S3SUPPLEMENTARY TEXT

METHODS

Isolate selection and whole genome sequencing. This study included 334 male urethral *N. gonorrhoeae* isolates from the Gonococcal Isolate Surveillance Project (GISP) that were collected between January and December of 2016. Species identification and microbiological isolation was performed at one of the GISP regional laboratories according to methods established by GISP (1). Isolates were cultivated on GC medium base agar supplemented with 1% hemoglobin and/or 1% IsoVitaleX (BD, Franklin Lakes, NJ, USA), and were incubated at 36 ± 1ºC, 5% CO_2. Susceptibility to AZM was determined using the agar dilution method performed according to CLSI guidelines (2) at the GISP regional laboratories, and isolates with MIC values ≥ 16 µg/mL were endpointed using the ETEST® method (bioMerieux, France) at CDC. Due to the absence of a CLSI MIC breakpoint for AZM in 2016, AZM MIC values were interpreted as susceptible (MIC ≤1 µg/mL) or reduced susceptible (MIC ≥2 µg/mL) based on the previously established GISP alert value (1).

Isolates selected for sequencing after confirmatory antimicrobial susceptibility testing either were shipped to a partnering state public health lab for DNA extraction, library preparation, and sequencing or were processed at CDC. DNA extraction and library preparation were performed according to established methods (3). Genomic DNA was sequenced (paired-end, 2 × 250-bp read length) on Illumina MiSeq sequencers (Illumina Denmark ApS, Copenhagen, Denmark).

Phylogenomic analysis. Quality assessment was performed using FastQC 0.10.1 (4), and de novo assembly was conducted using SPAdes 2.5.1 (5). KmerGenie was used when the number of contigs exceeded 150 bp (6), and contaminants were identified using Kraken 0.10.5 (7). Quality trimming was conducted using Cutadapt v. 1.8.3 (8). The core genome single-nucleotide polymorphism (SNP) alignment was generated using Parsnp 1.2 (9) with default parameters, and the FA19 genome (10) was used as the reference. To control for the effects of recombination, Gubbins v. 2.3.1 (11) was used to

identify and remove SNPs found in recombinant regions, resulting in a core genome SNP alignment of 20,665 bp. The maximum-likelihood phylogeny was reconstructed based on the resulting core genome SNP alignment using RAxML version 8.2.12 (12) with 1,000 bootstrap replicates, and the phylogenetic tree was visualized using the ETE Toolkit (13). Clusters were identified using the R package fastbaps (14). The SNP distance matrix was calculated using snp-dists v0.4 (15) and the statistical analysis of the distance matrix was performed using Python 3.6.

Molecular typing. WGS data were used to determine the multilocus sequencing typing (MLST) and *N. gonorrhoeae*-multiantigen sequence typing (NG-MAST) allelic profiles for the selected isolates. Briefly, MLST sequence types (STs) were obtained using the program stringMLST v0.3.6 (16), while NG-MAST STs were obtained using NGMASTER v0.4 (17). Isolates with novel sequences were submitted to the MLST database (https://pubmlst.org/neisseria/) or to the NG-MAST database (https://www.ng-mast.net) to obtain new STs. New STs obtained for this study are MLST STs 13526, 13532, 13536-13539, 13542-13543, 13547-13549, and NG-MAST STs 16484, 16503, 17093, 17341, 17495, 17632, 17635, 17636, 17640, 17643-17645, 17647-17649, 17659, 17661, 17662, 17668, 17671, 17673, 17683-17690, 17693-17746, 17905, 17909, 17916, 17951, 17973, 17976-17985.

Antimicrobial resistance gene determination. The genetic determinants listed in Tables 1, 2, and Supplementary Dataset S1 were extracted from the WGS data using a custom analysis pipeline written in Python. Individual raw reads and de novo assembled genomes were used as input for analysis. Similar to the phylogenomic analysis, raw reads are first analyzed using FastQC (4) then filtered and trimmed using trim_galore v0.3.7 (18). The trim_galore output was used as the input for reference mapping in breseq v0.30 (19), with FA19 (10) as the reference genome. To extract variants in the *N. gonorrhoeae* 23S rRNA and determine the copy number in the genome, breseq was run a second time using the sequence for a single allele of the 23S rRNA gene from strain MS11 (20) as the mapping

reference with the polymorphism-prediction option turned on. Variant extraction and data formatting from the breseq output utilized the Pandas and Biopython packages (21, 22).

