Genome-wide association identifies antibiotic resistance and susceptibility mutations in Neisseria gonorrhoeae

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Committee on Higher Degrees in Biological Sciences in Public Health

have examined a dissertation entitled

Genome-wide association identifies antibiotic resistance
and susceptibility mutations in *Neisseria gonorrhoeae*

presented by Kevin Chen Ma

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Marc Lipsitch

Signature _____

Typed name: Prof. Daniel Neafsey

Signature _____

Typed name: Prof. Eric Rubin

Signature _____

Typed name: Prof. Zamin Iqbal

Date: March 16, 2022

Genome-wide association identifies antibiotic resistance and susceptibility mutations in *Neisseria gonorrhoeae*

A dissertation presented

by

Kevin Chen Ma

to

The Committee on Higher Degrees in Biological Sciences in Public Health

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological Sciences in Public Health

Harvard University

Cambridge, Massachusetts

March 2022

Dissertation Advisor: Yonatan Hagai Grad Kevin Chen Ma

Genome-wide association identifies antibiotic resistance and susceptibility mutations in *Neisseria gonorrhoeae*

Abstract

The emergence of resistance to multiple frontline antibiotics complicates treatment of *Neisseria gonorrhoeae*, the etiologic agent of gonorrhea. Characterizing the molecular mechanisms of resistance will inform the development of sequence-based diagnostics and improve our understanding of how gonococci respond to host and drug pressures. Here, we conduct several bacterial genome-wide association studies (GWASes) that incorporate pan-genomic diversity and stringent control for population structure on the largest assembled dataset to date of gonococcal genome sequences and antibiotic testing data. In Chapter 2, we describe how we identified a single mutation associated with increased susceptibility to multiple antibiotics using GWAS and then characterized the environments it arises in using host metadata. In Chapter 3, we use genomic epidemiology in combination with insights from structural biology to identify clinical isolates with rare mutations conferring increased antibiotic resistance. In Chapter 4, we demonstrate how extending standard GWAS models by adjusting for the effect of known resistance genes results in substantially decreased false positives and increased power to detect new resistance genes. The results of these studies will advance our understanding of the evolution of *N. gonorrhoeae* and provide new strategies for interrogating the genetic basis of antibiotic resistance and susceptibility broadly in microbial pathogens.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I've been lucky to have the support of an incredible community throughout my dissertation. Thank you first to my advisor, Dr. Yonatan Grad, who has been an endless source of scientific inspiration, life advice, and puns. I'm also beyond grateful to all of the members of this incredible lab. Special shoutout to Grad Lab grad students Dan, Daphne, and Allie for all of your camaraderie and kvetching, and to our brilliant and patient postdocs Tatum, Sam, and Stephen for teaching me everything there is to know about genomics, experiments, modeling, and beyond. I'm also thrilled that the lab has grown so much and that there are so many new faces — I know you'll all have a blast.

To my committee, Dr. Megan Murray, Dr. Ashlee Earl, and Dr. Marc Lipsitch — thank you so much for all of your insights, scientific guidance, and thoughtfulness over the years. Thank you as well to Dr. Daniel Neafsay, Dr. Eric Rubin, and Dr. Zamin Iqbal both for serving as my dissertation examiners and for your advice along the way. Finally, a particular shoutout to Dr. Barry Bloom for your continued mentorship and for getting me started in the first place — in infectious diseases through LS120, and by helping me find the perfect thesis lab. Here's to many more late-night science discussions over Chinese food.

I've had the opportunity to be a part of several amazing student communities at Harvard, including my PhD program in Biological Sciences in Public Health; thank you to my wonderful cohort, especially Stacey and Francesca, and to Eric and Tatevik for all you do to support the students. In the Center for Communicable Disease Dynamics, I learned a ton about mathematical modeling, both its methodologies and the importance of supporting a diverse future generation of modelers. Thank you in particular to Marc, Tia, and the rest of CCDD

leadership and staff, as well as Tiggy, Madikay, Jess, Xueting, Nancy, and Rebecca for being so great to work with and learn from. Finally, thank you to Dr. Dyann Wirth and Dr. Sarah Fortune for your leadership in Immunology and Infectious Diseases, and to all of the departmental staff for creating a supportive environment for your trainees.

To all of the friends I've made in graduate school — Bolu, Megan, Harry, and more — I'm excited to see what you all will accomplish next and thrilled we had the chance to learn together. A big thank you as well to my sports teams friends for all of your support both on the field and off, in particular Vince, Jason, and Daniela. To Malek, Kevin, Andy, Cindy, Thom, Jack, and the DS team, thank you all for supporting me in your own ways, whether it be through sharing memes or pocketing me in Overwatch. And finally, thank you to my perpetually supportive and loving family, and my partner Eunice for always being willing to listen, laugh, and try out all the new foods together.

Chapter 1: An introduction to antibiotic resistance in *Neisseria gonorrhoeae*

Introduction

Neisseria gonorrhoeae is a Gram-negative obligate human bacterial pathogen and the etiologic agent of the sexually transmitted disease gonorrhea, which affects roughly 100 million people globally each year[1]. The emergence of multidrug-resistant strains of *N. gonorrhoeae* presents an urgent challenge to control gonorrhea and its complications, including pelvic inflammatory disease, infertility, and disseminated infection. In the US, the Centers for Disease Control and Prevention (CDC) recommends a change in empiric treatment protocols when the prevalence of drug resistance surpasses five percent. Resistance to penicillin and tetracycline met this threshold in the late 1980s and resistance to ciprofloxacin in 2007, leading to discontinuation of use of these drugs. The cephalosporins (e.g., oral cefixime or intramuscular ceftriaxone) were the only remaining alternatives, but the observation of cefixime treatment failures in Europe and Japan coupled with rising rates of cefixime resistance in CDC surveillance isolates led to cefixime use being discontinued as well[2,3].

In response to these trends, in 2012 the CDC recommended a combination regimen of the macrolide azithromycin (AZI), which targets the 50S ribosome, and ceftriaxone (CRO), which targets the penicillin-binding proteins responsible for cell wall cross-linking. Despite use of dual therapy, resistance has now emerged to both agents. Increased minimum inhibitory concentrations (MICs), a quantitative measure of resistance levels, to AZI are partially attributable to mutations in the multiple transferable resistance (i.e., *mtr*) efflux pump locus comprising the *mtrR* repressor and the *mtrCDE*-encoded pump. Disruption of *mtrR* due to active site or promoter mutations can raise MICs by nearly 10-fold in some strain backgrounds but alone are not sufficient to confer reduced susceptibility (defined by the CDC as an MIC ≥2.0 μg/ml)[4]. Recombination across the entire *mtr* locus can result in reduced susceptibility via a combination of transcriptional upregulation and structural changes to the macrolide-binding

residues of the pump[5,6]. Higher-level resistance arises through mutations in 23S rRNA, a ribosomal component, at the macrolide binding site. The commonly observed C2611T substitution results in MICs of 8-32 μg/mL, depending on the number of mutated alleles, whereas the A2059G substitution can result in even higher resistance levels when all four copies are mutated (MICs of ≥256 μg/mL)[7,8].

Reduced susceptibility to ceftriaxone (CRO) (defined by the CDC as an MIC ≥0.125 μg/ml) is also dependent, in part, on mutations abrogating function of the MtrR repressor, as well as synergistic mutations in the highly expressed outer membrane porin PorB[9]. These mutations alone are not sufficient to confer elevated MICs: target-site modification of the peptidoglycan cross-linking enzyme PBP2 encoded by the *penA* gene is also required. In Europe and North America, reduced susceptibility is associated with "mosaic" *penA* alleles that arise through recombination from commensal Neisserial donors, resulting in decreased binding affinity for beta-lactams[3,10-15]. The most prevalent and successful allele to date is the mosaic *penA* XXXIV allele, which has disseminated clonally, driven by transmission in men who have sex with men[14-16]. *penA* substitutions are considered to be the primary mechanism of reduced susceptibility, in contrast to other Gram-negative bacteria where acquisition of extended-spectrum beta-lactamases is a more common pathway to cephalosporin resistance[17].

Population surveillance of US clinical gonococcal isolates initially indicated that AZI and CRO resistance was non-overlapping[14], but case reports have now identified strains associated with dual therapy treatment failure that contain both mosaic *penA* and 23S rRNA A2059G mutations[18-20]. The emergence of these multidrug-resistant strains raises the possibility of widespread treatment failure, and the remaining options in the antibiotic arsenal are limited. The aminocyclitol spectinomycin targets the 30S ribosome and is used as frontline therapy in South Korea, but is currently unavailable in many countries[21]. Other drugs that may be repurposed for

gonococcal treatment include the carbapenem ertapenem and the aminoglycoside gentamicin, but current data are sparse on clinical efficacy, and the potential risks of bystander selection from widespread usage of carbapenems to treat this prevalent STD are considerable[21]. A fundamental shift in our approach to selecting antibiotics for treating *N. gonorrhoeae* is necessary to manage the long-term consequences of drug resistance.

The development of a molecular diagnostic that could provide rapid prediction of antibiotic susceptibility would slow the spread of resistance. While drug-resistant strains have emerged, over 60% of surveyed gonococcal isolates in the US remain pan-susceptible to all antibiotics, including antibiotics previously abandoned for empiric therapy[22]. By decreasing the use of current frontline therapy, mathematical modeling indicates that a point-of-care diagnostic could decrease the selective pressure for resistance to emerge to newer agents and thus prolong the effective lifespan of frontline antibiotics[23,24]. However, conventional approaches for developing diagnostics that can predict antibiotic susceptibility – such as nucleic acid amplification tests – require a deep understanding of the genotypic basis of resistance. This requirement also holds true for diagnostics in development that employ whole-genome sequencing (WGS) or CRISPR for nucleic acid detection[25,26].

While some gonococcal resistance mechanisms are known, unexplained variation in MICs exists for all drugs, including azithromycin and ceftriaxone. For azithromycin, a population genomic surveillance of 1102 US CDC clinical isolates indicated that the basis for resistance in over one-third of strains was unclear[14]. In this same dataset, while the mosaic *penA* XXXIV allele had high sensitivity for predicting ceftriaxone resistance, a small proportion of resistance likewise remained unexplained. Follow-up case reports confirmed that some of these strains harbored no known resistance mutations in *penA*, and instead had developed novel pathways to cephalosporin resistance[27]. Within the subset of strains that do have mosaic *penA* XXXIV, MICs

can vary by as much as fifteen-fold, indicating the presence of unknown epistatic factors that modulate the degree of resistance conferred by *penA* alleles. Accurate prediction of MICs for diagnostic purposes therefore rests critically on identifying these unknown drivers of antibiotic resistance and susceptibility.

Genome-wide association studies (GWASes) are a powerful statistical approach for discovering novel loci associated with antibiotic resistance. Previous approaches for uncovering resistance markers have relied on extensive characterization of a small number of candidate genes, such as the *mtr* efflux pump or the porin *porB*. These approaches will fail to uncover variants in novel genes. The approach of GWAS is to instead systematically search for variants across the genome that are over-represented in resistant compared to susceptible strains. GWASes have been successfully applied to uncover microbial resistance mutations for drug-resistant tuberculosis and other bacteria[28,29]; however, care must be taken to ensure that false positives do not arise from population structure, defined as the existence of subpopulations of related isolates that differ in the phenotype of interest[30]. Population structure can lead to spurious associations between non-causal alleles and the phenotype of interest if both variant and phenotype are similarly structured due to, for instance, the clonal expansion of resistant lineages. Statistical approaches to address this confounder vary, but recent studies demonstrate that a linear mixed model (LMM) can substantially control for population structure in microbial GWASes[31].

In an LMM-based GWAS, linear regression is run independently for every variant (Figure 1.1). For the phenotype, we use antibiotic MICs, providing a continuous measure of resistance levels. Variants can be either single nucleotide polymorphisms or, in our case, unitigs generated from *de novo* assemblies, which provide a flexible approach to capture variation in bacterial

pangenomes[32]. Unitigs capture variants not easily analyzed by conventional read mapping-based approaches such as genes on plasmids, accessory genes, and indels.

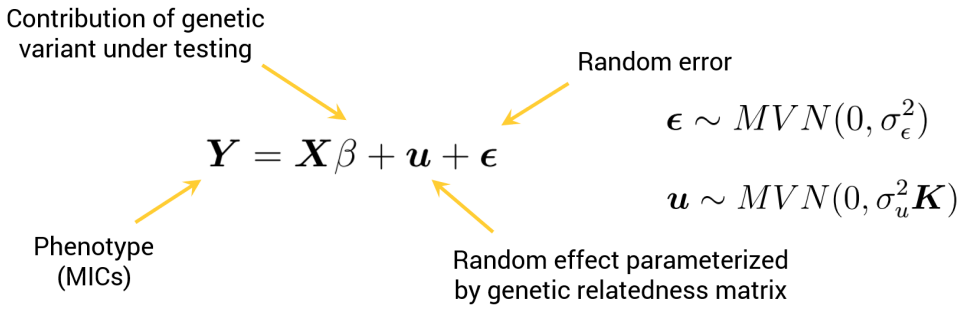


Figure 1.1 – Components of the linear mixed model GWAS. Here, $\mathbf{Y}$ is the vector of antibiotic MICs, $\mathbf{X}$ is the genetic variant (e.g., unitig) under consideration and β its fixed effect, $\boldsymbol{u}$ is a random effect parameterized with population structure matrix $\boldsymbol{K}$ and additive genetic variance σ^2_u, and $\boldsymbol{\epsilon}$ is a random effect that models the non-genetic effects parameterized with variance σ^2_e.

The LMM regression model controls for population structure by including a random effects term $\boldsymbol{u}$ that has variance parameterized by a genetic relatedness matrix $\boldsymbol{K}$, which allows for genetically related strains to have similar phenotypes. Numerous software packages have been developed to estimate LMMs, and I primarily use the Pyseer software package designed in particular for conducting microbial GWAS[33].

The field of GWAS as applied to *N. gonorrhoeae* is still nascent: Schubert et al. published the first LMM-based GWAS conducted on gonococcal antimicrobial resistance using an exploratory dataset of 1102 isolates and a validation dataset of 495 isolates on five antibiotics (penicillin, tetracycline, cefixime, ciprofloxacin, and azithromycin) in 2019[34]. They observed highly significant associations in known resistance markers for three of the five drugs, highlighting the potential for GWAS in *N. gonorrhoeae*. However, because only SNP-based variants were considered, accessory genome and plasmid-mediated resistance elements were missed.

Additionally, known variants with lower effect sizes were not detected, which could potentially be addressed by using larger sample sizes to increase the power of GWAS.

Here, we identify genetic markers of antibiotic resistance and susceptibility in *N. gonorrhoeae* by running microbial GWAS on the largest dataset of gonococcal genomes to date. This meta-analysis dataset comprises nearly five thousand WGS strains with associated MIC metadata for six antibiotics, collected over five decades across thirteen studies (Table 1.1). The magnitude of this sample size and the genetic and phenotypic diversity present in this global collection enhances the statistical power of microbial GWAS approaches[35], and we experimentally validate several of the most significant variants. In Chapter 2, we conduct GWAS on azithromycin, ceftriaxone, and ciprofloxacin, identifying a loss-of-function (LOF) mutation in the efflux pump MtrCDE as the most significant novel variant across all three antibiotics. This variant confers increased multi-drug susceptibility, and we demonstrate that these and other efflux pump loss mutations are overrepresented in cervical isolates, providing evidence that efflux pump expression in the female urogenital tract results in a fitness cost for pathogenic *Neisseria*. In Chapter 3, we conduct further targeted genomic epidemiology analysis of the MtrD efflux pump gene, leveraging data from structural biology studies on residues involved in drug binding. We identify clinical isolates in our meta-analysis dataset with rare mutations that are associated with azithromycin non-susceptibility. In Chapter 4, we develop a straightforward statistical method for improving the performance of GWAS in *N. gonorrhoeae* to identify resistance mutations with lower effect sizes. Compared to a single-variant GWAS, conducting GWAS conditioning on known resistance mutations with high effect sizes reduces the number of spurious associations and identifies a mutation in the RplD ribosomal protein as significantly associated with increased azithromycin resistance. Through targeted genomic analyses, we identify additional rare RplD mutations and show that in aggregate, these RplD resistance-associated mutations are geographically widespread and prevalent. Together, these results

highlight the potential for population-level genetic association and genomic epidemiology studies to uncover new pathways underlying antimicrobial resistance and susceptibility in *N. gonorrhoeae.*

Table 1.1 – Datasets included for the GWASes and genomic epidemiology studies[a].

Publication	Study summary	Timespan	Number
Mortimer et al., 2020[36]	Transmission and AMR surveillance in New York City, USA	2011-2015	888
Sánchez-Busó et al., 2019[37]	Worldwide phylogeography and evolution of gonococcus	1979-2012	408
Yahara et al., 2018[38]	AMR surveillance in Kyoto and Osaka, Japan	1996-2015	260
Ryan et al., 2018[39]	AMR surveillance in Ireland	2012-2016	39
Harris et al., 2018[15]	Genomic survey across 20 European Euro-GASP participant countries	2013	1048
Fifer et al., 2018[40]	High-level azithromycin resistance outbreak in UK	2004-2017	50
Lee et al., 2018[41]	Genomic epidemiology in New Zealand	2014-2015	397
Kwong et al., 2017[42]	Transmission among MSM in Melbourne, Australia	2005-2014	94
Eyre et al., 2017[43] and De Silva et al., 2016[44]	Transmission in Brighton, UK	2004-2011	231
Grad et al., 2016[14] and 2014[16]	AMR surveillance across CDC GISP clinics, USA	2000-2013	1100
Demczuk et al., 2016[45]	Azithromycin resistance surveillance in Canada	1991-2014	199
Demczuk et al., 2015[46]	Cephalosporin decreased susceptibility surveillance in Canada	1989-2013	114
Ezewudo et al., 2015[47]	Population structure and AMR surveillance	1982-2011	54

[a]In total, the datasets comprise 4852 gonococcal strains with genome sequencing data, antibiotic MICs, and patient site of infection and metadata.

References

1 Rubin, D. H. F., Ross, J. D. C. & Grad, Y. H. The frontiers of addressing antibiotic resistance in Neisseria gonorrhoeae. Transl Res 220, 122–137 (2020).

