



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Antibiotic treatment regimes as a driver of the global population dynamics of a major gonorrhoea lineage

Citation for published version:

Osnes, MO, van Dorp, L, Brynildsrud, OB, Alfsnes, K, Schneiders, T, Templeton, K, Yahara, K, Balloux, F, Caugant, DA & Eldholm, V 2020, *Antibiotic treatment regimes as a driver of the global population dynamics of a major gonorrhoea lineage*. *Molecular Biology and Evolution*. <https://doi.org/10.1093/molbev/msaa282>

Digital Object Identifier (DOI):

[10.1093/molbev/msaa282](https://doi.org/10.1093/molbev/msaa282)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Antibiotic treatment regimes as a driver of the global population dynamics of a major gonorrhea lineage

Magnus N. Osnes¹, Lucy van Dorp², Ola B. Brynildsrud¹, Kristian Alfsnes¹, Thamarai Schneiders³, Kate E. Templeton⁴, Koji Yahara⁵, Francois Balloux², Dominique A. Caugant¹, Vegard Eldholm^{1*}

¹ Division of Infectious Disease Control and Environmental Health, Norwegian Institute of Public Health, Oslo Norway

² UCL Genetics Institute, University College London, London WC1E 6BT, United Kingdom

³ Division of Infection Medicine, University of Edinburgh, 49 Little France Crescent, Edinburgh, EH16 4SB, United Kingdom.

⁴ Department of Laboratory Medicine, Royal Infirmary of Edinburgh, NHS Lothian, Edinburgh, United Kingdom

⁵ Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, Tokyo, Japan

* corresponding author

Keywords - *Neisseria gonorrhoeae*, antibiotic resistance, evolution, phylogeography

Abstract

The *Neisseria gonorrhoeae* multilocus sequence type (ST) 1901 is among the lineages most commonly associated with treatment failure. Here we analyze a global collection of ST-1901 genomes to shed light on the emergence and spread of alleles associated with reduced susceptibility to extended-spectrum cephalosporins (ESCs).

The genetic diversity of ST-1901 falls into a minor and a major clade, both of which were inferred to have originated in East Asia. The dispersal of the major clade from Asia happened in two separate waves expanding from around 1987 and 1996, respectively. Both waves first reached North America, and from there spread to Europe and Oceania, with multiple secondary reintroductions to Asia.

The ancestor of the second wave acquired the *penA* 34.001 allele, which significantly reduces susceptibility to extended-spectrum cephalosporins (ESCs). Our results suggest that the acquisition of this allele granted the second wave a fitness advantage at a time when ESCs became the key drug class used to treat gonorrhea. Following its establishment globally, the lineage has served as a reservoir for the repeated emergence of clones fully resistant to the ESC ceftriaxone, an essential drug for effective treatment of gonorrhea.

We infer that the effective population sizes of both clades went into decline as treatment schemes shifted from fluoroquinolones via ESC monotherapy to dual therapy with ceftriaxone and azithromycin in Europe and the USA. Despite the inferred recent population size decline, the short evolutionary path from the *penA* 34.001 allele to alleles providing full ceftriaxone resistance is a cause of concern.

Introduction

Gonorrhoea, caused by the gram-negative bacterium *Neisseria gonorrhoeae* (the gonococcus), is a sexually transmitted disease re-emerging in large parts of the world. Globally, gonococci have acquired resistance to every drug available to treat the disease. In recent years, an increasing number of infections resistant to first-line treatments has been reported, in some instances leading to treatment failure (Unemo et al. 2012; Fifer et al. 2016). Currently, the World Health Organization recommends dual treatment with the extended-spectrum cephalosporin (ESC) ceftriaxone and the macrolide azithromycin for empirical therapy (the World Health Organization 2016). Given widespread resistance to drugs historically used to treat gonorrhoea, successful treatment relies heavily on these two cornerstone drugs. So far, infections caused by strains resistant to both drugs have been sporadic and secondary transmission of these remains limited. Yet, the widespread circulation of alleles associated with reduced susceptibility to each of these drugs serves as a stern reminder that this picture could change.

The multilocus sequence type (ST) 1901 is amongst the most common causes of gonococcal infections exhibiting reduced ESC susceptibility (Shimuta et al. 2015; Harris et al. 2018). Within ST-1901, the most frequent *penA* alleles associated with resistance or reduced susceptibility (RRS) to ESCs are 10.001 and 34.001 (Yahara et al. 2018). Importantly, the *penA* 34.001 allele is only a single mutation away from full ceftriaxone resistance, as demonstrated e.g. by a recent case of treatment failure in France (Unemo et al. 2012). In addition, fluoroquinolone resistance and reduced susceptibility to the second cornerstone drug, azithromycin, is widespread within the ST (Unemo et al. 2012). ST-1901 has a global distribution, but its prevalence varies between regions: it is among the most commonly observed *N. gonorrhoeae* strains in Europe, the USA, and Japan, but seems to be less frequent in Africa, with none of the current 197 African isolates in the pubMLST database (Jolley et al. 2018) belonging to the ST-1901 lineage.

The high global incidence of ST-1901 displaying reduced susceptibility to key antibiotics prompted us to perform an in-depth genomic investigation. Here we analyze a global collection of 741 ST-1901 genomes to shed light on the historic spread of the sequence type, with particular focus on the emergence and spread of ESC RRS alleles. Our analyses support an East Asian origin of ST-1901, but identified North America as a central hub in the global dispersal of the strain. We also find that ceftriaxone resistant clones in Europe evolved locally from a background of reduced ESC susceptibility. Our analyses reveal consistent trends both in the routes of dispersal and in the emergence of alleles yielding reduced antibiotic susceptibility.

Results

Time-resolved phylogenetic inferences

A recombination-corrected ST-1901 whole-genome phylogeny was used to generate a time-resolved phylogeny, calibrated by sample collection dates. The estimated evolutionary rate was 6.46 mutations per genome per year (CI: 5.81 to 7.29), translating to 2.91×10^{-6} (CI: 2.61×10^{-6} to 3.28×10^{-6}) mutations/site/year, which falls in line with previously estimated rates for *N. gonorrhoeae* (Didelot et al. 2016; Osnes et al. 2020). A deep split at the root of the phylogeny separated ST-1901 into a major (638 out of 741 genomes) and a minor clade (Fig. 1), which we termed Clade A and Clade B, respectively. These clades are inferred to have diverged sometime around the 18th century (CI: 1626 to 1898). The vast majority of Clade B isolates did not carry *penA* alleles associated with reduced ESC susceptibility. To assess whether the substitution rates differed between Clades A and B, we created separate dated phylogenies for the two clades, which resulted in very similar evolutionary rate estimates (6.49 [CI: 5.76 - 7.30] and 6.51 [CI: 4.37 - 9.13] mutations per genome per year for Clades A and B respectively). Next, to investigate whether ST-1901 was truly monophyletic at the genome level, we generated a phylogenetic tree from 9535 genomes from a global in-house database. When ST-classifications were annotated on the tree, it was clear that Clade B isolates were only distantly related to Clade A and that the overlapping MLST profiles were the result of convergent evolution and recombination (Fig. S1).

Clade A isolates were by far the more prevalent and were also associated with higher ESC minimum inhibitory concentrations (MICs). We thus chose to focus our analysis on this clade. Within the clade, East Asian isolates were typically found on deeper branches of the phylogeny compared to isolates from the rest of the world. Indeed, the majority of European and North American isolates belonged to a single recently expanded clade with a common ancestor inferred to around 1996 (CI: 1992 to 1999), identifiable by the red node in Fig. 1.

Within Clade A, fluoroquinolone resistance was found to be universal, with 631 out of 638 isolates carrying a *gyrA* S91F in addition to either D95A or D95G and the *parC* S87R mutation. The remaining seven isolates harbored a *gyrA* S91Y mutation. The *mtrR* -35Adel promoter deletion, associated with reduced susceptibility to azithromycin and extended-spectrum cephalosporins, was found in nearly all isolates of both clades, as was the *rpsJ* V57M mutation associated with tetracycline resistance (Hu et al. 2005). In addition, the 23S 2611 C>T mutation associated with azithromycin resistance, mutations in *rpID* codons 69 and 70, possibly associated with reduced susceptibility (Grad et al. 2016), and other mutations in the coding portion (G45D) and promoter region of *mtrR*, were identified in a handful of isolates across the clades (Fig. S2 and extended isolate data available at [10.6084/m9.figshare.12423596](https://doi.org/10.6084/m9.figshare.12423596)).

Alleles associated with reduced ESC susceptibility.