To determine the nucleotide present in the 23S rRNA sequence at *E. coli* positions 2059 or 2611, the output from breseq after mapping to reference sequence X67293 was searched for Read Alignment (RA) evidence at positions 2047 (*E. coli* position 2059) and 2599 (*E. coli* position 2611). The frequency of the base within the mapped reads at specific positions was calculated using the polymorphism prediction feature of breseq, and the frequency was used to estimate the number of alleles in the *N. gonorrhoeae* genome as described (23).

To determine the presence of the full-length MtrR protein in the tested isolate, the *mtrR* nucleotide sequence was extracted from the *de novo* assembly using blastn and the *mtrR* nucleotide sequence from FA19. This nucleotide sequence was then translated and compared to the length of wild type MtrR.

Determination of the presence of mosaic-like sequence in the *mtrR* gene (i.e., mosaic-like *mtrR*) was made by extracting the nucleotide sequence of *mtrR* as described above. The nucleotide sequence was compared to a known mosaic-like *mtrR* sequence (24) using blastn. An identity cutoff of 98% was established based on pairwise comparisons of various mosaic-like *mtrR* loci to themselves and to wild type sequences. If the extracted sequence was ≥ 98% identical to the mosaic-like sequence, the extracted sequence was denoted as mosaic.

Determination of amino acid mutations in the coding sequence of *mtrR* and the presence of a disruption within the inverted repeat in the *mtrR* promoter region was accomplished using the method described above for calling nucleotide mutations in 23S rRNA genes. After mapping the raw reads to the reference genome, breseq was used for variant calling and to output amino acid mutations and annotations. This output was then searched for RA evidence within the *mtrR* coding sequence at amino acid positions 39, 44, 45, and 47. If evidence was found, that amino acid was returned. However, if no evidence was found for a position, the isolate was assumed to have the same amino acid as FA19. The

presence of a promoter disruption was determined by searching for RA evidence at FA19 genome positions 1110844–1110848. Promoter mutations were confirmed by performing a local blastn search (ncbi-blast+ v2.6.0 (25)) using the *mtrR-CDE* promoter sequence from FA19 as the query against the isolate *de novo* assembly.

The presence of *ermB*, *ermC* (reference from *Neisseria meningitidis* MC58), *ermF* (reference from *Bacteriodes fragilis*), or *mefA* was determined by mapping the trimmed reads to a reference sequence for each gene using breseq (26-29). The gene was called as present if the average read coverage across the locus was greater than 2.0.

Statistical analyses. $\chi 2$ tests with a Yates correction were used to determine associations between mutational patterns with respect to reduced susceptibility to AZM (30). A Bonferonni correction was applied to account for multiple tests. P values greater than or equal to 0.05 were considered not significant.

Data availability. Raw whole genome sequence data are available from the National Center for Biotechnology Information (31). Accession numbers for each isolate can be found in Supplementary Dataset S1.