2 Centers for Disease, C. & Prevention. 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 (2012).

3 Unemo, M. *et al.* 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, doi:10.1128/AAC.05760-11 (2012).

4 Zarantonelli, L., Borthagaray, G., Lee, E. H. & Shafer, W. M. Decreased azithromycin susceptibility of Neisseria gonorrhoeae due to mtrR mutations. *Antimicrob Agents Chemother* **43**, 2468-2472 (1999).

5 Rouquette-Loughlin, C. E. *et al.* Mechanistic Basis for Decreased Antimicrobial Susceptibility in a Clinical Isolate of Neisseria gonorrhoeae Possessing a Mosaic-Like mtr Efflux Pump Locus. *MBio* **9**, doi:10.1128/mBio.02281-18 (2018).

6 Wadsworth, C. B., Arnold, B. J., Sater, M. R. A. & Grad, Y. H. Azithromycin Resistance through Interspecific Acquisition of an Epistasis-Dependent Efflux Pump Component and Transcriptional Regulator in Neisseria gonorrhoeae. *MBio* **9**, doi:10.1128/mBio.01419-18 (2018).

7 Zhang, J. & van der Veen, S. Neisseria gonorrhoeae 23S rRNA A2059G mutation is the only determinant necessary for high-level azithromycin resistance and improves in vivo biological fitness. *J Antimicrob Chemother* **74**, 407-415, doi:10.1093/jac/dky438 (2019).

8 Jacobsson, S. *et al.* WGS analysis and molecular resistance mechanisms of azithromycin-resistant (MIC >2 mg/L) Neisseria gonorrhoeae isolates in Europe from 2009 to 2014. *J Antimicrob Chemother* **71**, 3109-3116, doi:10.1093/jac/dkw279 (2016).

9 Shafer, W. M. & Folster, J. P. Towards an understanding of chromosomally mediated penicillin resistance in Neisseria gonorrhoeae: evidence for a porin-efflux pump collaboration. *J Bacteriol* **188**, 2297-2299, doi:10.1128/JB.188.7.2297-2299.2006 (2006).

10 Ohnishi, M. *et al.* Is Neisseria gonorrhoeae initiating a future era of untreatable gonorrhea?: detailed characterization of the first strain with high-level resistance to ceftriaxone. *Antimicrob Agents Chemother* **55**, 3538-3545, doi:10.1128/AAC.00325-11 (2011).

11 Ameyama, S. *et al.* Mosaic-like structure of penicillin-binding protein 2 Gene (penA) in clinical isolates of Neisseria gonorrhoeae with reduced susceptibility to cefixime. *Antimicrob Agents Chemother* **46**, 3744-3749 (2002).

12 Nakayama, S. *et al.* New Ceftriaxone- and Multidrug-Resistant Neisseria gonorrhoeae Strain with a Novel Mosaic penA Gene Isolated in Japan. *Antimicrob Agents Chemother* **60**, 4339-4341, doi:10.1128/AAC.00504-16 (2016).

13 Whiley, D. M., Limnios, E. A., Ray, S., Sloots, T. P. & Tapsall, J. W. Diversity of penA alterations and subtypes in Neisseria gonorrhoeae strains from Sydney, Australia, that are less susceptible to ceftriaxone. *Antimicrob Agents Chemother* **51**, 3111-3116, doi:10.1128/AAC.00306-07 (2007).

14 Grad, Y. H. *et al.* Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United States, 2000-2013. *J Infect Dis* **214**, 1579-1587, doi:10.1093/infdis/jiw420 (2016).

15 Harris, S. R. *et al.* Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a genomic survey. *Lancet Infect Dis* **18**, 758-768, doi:10.1016/S1473-3099(18)30225-1 (2018).

16 Grad, Y. H. *et al.* Genomic epidemiology of Neisseria gonorrhoeae with reduced susceptibility to cefixime in the USA: a retrospective observational study. *Lancet Infect Dis* **14**, 220-226, doi:10.1016/S1473-3099(13)70693-5 (2014).

17 Paterson, D. L. Resistance in gram-negative bacteria: enterobacteriaceae. *Am J Med* **119**, S20-28; discussion S62-70, doi:10.1016/j.amjmed.2006.03.013 (2006).

18 Eyre, D. W. *et al.* Gonorrhoea treatment failure caused by a Neisseria gonorrhoeae strain with combined ceftriaxone and high-level azithromycin resistance, England, February 2018. *Euro Surveill* **23**, doi:10.2807/1560-7917.ES.2018.23.27.1800323 (2018).

19 Golparian, D. *et al.* Multidrug-resistant Neisseria gonorrhoeae isolate, belonging to the internationally spreading Japanese FC428 clone, with ceftriaxone resistance and intermediate resistance to azithromycin, Ireland, August 2018. *Euro Surveill* **23**, doi:10.2807/1560-7917.ES.2018.23.47.1800617 (2018).

20 Katz, A. R. *et al.* Cluster of Neisseria gonorrhoeae Isolates With High-level Azithromycin Resistance and Decreased Ceftriaxone Susceptibility, Hawaii, 2016. *Clin Infect Dis* **65**, 918-923, doi:10.1093/cid/cix485 (2017).

21 Unemo, M. Current and future antimicrobial treatment of gonorrhoea - the rapidly evolving Neisseria gonorrhoeae continues to challenge. *BMC Infect Dis* **15**, 364, doi:10.1186/s12879-015-1029-2 (2015).

22 Kirkcaldy, R. D. *et al.* Neisseria gonorrhoeae Antimicrobial Susceptibility Surveillance - The Gonococcal Isolate Surveillance Project, 27 Sites, United States, 2014. *MMWR Surveill Summ* **65**, 1-19, doi:10.15585/mmwr.ss6507a1 (2016).

23 Fingerhuth, S. M., Low, N., Bonhoeffer, S. & Althaus, C. L. Detection of antibiotic resistance is essential for gonorrhoea point-of-care testing: a mathematical modelling study. *BMC Med* **15**, 142, doi:10.1186/s12916-017-0881-x (2017).

24 Turner, K. M. *et al.* Analysis of the potential for point-of-care test to enable individualised treatment of infections caused by antimicrobial-resistant and susceptible strains of Neisseria gonorrhoeae: a modelling study. *BMJ Open* **7**, e015447, doi:10.1136/bmjopen-2016-015447 (2017).

25 Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438-442, doi:10.1126/science.aam9321 (2017).

26 Consortium, C. R. *et al.* Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *N Engl J Med* **379**, 1403-1415, doi:10.1056/NEJMoa1800474 (2018).

27 Abrams, A. J. *et al.* A Case of Decreased Susceptibility to Ceftriaxone in Neisseria gonorrhoeae in the Absence of a Mosaic Penicillin-Binding Protein 2 (penA) Allele. *Sex Transm Dis* **44**, 492-494, doi:10.1097/OLQ.0000000000000645 (2017).

28 Hicks, N. D. *et al.* Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance. *Nat Microbiol* **3**, 1032-1042, doi:10.1038/s41564-018-0218-3 (2018).

29 Coll, F. *et al.* Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. *Nat Genet* **50**, 307-316, doi:10.1038/s41588-017-0029-0 (2018).

30 Chen, P. E. & Shapiro, B. J. The advent of genome-wide association studies for bacteria. *Curr Opin Microbiol* **25**, 17-24, doi:10.1016/j.mib.2015.03.002 (2015).

31 Earle, S. G. *et al.* Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat Microbiol* **1**, 16041, doi:10.1038/nmicrobiol.2016.41 (2016).

32 Jaillard, M. *et al.* A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS Genet.* **14**, e1007758, doi:10.1371/journal.pgen.1007758 (2018).

33 Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**, 4310-4312, doi:10.1093/bioinformatics/bty539 (2018).

34 Schubert, B., Maddamsetti, R., Nyman, J., Farhat, M. R. & Marks, D. S. Genome-wide discovery of epistatic loci affecting antibiotic resistance in Neisseria gonorrhoeae using evolutionary couplings. *Nat Microbiol* **4**, 328-338, doi:10.1038/s41564-018-0309-1 (2019).

35 Lees, J. A., Tien Mai, T., Galardini, M., Wheeler, N. E. & Corander, J. Improved inference and prediction of bacterial genotype-phenotype associations using pangenome-spanning regressions. *bioRxiv*, doi:10.1101/852426 (2019).

36 Mortimer, T. D. *et al.* The distribution and spread of susceptible and resistant *Neisseria gonorrhoeae* across demographic groups in a major metropolitan center. *medRxiv*, 2020.2004.2030.20086413, doi:10.1101/2020.04.30.20086413 (2020).

37 Sanchez-Buso, L. *et al.* The impact of antimicrobials on gonococcal evolution. *Nat Microbiol*, doi:10.1038/s41564-019-0501-y (2019).

38 Yahara, K. *et al.* Genomic surveillance of Neisseria gonorrhoeae to investigate the distribution and evolution of antimicrobial-resistance determinants and lineages. *Microb Genom* **4**, doi:10.1099/mgen.0.000205 (2018).

39 Ryan, L. *et al.* Antimicrobial resistance and molecular epidemiology using whole-genome sequencing of Neisseria gonorrhoeae in Ireland, 2014-2016: focus on extended-spectrum cephalosporins and azithromycin. *Eur J Clin Microbiol Infect Dis*, doi:10.1007/s10096-018-3296-5 (2018).

40 Fifer, H. *et al.* Sustained transmission of high-level azithromycin-resistant Neisseria gonorrhoeae in England: an observational study. *Lancet Infect Dis* **18**, 573-581, doi:10.1016/S1473-3099(18)30122-1 (2018).

41 Lee, R. S. *et al.* Genomic epidemiology and antimicrobial resistance of Neisseria gonorrhoeae in New Zealand. *J Antimicrob Chemother* **73**, 353-364, doi:10.1093/jac/dkx405 (2018).

42 Kwong, J. C. *et al.* 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, doi:10.1136/sextrans-2017-053287 (2018).

43 Eyre, D. W. *et al.* WGS to predict antibiotic MICs for Neisseria gonorrhoeae. *J Antimicrob Chemother* **72**, 1937-1947, doi:10.1093/jac/dkx067 (2017).

44 De Silva, D. *et al.* Whole-genome sequencing to determine transmission of Neisseria gonorrhoeae: an observational study. *Lancet Infect Dis* **16**, 1295-1303, doi:10.1016/S1473-3099(16)30157-8 (2016).

45 Demczuk, W. *et al.* Genomic Epidemiology and Molecular Resistance Mechanisms of Azithromycin-Resistant Neisseria gonorrhoeae in Canada from 1997 to 2014. *J Clin Microbiol* **54**, 1304-1313, doi:10.1128/JCM.03195-15 (2016).

46 Demczuk, W. *et al.* Whole-genome phylogenomic heterogeneity of Neisseria gonorrhoeae isolates with decreased cephalosporin susceptibility collected in Canada between 1989 and 2013. *J Clin Microbiol* **53**, 191-200, doi:10.1128/JCM.02589-14 (2015).

47 Ezewudo, M. N. *et al.* Population structure of Neisseria gonorrhoeae based on whole genome data and its relationship with antibiotic resistance. *PeerJ* **3**, e806, doi:10.7717/peerj.806 (2015).

Chapter 2: Adaptation to the cervical environment is associated with increased antibiotic susceptibility in *Neisseria gonorrhoeae*

Authors

Kevin C Ma[1¶], Tatum D Mortimer[1¶], Allison L Hicks[1], Nicole E Wheeler[2], Leonor Sánchez-Busó[2], Daniel Golparian[3], George Taiaroa[4], Daniel HF Rubin[1], Yi Wang[1], Deborah A Williamson[4], Magnus Unemo[3], Simon R Harris[5], Yonatan H Grad[1,6]

[1] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, USA

[2] Centre for Genomic Pathogen Surveillance, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom

[3] WHO Collaborating Centre for Gonorrhoea and other STIs, Swedish Reference Laboratory for STIs, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

[4] Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

[5] Microbiotica Ltd, Biodata Innovation Centre, Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom

[6] Division of Infectious Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, USA

¶ These authors contributed equally to this work.

Author Contributions

KCM, TDM, ALH, NEW, and LSB performed and interpreted genomic analyses. KCM, DHFR, and YW performed experimental analyses. DG and MU provided data and conducted genomic analyses on historical isolates. GT and DAW provided data and interpreted results for the validation dataset. SRH, MU, DAW, and YHG supervised the project. KCM, TDM, and YHG wrote the paper with contributions from all authors.

Publication

This chapter appears as “Adaptation to the cervical environment is associated with increased antibiotic susceptibility in *Neisseria gonorrhoeae*” in *Nature Communications*, Volume 11, Article 4126, 2020.

Abstract

Neisseria gonorrhoeae is an urgent public health threat due to rapidly increasing incidence and antibiotic resistance. In contrast with the trend of increasing resistance, clinical isolates that have reverted to susceptibility regularly appear, prompting questions about which pressures compete with antibiotics to shape gonococcal evolution. Here, we used genome-wide association to identify loss-of-function (LOF) mutations in the efflux pump *mtrCDE* operon as a mechanism of increased antibiotic susceptibility and demonstrate that these mutations are overrepresented in cervical relative to urethral isolates. This enrichment holds true for LOF mutations in another efflux pump, *farAB*, and in urogenitally-adapted versus typical *N. meningitidis*, providing evidence for a model in which expression of these pumps in the female urogenital tract incurs a fitness cost for pathogenic *Neisseria*. Overall, our findings highlight the impact of integrating microbial population genomics with host metadata and demonstrate how host environmental pressures can lead to increased antibiotic susceptibility.

Introduction

Neisseria gonorrhoeae is the causative agent of the sexually transmitted disease gonorrhea. Antibiotics have played a key role in shaping gonococcal evolution[1-3], with *N. gonorrhoeae* gaining resistance to each of the first line antibiotics used to treat it[4-6]. As *N. gonorrhoeae* is an obligate human pathogen, the mucosal niches it infects—most commonly including the urethra, cervix, pharynx, and rectum—must also influence its evolution[7]. The gonococcal phylogeny suggests the interaction of these factors, with an ancestral split between a drug-susceptible lineage circulating primarily in heterosexuals and a drug-resistant lineage circulating primarily in men who have sex with men[3].

Despite the deeply concerning increase in antibiotic resistance reported in gonococcal populations globally[8], some clinical isolates of *N. gonorrhoeae* have become more susceptible to antibiotics[9,10]. This unexpected phenomenon prompts questions about which environmental pressures could be drivers of increased susceptibility and the mechanisms by which suppression or reversion of resistance may occur. To address these questions, we analyzed the genomes of a global collection of clinical isolates together with patient demographic and clinical data to identify mutations associated with increased susceptibility and define the environments in which they appear. We find that loss-of-function mutations in the efflux pump component *mtrC* are significantly associated with both antibiotic susceptibility and cervical infections, demonstrating how antibiotic and mucosal niche selective pressures intersect. In support of a model in which efflux pump expression incurs a cost in this niche, we also observe enrichment of loss-of-function mutations in cervical isolates in another efflux pump in *N. gonorrhoeae* and in urogenitally-adapted *N. meningitidis*. Our findings demonstrate how shifts in environmental pressures experienced by pathogenic *Neisseria* can lead to loss of efflux pump function and suppression of antibiotic resistance.

Results

Unknown genetic loci influence antibiotic susceptibility. We first assessed how well variation in antibiotic resistance phenotype was captured by the presence and absence of known resistance markers. To do so, we assembled and examined a global dataset comprising the genomes and minimum inhibitory concentrations (MICs) of 4852 isolates collected across 65 countries and 38 years (Figure 2.1, Supplementary Table S1.1). We modeled log-transformed MICs using multiple regression on a panel of experimentally characterized resistance markers for the three most clinically relevant antibiotics[5,11,12] (Supplementary Data S1.1). This enabled us to make quantitative predictions of MIC based on known genotypic markers and to assess how well these markers predicted true MIC values. For the macrolide azithromycin, we observed that 434/4505 (9.63%) isolates had predicted MICs that deviated by two dilutions or more from their reported phenotypic values. The majority (59.4%) of these isolates had MICs that were lower than expected, indicative of increased susceptibility unexplained by the genetic determinants in our model. Overall MIC variance explained by known resistance mutations was relatively low (adjusted R^2 = 0.667), in agreement with prior studies that employed whole-genome supervised learning algorithms to predict azithromycin resistance[13]. MIC variance explained by known resistance mutations was also low for ceftriaxone (adjusted R^2 = 0.674) but higher for ciprofloxacin (adjusted R^2 = 0.937), with 2.02% and 2.90% of strains, respectively, exhibiting two dilutions or lower reported MICs compared to predictions, similarly indicating unexplained susceptibility. The predictive modeling results therefore suggested unknown modifiers that promote susceptibility for multiple classes of antibiotics in *N. gonorrhoeae*.

