clonal dated

812   phylogeny at the left, the two branches of interest examined in the main text are colored cyan

813   and green, with 95% confidence interval of the two evolutionary events (acquisition of *penA*-

814   10 and simultaneous amino acid substitutions at GyrA 95 and ParC 87−88) colored purple.

815

816   **Fig 3. Schematic depiction of nucleotide sequence alignment of *penA* and its downstream**

817   **sequence.**

-31-

818   (A) ST-1901-associated core-genome group. *penA*-5 (top), 10 (middle), and 34 (bottom). (B)

819   ST-7363-associated core-genome group. *penA*-2 (top) and 10 (bottom). The coding sequence

820   (CDS) of *penA* is shown as a rectangle. The recombined sequences in *penA*-10 and 34 are

821   indicated by the red and orange lines, respectively.

822

823   **Fig 4. Schematic depiction summarizing the results regarding the origin of mosaic**
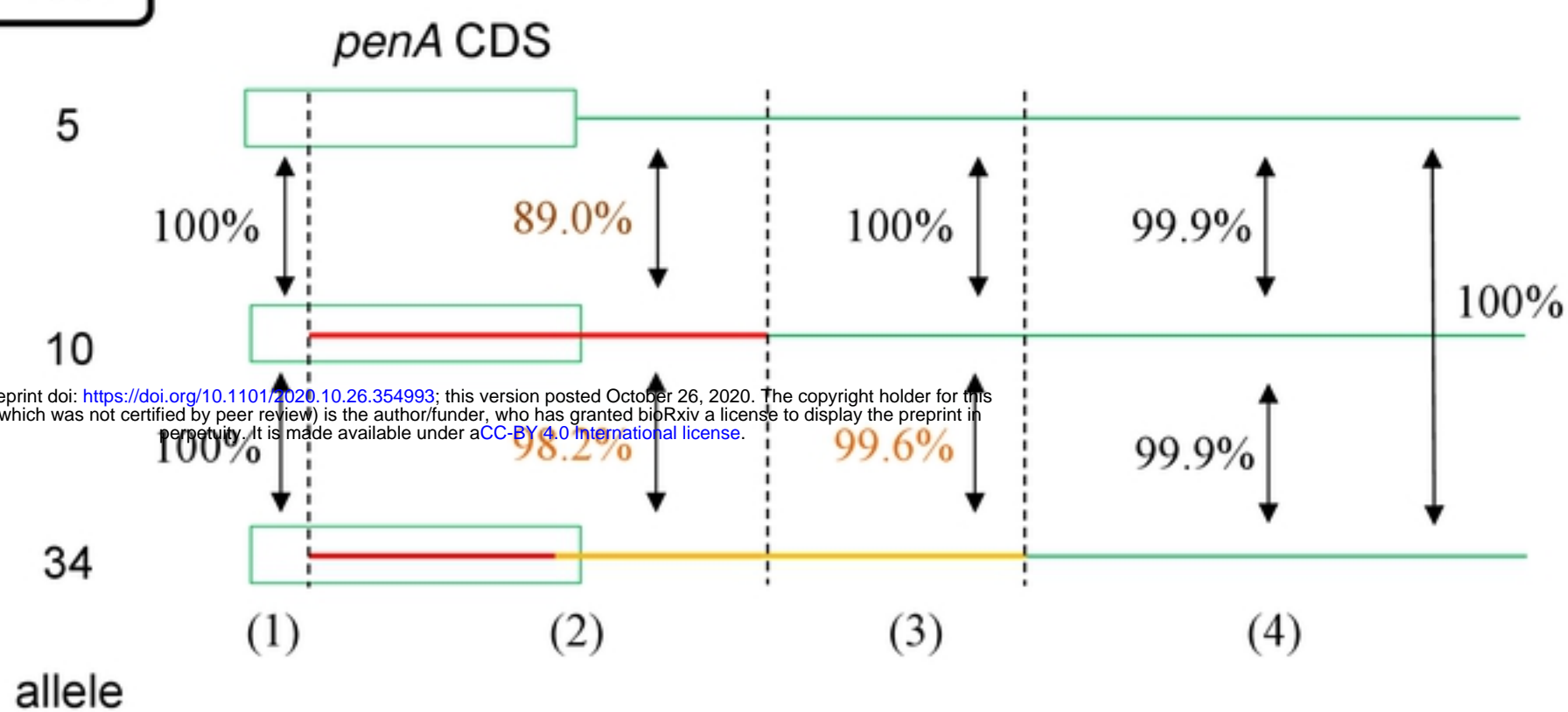
824   ***penA*-10 and 34.**