In order to investigate the distribution and spread of alleles contributing to reduced ESC susceptibility, we identified and annotated *penA*, *ponA* and *porB* alleles across the dataset. We identified a diverse set of *penA* alleles, with a dominance of *penA* 34.001 (Fig. 1). ESC MIC

values were available for 609 of the 741 isolates. Based on the observed MIC measurements, we defined alleles 10.001, 34.001, 34.009, 42.001, 71.001 and 72.001 as RRS *penA* alleles (Fig. 2 A; note that the 34.009 allele was identified in a single isolate and thus excluded from the figure). In addition, we considered the allele 13.001 as an ESC RRS allele based on published findings (Liao et al. 2011; Osnes et al. 2020), despite the lack of MIC data for isolates carrying it in this study.

We found significant variation in MIC distributions within groups of isolates carrying the same *penA* allele. Isolates carrying RRS *penA* alleles typically exhibited MICs around the resistance breakpoint for the ESC cefixime (EUCAST resistance breakpoint ≥ 0.25 $\mu\text{g}/\text{mL}$ for both cefixime and ceftriaxone (EUCAST 2020)). For ceftriaxone, very few isolates reached the breakpoint, with a total of four fully resistant isolates carrying 34.009 ($n = 1$) and 42.001 ($n = 3$) as notable exceptions.

In addition to the major ESC susceptibility determinant *penA*, mutations in *ponA* encoding the penicillin-binding protein 1 and the porin-encoding *porB* can modulate ESC susceptibility (Demczuk et al. 2020). In our ST-1901 collection, all but seven isolates carried the *ponA* L421P mutation (not shown, see supplementary isolate information). In *porB*, mutations in codons 120 and 121 have been shown to confer decreased ceftriaxone susceptibility (Liao et al. 2011; Osnes et al. 2020). To investigate the effect of different *porB* alleles in ESC RRS *penA* backgrounds, we stratified isolates carrying *penA* 10.001 and 34.001 by *porB* type. We restricted the analyses to these two *penA* alleles as they were the only ESC RRS alleles found in sufficient numbers across multiple *porB* allele backgrounds in our dataset. This analysis confirmed that *porB* alleles with mutations in both codons 120 and 121 were associated with higher ceftriaxone MICs (Mann-Whitney-Wilcoxon $p < 0.05$) in both *penA* backgrounds (Fig 2 B). No significant associations were found for cefixime.

Geographic dispersal patterns

To reconstruct the spatiotemporal dispersal history of ST-1901, the continent of isolation of each isolate was treated as a discrete trait for phylogeographic inference employing stochastic character mapping of discrete traits on phylogenies (SIMMAP) (Bollback 2006). The phylogeographic stochastic character mapping supports an East Asian origin for both Clade A and B (Fig. 3), with 71.3% and 63.5% posterior probability, respectively. The uncertainty in the geographic mapping of the root (i.e. ancestor of the A and B clades) stems from the long basal branches in the tree, which allow for multiple different geographical transition histories (see Fig. S3).

For Clade A, we observed clear phylogeographic patterns pointing to the direction of global transmission events. The deeper ancestral nodes in the clade were mapped to Asia with high confidence, whereas non-Asian isolates were almost exclusively found to be descendants of two independent expansions out of Asia. Both of these waves spread to North America first, followed by subsequent exports from North America to Europe. Lineages spawned by both the first and second waves successfully established themselves in North America and Europe and have remained in endemic circulation since. Apart from these two major waves, we only

observe two isolates representing independent introductions to Europe (Norway) and North America (Canada) from Asia for Clade A lineages, with no evidence of further spread (Fig. 3, event number 1 and 2). Introductions to Oceania were most often seeded from Europe, but on some occasions from North America.

We estimated that the first wave of Clade A out of Asia originated around 1987 (CI: 1980 to 1992). The founder of Wave 1 (green node in Fig. 1, and event number 3 in Figs. 3 and 4) harbored a 5.002 allele, which was retained in almost all descendants in Europe and North America. Interestingly, two successful reintroductions from North America or Europe to Asia were inferred from this wave, both of which spawned transmission chains in Japan (Figs. 3 and 4, event number 4 and 5). In one case, the strain (introduced around 2009, Fig. 3 event number 4) acquired a *penA* 10.001 allele around 2012 (CI: 2011, 2013), associated with cefixime resistance (Fig. 4, event number 4). Wave 1 also comprises lineages having independently acquired the *penA* 13.001 and 34.001 alleles in Bulgaria and USA around 2008 and 2010, respectively, though with no evidence of onward transmission (Figs. 3 and 4, event number 6 and 7). The inferred effective population size of the Wave 1 clade was found to peak around 2005 before starting to decline in approximately 2007 (Fig. 5 B).

The second wave of Clade A ST-1901 started with a transition from Asia to North America in 1996 (CI: 1992 to 1999), highlighted in Fig. 1 with a red node and in Figs. 3 and 4 (event number 8). Concomitant with the expansion to North America, the lineage acquired the *penA* 34.001 allele. From this point on, the lineage rapidly spread across Europe and North America. We infer multiple acquisitions of other RRS alleles from this state, most likely driven by point mutations. Of note, the *penA* 34.001 allele is only one mutation away from either of the alleles 72.001, 34.009, and 42.001, the latter two being associated with full ceftriaxone resistance.

A reintroduction to Japan from North America around 2006 (CI: 2005 to 2007) is marked by the acquisition of a *penA* 72.001 allele (probably via a point mutation) and successful establishment of the lineage (Figs. 3 and 4, event number 9). The median cefixime MIC for isolates harboring allele 72.001 is moderately higher than for those harboring 34.001 among our samples (0.19 µg/ml vs 0.12 µg/ml, respectively, see Fig. 2), whereas the median MIC for ceftriaxone was identical between the two groups. Additional independent RRS acquisition events in this wave occurred in tandem with introductions from North America to Europe. Notable events include the acquisition of full ceftriaxone resistance via the acquisition of allele 42.001 in the second half of 2010, leading to two diagnosed cases in the UK and France (Figs. 3 and 4, event number 10) and independent acquisitions of alleles 34.009 and 42.001 in Germany and Spain in patients diagnosed in 2011 and 2012, respectively (Figs. 3 and 4, event number 11 and 12). We inferred that the effective population size of Wave 2 grew steadily until it peaked around 2011, following which the effective population size declined (Fig. 5B).

To ensure that the above inferences were not overly shaped by biased sampling, we repeated the analyses on a significantly downsampled, but more geographically balanced dataset (see Materials and Methods). The inferences from this dataset largely recapitulated our findings above, with the exception of Wave 1, which was inferred to be the result of two export events from Asia (Fig. S4).

Pangenomic features of Clade A isolates

In addition to resistance-associated genes, we investigated whether any other associated genes or alleles may have contributed to the successful expansion of Clade A, particularly the second wave. A pangenome analysis of all ST-1901 isolates, employing a 98% identity threshold identified 1953 out of a total of 2479 genes (79%) present in at least 95% of all analysed genomes. Altering the identity threshold for pangenome clustering from 98% to 95% produced only marginally different summary statistics (2471 genes in total with 1963 present in at least 95% of the genomes); we thus considered the output from the run employing a 98% identity threshold.

Despite the recent emergence, two thirds (472 out of 741) of the ST-1901 isolates were part of the second wave of international expansion. In an effort to identify genes possibly involved in the successful expansion of Wave 2, beyond those known to be associated with drug susceptibility, the gene presence / absence matrix was scored to identify elements enriched or underrepresented in Wave 2 isolates. It was clear that the majority of identified non-core genes represented divergent alleles of the same genes rather than actual differences in gene content (Table S7). Elements differentially present in Wave 2 included both various *penA* and *porB* alleles, in addition to multiple variants of known hyper-variable genes such as *pilE* and *pilC*, encoding a pilus assembly protein involved in adherence and an outer membrane protein, respectively. When reducing the significant hits to only those present in $\geq 95\%$ of Wave 2 isolates and $\leq 5\%$ of non-Wave 2 isolates or vice versa, only two genes remained, namely a *pilC* allele overrepresented in Wave 2 and a multi-species hypothetical gene (Panaroo: "group_527"), homologous to the *Neisseria meningitidis* locus NEIS0608, underrepresented in the clade.

Overall, the pangenome reconstruction revealed that there was limited variation in gene content between isolates within ST-1901. On the other hand, the circulation of highly divergent alleles, probably resulting from incorporation of DNA sequences from other *Neisseria* species (Ezewudo et al. 2015), constitutes a rich source of genetic variation, including determinants of drug susceptibility.