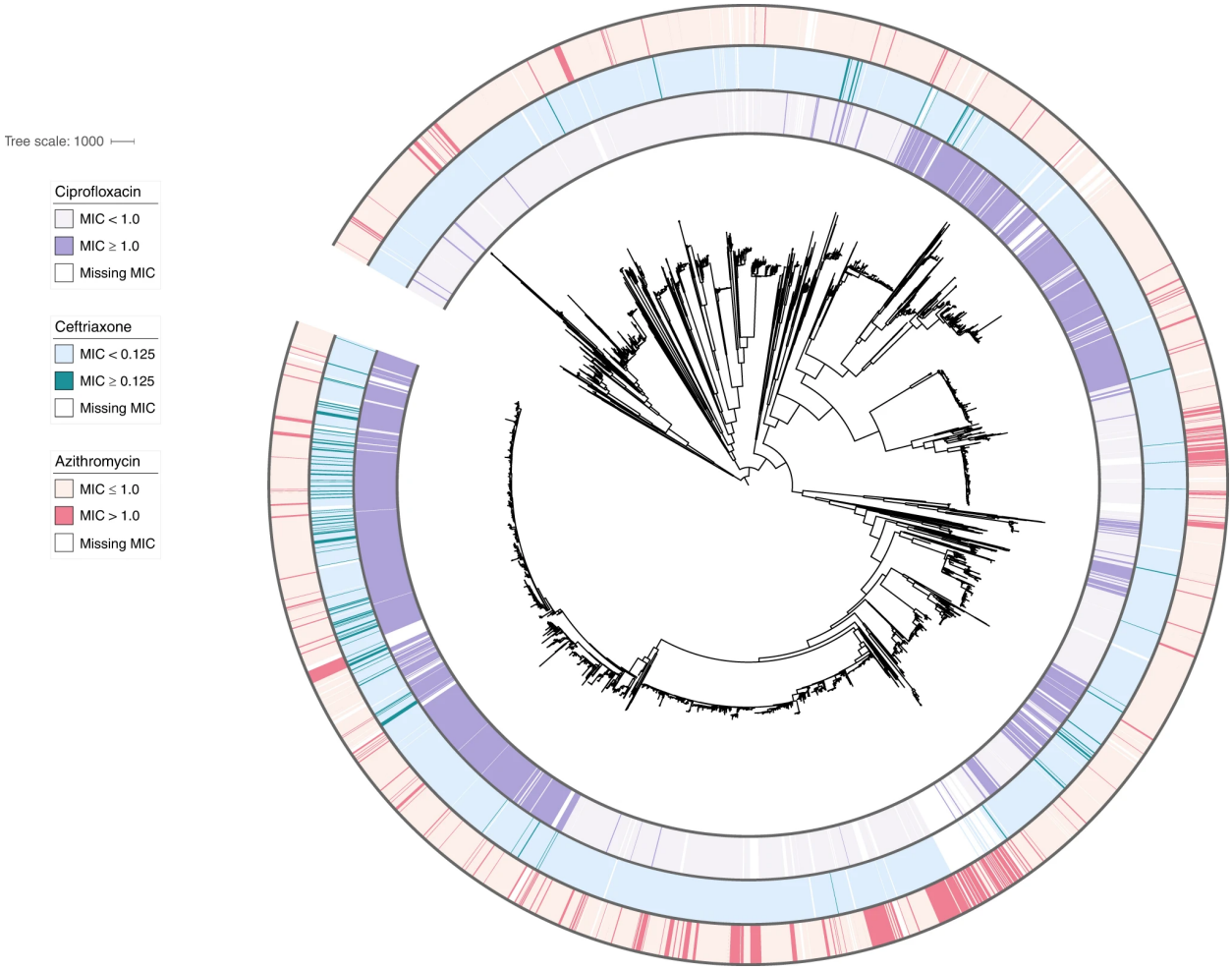


Figure 2.1 – Population structure and susceptibility profile of the *N. gonorrhoeae* global meta-analysis collection. A midpoint rooted recombination-corrected maximum likelihood phylogeny of 4852 genomes based on 68697 SNPs (Supplementary Table S1.1) was annotated with binarized resistance (ciprofloxacin) or decreased susceptibility (azithromycin, ceftriaxone) values. Annotation rings are listed in order of ciprofloxacin, ceftriaxone, and azithromycin from innermost to outermost. For ciprofloxacin, MIC < 1 ug/ml is light purple, and MIC ≥ 1 ug/ml is dark purple. For ceftriaxone, MIC < 0.125 ug/ml is light blue, and MIC ≥ 0.125 ug/ml is dark blue. For azithromycin, MIC ≤ 1 ug/ml is light pink, and MIC > 1 ug/ml is dark pink. Branch length represents total number of substitutions after removal of predicted recombination.

GWAS identifies a susceptibility-associated variant in *mtrC.* To identify novel antibiotic susceptibility loci in an unbiased manner, we conducted a bacterial genome-wide association study (GWAS). We used a linear mixed model framework to control for population structure, and we used unitigs constructed from genome assemblies to capture SNPs, indels, and accessory genome elements (see methods)[14-16]. Unitigs are a flexible representation of the genetic variation across a dataset that are constructed using compacted de Bruijn graphs, and have been previously applied as markers for microbial GWAS[16]. We performed a GWAS on the sequences of 4505 isolates with associated azithromycin MICs using a Bonferroni-corrected significance threshold of 3.38×10^{-7}. The linear mixed model adequately controlled for population structure (Supplementary Figure S1.1), and the proportion of phenotypic MIC variance attributable to genotype (i.e., narrow-sense heritability) estimated by the linear mixed model was high ($h^2 = 0.97$). In line with this, we observed highly significant unitigs with positive effect sizes corresponding to the known resistance substitutions C2611T and A2059G (*E. coli* numbering) in the 23S ribosomal RNA gene (Figure 2.2)[17]. The next most significant variant was a unitig associated with increased susceptibility that mapped to *mtrC* (β, or effect size on the $\log_2$-transformed MIC scale, = -2.82, 95% CI [-3.06, -2.57]; *p*-value = 2.81×10^{-108}). Overexpression of the *mtrCDE* efflux pump operon has been shown to decrease gonococcal susceptibility to a range of hydrophobic acids and antimicrobial agents[4,18], and conversely, knockout of the pump results in multi-drug hypersusceptibility[19]. To assess whether this *mtrC* variant was associated with increased susceptibility to other antibiotics, we performed GWAS for ceftriaxone (for which MICs were available from 4497 isolates) and for ciprofloxacin (4135 isolates). We recovered known ceftriaxone resistance mutations including recombination in the *penA* gene and ciprofloxacin resistance substitutions in DNA gyrase (*gyrA*). In agreement with the known pleiotropic effect of the MtrCDE efflux pump[19], we observed the same *mtrC* unitig at genome-wide significance associated with increased susceptibility to both ceftriaxone (β = -1.18, 95% CI [-1.34, -1.02]; *p*-value = 2.00×10^{-44}) and ciprofloxacin (β = -1.29, 95% CI [-1.54, -1.04]; *p*-value =

1.87×10^{-23}) (Figure 2.2). Across all three drugs, heritability estimates for this *mtrC* variant were comparable to that of prevalent major resistance determinants (azithromycin h^2: 0.323; ceftriaxone h^2: 0.208; ciprofloxacin h^2: 0.155), indicating that unexplained susceptibility in our model could be partially addressed by inclusion of this mutation.

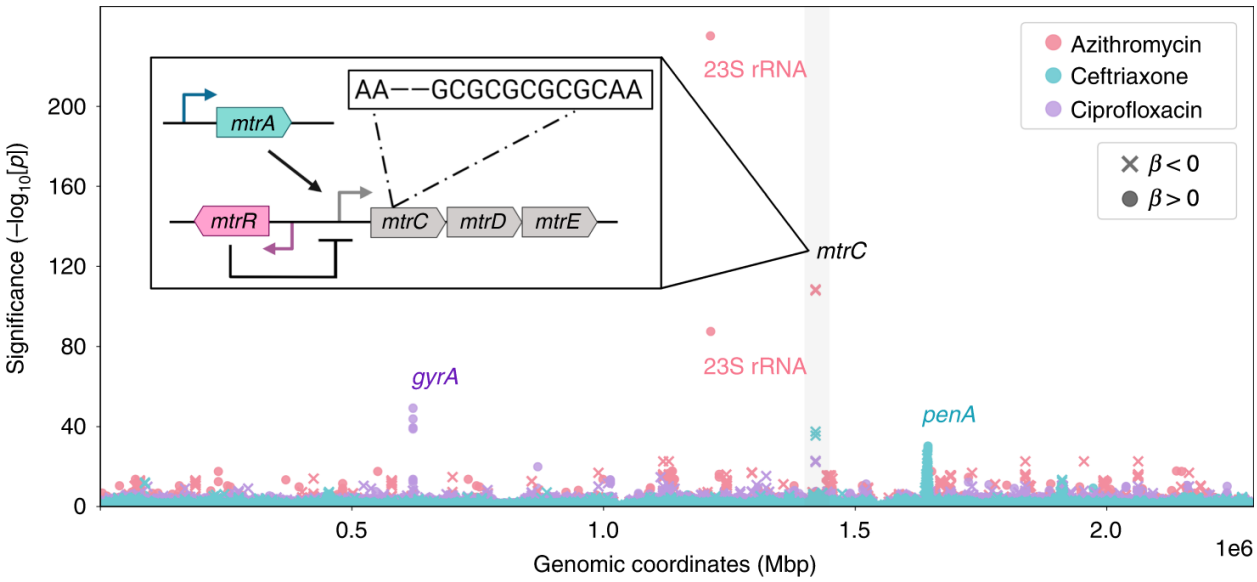


Figure 2.2 – GWAS identifies a variant mapping to *mtrC* associated with increased susceptibility. The Manhattan plot shows negative $\log_{10}$-transformed *p*-values (calculated using likelihood-ratio tests in the GWAS) for the association of unitigs with MICs to azithromycin (pink, n=4505), ceftriaxone (blue, n=4497), and ciprofloxacin (purple, n=4135). The sign of the GWAS regression coefficient β (with positive indicating an association with increased resistance and negative indicating an association with increased susceptibility) is indicated by an X for β < 0 and a dot for β > 0. Labels indicate known influential resistance determinants, and the *mtrC* variant associated with increased susceptibility was highlighted in gray. A full list of the annotated significant unitigs for each antibiotic can be found in Supplementary Data S1.2. Inset: schematic of the *mtr* genetic regulon including structural genes *mtrCDE*, the activator *mtrA*, and the repressor *mtrR*. The approximate genomic location within *mtrC* and specific nucleotide change of the *mtrC* GWAS variant relative to the gonococcal NCCP11945 reference genome (i.e., a two base pair deletion in a 'GC' dinucleotide repeat) is shown.

Annotation of the *mtrC* unitig revealed that it represented a two base pair deletion in a 'GC' dinucleotide hexarepeat, leading to early termination of *mtrC* translation and probable loss of MtrCDE activity[20] (Figure 2.2 inset). We also checked whether the two base pair deletion would

affect recognition by any of the gonococcal methylases[21], but no methylase target motif sites mapped to the hexarepeat or its direct surrounding sequences. A laboratory-generated gonococcal mutant with a four base pair deletion in this same *mtrC* dinucleotide hexarepeat exhibited multi-drug susceptibility[20], and clinical gonococcal isolates hypersensitive to erythromycin were shown to have mutations mapping to this locus[22]. To directly test the hypothesis that the two base pair deletion also contributed to increased susceptibility for the panel of antibiotics we examined, we complemented the mutation in a clinical isolate belonging to the multidrug-resistant lineage ST-1901[23] and observed significant increases in MICs for all three antibiotics, as predicted by the GWAS (Supplementary Table S1.2).

We searched for additional *mtrC* loss-of-function (LOF) mutations and found six clinical isolates with genomes encoding indels outside of the dinucleotide hexarepeat that also were associated with increased susceptibility (Supplementary Figure S1.2a). Ten isolates that had acquired the two base pair deletion also have a two base pair insertion elsewhere in *mtrC* that restores the original coding frame, suggesting that loss of MtrC function may be reverted by further mutation or recombination (Supplementary Figure S1.2a). In line with this, *mtrC* LOF mutations have emerged numerous times throughout the phylogeny (Supplementary Figure S1.3), indicative of possible repeated losses of a dinucleotide in the hexarepeat region due to DNA polymerase slippage, which may occur at a higher rate than single nucleotide nonsense mutations[24]. In total, including all strains with *mtrC* frameshift mutations and excluding revertants, we identified 185 isolates (3.82%) that encoded a LOF allele of *mtrC* (Supplementary Table S1.3). Presence of the *mtrC* LOF mutation in isolates with known resistance markers was correlated with significantly reduced MICs (Supplementary Figure S1.4), and inclusion of *mtrC* LOF mutations in our linear model increased adjusted R^2 values (azithromycin: 0.667 to 0.704; ceftriaxone: 0.674 to 0.690; ciprofloxacin: 0.937 to 0.939), decreased the proportion of strains with unexplained susceptibility (azithromycin: 5.73% to 3.88%; ceftriaxone: 2.02% to 1.73%;

ciprofloxacin: 2.90% to 2.42%), and significantly improved model fit (p-value < 2.2×10^{-16} for all three antibiotics; Likelihood-ratio χ^2 test for nested models). *mtrC* LOF strains were identified in 28 of the 66 countries surveyed and ranged in isolation date from 2002 to 2017. Because most strains in this dataset were collected within the last two decades, we also examined a dataset of strains collected in Denmark from 1928 to 2013 to understand the historical prevalence of *mtrC* LOF mutations[25]. We observed an additional 10 strains with the 'GC' two base pair deletion ranging in isolation date from 1951-2000, indicating that *mtrC* LOF strains have either repeatedly arisen or persistently circulated for decades. Our results demonstrate that a relatively common mechanism of gonococcal acquired antibiotic susceptibility is a two base pair deletion in *mtrC* and that such mutations are globally and temporally widespread.

Loss of MtrCDE pump is associated with cervical infection. The MtrCDE pump has been demonstrated to play a critical role in gonococcal survival in the presence of human neutrophils and in the female murine genital tract model of gonococcal infection, and overexpression of *mtrCDE* results in substantial fitness benefits for dealing with both antimicrobial and environmental pressures[26-29]. The relative frequency of the *mtrC* LOF mutations we observe (occurring in approximately 1 in every 25 isolates) thus seems unusual for a mutation predicted to be deleterious for human infection. *mtrC* LOF strains do not grow more or less quickly *in vitro* than *mtrC* wild-type strains, indicating that this mutation does not confer a simple fitness benefit due to reduced energetic cost[22,26,30]. Instead, we hypothesized that there are unique environments that select for non-functional efflux pump.

We aggregated patient-level metadata across included studies on sex partner preferences and anatomical site of infection. Sexual behavior and *mtrC* genotypic information was available for 1975 isolates from individual patients. There was a significant association between *mtrC* LOF and sexual behavior (p-value = 0.04021; two-sided Fisher's exact test) (Figure 2.3a,

Supplementary Table S1.4), and *mtrC* LOF occurred more often in isolates from men who have sex with women (MSW) (28/626, 4.47%) compared to isolates from men who have sex with men (MSM) (31/1189, 2.61%) (OR = 1.75, 95% CI [1.00-3.04], *p*-value = 0.037; two-sided Fisher's exact test). To understand whether anatomical selective pressures contributed to this enrichment, we analyzed the site of infection and *mtrC* genotypic information available for 2730 isolates. *mtrC* LOF mutations were significantly associated with site of infection (*p*-value = 6.49×10^{-5}; two-sided Fisher's exact test) and were overrepresented particularly in cervical isolates: 16 out of 129 (12.4%) cervical isolates contained an *mtrC* LOF mutation compared to 82 out of 2249 urethral isolates (3.65%; OR = 3.74, 95% CI [1.98-6.70], *p*-value = 4.71×10^{-5}; two-sided Fisher's exact test), 3 out of 106 pharyngeal isolates (2.83%; OR = 4.83, 95% CI [1.33-26.63], *p*-value = 0.00769; two-sided Fisher's exact test), and 4 out of 246 rectal isolates (1.63%; OR = 8.52, 95% CI [2.67-35.787], *p*-value = 2.39×10^{-5}; two-sided Fisher's exact test) (Figure 2.3b, Supplementary Table S1.5). Because our meta-analysis collection comprises datasets potentially biased by preferential sampling for drug-resistant strains, we validated our epidemiological associations on a set of 2186 sequenced isolates, corresponding to all cultured isolates of *N. gonorrhoeae* in the state of Victoria, Australia in 2017[31]. We again observed significant associations between *mtrC* LOF and sexual behavior (*p*-value = 0.0180; two-sided Fisher's exact test) as well as anatomical site of infection (*p*-value = 0.0256; two-sided Fisher's exact test) (Supplementary Figure S1.5, Supplementary Tables S1.6-7). *mtrC* LOF mutations were again overrepresented in cervical isolates: 9 out of 227 (3.96%) cervical isolates contained an *mtrC* LOF mutation compared to 15 out of 882 urethral isolates (1.70%; OR = 2.38, 95% CI [0.91-5.91], *p*-value = 0.0679; two-sided Fisher's exact test), 3 out of 386 pharyngeal isolates (0.78%; OR = 5.26, 95% CI [1.29-30.51], *p*-value = 0.0117; two-sided Fisher's exact test), and 7 out of 632 rectal isolates (1.11%; OR = 3.68, 95% CI [1.20-11.78], *p*-value = 0.0173; two-sided Fisher's exact test). These results indicate that environmental pressures unique to female

urogenital infection may select for loss of the primary gonococcal efflux pump resulting in broadly increased susceptibility to antibiotics and host-derived antimicrobial peptides.