SUPPLEMENTARY FIGURES

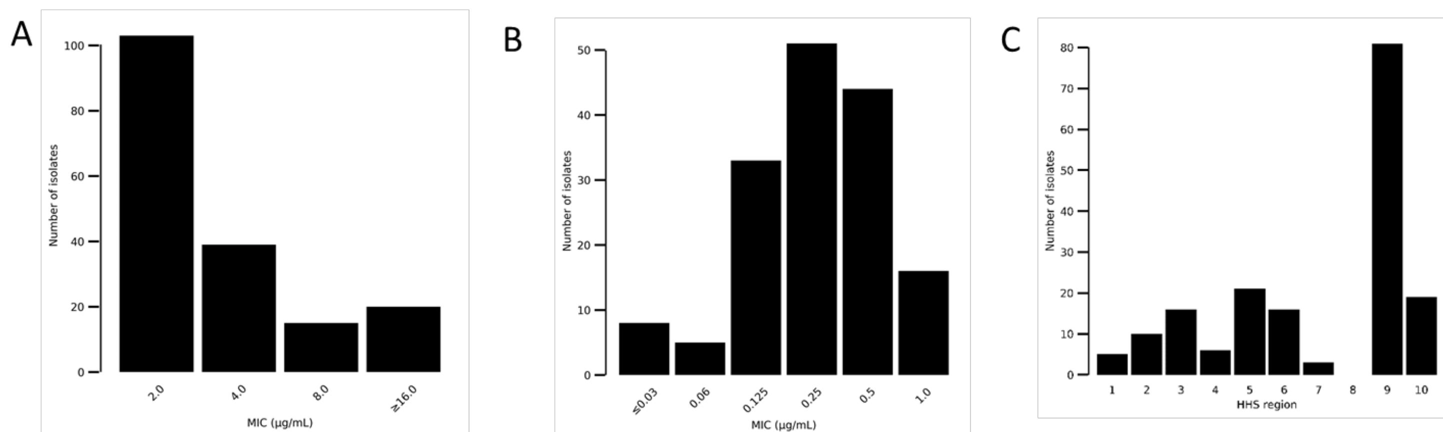

FIGURE S1 Distribution of isolates by azithromycin MIC and geography. A) MIC distribution among included isolates with an AZM MIC ≥ 2 µg/mL (n = 177). B) MIC distribution among included isolates with an AZM MIC ≤ 1 µg/mL (n = 157). C) HHS region distribution of the number of included isolates with an AZM MIC ≥ 2 µg/mL. See Table S1 for the distribution of GISP sites per HHS region.

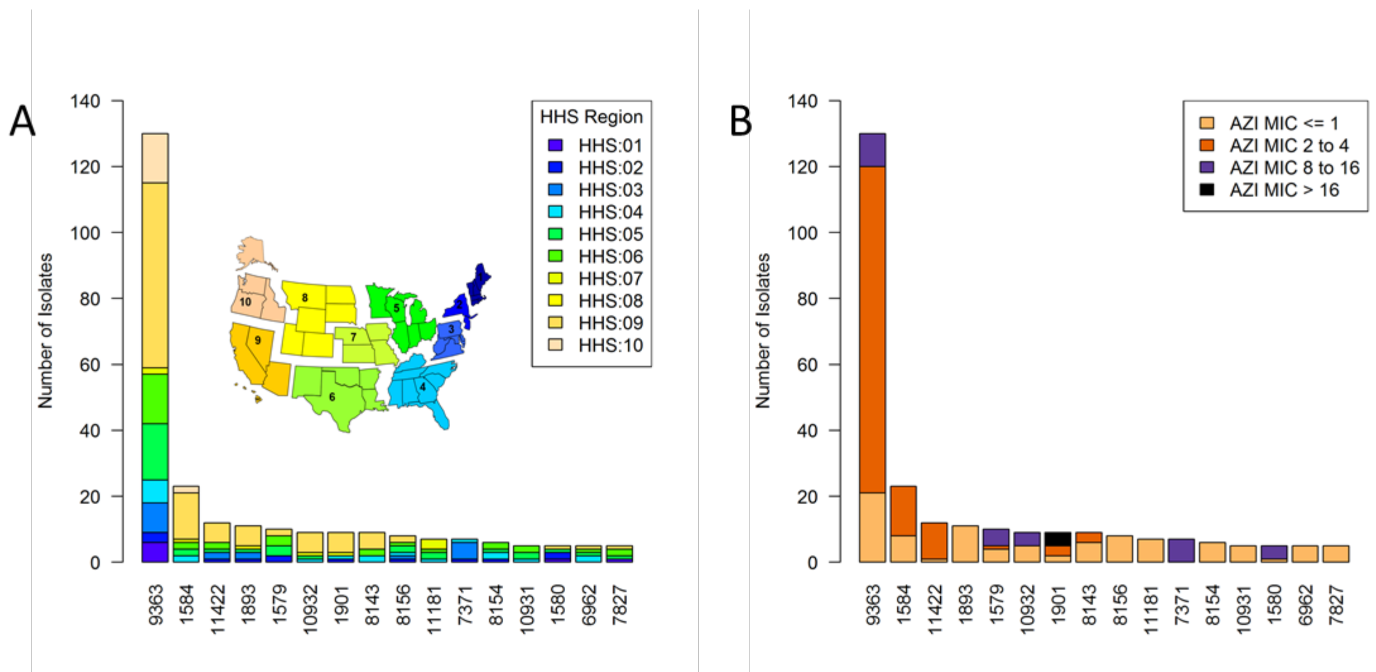

FIGURE S2 Distribution of MLST STs by HHS region or AZM MIC. A) Geographic distribution of MLST STs according to HHS region. B) Histogram of MLST STs showing AZM MIC (µg/mL) categorical breakdown. MLST STs represented by five or more isolates are included in the graphs.