Discussion

We found that the global distribution of ST-1901 Clade A, one of the most successful lineages of *N. gonorrhoeae* worldwide, was largely shaped by two waves of dispersal out of East Asia. Despite the East Asian origins of these clades, our analyses suggest that North America has played a pivotal role as a transit hub for further spread to Europe and, on occasion, back to Asia. Each of the waves was likely spawned by two singular exports out of Asia, the details of which are unknown. The successful expansions of the two clades however, were likely aided by resistance to key antibiotics.

The first wave of ST-1901 spread out of East Asia, probably in the late 1980s (1980-1992) from a background of fluoroquinolone and tetracycline resistance (mediated by chromosomal

gyrA, *parC* mutations and *rpsJ* V57M, respectively). In addition, the ancestor of Wave 1 exhibited reduced azithromycin susceptibility mediated by the *mtrR* -35Adel promoter mutation in combination with reduced susceptibility or resistance to penicillins brought about by the *ponA* L421P mutation in combination with the *penA* 5.001 allele (A517G & G543S) (Demczuk et al. 2020) (Figs. 3 and 4, event number 3). In the 1980s, tetracycline and fluoroquinolones were widely used to treat gonorrhea. As a result of widespread resistance, tetracycline treatment was gradually discontinued from the late 1980s, whereas fluoroquinolones were used up until the late 1990s in many East Asian countries. In the USA, fluoroquinolone treatment was discontinued for men who have sex with men in 2004, following widespread resistance in this group, and for all patients from 2007 (Unemo and Shafer 2014).

It is thus overwhelmingly likely that the first wave of international ST-1901 dissemination was aided by resistance to fluoroquinolones. Interestingly, in one instance, the reintroduction of the strain to Japan was rapidly followed by the acquisition of a *penA* 10.001 allele around 2013 (Figs. 3 and 4, event number 4), associated with cefixime resistance. This observation may suggest the presence of more favorable conditions for the evolution and transmission of cefixime resistance in East Asia compared to Europe and North America at the time. We found that the effective population size of the Wave 1 clade started declining around the time of the shift from fluoroquinolone to ESC use in the USA, consistent with new treatment regimens having played a role in hampering the spread of this clade.

The second and largest wave of global dispersal (Wave 2) started out of East Asia in the 1990s (1992-1999), with the majority of ST-1901 isolates globally descending from this event. The ancestor of Wave 2 acquired the *penA* 34.001 allele, resulting in reduced susceptibility to ESCs (Figs. 3 and 4, event number 8). ESCs became the favored drugs following the shift away from fluoroquinolone use. In Japan, oral ESCs such as cefixime were used as monotherapy, often at relatively low concentrations (Unemo and Shafer 2014), which was probably a driver for the emergence and spread of *penA* RRS alleles. We found that the second wave of global ST-1901 dispersal was rapidly established in North America followed by multiple exports to Europe. From Europe, the strain spread further to Oceania on at least three occasions.

Similar analyses performed on a downsampled geographically balanced dataset (see Materials and Methods) largely recapitulated the above findings: The credibility intervals for key node ages (Table S6) overlapped with our original inference, despite a weaker temporal signal in the downsampled dataset (Root-to-tip R^2 of 0.12, where higher R^2 is indicative of a stronger correlation between sampling times and genetic divergence). In addition, the spatio-temporal history of Wave 2 was perfectly recapitulated, whereas the inferred first wave was split in two out-of-Asia events in the downsampled dataset with independent introductions to Europe and America (Fig. S4). This split seems to be the result of the loss of informative European and American genomes. When also taking into account the stronger temporal signal of the full dataset, we believe the inferences from the full dataset to be more robust.

In Europe, Wave 2 representatives also evolved full-blown ceftriaxone resistance on at least three occasions (Figs. 3 and 4, event number 10, 11 and 12), but these clones seem to have failed to generate significant onward transmission. The limited spread of these clones

recapitulates earlier findings (Ohnishi et al. 2011) and might suggest that *penA* alleles yielding ceftriaxone resistance have a fitness disadvantage. In a pattern consistent with our observations for Wave 1, a reintroduction around 2006 of the Wave 2 strain to Japan from North America was followed by the acquisition of a *penA* 72.001 allele from a 34.001 background (Figs. 3 and 4, event number 9). Allele 72.001 is associated with further increases in cefixime MICs compared to 34.001, and the successful establishment of the strain is again indicative of stronger selection for cefixime resistance in Japan compared to Europe and North America.

In the USA, monotherapy with cefixime or ceftriaxone was widespread in the years following the discontinuation of fluoroquinolone use. In 2010, the recommended regimen was changed to dual therapy with a cephalosporin and azithromycin or doxycycline (Anon 2019). After observing both increasing cefixime MICs in the USA and cases of treatment failure abroad, the Centers for Disease Control (CDC) specified ceftriaxone as the ESC of choice in 2012 (Centers for Disease Control and Prevention (CDC) 2012). This timeline was similar in Europe, where ESC monotherapy was recommended for empirical therapy from 2009, and dual therapy mirroring the CDC recommendations from 2012 (Bignell et al. 2013). Our inferences suggest that the effective population size of Wave 2 peaked in 2011 and has since declined steeply. This suggests that the introduction of a dual treatment scheme, including ceftriaxone, succeeded in slowing the transmission of the second wave of ST-1901. In fact, the few cases of full ceftriaxone resistance we detected were inferred to have evolved in Europe during the period of ESC monotherapy (2009-2012).

Our reconstruction of the global dissemination of ST-1901 Clade A points to a central role of antibiotic-induced selection in shaping the evolution and spread of this strain, in line with other recent studies of the gonococcus (Sánchez-Busó et al. 2019; Golparian et al. 2020). We cannot rule out the existence of additional drivers of the expansion of Clade A, such as prevalence in high-risk groups, as the available metadata is lacking in this regard. In terms of genome dynamics, Clade A does not seem to stand out, as we infer nearly identical evolutionary rates and rates of recombination (Fig. S5) between Clade A and the less dispersed Clade B. Our phylogeographic analyses revealed clear, non-random patterns in the global movement of this ST, recapitulated by two temporally distinct waves (Fig. 5). After its introduction to North America from East Asia, the second of these waves involved a large number of exports from North America to Europe. ST-1901 has spread efficiently in the Western world, and is among the most frequently observed strains in both Europe and the USA to this day. This is particularly concerning given its association with reduced ESC susceptibility (Grad et al. 2014; Harris et al. 2018).

Conclusions

The global distribution of *N. gonorrhoeae* ST-1901 is mainly the result of two discrete waves of transmission originating in East Asia, with North America playing a central role as a transit hub for international dissemination. Strains from the largest and most recent second wave, carry a *penA* 34.001 allele which confers reduced ESC susceptibility in and of itself, but is also only a single point mutation removed from full ceftriaxone resistance. We show that

resistance has evolved independently from the 34.001 background on at least three occasions in Europe, but that these events did not result in significant onward-transmission. The lack of onward-transmission of ceftriaxone resistant clones is suggestive of reduced fitness. However, compensatory mutations alleviating this fitness reduction were shown to evolve in a mouse model (Vincent et al. 2018), and it remains to be seen if similar mutants will evolve among naturally circulating gonococci. The second wave of ST-1901 dispersal led to the successful global establishment of a strain exhibiting reduced susceptibility or resistance to most drugs appropriate for gonorrhea treatment. However, we infer that the effective population size of the Wave 2 clade went into decline coinciding with the introduction of dual therapy with ceftriaxone and azithromycin in Europe and the USA. The short evolutionary path from *penA* 34.001 to full ceftriaxone resistance suggests that resistance is likely to evolve repeatedly also in the future. It will be of utmost importance to keep such clones contained.

Materials and Methods

Sample collection

Genomes from the following sources were included: (1) Isolates received at the Norwegian Institute of Public Health between 2016-2018 that matched our typing-specific inclusion criteria (see below); (2) Isolates obtained primarily from Scotland, retained within the Scottish Bacterial STI laboratory as ST-1901 from the period 2015-2016; (3) Isolates from a number of published studies (Grad et al. 2014; Kwong et al. 2016; Kwong et al. 2018; Lee et al. 2018; Yahara et al. 2018; Peng et al. 2019; Sánchez-Busó et al. 2019). When typing information was not available, the genomes were screened, and those that matched our inclusion criteria retained. NCBI BioProject accessions for these are PRJNA431691, PRJDB6496, PRJNA394216, PRJNA266539, PRJEB4024, PRJEB2999 and PRJEB17738; (4) Isolates available at PathogenWatch (<https://pathogen.watch/>) as of May 2019 that matched our inclusion criteria.