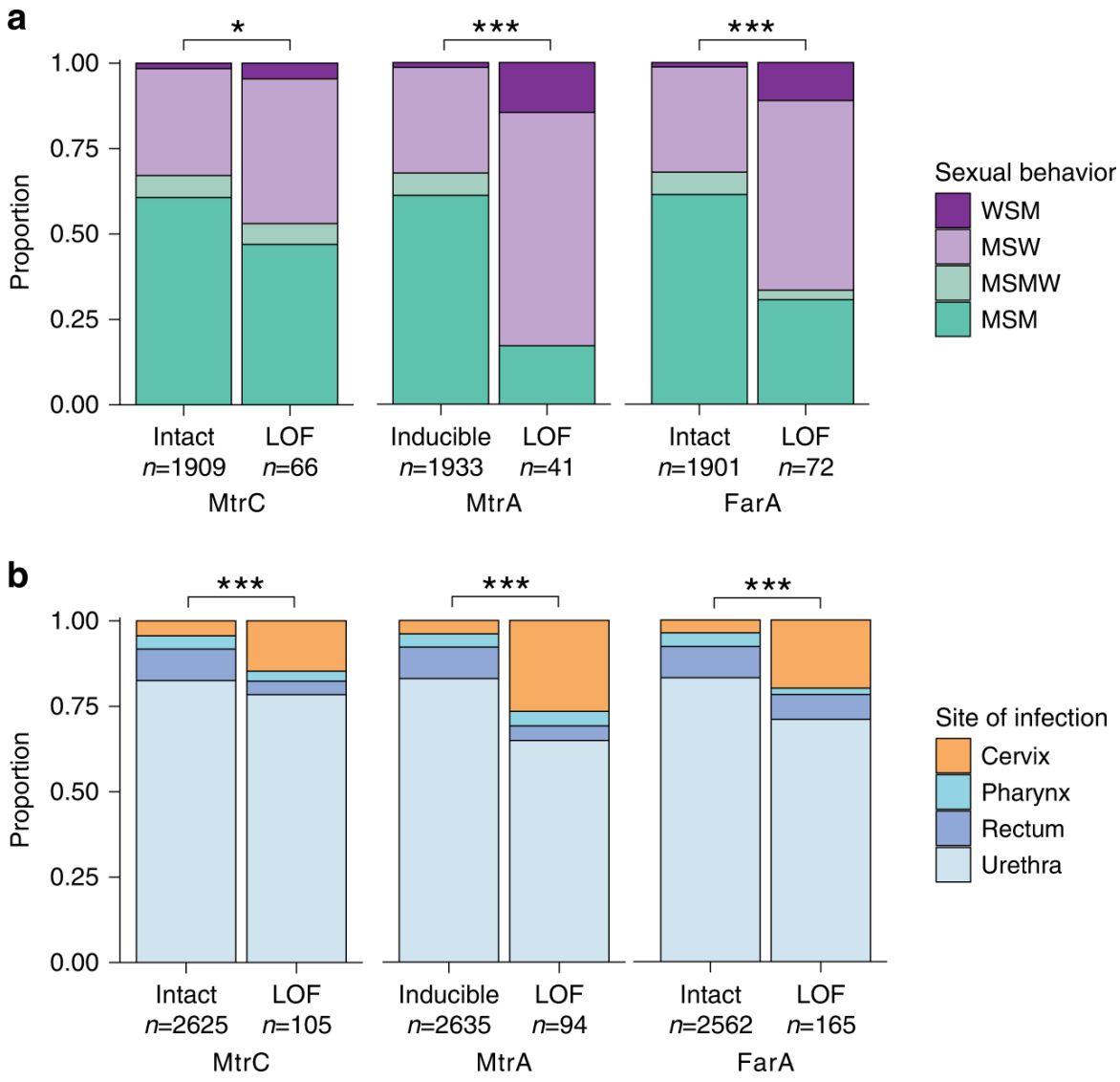


Figure 2.3 – Gonococcal *mtrC*, *mtrA*, and *farA* LOF mutations are associated with cervical infection. a) Sexual behavior of patients infected with isolates with either intact or LOF alleles of *mtrC* (left), *mtrA* (middle), or *farA* (right). b) Site of infection in patients infected with isolates with either intact or LOF alleles of *mtrC* (left), *mtrA* (middle), or *farA* (right). *mtrA* alleles were predicted as LOF only in the absence of other epistatic Mtr overexpression mutations. Statistical significance between genotype and patient metadata was assessed by two-sided Fisher's exact test: * $p<0.05$, ** $p<0.01$, and *** $p<0.001$. Exact *p*-values from left to right for analyses in a) were 0.04021, 1.81×10^{-11}, 5.06×10^{-10} and for b) were 6.49×10^{-5}, 1.64×10^{-12}, 1.78×10^{-12}. WSM = women who have sex with men, MSW = men who have sex with women, MSMW = men who have sex with men and women, MSM = men who have sex with men.

***mtrA* LOF offers an additional level of adaptive regulation.** The association of *mtrC* LOF mutations with cervical specimens suggests that other mutations that downregulate expression of the *mtrCDE* operon should also promote adaptation to the cervical niche. The MtrCDE efflux pump regulon comprises the MtrR repressor and the MtrA activator (Figure 2.2 inset), the latter of which exists in two allelic forms: a wild-type functional gene capable of inducing *mtrCDE* expression and a variant rendered non-functional by an 11-bp deletion near the 5' end of the gene[32] (Supplementary Figure S1.2b). Knocking out *mtrA* has a detrimental effect on fitness in the gonococcal mouse model, and epistatic *mtrR* mutations resulting in overexpression of *mtrCDE* compensate for this fitness defect by masking the effect of the *mtrA* knockout[27]. Prior work assessing the genomic diversity of *mtrA* in a set of 922 primarily male urethral specimens found only four isolates with the 11-bp deletion (0.43%)[33], seemingly in agreement with the *in vivo* importance of *mtrA*. However, in our global meta-analysis dataset, 362/4842 isolates (7.48%) were predicted to be *mtrA* LOF, of which the majority (357/362, 98.6%) were due to the 11-bp deletion. Of the 4842 isolates, 268 (5.53%) had *mtrA* LOF mutations in non-*mtrCDE* overexpression backgrounds (as defined by the absence of known *mtrR* promoter or coding sequence mutations or *mtrCDE* mosaic alleles) and therefore not epistatically masked (Supplementary Table S1.8). We repeated our epidemiological associations on these *mtrA* LOF strains without concurrent overexpression mutations and observed highly significant associations with reported patient sexual behavior (p-value = 1.81×10^{-11}; two-sided Fisher's exact test) and site of infection (p-value = 1.64×10^{-12}; two-sided Fisher's exact test) (Figure 2.3, Supplementary Tables S1.9-10). As with *mtrC* LOF mutations, *mtrA* LOF mutations were significantly overrepresented in cervical isolates: 25 out of 129 (19.4%) cervical isolates contained an *mtrA* LOF mutation compared to 61 out of 2248 urethral isolates (2.71%; OR = 8.60, 95% CI [4.96-14.57], p-value = 4.60×10^{-13}; two-sided Fisher's exact test), 4 out of 106 pharyngeal isolates (3.78%; OR = 6.09, 95% CI [2.00-24.93], p-value = 0.000240; two-sided

Fisher's exact test), and 4 out of 246 rectal isolates (1.63%; OR = 14.43, 95% CI [4.81-58.52], *p*-value = 3.00×10^{-9}; two-sided Fisher's exact test). In the Australian validation cohort[31], the majority (81/85, 95.3%) of *mtrA* LOF strains had concurrent *mtrCDE* overexpression mutations, so it was not possible to test for these associations (Supplementary Table S1.11). In such genetic backgrounds where overexpression mutations mask the effect of *mtrA* LOF, *mtrC* LOF is the preferred method of efflux pump downregulation: the majority of *mtrC* LOF mutations in both the global dataset (174/180, 96.7%) and the Australian cohort (33/35, 94.3%) occurred in backgrounds with known *mtr* overexpression mutations (Supplementary Tables S1.12-13). Phylogenetic analysis showed that the distribution of *mtrA* LOF differed from that of *mtrC* LOF with fewer introductions but more sustained transmission and that the two mutations were largely non-overlapping (Supplementary Figure S1.3); only four strains had both *mtrA* and *mtrC* LOF mutations. Our results indicate that multiple adaptive paths for MtrCDE efflux pump downregulation exist depending on genetic interactions with other concurrent mutations in the *mtrCDE* regulon.

FarAB efflux pump LOF is associated with cervical infection. The associations we observed in the *mtrCDE* regulon raised the question of the mechanism by which the cervical environment could select for pump downregulation. Recent work on *Pseudomonas* suggested one possible model: overexpression of homologous *P. aeruginosa* efflux pumps belonging to the same resistance/nodulation/cell division (RND) proton/substrate antiporter family as MtrCDE results in a fitness cost due to increased cytoplasmic acidification[34]. This fitness cost was only observed in anaerobic conditions, where aerobic respiration cannot be used to dissipate excess protons efficiently[34]. Analogous conditions in the female urogenital tract, potentially augmented by environmental acidity, could create a similar selective pressure during human infection that leads to pump downregulation or loss.

This model predicts that adaptation to these conditions would similarly result in downregulation of FarAB, the other proton-substrate antiporter efflux pump in *N. gonorrhoeae*. FarAB is a member of the major facilitator superfamily (MFS) of efflux pumps and effluxes long-chain fatty acids[35,36]. In our global dataset, 332/4838 (6.86%) of isolates were predicted to have *farA* LOF mutations, of which the majority (316/332; 95.2%) were due to indels in a homopolymeric stretch of eight 'T' nucleotides near the 5' end of the gene (Supplementary Figure S1.2c). *farA* LOF mutations were associated with patient sexual behavior (p-value = 5.06×10^{-10}; two-sided Fisher's exact test) and site of infection (p-value = 1.78×10^{-12}; two-sided Fisher's exact test) and overrepresented in cervical isolates: 33 out of 129 (25.6%) cervical isolates contained a *farA* LOF mutation compared to 117 out of 2246 urethral isolates (5.21%; OR = 6.25, 95% CI [3.90-9.83], p-value = 3.24×10^{-13}; two-sided Fisher's exact test), 3 out of 106 pharyngeal isolates (2.83%; OR = 11.70, 95% CI [3.50-61.61], p-value = 3.80×10^{-7}; two-sided Fisher's exact test), and 12 out of 246 rectal isolates (4.88%; OR = 6.66, 95% CI [3.19-14.80], p-value = 1.57×10^{-8}; two-sided Fisher's exact test) (Figure 2.3, Supplementary Table S1.14-15). *farA* LOF mutations were prevalent also in our Australian validation dataset[31] (225/2180; 10.32%) and again associated with sexual behavior (p-value < 2.20×10^{-16}; two-sided Fisher's exact test) and site of infection (p-value < 2.20×10^{-16}; two-sided Fisher's exact test) (Supplementary Figure S1.5, Supplementary Table S1.16-17). The phylogenetic distribution of *farA* LOF indicated sustained transmission (Supplementary Figure S1.3) and overlapped with that of *mtrA* LOF mutations (48.9% of isolates with *mtrA* LOF mutations also had *farA* LOF mutations), potentially indicating additive contributions to cervical adaptation. Furthermore, MtrR activates *farAB* expression by repressing the *farR* repressor[37]. This cross-talk between the two efflux pump operons indicates that in *mtrCDE* overexpression strains where MtrR activity is impaired, the effect of *farA* LOF – like *mtrA* LOF – may be masked. We did not observe frequent LOF mutations in the sodium gradient-dependent MATE family efflux pump NorM[38] or in the ATP-dependent ABC family pump MacAB[39] (Supplementary Table S1.3). The prevalence and cervical enrichment of *farA*

LOF mutations and the relative rarity of LOF mutations in other non-proton motive force-driven pumps suggests that cytoplasmic acidification may be a mechanism by which the female urogenital tract selects for efflux pump loss.

Meningococcal *mtrC* LOF is driven by urogenital adaptation. *N. meningitidis,* a species closely related to *N. gonorrhoeae,* colonizes the oropharyngeal tract and can cause invasive disease, including meningitis and septicemia[40]. We characterized *mtrC* diversity in a collection of 14,798 *N. meningitidis* genomes, reasoning that the cervical environmental pressures that select for efflux pump LOF in the gonococcus will be rarely encountered by the meningococcus. In agreement with this, the 'GC' hexarepeat associated with most gonococcal *mtrC* LOF mutations was less conserved in *N. meningitidis*; only 9684/14798 (65.4%) isolates contained an intact hexarepeat compared to 4644/4847 (95.8%) of *N. gonorrhoeae* isolates (*p*-value < 2.2×10^{-16}; two-sided Fisher's exact test). In this same collection, we observed *mtrC* LOF due to deletions in the hexarepeat region in only 82 meningococcal isolates (0.55%), with a similar frequency of 25/4059 (0.62%) in a curated dataset comprising all invasive meningococcal disease isolates collected in the UK from 2009-2013[41]. The observed interruption of 'GC' dinucleotide repeats, predicted to result in a lower mutation rate[42], and the relative rarity of *mtrC* LOF mutations suggests that efflux pump loss is not generally adaptive in *N. meningitidis.* However, a urogenitally-adapted meningococcal lineage has recently emerged in the US associated with outbreaks of non-gonococcal urethritis in heterosexual patients[43-45]. In isolates from this lineage, the prevalence of *mtrC* LOF mutations was 18/207 (8.70%), substantially higher than typical *N. meningitidis* and comparable to the prevalence of gonococcal *mtrC* LOF mutations in MSW in our global dataset (4.47%). We compared the frequency of *mtrC* LOF mutations in the urogenital lineage to geographically and genetically matched isolates (i.e., all publicly available n=456 PubMLST ST-11 North American isolates) and observed a significant difference in prevalence (18 out of 207 or 8.70% versus 2 out of 249 or 0.80%; OR = 11.71,

95% CI [2.75-105.37], *p*-value = 3.31×10^{-5}; two-sided Fisher's exact test). Most *mtrC* LOF mutations occurred due to the same hexarepeat two base pair deletion that we previously observed for *N. gonorrhoeae*, and in line with this, *mtrC* LOF in this urogenital lineage arose multiple times independently similarly to gonococcal *mtrC* LOF mutations (Figure 2.4, Supplementary Figure S1.3). *farA* LOF mutations were not observed in this meningococcal lineage. We conclude that MtrCDE efflux pump LOF is rare in typical meningococcal strains that inhabit the oropharynx but elevated in frequency in a unique urogenitally-adapted lineage circulating in heterosexuals, indicative of potential ongoing adaptation to the cervical niche. Our results suggest that efflux pump loss is broadly adaptive for cervical colonization across pathogenic *Neisseria*.

Discussion

In an era in which widespread antimicrobial pressure has led to the emergence of extensively drug-resistant *N. gonorrhoeae*[46], isolates that appear to have reverted to susceptibility still arise[9,10], demonstrating that antibiotic and host environmental pressures interact to shape the evolution of *N. gonorrhoeae*. Here, we showed that frameshift-mediated truncations in the *mtrC* component of the MtrCDE efflux pump are the primary mechanism for epistatic increases in antibiotic susceptibility across a global collection of clinical gonococcal isolates, as suggested by prior work[20,22]. *mtrC* LOF mutations are enriched in cervical isolates and a frameshifted form of the pump activator MtrA exhibits similar trends, supporting a model in which reduced or eliminated *mtrCDE* efflux pump expression contributes to adaptation to the female genital tract. We hypothesized that the mechanism by which this occurs is through increased cytoplasmic acidification in anaerobic conditions[34] and demonstrated that LOF mutations in *farA*, encoding a subunit of the other proton motive force-driven pump FarAB, were likewise enriched in cervical isolates.

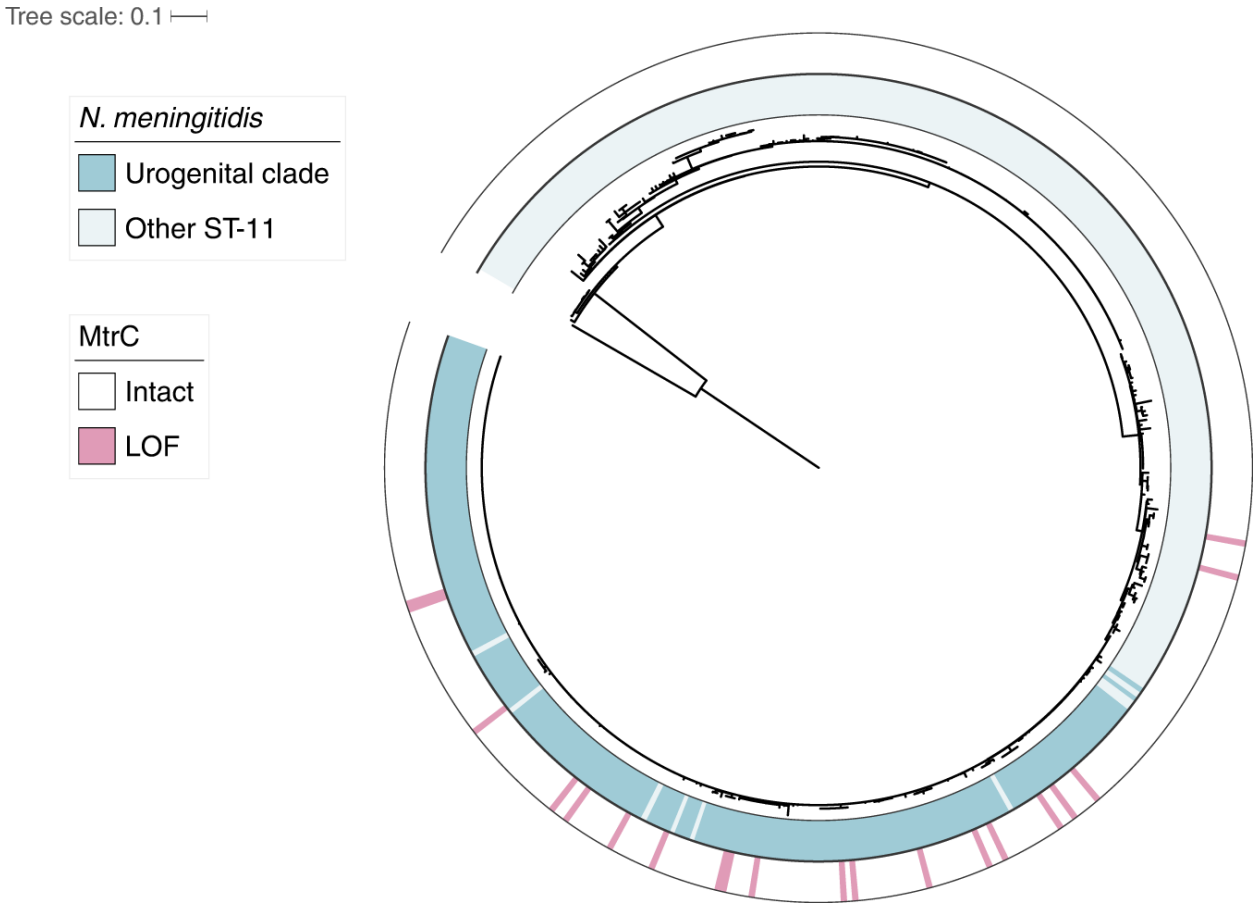


Figure 2.4 – *mtrC* LOF mutations are enriched in a lineage of ST-11 urogenitally-adapted *N. meningitidis*. A core-genome maximum likelihood phylogeny based on 25045 SNPs was estimated of all North American ST-11 *N. meningitidis* strains from PubMLST (n=456; accessed 2019-09-03) rooted with meningococcal reference genome MC58. Membership in ST-11 urogenital clade (dark blue) defined as in Retchless et al., 2018[44]. Genomes with *mtrC* LOF mutations are indicated in pink. Branch length represents substitutions per site.