825   Rectangles indicate genes, second of which is *penA*. In ST-7363 (top) and ST-1901 (bottom),

826   the upper part shows the ancestral sequences (*penA*-2 and 5, respectively) while the lower
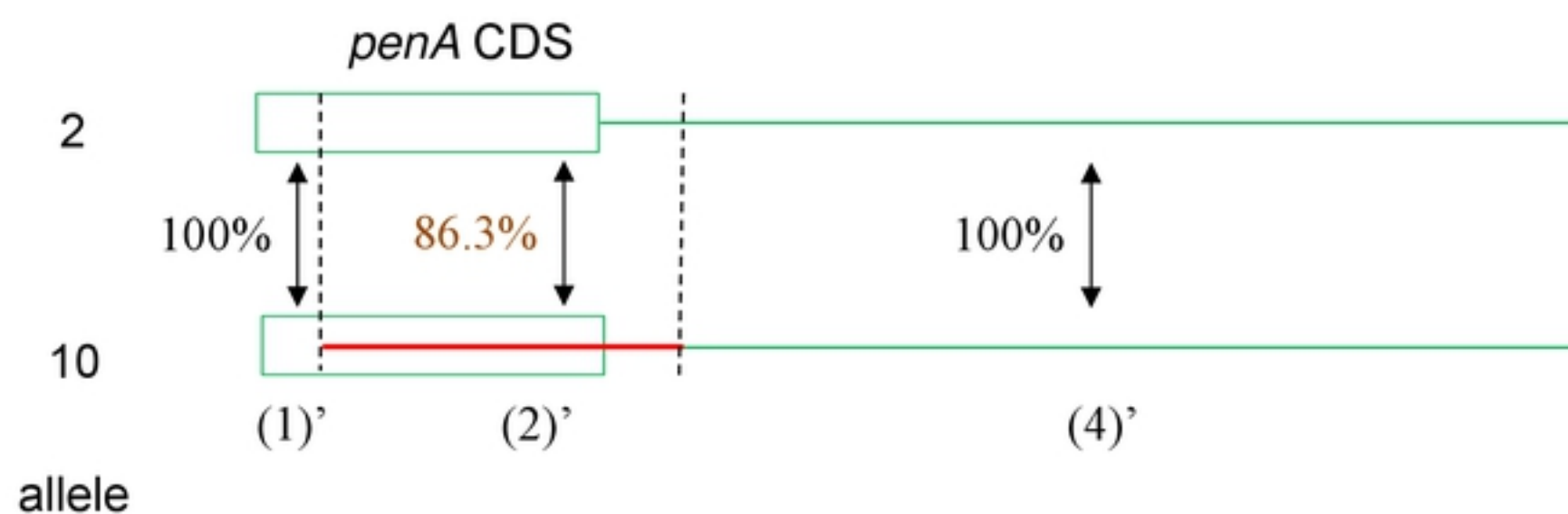
827   part shows the recombined sequences.