Only isolates with provided collection year and country of isolation were included. In addition to isolates matching the ST-1901 MLST profile, we included isolates belonging to ST-1579, ST-7360, ST-13590 and ST-8137. The first two STs have been described as single-locus variants of ST-1901 (Yahara et al. 2018) and all four STs were found embedded within the ST-1901 genome-based phylogeny in a previous analysis of Norwegian gonococcal genomes (Alfsnes et al. 2020). Multilocus sequence typing (MLST) was performed *in silico* (<https://github.com/tseemann/mlst>) using the pubMLSTdatabase (Jolley et al. 2018) to determine the sequence type. Extended data on all isolates is available in the Figshare repository [10.6084/m9.figshare.12423596](https://doi.org/10.6084/m9.figshare.12423596).

In total, the study included 139 genomes from East Asia, 414 genomes from Europe, 160 genomes from the USA and 26 genomes from Oceania, in addition to one isolate from South America and one from the Philippines.

Ethical approval was not required as the study was initiated within the legal mandate of the Norwegian Institute of Public Health (NIPH) to investigate and report on infectious disease

outbreaks. The study is restricted to the genomic analysis of microbial data and did not include patient-level data beyond country and date (year) of specimen collection.

Genome analyses

Raw Illumina reads were assembled *de novo* using SPAdes v3.13.0 (Bankevich et al. 2012) employing the 'careful' mode. The assemblies were further filtered to remove contigs with a *k*-mer-coverage < 3 and length < 500 nucleotides. The genomes were subsequently screened with Mash v2.2.2 (Ondov et al. 2016) for the presence of contigs from other genera. Assemblies exhibiting signs of significant contamination (> 100/1000 shared hashes) with sequences from bacterial taxa other than *N. gonorrhoeae* were subsequently excluded from downstream analyses.

For comparative analyses, a high-quality reference genome (isolate 600751) was generated by hybrid assembly. Long-read data were produced on the Oxford Nanopore GridION platform as described previously (Brynildsrud et al. 2018) and short-read data was generated on the Illumina MiSeq platform. Unicycler v0.4.7 (Wick et al. 2017) was used to *de novo* assemble a high-quality reference genome using jointly the long Oxford Nanopore reads and short accurate Illumina reads. The resulting assembly resolved a circular genome of 2,222,926 base pairs. All short read assemblies were aligned to this closed reference genome using Parsnp v1.2 (Treangen et al. 2014). A whole-genome multifasta was generated by retaining the reference nucleotide position for all sites by filling all non-core regions with reference nucleotides using an in-house script (<https://github.com/krepsen/parsnp2fasta>). In total, 1,716,165 sites were retained by Parsnp, and the remaining sites filled with reference nucleotides.

Assembled genomes were annotated using Prokka v1.14.6. Prior to pangenome reconstruction, the annotated genomes were again screened with Mash (Ondov et al. 2016) as implemented through Panaroo v1.1.2 (Tonkin-Hill et al. 2020). A single genome (SRR6765327) was removed from downstream analyses on the basis of it being assembled into a high number of contigs (*n*=397, all other assemblies were assembled into < 300 contigs). In terms of genome size and gene content there was nothing suspect about the assembly, but as contig ends are generally challenging for pangenome analyses, we chose to exclude this genome. The Prokka-annotated genomes were used as input for pangenome analysis using Panaroo, employing default parameters in 'strict' mode, apart from the clustering identity threshold for which two different limits were tested: 0.98 (default) and 0.95. Finally, Scoary (Brynildsrud et al. 2016) was used to identify genes with a non-random distribution in the clade corresponding to the second wave compared to non-Wave 2 isolates (that is, genes over- or under-represented in Wave 2 isolates). Output from the Panaroo (Table S8) and Scoary (Table S7) analyses are available as supplementary tables.

Genome clustering was performed on a large in-house global collection of 9535 assembled genomes, using PopPUNK (Lees et al. 2019). A Neighbor-joining tree was generated using RapidNJ as implemented in PopPUNK (Simonsen and Pedersen 2011). The phylogeny was annotated with MLST profiles generated *in silico* (<https://github.com/tseemann/mlst>).

Temporal phylogenetic analyses

A maximum likelihood phylogeny was reconstructed from the whole-genome alignments using IQ-tree v1.6.9 (Nguyen et al. 2015) with 1000 bootstrap replicates (option *-bb*). The resulting phylogeny was used as input for ClonalFrameML v1.2 (Didelot and Wilson 2015) in order to identify recombination tracts and homoplasies across the whole-genome alignments. The ratio of recombination rate to mutation rate (R/θ) was estimated at 0.42. A phylogeny annotated with identified recombination blocks is available as a supplementary figure (Fig. S5). The output from ClonalFrameML was loaded directly to BactDating v1.0.6 (Didelot et al. 2018) which accounts for branch-specific recombination rates, rather than simply ignoring recombinant regions. Root-to-tip regression with concomitant inference of the best root location ($R^2=0.17$) and tip-date-randomization performed within BactDating demonstrated a clear temporal signal in the data (Figs. S6 and S7). 200 million Markov chain Monte Carlo (MCMC) steps were performed to generate a time-resolved tree using the default gamma model for clock rate. Post removal of the first 20 million steps, the resulting evolutionary rate was estimated to 6.54 substitutions/genome/year (95% CI: 5.76 - 7.34).

To verify the temporal inferences from the above analyses, a separate workflow was employed: Gubbins (Croucher et al. 2015) was run on the Clade A portion of the dataset, followed by least squares dating using LSD2 (To et al. 2016) with 1000 bootstrap replicates for temporal inference. The resulting TMRCA estimate for Clade A was 1838 (CI: 1810 to 1861), which was well aligned with our original BactDating-inferred TMRCA in 1887 (CI: 1798 to 1938).

Effective population size estimates

To describe the growth and decline of the waves out of Asia we estimated their effective population size over time using the *Skygrowth* R-package v0.2.0 (Volz and Didelot 2018). *Skygrowth* uses a nonparametric Bayesian model to estimate the effective population size over time given the estimated genealogies. The model places a prior on the change in the growth rate of the effective pathogen population size, not on the change in the logarithm of the effective population size, like the commonly used *skyride* models (Minin et al. 2008). For sections of the phylogenies where data are sparse this has the advantage of using the information on the growth rate learned from the previous parts of the phylogeny, whereas the *skyride* models has a tendency to estimate stabilizing effective population sizes in such situations. We used the MCMC implementation in the R-package to fit the *Skygrowth* model. We ran one chain with 10 million steps and discarded the first 20% of the iterations as burn-in. The results were summarized using the median estimate and credibility intervals covering the 0.025 to 0.975 percentiles of the posterior distribution.

Identification of resistance alleles

To identify mutations and alleles associated with reduced susceptibility or resistance to relevant antibiotics, whole genome assemblies were aligned against gene-specific databases

downloaded from pubMLST (Jolley et al. 2018) using BLAST (Altschul et al. 1990). This was done for the following genes/segments: *penA* (NEIS1753), *ponA* (NG_ponA), *porB* (NG_porB), *mtrR* ('mtrR - includes promoter region) 23S (NG_23S), *parC* (NG_parC), *gyrA* (NG_gyrA), *rplD* (NEIS0133) and *rpsJ* (NEIS0129). In addition, to identify mosaic *mtrD* alleles associated with azithromycin resistance (Wadsworth et al. 2018; Williamson et al. 2019), the assemblies were blasted against the *mtrD* gene of FA1090, and the alignments visually inspected. For *penA* alleles, we used NG-STAR nomenclature (Demczuk et al. 2017).

Phylogeography and mutational mapping

Geographical transition patterns over time were inferred for the time-calibrated phylogeny using stochastic character mapping (Bollback 2006) as implemented in the R-package Phytools v0.6-99 (Revell 2012). We simulated 1000 stochastic character mappings conditional on the observed geographical states on the tips. We used four continents as categories: "North America", "Europe", "Asia" and "Oceania", to capture large distance transitions that occurred in small time intervals. The posterior probability of the geographical state of a node was calculated as the fraction of the inferred character mappings for which that node was in a given character state (continent). The times of geographical transition events were considered to be the times when the transitions happened on the branches leading to the downstream nodes with posterior probabilities higher than 50% for new locations. Since not all simulated character histories yielded downstream nodes in the same location, the times were calculated as the average time of the simulated character histories that were compatible with the events we were interpreting (so based on at least 50% of the geographical mappings, by the criteria above). We termed a geographical transition event a "successful establishment" if it produced more than one successive downstream observation in the new location.