The LOF mutations we observed in *mtrC* and *farA* primarily occurred in short homopolymeric sequences (though with low numbers of repeated units) and thus may occur at higher rates than insertions or deletions in non-repetitive regions or nonsense mutations, similar to other resistance suppressor mutations[47], though this will need to be confirmed in future experiments. In total, 42.6% of cervical isolates in the global dataset and 32.6% in the validation dataset contained a LOF mutation in either *mtrC, farA,* or *mtrA*, indicating that efflux pump downregulation via multiple genetic mechanisms is prevalent in cervical infection. These results complement prior studies suggesting that *mtrR* LOF resulting in increased resistance to fecal lipids plays a critical role in gonococcal adaptation to the rectal environment[48,49], and taken together suggest a model in which the fitness benefit of efflux pump expression is highly context dependent.

Other selective forces could also have contributed to the observed enrichment of LOF mutations in cervical isolates. For instance, iron levels modulate *mtrCDE* expression through Fur (the ferric uptake regulator) and MpeR[50]. Iron limitation results in increased expression of *mtrCDE*, and conversely, iron enrichment result in decreased expression, suggesting a fitness cost for *mtrCDE* expression during high iron conditions. Variation in environmental iron levels, such as in the menstrual cycle, may provide another selective pressure for LOF mutations to arise particularly when MtrR function is impaired through active site or promoter mutations. Differing rates of antibiotic use for gonorrhea in men and women due to increased asymptomatic infection in women might also select for *mtrC* LOF mutations, but this would not explain the associations we observed for the non-antibiotic substrate efflux pump *farAB* or the increased frequency of *mtrC* LOF mutations in urogenitally-adapted meningococci. RNA sequencing from men and women infected with gonorrhea demonstrated a 4-fold lower

expression of *mtrCDE* in women, re-affirming the idea that efflux pump expression in the female genital tract incurs a fitness cost[51].

Despite significant associations, only a proportion of cervical isolates exhibited these LOF genotypes, suggesting variation in cervix-associated pressures or indicating that cervical culture specimens were obtained before niche pressures could select for pump downregulation. This variation could also lead to mixed populations of efflux pump wild-type WT and LOF strains; however, because only one clonal isolate per site per patient is typically sequenced in clinical surveillance studies, we would be unable to detect this intra-host patient diversity. Targeted amplicon sequencing of LOF loci directly from patients in future studies would help to assess whether this intra-host diversity plays a role in infection and transmission. In particular, this intra-host pathogen diversity could facilitate transmission from the female genital tract to other sites of infection, where efflux pump activity incurs less of a fitness cost. In those new sites, isolates with wild-type efflux pump loci in the mixed population could selectively expand relative to LOF efflux pump strains and also serve as possible recombination donors of wild-type alleles. This standing genetic variation would therefore facilitate gonococcal adaptation across different mucosal niches. Additionally, while the cervix is the primary site of infection and source for culture in women, the selective pressures at play may include other sites more broadly in the female genital tract and may be influenced by the presence of other microbial species both pathogenic and commensal.

Our model extended to the other pathogenic *Neisseria* species, *N. meningitidis*, in that a urogenital clade transmitting in primarily heterosexual populations appeared to be undergoing further urogenital adaptation via the same *mtrC* frameshift mutation that was most commonly observed for *N. gonorrhoeae*. In the absence of data on cases of cervicitis, we hypothesized

that for this meningococcal lineage, efflux pump LOF emerged in the female urogenital tract and was transmitted to heterosexual men resulting in the enrichment we observed. Efflux pumps are common across Gram-negative bacteria[52], and their loss may be a general adaptive strategy for species that face similar pressures as *N. gonorrhoeae* and urogenitally-adapted *N. meningitidis*. In support of this, clinical isolates of *Pseudomonas aeruginosa* with truncations in genes homologous to *mtrC*[53,54] and exhibiting antibiotic hypersensitivity have been obtained from cystic fibrosis patients, in whom the thick mucus in airway environments can likewise exhibit increased acidity and decreased oxygen availability[55,56].

Our results also suggest potential therapeutic avenues for addressing the emergence of multidrug-resistant gonococcal strains. Selective knockdown of MtrCDE homologs in other bacteria via antisense RNA[57] and bacteriophages[58] has successfully re-sensitized resistant strains and enhanced antibiotic efficacy, and ectopic expression in *N. gonorrhoeae* of the *mtrR* repressor in a cephalosporin-resistant strain enhances gonococcal killing by β-lactam antibiotics in the mouse model[59]. Our population-wide estimated effect sizes for *mtrC* LOF mutations provide a prediction for the re-sensitization effect of MtrCDE knockdown across multiple genetic backgrounds and suggest particularly strong effects for the macrolide azithromycin (Supplementary Figure S1.4). Because the correlation between MIC differences and clinical efficacy is still not well understood[60,61], follow up studies to assess treatment efficacy differences in patients with and without *mtrC* LOF strains can help to quantify the expected effect of MtrCDE knockdown in the clinical context.

In summary, by analysis of population genomics and patient clinical data, we have shown that pathogenic *Neisseria* can use multiple avenues of efflux pump perturbation as an adaptive

strategy to respond to host environmental pressures and illustrate how these host pressures may result in increased antibiotic susceptibility in *N. gonorrhoeae*.

Methods

Genomics pipeline. Reads for isolates with either associated azithromycin, ciprofloxacin, or ceftriaxone MIC metadata were downloaded from datasets listed in Supplementary Table S1.1. Reads were inspected using FastQC version 0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and removed if GC content diverged notably from expected values (approximately 52-54%) or if base quality was systematically poor. We mapped read data to the NCCP11945 reference genome (RefSeq accession: NC_011035.1) using BWA-MEM (version 0.7.17-r1188)[62,63] and deduplicated reads using Picard (version 2.8.0) (https://github.com/broadinstitute/picard). BamQC in Qualimap (version 2.2.1)[64] was run to identify samples with less than 70% of reads aligned or samples with less than 40X coverage, which were discarded. We used Pilon (version 1.16)[65] to call variants with mindepth set to 10 and minmq set to 20 and generated pseudogenomes from Pilon VCFs by including all PASS sites and alternate alleles with AF > 0.9; all other sites were assigned as 'N'. Samples with greater than 15% of sites across the genome missing were also excluded. We created *de novo a*ssemblies using SPAdes (version 3.12.0 run using 8 threads, paired end reads where available, and the --careful flag set)[66] and quality filtered contigs to ensure coverage greater than 10X, length greater than 500 base pairs, and total genome size approximately equal to the FA1090 genome size (2.0 to 2.3 Mbp). We annotated assemblies with Prokka (version 1.13)[67], and clustered core genes using Roary (version 3.12)[68] (flags -z -e -n -v -s -i 92) and core intergenic regions using piggy (version 1.2)[69]. A recombination-corrected phylogeny of all isolates was constructed by running Gubbins (version 2.3.4) on the

aligned pseudogenomes and visualized in iTOL (version 4.4.2)[70-72]. All isolates with associated metadata and accession numbers are listed in Supplementary Data 3 and 4.

Resistance allele calling. Known resistance determinants in single-copy genes were called by identifying expected SNPs in the pseudogenomes. For categorizing mosaic alleles of *mtr*, we ran BLASTn (version 2.6.0)[73] on the *de novo* assemblies using a query sequence from FA1090 (Genbank accession: NC_002946.2) comprising the *mtr* intergenic promoter region and *mtrCDE*. BLAST results were aligned using MAFFT (version 7.450)[74] and clustered into distinct allelic families using FastBAPS (version 1.0.0)[75]. We confirmed that horizontally-transferred *mtr* alleles associated with resistance from prior studies[5] corresponded to distinct clusters in FastBAPS. A similar approach was used to cluster *penA* alleles after running BLASTn with a *penA* reference sequence from FA1090. Variant calling in the multi-copy 23S rRNA locus was done by mapping to a modified NCCP11945 reference genome containing only one copy of the 23S rRNA and analyzing variant allele frequencies[76]. We identified truncated MtrR proteins using Prokka annotations, and mutations in the *mtr* promoter region associated with upregulation of *mtrCDE* (A deletion and TT insertion in inverted repeat, *mtr* 120) using an alignment of the *mtr* promoter from piggy output.

Phenotype processing and linear models. We doubled GISP azithromycin MICs before 2005 to account for the GISP MIC protocol testing change[77]. Samples with binary resistance phenotypes (i.e., “SUS” and “RES”) were discarded. For samples with MICs listed as above or below a threshold (indicated by greater than or less than symbols), the MIC was set to equal the provided threshold. MICs were $\log_2$-transformed for use as continuous outcome variables in linear modeling and GWAS. We modeled transformed MICs using a panel of known resistance markers[11,12] and included the recently characterized mosaic *mtrCDE* alleles[5] and

rplD G70D substitution[78] conferring azithromycin resistance, as well as isolate country of origin. Formulas called by the lm function in R (version 3.5.1) for each drug were (with codon or nucleotide site indicated after each gene or rRNA, respectively):

Azithromycin: Log_AZI ~ Country + MtrR 39 + MtrR 45 + MtrR LOF + *mtrR* promoter + *mtrRCDE* BAPS + RplD G70D + 23S rRNA 2059 + 23S rRNA 2611

Ceftriaxone: Log_CRO ~ Country + MtrR 39 + MtrR 45 + MtrR LOF + *mtrR* promoter + *penA* BAPS + PonA 421 + PenA 501 + PenA 542 + PenA 551 + PorB 120 + PorB 121

Ciprofloxacin: Log_CIP ~ Country + MtrR 39 + MtrR 45 + MtrR LOF + *mtrR* promoter + GyrA 91 + GyrA 95 + ParC 86 + ParC 87 + ParC 91 + PorB 120 + PorB 121

To visualize the continuous MICs using thresholds as on Figure 2.1, we binarized MICs using the CLSI resistance breakpoint for ciprofloxacin, the CLSI non-susceptibility breakpoint for azithromycin, and the CDC GISP surveillance breakpoint for ceftriaxone.

GWAS and unitig annotation. We used a regression-based GWAS approach to identify novel susceptibility mutations. In particular, we employed a linear mixed model with a random effect to control for the confounding influence of population structure and a fixed effect to control for isolate country of origin. Though the outcome variable (log_2-transformed MICs) is the same, in contrast to the linear modeling approach described above, which models the linear, additive effect of multiple, known resistance mutations, regression in a GWAS is usually run independently and univariately on each variant for all identified variants in the genome, providing a systematic way to identify novel contributors to the outcome variable. Linear mixed model GWAS was run using Pyseer (version 1.2.0 with default allele frequency filters) on the 480,902 unitigs generated from GATB (version 1.3.0); the recombination-corrected phylogeny from Gubbins was used to parameterize the Pyseer population structure random effects term

and isolate country of origin was included as a fixed effect covariate. To create the Manhattan plot, we mapped all unitigs from the GWAS using BWA-MEM (modified parameters: -B 2 and -O 3) to the pan-susceptible WHO F strain reference genome (Genbank accession: GCA_900087635.2) edited to contain only one locus of the 23S rRNA. Significant unitigs were annotated using Pyseer's annotation pipeline. Unitigs mapping to multiple sites in the genome and in or near the highly variable *pilE* (encoding pilus subunit) or *piiC* (encoding opacity protein family) genes were excluded, as were unitigs less than twenty base pairs in length. Due to redundancy and linkage, variants will be spanned by multiple overlapping unitigs with similar frequencies and *p*-values. For ease of interpretation, we grouped unitigs within 50 base pairs of each other and represented each cluster by the most significant unitig. Unitigs with allele frequency greater than 50% were also excluded as they represented the majority allele. Unitig clusters were then annotated by gene or adjacent genes for unitigs mapping to intergenic regions and further analyzed for predicted functional effect relative to the WHO F reference genome in Geneious Prime (version 2019.2.1, https://www.geneious.com).

Identifying LOF and upregulation alleles. To identify predicted LOF alleles of efflux pump proteins, we ran BLASTn on the *de novo* assemblies using a query sequence from FA1090 (reference genome FA19 was used for *mtrA*). Sequences that were full-length or approximately full-length (+/- 5 nucleotides) beginning with expected start codons were translated using Python (version 3.6.5) and Biopython (version 1.69)[79]. Peptides shorter than 90% of the expected full-length size of the protein were further analyzed using Geneious Prime (version 2019.2.1, https://www.geneious.com) to identify the nucleotide mutations resulting in predicted LOF by alignment of the nucleotide sequences. We called *mtrCDE* overexpression status by identifying the presence of any of the known *mtrR* promoter mutations, MtrR coding sequence mutations, and mosaic *mtrCDE* alleles.

Experimental validation. *N. gonorrhoeae* culture was conducted on GCB agar (Difco) plates supplemented with 1% Kellogg's supplements[80] at 37°C in a 5% CO2 atmosphere. Antimicrobial susceptibility testing was conducted on GCB agar supplemented with 1% IsoVitaleX (Becton Dickinson) using Etests (bioMérieux) at 37°C in a 5% CO2 atmosphere. We selected a clinical isolate (NY0195[81]) from the multidrug-resistant lineage ST-1901[23] that contained an *mtrC* LOF mutation mediated by a two base pair hexarepeat deletion and confirmed via Etests that its MIC matched, within one dilution, its reported MIC. Isolate NY0195 contained mosaic *penA* allele XXXIV conferring cephalosporin reduced susceptibility and the *gyrA* S91F substitution conferring ciprofloxacin resistance[9]. We complemented the *mtrC* LOF mutation in this strain by transforming it via electroporation[80] with a 2kb PCR product containing a *Neisserial* DNA uptake sequence and an in-frame *mtrC* allele, obtained by colony PCR from a neighboring isolate (GCGS0759). After obtaining transformants by selecting on an azithromycin 0.05 μg/mL GCB plate supplemented with Kellogg's supplement, we confirmed successful transformation by Sanger sequencing of the *mtrC* gene. No spontaneous mutants on azithromycin 0.05 μg/mL plates were observed after conducting control transformations in the absence of GCGS0759 *mtrC* PCR product. We conducted antimicrobial susceptibility testing in triplicate using Etests, assessing statistical significance between parental and transformant MICs by a two-sample t-test.

Metadata analysis. Patient metadata were collected from the following publications from Supplementary Table S1.1 that had information on site of infection: Demczuk et al., 2015, Demczuk et al., 2016, Ezewudo et al., 2015, Grad et al., 2014 and 2016, Kwong et al., 2017, Lee et al., 2018, and Mortimer et al., 2020. Sites of infection were standardized across datasets using a common ontology (i.e., specified as urethra, rectum, pharynx, cervix, or

other). Two-sided Fisher's exact test in R (version 3.5.1) was used to infer whether there was nonrandom association between *mtrC* LOF presence and either anatomical site of infection or sexual behavior. For sexual behavior analysis, isolates cultured from multiple sites on the same patient were counted as only one data point.

Meningococcal *mtrC* analysis. *mtrC* alleles from *N. meningitidis* assembled genomes were downloaded from PubMLST (n=14798; accessed 2019-09-03) by setting (species = "Neisseria meningitidis"), filtering by (Sequence bin size >= 2 Mbp), and exporting sequences for Locus "NEIS1634"[82]. *mtrC* LOF alleles were identified as described above. We generated a core-genome maximum likelihood phylogeny of all North American ST-11 *N. meningitidis* strains from PubMLST (n=456; accessed 2019-09-03) rooted with meningococcal reference genome MC58 (Genbank accession: AE002098.2) using Roary (version 3.12) (flags -z -e -n -v -s -i 92) and annotated it using metadata from Retchless et al., 2018[44] (see Supplementary Data 5 for PubMLST IDs). Overrepresentation of *mtrC* LOF alleles in the US urogenital lineage compared to selected control datasets was assessed using two-sided Fisher's exact test in R (version 3.5.1).

Data Availability

In Supplementary Data S1.3-4, we have included accession numbers (via publicly hosted database NCBI SRA) for accessing all raw sequence data used for *N. gonorrhoeae* analyses. Intermediate outputs from the genomics pipeline (e.g., *de novo* assemblies) may also be available from the authors upon request. In Supplementary Data S1.5, we have included accession numbers (via publicly hosted database PubMLST: https://pubmlst.org/neisseria/) for accessing all sequence data used for *N. meningitidis* analyses. Source data underlying all

figures are available in Supplementary Data S1.1-2 or at https://github.com/gradlab/mtrC-GWAS.

Code Availability

Code to reproduce the analyses and figures is available at https://github.com/gradlab/mtrC-GWAS or from the authors upon request.

References

1 Olesen, S. W. *et al.* Azithromycin Susceptibility Among Neisseria gonorrhoeae Isolates and Seasonal Macrolide Use. *J Infect Dis* **219**, 619-623, doi:10.1093/infdis/jiy551 (2019).

2 Olesen, S. W. & Grad, Y. H. Deciphering the Impact of Bystander Selection for Antibiotic Resistance in Neisseria gonorrhoeae. *J Infect Dis* **221**, 1033-1035, doi:10.1093/infdis/jiz156 (2020).

3 Sanchez-Buso, L. *et al.* The impact of antimicrobials on gonococcal evolution. *Nat Microbiol*, doi:10.1038/s41564-019-0501-y (2019).

4 Unemo, M. & Shafer, W. M. Antimicrobial resistance in Neisseria gonorrhoeae in the 21st century: past, evolution, and future. *Clin Microbiol Rev* **27**, 587-613, doi:10.1128/CMR.00010-14 (2014).

5 Wadsworth, C. B., Arnold, B. J., Sater, M. R. A. & Grad, Y. H. Azithromycin Resistance through Interspecific Acquisition of an Epistasis-Dependent Efflux Pump Component and Transcriptional Regulator in Neisseria gonorrhoeae. *MBio* **9**, doi:10.1128/mBio.01419-18 (2018).

6 Rouquette-Loughlin, C. E. *et al.* Mechanistic Basis for Decreased Antimicrobial Susceptibility in a Clinical Isolate of Neisseria gonorrhoeae Possessing a Mosaic-Like mtr Efflux Pump Locus. *MBio* **9**, doi:10.1128/mBio.02281-18 (2018).

7 Quillin, S. J. & Seifert, H. S. Neisseria gonorrhoeae host adaptation and pathogenesis. *Nat Rev Microbiol* **16**, 226-240, doi:10.1038/nrmicro.2017.169 (2018).

8 Wi, T. *et al.* Antimicrobial resistance in Neisseria gonorrhoeae: Global surveillance and a call for international collaborative action. *PLoS Med* **14**, e1002344, doi:10.1371/journal.pmed.1002344 (2017).