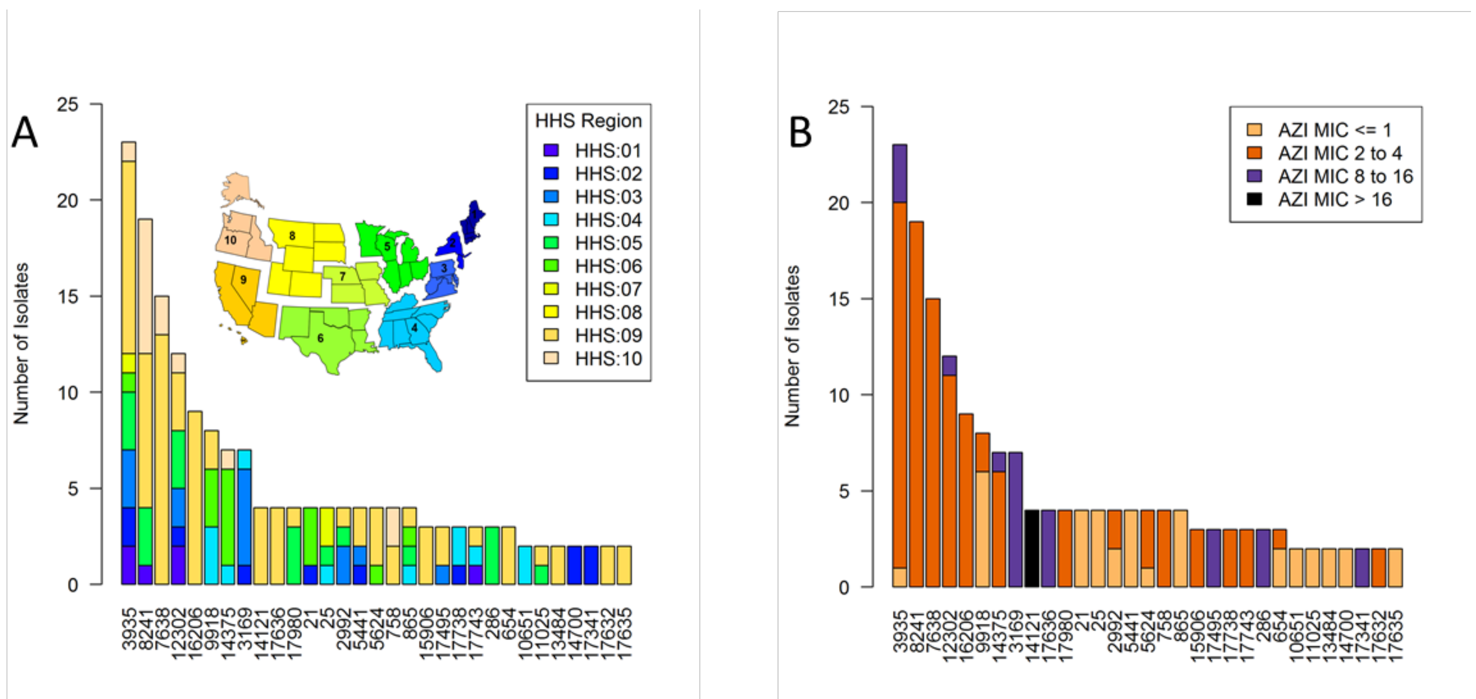

FIGURE S3 – Distribution of NG-MAST STs by HHS region or AZM MIC. A) Geographic distribution of NG-MAST STs according to HHS region. B) Histogram of NG-MAST STs showing AZM MIC (µg/mL) categorical breakdown. NG-MAST STs represented by two or more isolates are included in the graphs.