To describe the acquisition of the known *penA* alleles on the internal nodes of the phylogeny, we performed another stochastic character mapping analysis as previously described. Since some of the states are observed very few times in the phylogeny, leading to uncertain transition rate estimations, we defined a custom transition matrix that only estimated transition rates between groups of states with multiple observations. Noting that the rare alleles 34.009, 42.001, and 72.001 are only one point mutation away from the allele 34.001, and observing that all samples with a small genetic distance from these carried the 34.001 allele, we only allowed transitions from 34.001 to these rare alleles. The method and transition matrix estimates are described in more detail in Appendix 1 in Supplementary Materials. The times of transitions between allele states were calculated in the same way as for the geographical states.

Downsampling for a balanced global representation

To rule out that key spatio-temporal inferences were the result of unbalanced sampling across regions and phenotypic drug susceptibility, we generated a downsampled dataset and re-run temporal and phylogeographic analyses. The downsampled dataset included 90 genomes from each of the regions North America, Europe and East Asia, as well as all the genomes

from Oceania (n=26), the Philippines, India and Chile (n=4 in total). The European and East Asian collections were downsampled in a random fashion, whereas the North American collection was downsampled in a targeted manner to minimize the effect of targeted sampling of resistant isolates from the USA (Grad et al. 2014): We thus included all 46 phenotypically cefixime susceptible isolates in addition to 30 resistant isolates from the USA. These were complemented by 14 Canadian genomes from isolates with unknown MICs. Parsnp, Gubbins and BactDating were employed to generate a dated phylogeny on which phylogeographic mapping was performed using SIMMAP as described above.

Data access

Extended data on all isolates, including accession codes, is available in the Figshare repository [10.6084/m9.figshare.12423596](https://doi.org/10.6084/m9.figshare.12423596)

Acknowledgements

The authors would like to thank our colleagues responsible for culturing and whole genome sequencing of gonococci at the Norwegian Institute of Public Health. We would also like to thank colleagues from The Scottish Bacterial STI reference lab for facilitating access and provision of the ST1901 isolates and Cecile Minder for the propagation and DNA extraction from these strains.

Magnus N. Osnes is supported by The National Graduate School in Infection Biology and Antimicrobials (IBA) hosted by the University of Oslo (IBA project 249062). L.v.D and F.B. acknowledge financial support from the Newton Fund UK-China NSFC initiative (grant MR/P007597/1) and the BBSRC (equipment grant BB/R01356X/1).

No funding bodies were involved in the design of the study, the collection, analysis, and interpretation of data and in writing the manuscript.

References

- Alfsnes K, Eldholm V, Olsen AO, Brynildsrud OB, Bohlin J, Steinbakk M, Caugant DA. 2020. Genomic epidemiology and population structure of *Neisseria gonorrhoeae* in Norway, 2016-2017. *Microb Genom* 6:e000359.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Anon. 2019. Gonococcal Infections - 2015 STD Treatment Guidelines. Available from: <https://www.cdc.gov/std/tg2015/gonorrhea.htm>

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455–477.
- Bignell C, Unemo M, European STI Guidelines Editorial Board. 2013. 2012 European guideline on the diagnosis and treatment of gonorrhoea in adults. *Int. J. STD AIDS* 24:85–92.
- Bollback JP. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7:88.
- Brynildsrud OB, Eldholm V, Bohlin J, Uadiale K, Obaro S, Caugant DA. 2018. Acquisition of virulence genes by a carrier strain gave rise to the ongoing epidemics of meningococcal disease in West Africa. *Proc. Natl. Acad. Sci. U. S. A.* 115:5510–5515.
- Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17:238.
- Centers for Disease Control and Prevention (CDC). 2012. Update to CDC's Sexually transmitted diseases treatment guidelines, 2010: oral cephalosporins no longer a recommended treatment for gonococcal infections. *MMWR Morb. Mortal. Wkly. Rep.* 61:590–594.
- Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43:e15.
- Demczuk W, Martin I, Sawatzky P, Allen V, Lefebvre B, Hoang L, Naidu P, Minion J, VanCaeseele P, Haldane D, et al. 2020. Equations To Predict Antimicrobial MICs in *Neisseria gonorrhoeae* Using Molecular Antimicrobial Resistance Determinants. *Antimicrob. Agents Chemother.* 64:e02005–e02019.
- Demczuk W, Sidhu S, Unemo M. 2017. *Neisseria gonorrhoeae* sequence typing for antimicrobial resistance, a novel antimicrobial resistance multilocus typing scheme for tracking global dissemination of *N. gonorrhoeae* Strains. *J Clin Microbiol* 55:1454–1468.
- Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. 2018. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* 46:e134.
- Didelot X, Dordel J, Whittles LK, Collins C, Bilek N, Bishop CJ, White PJ, Aanensen DM, Parkhill J, Bentley SD, et al. 2016. Genomic Analysis and Comparison of Two Gonorrhea Outbreaks. *MBio* 7:e00525–16.
- Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* 11:e1004041.
- EUCAST. 2020. European Committee on Antimicrobial Susceptibility Testing Breakpoint tables for interpretation of MICs and zone diameters Version 10.0. http://www.eucast.org/clinical_breakpoints/
- Ezewudo MN, Joseph SJ, Castillo-Ramirez S, Dean D, Del Rio C, Didelot X, Dillon J-A, Selden RF, Shafer WM, Turingan RS, et al. 2015. Population structure of *Neisseria gonorrhoeae* based on whole genome data and its relationship with antibiotic resistance. *PeerJ* 3:e806.

- Fifer H, Natarajan U, Jones L, Alexander S, Hughes G, Golparian D, Unemo M. 2016. Failure of Dual Antimicrobial Therapy in Treatment of Gonorrhoea. *N. Engl. J. Med.* 374:2504–2506.
- Golparian D, Harris SR, Sánchez-Busó L, Hoffmann S, Shafer WM, Bentley SD, Jensen JS, Unemo M. 2020. Genomic evolution of *Neisseria gonorrhoeae* since the preantibiotic era (1928–2013): antimicrobial use/misuse selects for resistance and drives evolution. *BMC Genomics* 21:116.
- Grad YH, Harris SR, Kirkcaldy RD, Green AG, Marks DS, Bentley SD, Trees D, Lipsitch M. 2016. Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United States, 2000–2013. *J. Infect. Dis.* 214:1579–1587.
- Grad YH, Kirkcaldy RD, Trees D, Dordel J, Harris SR, Goldstein E, Weinstock H, Parkhill J, Hanage WP, Bentley S, et al. 2014. Genomic epidemiology of *Neisseria gonorrhoeae* with reduced susceptibility to cefixime in the USA: a retrospective observational study. *Lancet Infect. Dis.* 14:220–226.
- Harris SR, Cole MJ, Spiteri G, Sánchez-Busó L, Golparian D, Jacobsson S, Goater R, Abudahab K, Yeats CA, Bercot B, et al. 2018. Public health surveillance of multidrug-resistant clones of *Neisseria gonorrhoeae* in Europe: a genomic survey. *Lancet Infect. Dis.* 18:758–768.
- Hu M, Nandi S, Davies C, Nicholas RA. 2005. High-level chromosomally mediated tetracycline resistance in *Neisseria gonorrhoeae* results from a point mutation in the *rpsJ* gene encoding ribosomal protein S10 in combination with the *mtrR* and *penB* resistance determinants. *Antimicrob. Agents Chemother.* 49:4327–4334.
- Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3:124.
- Kwong JC, Chow EPF, Stevens K, Stinear TP, Seemann T, Fairley CK, Chen MY, Howden BP. 2018. Whole-genome sequencing reveals transmission of gonococcal antibiotic resistance among men who have sex with men: an observational study. *Sex. Transm. Infect.* 94:151–157.
- Kwong JC, Gonçalves da Silva A, Dyet K, Williamson DA, Stinear TP, Howden BP, Seemann T. 2016. NGMASTER:in silico multi-antigen sequence typing for *Neisseria gonorrhoeae*. *Microb Genom* 2:e000076.
- Lee RS, Seemann T, Heffernan H, Kwong JC, Gonçalves da Silva A, Carter GP, Woodhouse R, Dyet KH, Bulach DM, Stinear TP, et al. 2018. Genomic epidemiology and antimicrobial resistance of *Neisseria gonorrhoeae* in New Zealand. *J. Antimicrob. Chemother.* 73:353–364.
- Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. 2019. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res.* 29:304–316.
- Liao M, Gu W-M, Yang Y, Dillon J-AR. 2011. Analysis of mutations in multiple loci of *Neisseria gonorrhoeae* isolates reveals effects of PIB, PBP2 and MtrR on reduced susceptibility to ceftriaxone. *J. Antimicrob. Chemother.* 66:1016–1023.
- Minin VN, Bloomquist EW, Suchard MA. 2008. Smooth skyride through a rough skyline:

Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* 25:1459–1471.

- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32:268–274.
- Ohnishi M, Golparian D, Shimuta K, Saika T, Hoshina S, Iwasaku K, Nakayama S-I, Kitawaki J, Unemo M. 2011. Is *Neisseria gonorrhoeae* initiating a future era of untreatable gonorrhoea?: detailed characterization of the first strain with high-level resistance to ceftriaxone. *Antimicrob. Agents Chemother.* 55:3538–3545.
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17:132.
- Osnes MN, Didelot X, de Korne-Elenbaas J, Alfsnes K, Brynildsrud OB, Syversen G, Nilsen Ø, de Blasio BF, Caugant DA, Eldholm V. 2020. The sudden emergence of a *Neisseria gonorrhoeae* strain with reduced susceptibility to extended-spectrum cephalosporins, Norway. *BioRxiv* <https://doi.org/10.1101/2020.02.07.935825>.
- Peng J-P, Yin Y-P, Chen S-C, Yang J, Dai X-Q, Zheng H-P, Gu W-M, Zhu B-Y, Yong G, Zhong N, et al. 2019. A Whole-genome Sequencing Analysis of *Neisseria gonorrhoeae* Isolates in China: An Observational Study. *EClinicalMedicine* 7:47–54.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3:217–223.
- Sánchez-Busó L, Golparian D, Corander J, Grad YH, Ohnishi M, Flemming R, Parkhill J, Bentley SD, Unemo M, Harris SR. 2019. The impact of antimicrobials on gonococcal evolution. *Nat Microbiol* 4:1941–1950.
- Shimuta K, Watanabe Y, Nakayama S-I, Morita-Ishihara T, Kuroki T, Unemo M, Ohnishi M. 2015. Emergence and evolution of internationally disseminated cephalosporin-resistant *Neisseria gonorrhoeae* clones from 1995 to 2005 in Japan. *BMC Infect. Dis.* 15:378.
- Simonsen M, Pedersen CNS. 2011. Rapid computation of distance estimators from nucleotide and amino acid alignments. In: Proceedings of the 2011 ACM Symposium on Applied Computing. SAC '11. New York, NY, USA: Association for Computing Machinery. p. 89–93.
- Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A. 2020. Producing Polished Prokaryotic Pangenomes with the Panaroo Pipeline. *BioRxiv* <https://doi.org/10.1101/2020.01.28.922989>.
- To T-H, Jung M, Lycett S, Gascuel O. 2016. Fast Dating Using Least-Squares Criteria and Algorithms. *Syst. Biol.* 65:82–97.
- Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15:524.
- Unemo M, Golparian D, Nicholas R, Ohnishi M, Gallay A, Sednaoui P. 2012. High-level cefixime- and ceftriaxone-resistant *Neisseria gonorrhoeae* in France: novel penA mosaic allele in a successful international clone causes treatment failure. *Antimicrob. Agents*

Chemother. 56:1273–1280.

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

Vincent LR, Kerr SR, Tan Y, Tomberg J, Raterman EL, Dunning Hotopp JC, Unemo M, Nicholas RA, Jerse AE. 2018. In Vivo-Selected Compensatory Mutations Restore the Fitness Cost of Mosaic *penA* Alleles That Confer Ceftriaxone Resistance in *Neisseria gonorrhoeae*. *MBio* 9:e01905-17.

Volz EM, Didelot X. 2018. Modeling the Growth and Decline of Pathogen Effective Population Size Provides Insight into Epidemic Dynamics and Drivers of Antimicrobial Resistance. *Syst. Biol.* 67:719–728.

Wadsworth CB, Arnold BJ, Sater MRA, Grad YH. 2018. Azithromycin Resistance through Interspecific Acquisition of an Epistasis-Dependent Efflux Pump Component and Transcriptional Regulator in *Neisseria gonorrhoeae*. *MBio* 9:e01419–18.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595.

Williamson DA, Chow EPF, Gorrie CL, Seemann T, Ingle DJ, Higgins N, Easton M, Taiaroa G, Grad YH, Kwong JC, et al. 2019. Bridging of *Neisseria gonorrhoeae* lineages across sexual networks in the HIV pre-exposure prophylaxis era. *Nat. Commun.* 10:3988.

the World Health Organization. 2016. WHO guidelines for the treatment of *Neisseria gonorrhoeae*. WHO Available from: <https://www.who.int/reproductivehealth/publications/rtis/gonorrhoea-treatment-guidelines/en/>

Yahara K, Nakayama S-I, Shimuta K, Lee K-I, Morita M, Kawahata T, Kuroki T, Watanabe Y, Ohya H, Yasuda M, et al. 2018. Genomic surveillance of *Neisseria gonorrhoeae* to investigate the distribution and evolution of antimicrobial-resistance determinants and lineages. *Microb Genom* 4:e000205.

Figure legends

Figure 1. Dated genome-wide SNP phylogeny of ST-1901 annotated with region of sampling, *penA* and *porB* mutations as well as ESC MIC values for isolates where these were available. The parental nodes for two waves of expansion out of Asia are colored (Wave 1 in green and Wave 2 in red, see text for details). CFM=cefixime, CRO=ceftriaxone; EUR-W, EUR-S and EUR-E/C = Western, Southern and Eastern/Central Europe respectively; NAm-W and NAm-E = Western and Eastern North America respectively; SAm = South America; Asia-E and Asia-S = Eastern and Southern Asia respectively; Ocean = Oceania.

Figure 2. Extended-spectrum cephalosporin MICs as a function of *penA* and *porB* allele. A) Observed MIC values across all *penA* alleles, only including alleles with at least two MIC measurements (609 total observations) The dotted horizontal line marks the clinical breakpoint for cefixime and ceftriaxone (0.25 mg/μl). RRS alleles are colored dark-red. B) Impact of *porB* alleles on ESC MICs in RRS *penA* genetic backgrounds. For *penA* alleles 34.009, 42.001, 71.001 and 72.001 there were too few observations for statistical inference. Mann-Whitney-Wilcoxon tests were performed to assess whether MICs varied significantly by *porB* allele type.

Figure 3. Time-resolved phylogeny of ST-1901 Clade A annotated with mapped geographical locations indicated by pie charts on nodes and tips. The outer color strip contains *penA* allele annotations. Historic events referred to in the text are indicated by stippled lines ending in event numbers.

Figure 4. Stochastic character mapping of the different *penA* alleles on the grey subset in Figure 3 leaving out Clade B and some of the basal samples from Asia in Clade A. The clades with grey background show Wave 1 and Wave 2. The color strips aligned to the phylogeny shows geographical locations (continent) of the tips. Historic events referred to in the text are indicated by stippled lines ending in event numbers.

Figure 5. A) Map showing dispersal and successful establishments of Wave 1 and Wave 2. Wave1 emerged and spread before the acquisition of ESC RRS alleles. Wave 2 had acquired the *penA* RRS allele 34.001 prior to its global expansion. A founder event in America was important for its further and repeated introduction to Europe. B) Effective population size over time estimated using Skygrowth (Volz and Didelot 2018) in the clade termed Wave 1. The vertical dotted line shows the point in time where with the maximum effective population size. The colored bars illustrate the recommended antibiotic treatments over time in the USA and Europe. FLQ=fluoroquinolones, CEPH=unspecified cephalosporin, AZM=azithromycin, CRO=ceftriaxone. The background shading in the plot illustrates the predicted effectiveness of specific antibiotics and combinations of antibiotics based on the RRS alleles present in each clade. Light orange shading indicates partial effect, darker red shading indicates strong effect. C) Similar to panel B for Wave 2.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