9 Grad, Y. H. *et al.* Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United States, 2000-2013. *J Infect Dis* **214**, 1579-1587, doi:10.1093/infdis/jiw420 (2016).

10 Yahara, K. *et al.* Genomic surveillance of Neisseria gonorrhoeae to investigate the distribution and evolution of antimicrobial-resistance determinants and lineages. *Microb Genom* **4**, doi:10.1099/mgen.0.000205 (2018).

11 Eyre, D. W. *et al.* WGS to predict antibiotic MICs for Neisseria gonorrhoeae. *J Antimicrob Chemother* **72**, 1937-1947, doi:10.1093/jac/dkx067 (2017).

12 Demczuk, W. *et al.* Equations To Predict Antimicrobial MICs in Neisseria gonorrhoeae Using Molecular Antimicrobial Resistance Determinants. *Antimicrob Agents Chemother* **64**, doi:10.1128/AAC.02005-19 (2020).

13 Hicks, A. L. *et al.* Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. *PLoS Comput Biol* **15**, e1007349, doi:10.1371/journal.pcbi.1007349 (2019).

14 Earle, S. G. *et al.* Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat Microbiol* **1**, 16041, doi:10.1038/nmicrobiol.2016.41 (2016).

15 Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics*, bty539-bty539, doi:10.1093/bioinformatics/bty539 (2018).

16 Jaillard, M. *et al.* A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS Genet* **14**, e1007758, doi:10.1371/journal.pgen.1007758 (2018).

17 Mortimer, T. D. & Grad, Y. H. Applications of genomics to slow the spread of multidrug-resistant Neisseria gonorrhoeae. *Ann N Y Acad Sci*, doi:10.1111/nyas.13871 (2018).

18 Veal, W. L., Nicholas, R. A. & Shafer, W. M. Overexpression of the MtrC-MtrD-MtrE efflux pump due to an mtrR mutation is required for chromosomally mediated penicillin resistance in Neisseria gonorrhoeae. *J Bacteriol* **184**, 5619-5624, doi:10.1128/jb.184.20.5619-5624.2002 (2002).

19 Golparian, D., Shafer, W. M., Ohnishi, M. & Unemo, M. Importance of multidrug efflux pumps in the antimicrobial resistance property of clinical multidrug-resistant isolates of Neisseria gonorrhoeae. *Antimicrob Agents Chemother* **58**, 3556-3559, doi:10.1128/AAC.00038-14 (2014).

20 Veal, W. L. *et al.* Loss-of-function mutations in the mtr efflux system of Neisseria gonorrhoeae. *Microbiology* **144 (Pt 3)**, 621-627, doi:10.1099/00221287-144-3-621 (1998).

21 Sanchez-Buso, L., Golparian, D., Parkhill, J., Unemo, M. & Harris, S. R. Genetic variation regulates the activation and specificity of Restriction-Modification systems in Neisseria gonorrhoeae. *Sci Rep* **9**, 14685, doi:10.1038/s41598-019-51102-2 (2019).

22 Eisenstein, B. I. & Sparling, P. F. Mutations to increased antibiotic sensitivity in naturally-occurring gonococci. *Nature* **271**, 242-244, doi:10.1038/271242a0 (1978).

23 Shimuta, K. *et al.* Emergence and evolution of internationally disseminated cephalosporin-resistant Neisseria gonorrhoeae clones from 1995 to 2005 in Japan. *BMC Infect Dis* **15**, 378, doi:10.1186/s12879-015-1110-x (2015).

24 Bichara, M., Pinet, I., Schumacher, S. & Fuchs, R. P. Mechanisms of dinucleotide repeat instability in Escherichia coli. *Genetics* **154**, 533-542 (2000).

25 Golparian, D. *et al.* Genomic evolution of Neisseria gonorrhoeae since the preantibiotic era (1928-2013): antimicrobial use/misuse selects for resistance and drives evolution. *BMC Genomics* **21**, 116, doi:10.1186/s12864-020-6511-6 (2020).

26 Handing, J. W., Ragland, S. A., Bharathan, U. V. & Criss, A. K. The MtrCDE Efflux Pump Contributes to Survival of Neisseria gonorrhoeae From Human Neutrophils and Their Antimicrobial Components. *Front Microbiol* **9**, 2688, doi:10.3389/fmicb.2018.02688 (2018).

27 Warner, D. M., Folster, J. P., Shafer, W. M. & Jerse, A. E. Regulation of the MtrC-MtrD-MtrE efflux-pump system modulates the in vivo fitness of Neisseria gonorrhoeae. *J Infect Dis* **196**, 1804-1812, doi:10.1086/522964 (2007).

28 Warner, D. M., Shafer, W. M. & Jerse, A. E. Clinically relevant mutations that cause derepression of the Neisseria gonorrhoeae MtrC-MtrD-MtrE Efflux pump system confer different levels of antimicrobial resistance and in vivo fitness. *Mol Microbiol* **70**, 462-478, doi:10.1111/j.1365-2958.2008.06424.x (2008).

29 Folster, J. P. *et al.* MtrR modulates rpoH expression and levels of antimicrobial resistance in Neisseria gonorrhoeae. *J Bacteriol* **191**, 287-297, doi:10.1128/JB.01165-08 (2009).

30 Wang, S. *et al.* Gonococcal MtrE and its surface-expressed Loop 2 are immunogenic and elicit bactericidal antibodies. *J Infect* **77**, 191-204, doi:10.1016/j.jinf.2018.06.001 (2018).

31 Williamson, D. A. *et al.* Bridging of Neisseria gonorrhoeae lineages across sexual networks in the HIV pre-exposure prophylaxis era. *Nat Commun* **10**, 3988, doi:10.1038/s41467-019-12053-4 (2019).

32 Rouquette, C., Harmon, J. B. & Shafer, W. M. Induction of the mtrCDE-encoded efflux pump system of Neisseria gonorrhoeae requires MtrA, an AraC-like protein. *Mol Microbiol* **33**, 651-658, doi:10.1046/j.1365-2958.1999.01517.x (1999).

33 Vidyaprakash, E., Abrams, A. J., Shafer, W. M. & Trees, D. L. Whole-Genome Sequencing of a Large Panel of Contemporary Neisseria gonorrhoeae Clinical Isolates Indicates that a Wild-Type mtrA Gene Is Common: Implications for Inducible Antimicrobial Resistance. *Antimicrob Agents Chemother* **61**, doi:10.1128/AAC.00262-17 (2017).

34 Olivares Pacheco, J., Alvarez-Ortega, C., Alcalde Rico, M. & Martinez, J. L. Metabolic Compensation of Fitness Costs Is a General Outcome for Antibiotic-Resistant Pseudomonas aeruginosa Mutants Overexpressing Efflux Pumps. *MBio* **8**, doi:10.1128/mBio.00500-17 (2017).

35 Zgurskaya, H. I. Multicomponent drug efflux complexes: architecture and mechanism of assembly. *Future Microbiol* **4**, 919-932, doi:10.2217/fmb.09.62 (2009).

36 Lee, E. H. & Shafer, W. M. The farAB-encoded efflux pump mediates resistance of gonococci to long-chained antibacterial fatty acids. *Mol Microbiol* **33**, 839-845, doi:10.1046/j.1365-2958.1999.01530.x (1999).

37 Lee, E. H., Rouquette-Loughlin, C., Folster, J. P. & Shafer, W. M. FarR regulates the farAB-encoded efflux pump of Neisseria gonorrhoeae via an MtrR regulatory mechanism. *J Bacteriol* **185**, 7145-7152, doi:10.1128/jb.185.24.7145-7152.2003 (2003).

38 Rouquette-Loughlin, C., Dunham, S. A., Kuhn, M., Balthazar, J. T. & Shafer, W. M. The NorM efflux pump of Neisseria gonorrhoeae and Neisseria meningitidis recognizes antimicrobial cationic compounds. *J Bacteriol* **185**, 1101-1106, doi:10.1128/jb.185.3.1101-1106.2003 (2003).

39 Rouquette-Loughlin, C. E., Balthazar, J. T. & Shafer, W. M. Characterization of the MacA-MacB efflux system in Neisseria gonorrhoeae. *J Antimicrob Chemother* **56**, 856-860, doi:10.1093/jac/dki333 (2005).

40 Stephens, D. S., Greenwood, B. & Brandtzaeg, P. Epidemic meningitis, meningococcaemia, and Neisseria meningitidis. *Lancet* **369**, 2196-2210, doi:10.1016/S0140-6736(07)61016-2 (2007).

41 Hill, D. M. *et al.* Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study. *Lancet Infect Dis* **15**, 1420-1428, doi:10.1016/S1473-3099(15)00267-4 (2015).

42 Bichara, M., Schumacher, S. & Fuchs, R. P. Genetic instability within monotonous runs of CpG sequences in Escherichia coli. *Genetics* **140**, 897-907 (1995).

43 Bazan, J. A. *et al.* Notes from the Field: Increase in Neisseria meningitidis-Associated Urethritis Among Men at Two Sentinel Clinics - Columbus, Ohio, and Oakland County, Michigan, 2015. *MMWR Morb Mortal Wkly Rep* **65**, 550-552, doi:10.15585/mmwr.mm6521a5 (2016).

44 Retchless, A. C. *et al.* Expansion of a urethritis-associated Neisseria meningitidis clade in the United States with concurrent acquisition of N. gonorrhoeae alleles. *BMC Genomics* **19**, 176, doi:10.1186/s12864-018-4560-x (2018).

45 Tzeng, Y. L. *et al.* Emergence of a new Neisseria meningitidis clonal complex 11 lineage 11.2 clade as an effective urogenital pathogen. *Proc Natl Acad Sci U S A* **114**, 4237-4242, doi:10.1073/pnas.1620971114 (2017).

46 Whiley, D. M., Jennison, A., Pearson, J. & Lahra, M. M. Genetic characterisation of Neisseria gonorrhoeae resistant to both ceftriaxone and azithromycin. *Lancet Infect Dis* **18**, 717-718, doi:10.1016/S1473-3099(18)30340-2 (2018).

47 Kime, L. *et al.* Transient Silencing of Antibiotic Resistance by Mutation Represents a Significant Potential Source of Unanticipated Therapeutic Failure. *MBio* **10**, doi:10.1128/mBio.01755-19 (2019).

48 Morse, S. A. *et al.* Gonococcal strains from homosexual men have outer membranes with reduced permeability to hydrophobic molecules. *Infect Immun* **37**, 432-438 (1982).

49 Shafer, W. M., Balthazar, J. T., Hagman, K. E. & Morse, S. A. Missense mutations that alter the DNA-binding domain of the MtrR protein occur frequently in rectal isolates of Neisseria gonorrhoeae that are resistant to faecal lipids. *Microbiology* **141 (Pt 4)**, 907-911, doi:10.1099/13500872-141-4-907 (1995).

50 Mercante, A. D. *et al.* MpeR regulates the mtr efflux locus in Neisseria gonorrhoeae and modulates antimicrobial resistance by an iron-responsive mechanism. *Antimicrob Agents Chemother* **56**, 1491-1501, doi:10.1128/AAC.06112-11 (2012).

51 Nudel, K. *et al.* Transcriptome Analysis of Neisseria gonorrhoeae during Natural Infection Reveals Differential Expression of Antibiotic Resistance Determinants between Men and Women. *mSphere* **3**, doi:10.1128/mSphereDirect.00312-18 (2018).

52 Du, D. *et al.* Multidrug efflux pumps: structure, function and regulation. *Nat Rev Microbiol* **16**, 523-539, doi:10.1038/s41579-018-0048-6 (2018).

53 Vettoretti, L. *et al.* Efflux unbalance in Pseudomonas aeruginosa isolates from cystic fibrosis patients. *Antimicrob Agents Chemother* **53**, 1987-1997, doi:10.1128/AAC.01024-08 (2009).

54 Chalhoub, H. *et al.* Mechanisms of intrinsic resistance and acquired susceptibility of Pseudomonas aeruginosa isolated from cystic fibrosis patients to temocillin, a revived antibiotic. *Sci Rep* **7**, 40208, doi:10.1038/srep40208 (2017).

55 Garnett, J. P. *et al.* Hyperglycaemia and Pseudomonas aeruginosa acidify cystic fibrosis airway surface liquid by elevating epithelial monocarboxylate transporter 2 dependent lactate-H(+) secretion. *Sci Rep* **6**, 37955, doi:10.1038/srep37955 (2016).

56 Worlitzsch, D. *et al.* Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. *J Clin Invest* **109**, 317-325, doi:10.1172/JCI13870 (2002).

57 Ayhan, D. H. *et al.* Sequence-Specific Targeting of Bacterial Resistance Genes Increases Antibiotic Efficacy. *PLoS Biol* **14**, e1002552, doi:10.1371/journal.pbio.1002552 (2016).

58 Chan, B. K. *et al.* Phage selection restores antibiotic sensitivity in MDR Pseudomonas aeruginosa. *Sci Rep* **6**, 26717, doi:10.1038/srep26717 (2016).

59 Chen, S. *et al.* Could Dampening Expression of the Neisseria gonorrhoeae mtrCDE-Encoded Efflux Pump Be a Strategy To Preserve Currently or Resurrect Formerly Used Antibiotics To Treat Gonorrhea? *MBio* **10**, doi:10.1128/mBio.01576-19 (2019).

60 Cole, M. J. *et al.* Gentamicin, azithromycin and ceftriaxone in the treatment of gonorrhoea: the relationship between antibiotic MIC and clinical outcome. *J Antimicrob Chemother* **75**, 449-457, doi:10.1093/jac/dkz436 (2020).

61 Tapsall, J. W. *et al.* Failure of azithromycin therapy in gonorrhea and discorrelation with laboratory test parameters. *Sex Transm Dis* **25**, 505-508, doi:10.1097/00007435-199811000-00002 (1998).

62 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

63 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).

64 Garcia-Alcalde, F. *et al.* Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**, 2678-2679, doi:10.1093/bioinformatics/bts503 (2012).

65 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).

66 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

67 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).

68 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691-3693, doi:10.1093/bioinformatics/btv421 (2015).

69 Thorpe, H. A., Bayliss, S. C., Sheppard, S. K. & Feil, E. J. Piggy: a rapid, large-scale pan-genome analysis tool for intergenic regions in bacteria. *Gigascience* **7**, 1-11, doi:10.1093/gigascience/giy015 (2018).

70 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-W259, doi:10.1093/nar/gkz239 (2019).

71 Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**, e15, doi:10.1093/nar/gku1196 (2015).

72 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

73 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

74 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066, doi:10.1093/nar/gkf436 (2002).

75 Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res* **47**, 5539-5549, doi:10.1093/nar/gkz361 (2019).

76 Johnson, S. R., Grad, Y., Abrams, A. J., Pettus, K. & Trees, D. L. Use of whole-genome sequencing data to analyze 23S rRNA-mediated azithromycin resistance. *Int J Antimicrob Agents* **49**, 252-254, doi:10.1016/j.ijantimicag.2016.10.023 (2017).

77 Kersh, E. N. *et al.* Rationale for a Neisseria gonorrhoeae Susceptible Only Interpretive Breakpoint for Azithromycin. *Clin Infect Dis*, doi:10.1093/cid/ciz292 (2019).

78 Ma, K. C. *et al.* Increased power from bacterial genome-wide association conditional on known effects identifies *Neisseria gonorrhoeae* macrolide resistance mutations in the 50S ribosomal protein L4. Preprint at https://www.biorxiv.org/content/10.1101/2020.03.24.006650v1 (2020).

79 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423, doi:10.1093/bioinformatics/btp163 (2009).

80 Dillard, J. P. Genetic Manipulation of Neisseria gonorrhoeae. *Curr Protoc Microbiol* **Chapter 4**, Unit4A 2, doi:10.1002/9780471729259.mc04a02s23 (2011).

81 Mortimer, T. D. *et al.* The distribution and spread of susceptible and resistant *Neisseria gonorrhoeae* across demographic groups in a major metropolitan center. Preprint at https://www.medrxiv.org/content/10.1101/2020.04.30.20086413v1 (2020).

82 Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* **3**, 124, doi:10.12688/wellcomeopenres.14826.1 (2018).

Chapter 3: Efflux pump antibiotic binding site mutations are associated with azithromycin nonsusceptibility in clinical *Neisseria gonorrhoeae* isolates

Authors

Kevin C Ma[1¶], Tatum D Mortimer[1¶], Yonatan H Grad[1,2]

[1] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, USA

[2] Division of Infectious Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, USA

[¶] These authors contributed equally to this work.

Author Contributions

KCM and TDM performed and interpreted genomic analyses. KCM, TDM, and YHG wrote the paper.

Publication

This chapter appears as "Efflux pump antibiotic binding site mutations are associated with azithromycin nonsusceptibility in clinical *Neisseria gonorrhoeae* isolates" in *mBio*, Volume 11, Issue 4, Pages e01509-20, 2020.

Introduction

In the previous chapter, we used GWAS to identify mutations in the MtrCDE efflux pump associated with antibiotic susceptibility. Because GWAS approaches may suffer from limited power to detect rare causal mutations, we conducted follow-up analyses on the *mtrCDE* operon to identify rare mutations associated with increases in antibiotic MICs. To accomplish this, we used findings from a structural biology study conducted by Lyu and Moseng et al. to guide our genomic analyses by focusing on key amino acid residues predicted to be involved in drug binding.

Results

Lyu and Moseng et al. used cryo-electron microscopy to characterize key residues involved in drug binding by mosaic-like MtrD efflux pump alleles in *Neisseria gonorrhoeae*[1]. Isogenic experiments introducing key MtrD substitutions R714G and K823E increased macrolide MICs, leading the authors to predict that non-mosaic MtrD "gonococcal strains bearing both the *mtrR* promoter and amino acid changes at MtrD positions 714 or 823 could lead to clinically significant levels of Azi nonsusceptibility resistance". We tested this hypothesis by analyzing a global meta-analysis collection of 4852 *N. gonorrhoeae* genomes[2]. In support of their prediction, we identified clinical isolates with novel non-mosaic MtrD drug binding site substitutions across multiple genetic backgrounds associated with elevated azithromycin MICs (Table 3.1).