828

829

830

-32-
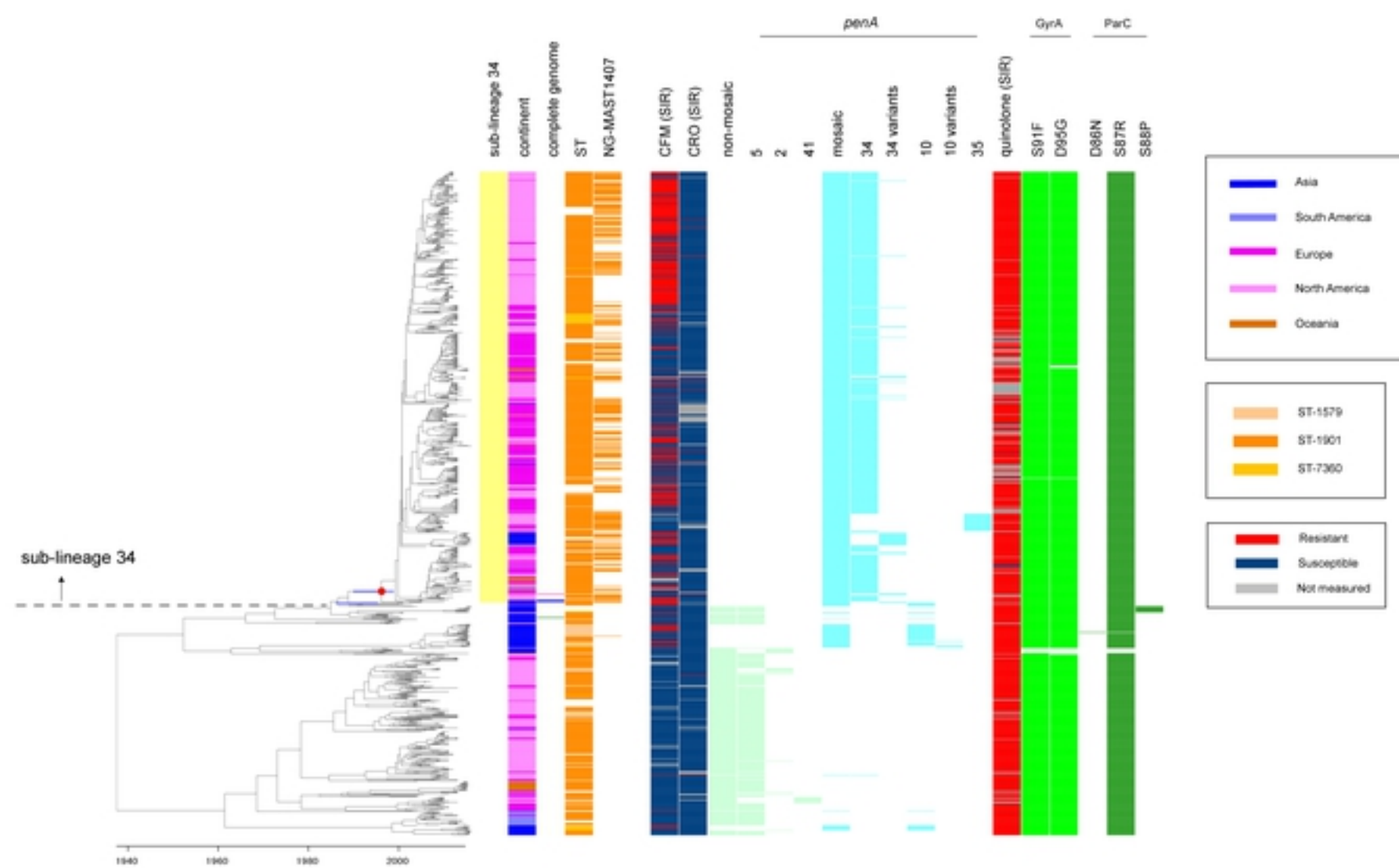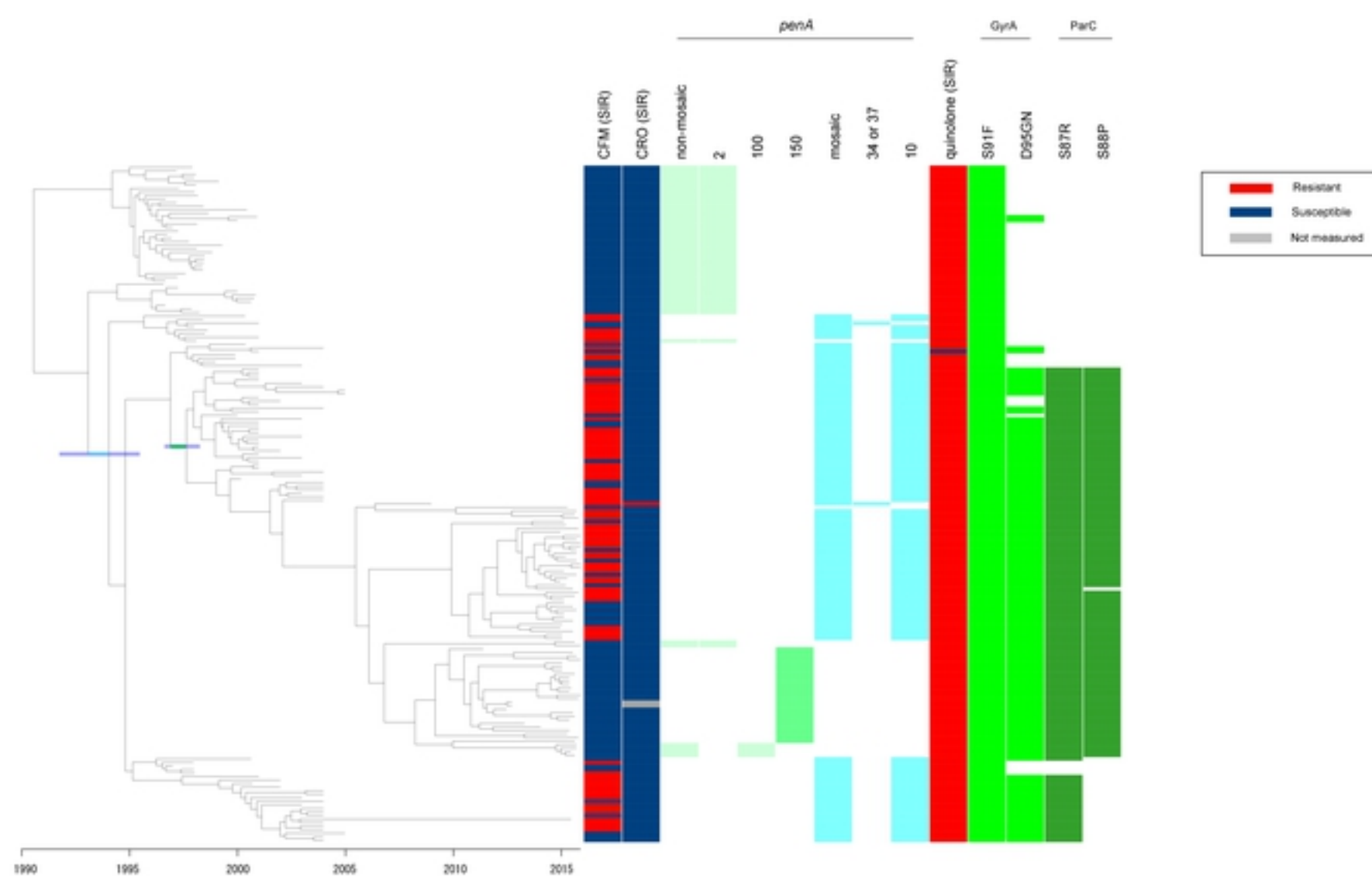
ST-7363

penA

2

1 -293   1749bp

mosaic ends at 390bp downstream

10

(generated between September 1991-May 1995)

another *Neisseria* species

ST-1901

penA

5

approximately 2.5kb fragment

1 (ST-1594)

1 -293   1749bp

mosaic ends at 887bp downstream

10

(generated between April 1986-August 1995)

34

(generated between May 1990-July 1999)