World region

- EUR-W
- EUR-S
- EUR-E/C
- NAm-W
- NAm-E
- SAm
- Asia-E
- Asia-S
- Ocean

ESCs

- >0.5
- >0.25
- >0.125
- >0.064
- ≥0.032
- <0.032
- nd

penA

- 10.001
- 34.001
- 13.001
- 42.001
- /34.009
- 72.001
- 5.002
- 12.001
- 12.004
- 9.001
- 35.001
- other

porB

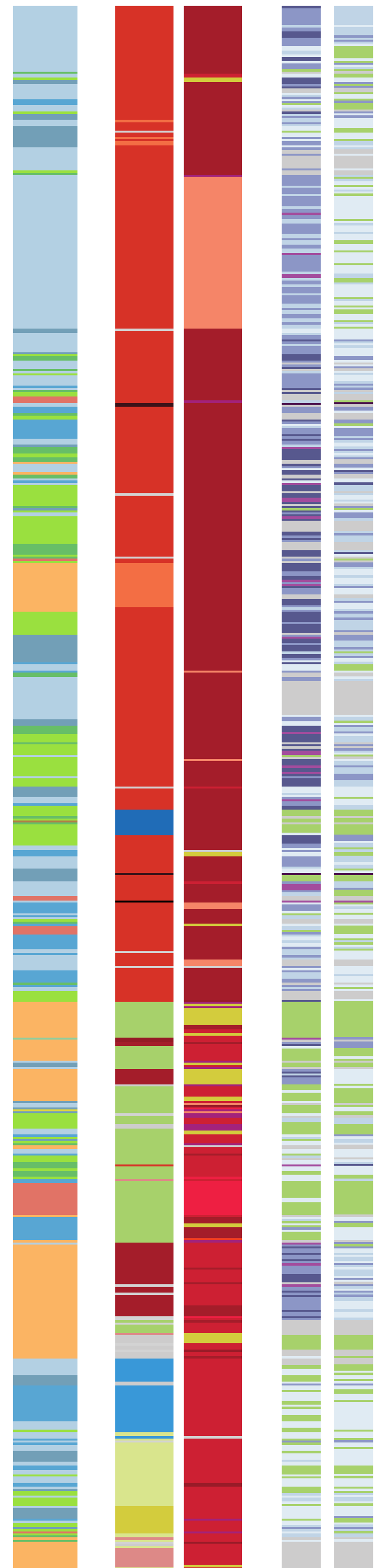
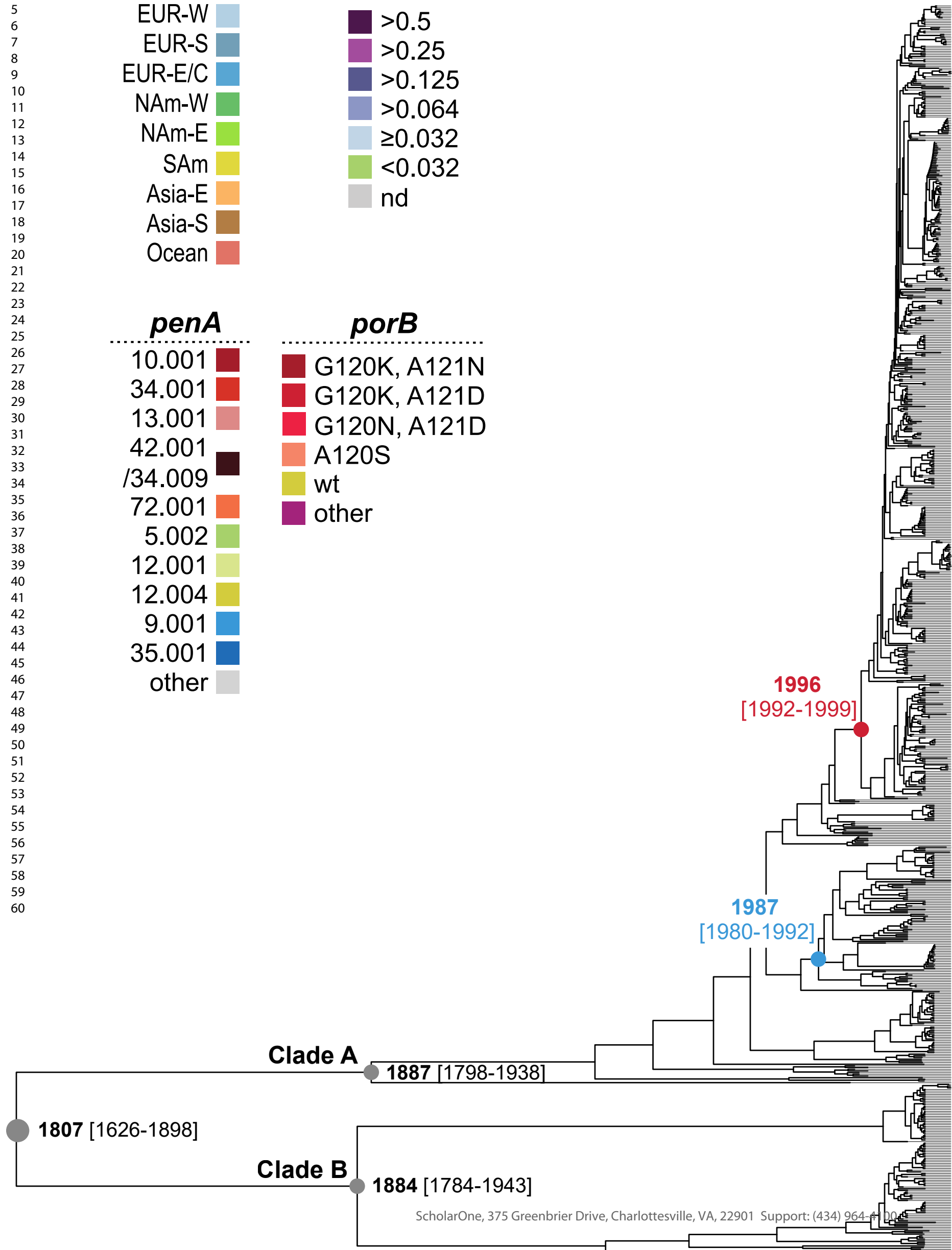
- G120K, A121N
- G120K, A121D
- G120N, A121D
- A120S
- wt
- other