Of the 4852 isolates, 12 isolates contained non-synonymous mutations at position R714 to amino acids H, L, or C and 7 isolates contained K823 mutations to E or N in the non-mosaic MtrD background. We did not observe substitutions at positions 174, 669, 821, and 825, in line with the authors' demonstration that isogenic mutants at these codons had identical or lowered macrolide MICs. The azithromycin geometric mean MICs of the clinical isolates with mutations

at R714 and K823 were 1.25 μg/mL and 2.12 μg/mL respectively, both of which are above the CLSI azithromycin non-susceptibility threshold (Figure 3.1a). There was a significant difference in mean MIC distributions comparing MtrD substitution strains with genetically matched controls ($p=0.0008$, mean $\log_2$(MIC) difference=1.86, paired samples Wilcoxon test; Table 3.2). There was also a significant difference in mean MIC distributions for ceftriaxone ($p=0.045$, mean $\log_2$(MIC) difference=0.56) but not for ciprofloxacin ($p=0.62$).

Nearly all MtrD substitution strains contained *mtrR* promoter mutations that increase MtrCDE pump expression (Table 3.1)[3]. The isolate with an MtrD R714H mutation and the lowest observed azithromycin MIC of 0.19 μg/mL did not have an *mtrR* promoter mutation, consistent with epistasis across the *mtrRCDE* operon[4]. Contributions from ribosomal mutations can also synergistically increase macrolide resistance: the isolate with an MtrD K823E substitution and the highest observed azithromycin MIC of 8.0 μg/mL contained an RplD G70S mutation previously implicated in macrolide resistance[5]. Seven MtrD isolates also had mosaic *penA* XXXIV alleles conferring cephalosporin reduced susceptibility, indicating a potential route to dual therapy resistance.

MtrD R714 and MtrD K823 substitutions were each acquired seven times across the phylogeny, suggesting acquisition of the mutation is possible in different genetic backgrounds (Figure 3.1b). Four of the MtrD K823 acquisitions were associated with more than one isolate descending from the same ancestor, suggesting that these strains are successfully transmitted. In line with this, non-recombinant SNP distances between isolates in each of the four clusters were all below 18 SNPs, with 3/4 clusters below the 10 SNP cutoff previously used as evidence for defining a transmission cluster[6,7].

Complementing the experimental and structural biology approach taken by Lyu and Moseng et al., we demonstrated using genomics that clinical isolates have acquired novel MtrD binding site mutations which, in combination with *mtrR* promoter and RplD mutations, can result in azithromycin non-susceptibility. As azithromycin resistant strains have been growing in prevalence[8-13], our data support the inclusion of MtrD binding site residues in future genomic surveillance and genotype-to-phenotype diagnostics and modeling studies for characterizing gonococcal resistance.

Methods

MtrD substitutions. We ran BLASTn on the assemblies using an *mtrD* reference sequence from gonococcal strain FA1090 (Genbank accession: NC_002946.2) and filtered out hits with lower than 99% identity to remove mosaic *mtrD* sequences. LOF alleles were also removed by identifying strains with predicted peptides 90% or shorter than the FA1090 allele. *mtrD* sequences were aligned using MAFFT (version 7.450)[14], translated, and diversity at residues 174, 669, 714, 821, 823, and 825 was characterized using Python (version 3.6.5) and Biopython (version 1.69)[15].

Nearest neighbor analysis. We calculated non-recombinant SNP distances using snp-dists (version 0.6.3, https://github.com/tseemann/snp-dists) from a recombination masked alignment of polymorphisms produced by a previous analysis of recombination using Gubbins (version 2.3.4)[16]. For each strain with a MtrD mutation at codon 714 or 823, we identified the most closely related strain in the dataset without mutations at these positions and with available azithromycin MICs. We used a paired samples Wilcoxon test to test for significant differences between log transformed azithromycin MICs and their nearest neighbor in R (version 3.6.1).

Non-recombinant SNP distances were also used to identify genetic distance between MtrD substitution strains within clusters.

Data availability

All code, metadata, and intermediate analyses files to replicate analyses are available at https://github.com/gradlab/mtrD-resistance/. An interactive and downloadable version of the phylogeny is hosted at https://itol.embl.de/tree/12810324163074215911078 15.

Figure 3.1 – MtrD mutations associated with increased azithromycin MICs have emerged across the *N. gonorrhoeae* phylogeny. a) Comparison of AZI MIC distributions for strains with and without non-mosaic MtrD substitutions at R714 and K823, and b) phylogenetic distribution of MtrD substitution strains in a recombination-corrected phylogeny of the 4852 strains from the global meta-analysis collection. In 1b, triangles indicate singleton strains and stars indicate clusters of two or more strains; cluster number labels correspond to cluster labels in Table 3.1.

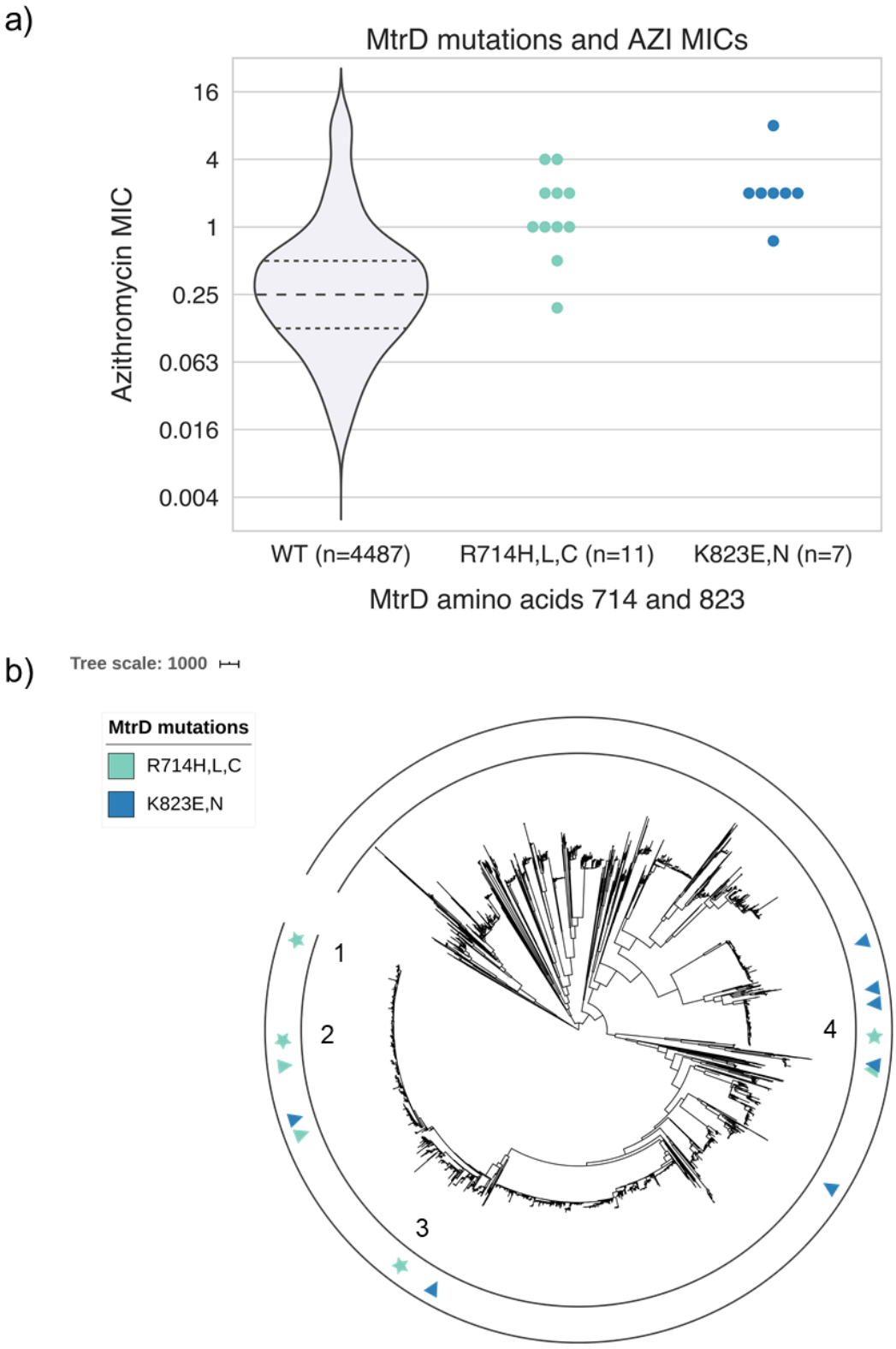


Figure 3.1 (continued)

Table 3.1 – MtrD substitution strains, associated metadata, and resistance allele genotypes.

Accession	Ref.	AZI MIC (µg/mL)	MtrD Allele	Cluster[a]	*mtrR* Promoter	RplD G70 Allele	23S rRNA	*penA* Allele
ERR1469714	[9]	1	R714H	1	Adel	WT	WT	XXXIV
ERR1528327	[9]	1	R714H	2	Adel	WT	WT	XXXIV
ERR1514686	[9]	NA	R714H	2	Adel	WT	WT	XXXIV
ERR1469709	[9]	1	R714H	1	Adel	WT	WT	XXXIV
SRR1661243	[10]	1	R714H	3	Adel	WT	WT	Non-mosaic
SRR2736280	[11]	2	R714H	NA	Adel	WT	WT	XXXIV
SRR2736175	[11]	2	R714H	3	Adel	WT	WT	Non-mosaic
SRR2736167	[11]	2	R714H	3	Adel	WT	WT	Non-mosaic
ERR349976	[12]	0.19	R714H	NA	WT	WT	WT	Non-mosaic
ERR854880	[13]	4	R714L	4	Adel	WT	WT	Non-mosaic
ERR855125	[13]	4	R714L	4	Adel	WT	WT	Non-mosaic
ERR855232	[13]	0.5	R714C	NA	Adel	WT	WT	XXXIV
ERR363653	[12]	0.75	K823E	NA	Adel	WT	WT	Non-mosaic
ERR855395	[13]	8	K823E	NA	Adel	R	WT	XXXIV
ERR855128	[13]	2	K823E	NA	Adel	WT	WT	Non-mosaic
ERR1067793	[13]	2	K823E	NA	Adel	WT	WT	Non-mosaic

SRR2736213	11	2	K823E	NA	Adel	S	WT	Non-mosaic
SRR2736281	11	2	K823E	NA	Adel	WT	WT	Non-mosaic
SRR2736124	11	2	K823N	NA	Adel	WT	WT	Non-mosaic

Table 3.1 (continued)

[a]Cluster number corresponds to cluster number labels in the Figure 3.1b phylogeny.

Table 3.2 – Comparison of AZI MICs of MtrD substitution strains and their nearest neighbors.

MtrD substitution strain	MtrD Mutation	Substitution Strain AZI MIC (µg/mL)	Matched MtrD WT strain	WT Strain AZI MIC (µg/mL)	Non-recombinant SNP distance
SRR1661243	mtrD_714	1	SRR1661226	0.25	4
SRR2736124	mtrD_823	2	SRR2736094	2	7
SRR2736167	mtrD_714	2	SRR1661226	0.25	15
SRR2736175	mtrD_714	2	SRR1661226	0.25	5
SRR2736213	mtrD_823	2	DRR124869	0.5	145
SRR2736280	mtrD_714	2	ERR855279	1	17
SRR2736281	mtrD_823	2	ERR855005	0.125	12
ERR855128	mtrD_823	2	ERR1560878	0.125	106
ERR1067793	mtrD_823	2	ERR1560869	1.5	22
ERR855395	mtrD_823	8	ERR191732	1	29
ERR855232	mtrD_714	0.5	ERR854884	0.5	9
ERR854880	mtrD_714	4	ERR1471127	0.25	24
ERR855125	mtrD_714	4	ERR1560869	1.5	28
ERR1469709	mtrD_714	1	ERR1528280	0.25	11
ERR1469714	mtrD_714	1	ERR1528280	0.25	11
ERR1514686	mtrD_714	NA	ERR1560893	0.5	31
ERR1528327	mtrD_714	1	ERR1560893	0.5	22
ERR349976	mtrD_714	0.19	ERR1560863	0.38	772
ERR363653	mtrD_823	0.75	ERR363634	0.25	71

References

1. Lyu M*, et al.* Cryo-EM Structures of a Gonococcal Multidrug Efflux Pump Illuminate a Mechanism of Drug Recognition and Resistance. *mBio* **11**, (2020).

2. Ma KC*, et al.* Increased antibiotic susceptibility in Neisseria gonorrhoeae through adaptation to the cervical environment. *bioRxiv*, 2020.2001.2007.896696 (2020).

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

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

5. Ma KC*, et al.* Increased power from bacterial genome-wide association conditional on known effects identifies Neisseria gonorrhoeae macrolide resistance mutations in the 50S ribosomal protein L4. *bioRxiv*, 2020.2003.2024.006650 (2020).

6. Williamson DA*, et al.* Bridging of Neisseria gonorrhoeae lineages across sexual networks in the HIV pre-exposure prophylaxis era. *Nat Commun* **10**, 3988 (2019).

7. Mortimer TD*, et al.* The distribution and spread of susceptible and resistant Neisseria gonorrhoeae across demographic groups in a major metropolitan center. *medRxiv*, 2020.2004.2030.20086413 (2020).

8. CDC. Sexually Transmitted Disease Surveillance 2018.) (2019).

9. Harris SR*, et al.* Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a genomic survey. *Lancet Infect Dis* **18**, 758-768 (2018).

10. Demczuk W*, et al.* Whole-genome phylogenomic heterogeneity of Neisseria gonorrhoeae isolates with decreased cephalosporin susceptibility collected in Canada between 1989 and 2013. *J Clin Microbiol* **53**, 191-200 (2015).

11. Demczuk W*, et al.* Genomic epidemiology and molecular resistance mechanisms of azithromycin resistant Neisseria gonorrhoeae in Canada from 1997 to 2014. *J Clin Microbiol*, (2016).

12. Sánchez-Busó L*, et al.* Antimicrobial exposure in sexual networks drives divergent evolution in modern gonococci. *bioRxiv*, 334847 (2018).

13. Grad YH*, et al.* Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United States, 2000-2013. *J Infect Dis* **214**, 1579-1587 (2016).

14. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-66.

15. 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.

16. 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.

Chapter 4: Increased power from conditional bacterial genome-wide association identifies macrolide resistance mutations in *Neisseria gonorrhoeae*

Authors

Kevin C Ma[1], Tatum D Mortimer[1], Marissa A Duckett[1], Allison L Hicks[1], Nicole E Wheeler[2], Leonor Sánchez-Busó[2], Yonatan H Grad[1,3]

[1] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, USA

[2] Centre for Genomic Pathogen Surveillance, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom

[3] Division of Infectious Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, USA

Author contributions

K.C.M., T.D.M., A.L.H., N.E.W., and L.S.B. performed and interpreted genomic analyses. K.C.M. and M.A.D. performed and interpreted experimental analyses. Y.H.G. supervised the project. K.C.M. and Y.H.G. wrote the paper with contributions from all authors.

Publication

This chapter appears as "Increased power from conditional bacterial genome-wide association identifies macrolide resistance mutations in *Neisseria gonorrhoeae*" in *Nature Communications*, Volume 11, Article 5374, 2020.

Abstract

The emergence of resistance to azithromycin complicates treatment of *Neisseria gonorrhoeae*, the etiologic agent of gonorrhea. Substantial azithromycin resistance remains unexplained after accounting for known resistance mutations. Bacterial genome-wide association studies (GWAS) can identify novel resistance genes but must control for genetic confounders while maintaining power. Here, we show that compared to single-locus GWAS, conducting GWAS conditioned on known resistance mutations reduces the number of false positives and identifies a G70D mutation in the RplD 50S ribosomal protein L4 as significantly associated with increased azithromycin resistance (*p*-value = 1.08×10^{-11}). We experimentally confirm our GWAS results and demonstrate that RplD G70D and other macrolide binding site mutations are prevalent (present in 5.42% of 4850 isolates) and widespread (identified in 21/65 countries across two decades). Overall, our findings demonstrate the utility of conditional associations for improving the performance of microbial GWAS and advance our understanding of the genetic basis of macrolide resistance.

Introduction

Increasing antibiotic resistance in *Neisseria gonorrhoeae*, the causative agent of the sexually transmitted disease gonorrhea, threatens effective control of this prevalent pathogen[1-3]. Current empiric antibiotic therapy in the US comprises a combination of the cephalosporin ceftriaxone and the macrolide azithromycin, but increasing prevalence of azithromycin resistance has led some countries, such as the UK, to instead recommend ceftriaxone monotherapy[4]. Rapid genotypic diagnostics for antimicrobial susceptibility have been proposed as a platform to tailor therapy and to extend the clinically useful lifespan of anti-gonococcal antibiotics[5,6]. These rapid diagnostics rest on robust genotype-to-phenotype predictions. For some antibiotics, such as ciprofloxacin, resistance is predictable by target site mutations in a single gene, *gyrA*[3,5]. However, recent efforts to predict azithromycin minimum inhibitory concentrations (MICs) using regression-based or machine-learning approaches have indicated that a substantial fraction of phenotypic resistance is unexplained, particularly among strains with lower-level resistance[3,7,8]. An improved understanding of the genetic mechanisms and evolutionary pathways to macrolide resistance will therefore be critical for informing the development of diagnostics.

Macrolides function by binding to the 50S ribosome and inhibiting protein synthesis[9]. Increased resistance can occur in *N. gonorrhoeae* through target site modification, primarily via 23S rRNA mutations C2611T[10] and A2059G[11], and through efflux pump upregulation. The main efflux pump associated with antibiotic resistance in the gonococcus is the Mtr efflux pump, comprising a tripartite complex encoded by the *mtrCDE* operon under the regulation of the MtrR repressor and the MtrA activator[1,12-17]. Active site or frameshift mutations in the coding sequence of *mtrR* and promoter mutations in the *mtrR* promoter upregulate *mtrCDE* and result in increased macrolide resistance[1,18]. Mosaic sequences originating from recombination with homologs from

commensal *Neisseria* donors can also result in structural changes to *mtrD* and increased expression of *mtrCDE*, which synergistically act to confer resistance[19,20].