SUPPLEMENTARY TABLE

TABLE S1 Distribution of GISP sites by HHS region in 2016

HHS region	GISP Sites (#)
1	1
2	2
3	1
4	3
5	6
6	3
7	1
8	0
9	8
10	2

REFERENCES

1. Centers for Disease Control and Prevention. 2016. GISP Protocol 2016.
2. Clinical and Laboratory Standards Institute. 2012. Methods for antimicrobial susceptibility testing of anaerobic bacteria. Clinical and Laboratory Standards Institute,, Wayne, PA.
3. Centers for Disease Control and Prevention. February 11, 2016 2016. PulseNet Methods & Protocols. https://www.cdc.gov/pulsenet/pathogens/wgs.html. Accessed July 30.
4. Andrews S. 2014. FastQC A Quality Control tool for High Throughput Sequence Data, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-77.
6. Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30:31-7.
7. Davis MP, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. 2013. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. Methods 63:41-9.
8. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17:3.
9. 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.
10. Abrams AJ, Trees DL, Nicholas RA. 2015. Data from "Complete Genome Sequences of Three Neisseria gonorrhoeae Laboratory Reference Strains." GenBank https://www.ncbi.nlm.nih.gov/nuccore/CP012026.1 (accession no. CP012026).

11. 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.

12. Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456-63.

13. Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol 33:1635-8.

14. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. 2019. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res 47:5539-5549.

15. Seemann T. 2018. Pairwise SNP distance matrix from a FASTA sequence alignment, v0.4. https://github.com/tseemann/snp-dists.

16. Gupta A, Jordan IK, Rishishwar L. 2017. stringMLST: a fast k-mer based tool for multilocus sequence typing. Bioinformatics 33:119-121.

17. Kwong JC, Goncalves 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.

18. Kreuger F. 2012. Trim Galore, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

19. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165-88.

20. Wolff K, Sperka S, Stern A. 1992. Data from "Phylogeny and nucleotide sequence of a 23S rRNA gene from Neisseria gonorrhoeae and Neisseria meningitidis." GenBank https://www.ncbi.nlm.nih.gov/nuccore/X67293 (accession no. X67293).

21. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422-3.

22. McKinney W. 2010. Data Structures for Statistical Computing in Python, p 51-56. *In* van der Walt S, Millman J (ed), Proceedings of the 9th Python in Science Conference.

23. Johnson SR, Grad Y, Abrams AJ, Pettus K, Trees DL. 2017. Use of whole-genome sequencing data to analyze 23S rRNA-mediated azithromycin resistance. Int J Antimicrob Agents 49:252-254.

24. Demczuk W, Martin I, Peterson S, Bharat A, Van Domselaar G, Graham M, Lefebvre B, Allen V, Hoang L, Tyrrell G, Horsman G, Wylie J, Haldane D, Archibald C, Wong T, Unemo M, Mulvey MR. 2016. Data from "Genomic epidemiology and molecular resistance mechanisms of azithromycin resistant Neisseria gonorrhoeae in Canada from 1997 to 2014." GenBank https://www.ncbi.nlm.nih.gov/nuccore/KT954125 (accession no. KT954125).

25. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

26. Cousin S, Whittington WLH, Roberts MC. 2003. Acquired Macrolide Resistance Genes in Pathogenic Neisseria spp. Isolated between 1940 and 1987. Antimicrobial Agents and Chemotherapy 47:3877-3880.

27. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Hickey EK, Haft DH, Salzberg SL, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parksey DS, Blair E, Cittone H, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Vamathevan J, Gill J, Scarlato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC. 2000. Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science 287:1809-15.

28. Smith CJ. 1987. Nucleotide sequence analysis of Tn4551: use of ermFS operon fusions to detect promoter activity in Bacteroides fragilis. J Bacteriol 169:4589-96.

29. Roberts MC, No DB. 2016. Data from "Neisseria gonorrhoeae isolate 492 rRNA methylase leader peptide (ermB-L) gene, complete cds; and rRNA methylase (ermB) gene, partial cds." GenBank https://www.ncbi.nlm.nih.gov/nuccore/EU048317.1 (accession no. EU048317.1).

30. Kim HY. 2017. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor Dent Endod 42:152-155.

31. Centers for Disease Control and Prevention. 2016. Data from "Neisseria gonorrhoeae Genome Sequencing." NCBI SRA https://www.ncbi.nlm.nih.gov/bioproject/PRJNA317462 (accession no. PRJNA317462).