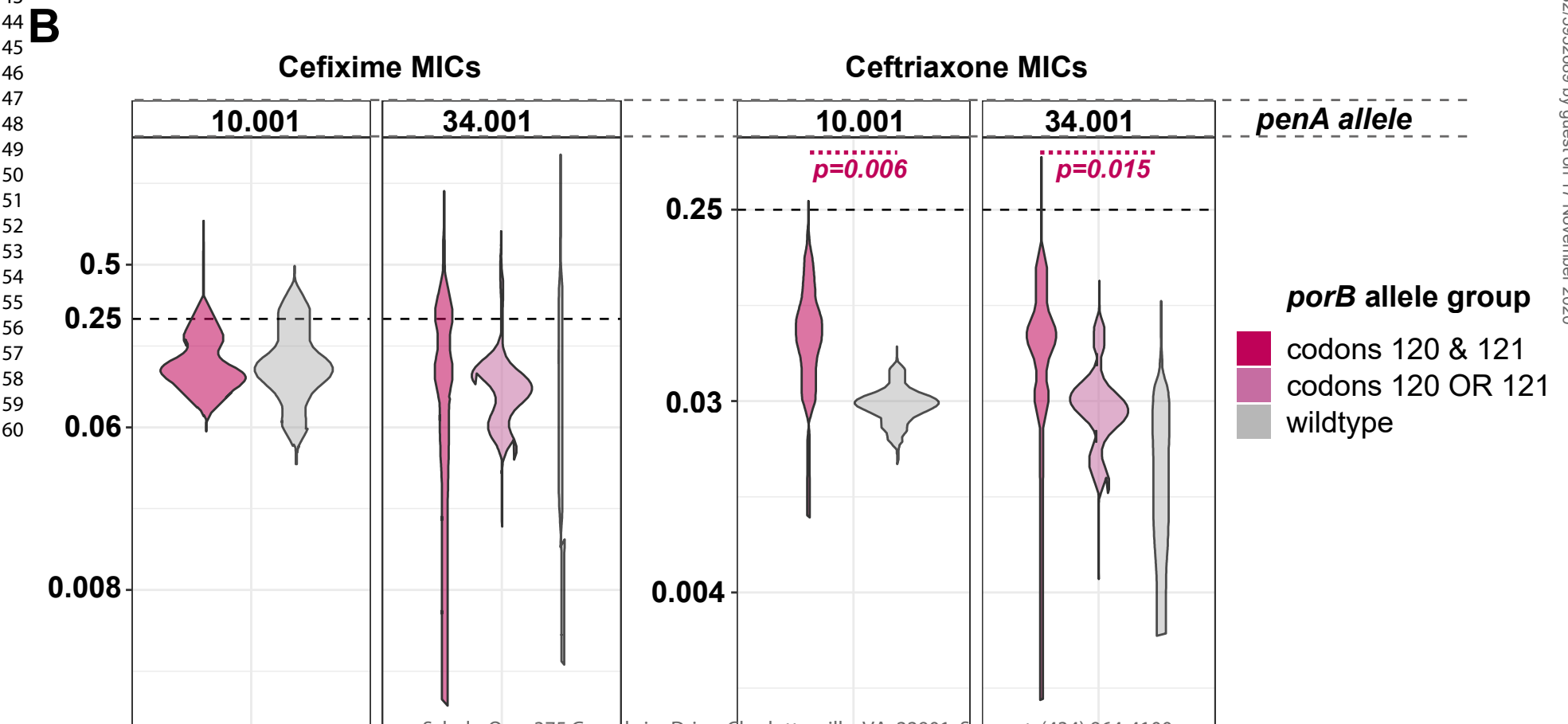
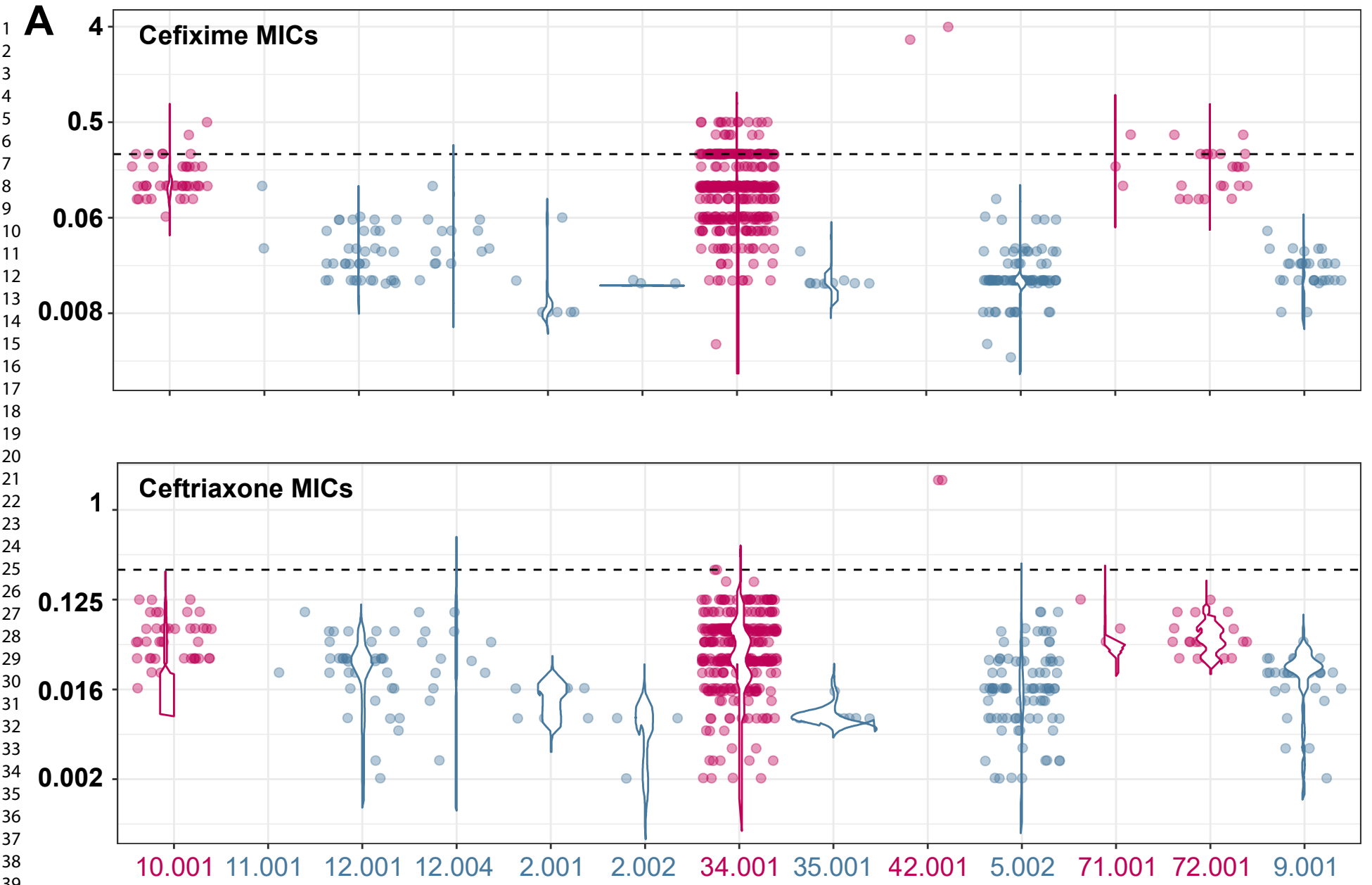
W-Region

penA

porB

CFM MIC
CRO MIC





Numbered events

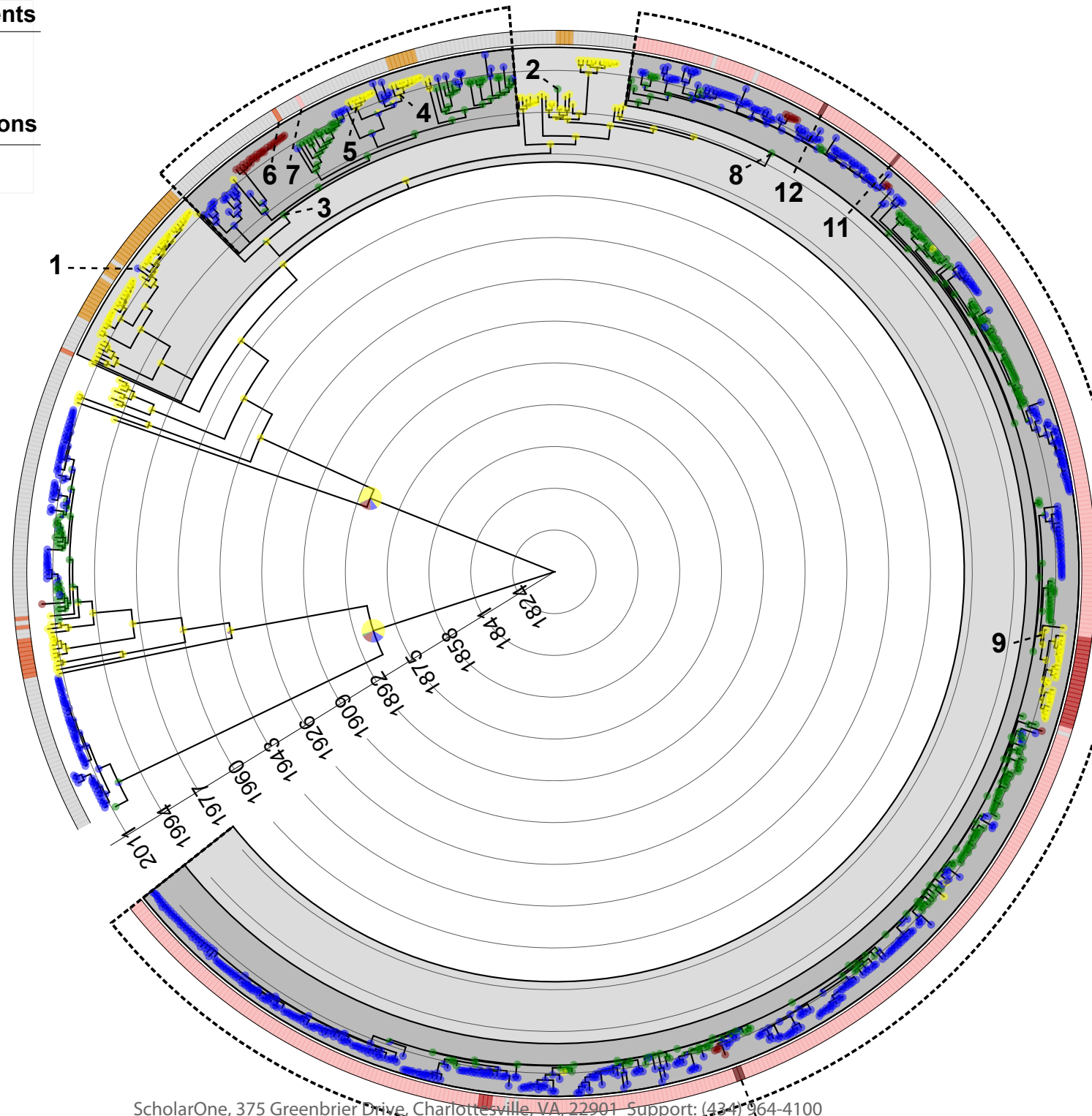
- 1
- 2
- 3
- 4

Mapped locations

- 5
- 6 ● America
- 7
- 8 ● Asia
- 9
- 10 ● Europe
- 11 ● Oceania
- 12
- 13
- 14

PenA alleles

- 15
- 16
- 17 ■ 10.001
- 18 ■ 13.001
- 19
- 20 ■ 34.001
- 21 ■ 34.009
- 22
- 23 ■ 42.001
- 24 ■ 72.001
- 25
- 26 ■ other
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49



Numbered events

- 1 -----
- 2 -----
- 3 -----

Mapped *penA* alleles

- 4 10.001
- 5 13.001
- 6 34.001
- 7 34.009
- 8 42.001
- 9 72.001
- 10 other

Continent

- 11 America
- 12 Asia
- 13 Europe
- 14 Oceania