Here, we use genome-wide association on a global meta-analysis dataset to identify additional genetic variants that confer increased azithromycin resistance in *N. gonorrhoeae.* We find that conventional single-locus bacterial GWAS approaches result in confounded results and reduced power, but conducting GWAS conditional on known resistance mutations in 23S rRNA reduces linkage-mediated confounding and increases power to recover mutations associated with lower-level resistance. We experimentally validate one such mutation in the 50S ribosomal protein RplD and identify other rare RplD variants associated with resistance, highlighting the ability of conditional bacterial GWAS to discover causal genes for polygenic microbial phenotypes.

Results

We previously conducted a linear mixed model GWAS using a global meta-analysis collection of 4852 *N. gonorrhoeae* isolates, collected across 15 studies and spanning 65 countries and 38 years[7]. After conducting GWAS on the 4505 isolates with associated azithromycin MICs, we identified highly significant unitigs (i.e., genetic variants generated from *de novo* assemblies) mapping to the 23S rRNA, associated with increased resistance, and to the efflux pump gene *mtrC*, associated with increased susceptibility and cervical infections[7]. These results highlighted the potential for GWAS to identify novel modifiers of resistance in *N. gonorrhoeae.* However, the characterized mutations did not fully explain azithromycin heritability and thus pointed towards unknown genetic variants.

Conditional GWAS identifies a resistance mutation in RplD.

To identify these variants, we re-analyzed the GWAS results focusing on the remaining unitigs, which had lower effect sizes and *p*-values closer to the Bonferroni-corrected *p*-value threshold, calculated using the number of unique patterns, of 3.38×10^{-7}. Numerous variants were significantly associated with increased MICs, many of which mapped to genes (e.g., *hprA*, WHO_F.1254, and *ydfG*) that had not previously been implicated in macrolide resistance in *Neisseria* (Supplementary Data S2.1). While these signals could represent novel causal resistance genes, we hypothesized that at least some of these variants could have been spuriously driven to association via genetic linkage with the highly penetrant (A2059G: β, or effect size on the $\log_2$-transformed MIC scale, = 7.14, 95% CI [6.44, 7.84]; C2611T: $\beta = 3.67$, 95% CI [3.46, 3.88]) and population-stratified 23S rRNA resistance mutations (Supplementary Figure S2.1). Supporting this hypothesis, r^2 – a measure of linkage ranging from 0 to 1 – between significant variants and 23S rRNA resistance mutations showed a bimodal distribution with a peak at 0.84 and at 0.04 (Supplementary Figure S2.2). The three significant variants that mapped to *hprA*, WHO_F.1254, and *ydfG* had elevated r^2 values of 0.16, 0.82, and 0.80 respectively; all three variants demonstrated clear phylogenetic overlap with 23S rRNA mutations (Supplementary Figure S2.1). Additionally, we did not observe unitigs associated with previously experimentally validated resistance mutations in the *mtrR* promoter[14] or the *mtrCDE* mosaic alleles[19,20], suggesting decreased power to detect known causal variants with lower effect sizes.

To control for the confounding effect of the 23S rRNA mutations, we conducted a conditional GWAS by incorporating additional covariates in our linear mixed model encoding the number of copies (ranging from 0 to 4) of the resistance-conferring 23S rRNA substitutions C2611T and A2059G. We also conditioned on isolate dataset of origin to address potential spurious hits

arising from study-specific sequencing methodologies. After conditioning, the previously significant genes linked to 23S rRNA ($r^2 > 0.80$) decreased below the significance threshold, indicating that they were indeed driven to significance by genetic linkage (Figure 4.1, Supplementary Data S2.2). The most significant variants after the previously reported *mtrC* indel[7] mapped to the *mtrR* promoter (β, or effect size, = –0.86, 95% CI [-1.05, -0.68]; *p*-value = 5.44×10^{-20}), encoding the complement of the *mtrR* promoter 1 bp deletion[21], and to *mtrC* (β = 1.23, 95% CI [0.93, 1.53]; *p*-value = 9.03×10^{-16}), in linkage with mosaic *mtr* alleles[19,20]. The increased significance of these known efflux pump resistance mutations suggested improved power to recover causal genes with lower effects. Conditioning on dataset did not substantially affect these results but helped to remove other spurious variants arising due to study-specific biases[22] (Supplementary Figure S2.3, Supplementary Data S2.3).

A glycine to glutamic acid substitution at site 70 of the 50S ribosomal protein L4 (RplD) was significantly associated with increased azithromycin MICs after conducting the conditional GWAS (β = 0.95, 95% CI [0.68, 1.23]; *p*-value = 1.08×10^{-11}) (Figure 4.1, Supplementary Data S2.2). Structural analysis of the *Thermus thermophilus* 50S ribosome complexed with azithromycin suggests that this amino acid is an important residue in macrolide binding (Supplemental Figure S2.4), and RplD substitutions at this binding site modulate macrolide resistance in other bacteria[23,24]. This substitution has previously been observed rarely in gonococcus and the association with azithromycin resistance versus susceptibility was non-significant[3,25,26]; as a result, the role of RplD mutations in conferring macrolide resistance was unclear.

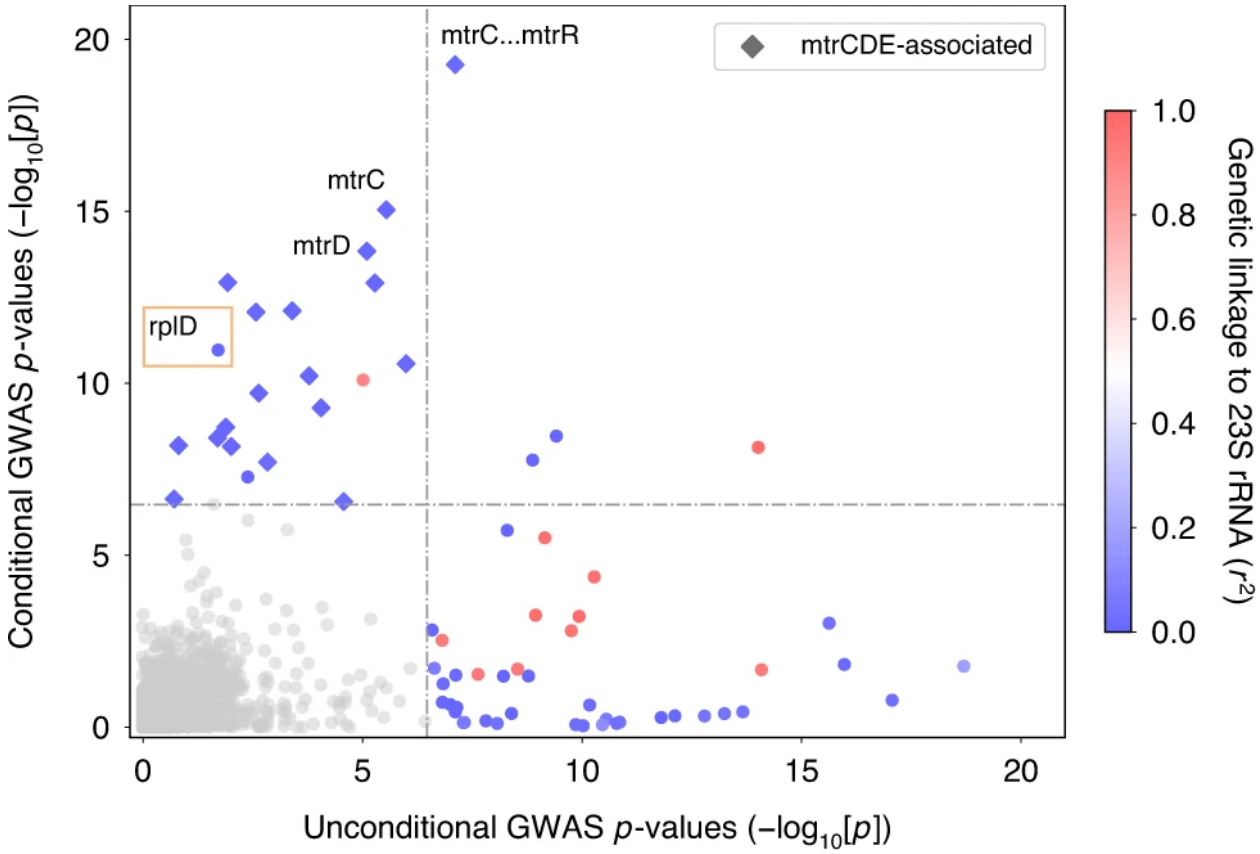


Figure 4.1 – GWAS conditional on 23S rRNA mutations and dataset demonstrates decreased confounding and increased power. Each variant is plotted using negative $\log_{10}$-transformed *p*-values, calculated using likelihood-ratio tests, for the association with azithromycin MICs in unconditional and conditional GWASes. Genetic linkage measured by r^2 to 23S rRNA mutations A2059G and C2611T is colored for significant variants as indicated on the right, ranging from 0 (blue) to 0.5 (white) to 1 (red). Variants associated with previously experimentally verified resistance mechanisms in the *mtrR* and *mtrCDE* promoters and coding regions are denoted using diamonds. Bonferroni thresholds, calculated using the number of unique patterns, for both GWASes are depicted using a gray dashed line at 3.38×10^{-7}. Plot axes are limited to highlight variants associated with lower-level resistance; as a result, the highly significant 23S rRNA substitutions and *mtrC* indel mutations[7] are not shown.

To assess the contribution of RplD mutations to continuous azithromycin MIC levels, we modeled MICs using a linear regression framework with known genetic resistance determinants as predictors (Supplementary Data S2.4 and S2.5)[7,27]. Compared to this baseline model, inclusion of the RplD G70D mutation decreased the number of strains with unexplained MIC variation (defined as absolute model error greater than one MIC dilution) from 1514 to 1463,

improved adjusted R^2 from 0.691 to 0.704, and significantly improved model fit (p-value < 10^{-10}; χ^2 test statistic = 288.51; Likelihood-ratio χ^2 test for nested models). These results indicate that RplD G70D is a strong candidate for addressing a portion of the unexplained azithromycin resistance in *N. gonorrhoeae*.

Genomic epidemiology of RplD macrolide binding site mutations.

We next assessed the population-wide prevalence and diversity of RplD-azithromycin binding site mutations. The RplD G70D mutation was present in 231 out of 4850 isolates (4.76%) with multiple introductions observed across varied genetic backgrounds (Figure 4.2). An additional 34 isolates contained mutations at amino acids 68 (G68D, G68C), 69 (T69I), and 70 (G70S, G70A, G70R, G70duplication) (Figure 4.3a). These other putative RplD binding site mutations were associated with significantly higher azithromycin MICs compared to both RplD G70D and RplD wild-type strains, indicating multiple avenues for disruption of macrolide binding (Figure 4.3b). Grouping all RplD binding site mutations together resulted in increased effect size (β = 1.02) and p-value (9.25×10^{-18}) in the conditional GWAS linear mixed model compared to the association with just RplD G70D (β = 0.95, p-value = 1.08×10^{-11}). Strains with RplD binding site mutations were identified from 21 countries from 1993 to 2015 with prevalence reaching over 10% in some datasets (New York City 2011-2015[28] and Japan 1996-2015[29]; Supplementary Table S2.1), in line with sustained transmission of RplD G70D strains (Figure 4.2). Our results suggest that macrolide binding to the 50S ribosome can be disrupted via multiple mutations and that these mutations are widespread contributors to azithromycin resistance in some populations.

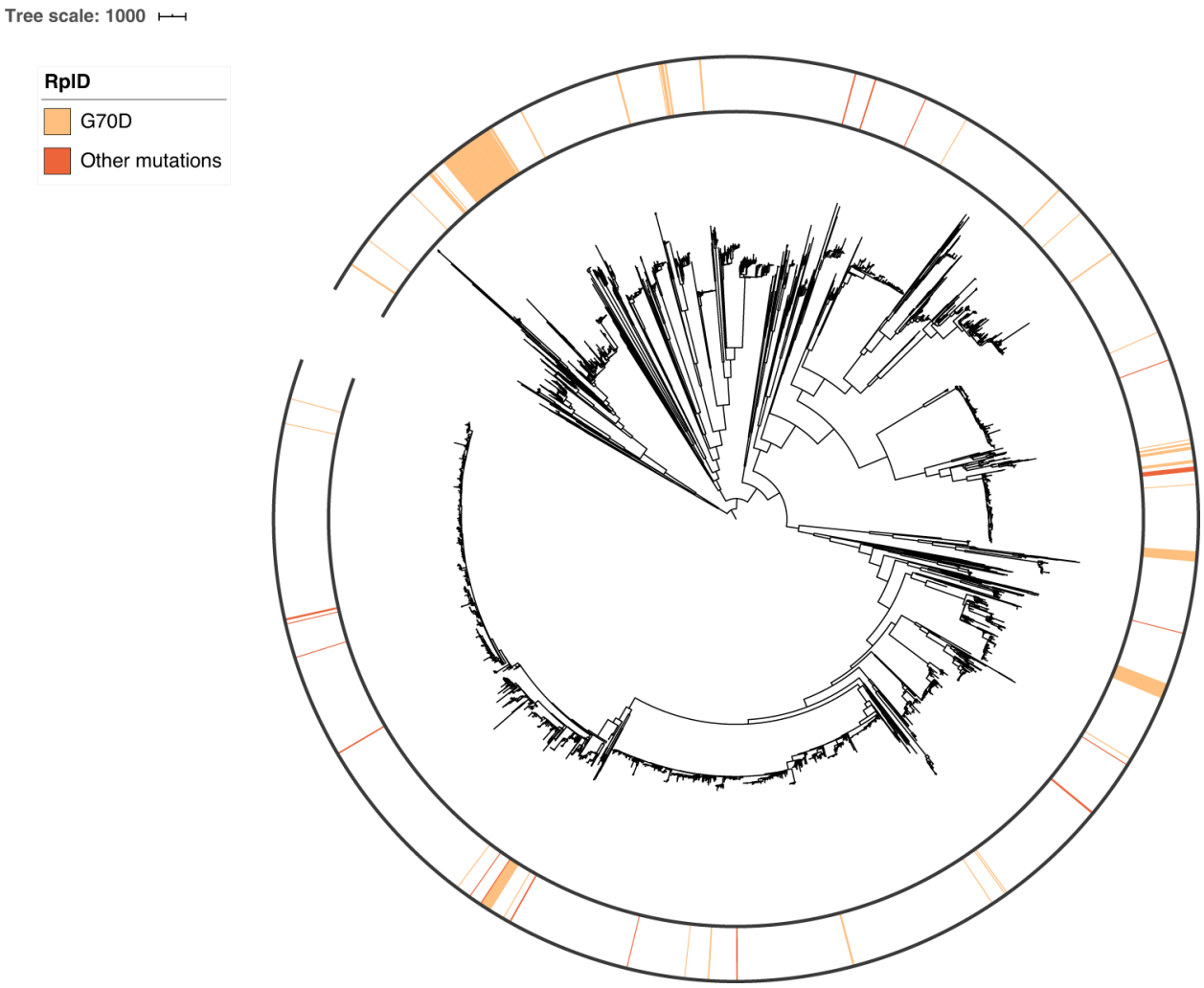


Figure 4.2 – Population structure of RplD binding site mutations in a global gonococcal meta-analysis dataset. A midpoint rooted recombination-corrected maximum likelihood phylogeny of 4852 genomes based on 68697 SNPs non-recombinant from Ma and Mortimer et al.[7] was annotated with the presence of RplD macrolide binding site mutations (orange for G70D and dark orange for other binding site mutations). Branch length represents total number of substitutions after removal of predicted recombination.

a

Mutation	Number (out of 4850 strains)	Geometric mean of azithromycin MIC
G70D[3,25]	231	0.41
G70S[36]	2	1.0
G70A	8	0.92
G70R	2	4.0
T69I[36]	4	0.52
G68D[3,26]	13	0.88
G68C	1	2.0
G68D and G70D	2	1.41
G68D and G70duplication	1	4
G70duplication	1	12

b

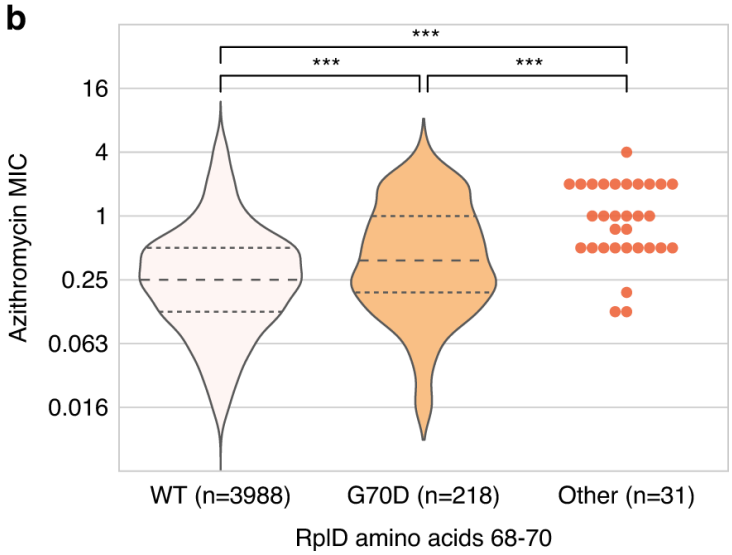


Figure 4.3 – Mean (a) and distribution (b) of azithromycin MICs for RplD macrolide binding site variants. Previously reported mutations are cited with the first reporting publications. Violin and swarm plots and statistical analyses were limited to isolates with MICs < 8 to exclude isolates with 23S rRNA mutations. Rare RplD mutations (denoted as "Other (n=31)") were grouped for visualization and statistical analysis, and thus were assumed to all have the same effect. Quartiles within violin plots are depicted using dotted lines. Statistical significance between RplD variants and RplD wildtype MIC distributions was assessed by two-sided Mann-Whitney U Test: * $p<0.05$, ** $p<0.01$, and *** $p<0.001$. Exact *p*-values from left to right were 3.09×10^{-7} (WT vs. G70D), 3.91×10^{-10} (WT vs. Other), 5.74×10^{-5} (G70D vs. Other).

Experimental validation and growth dynamics of RplD G70D strains.

To experimentally verify that RplD G70D contributes to macrolide resistance, we constructed an isogenic derivative of the laboratory strain 28BI with the G70D substitution (two biological replicates: C5 and E9) and tested for MIC differences across a panel of macrolides. Azithromycin and erythromycin MICs increased by three-fold, and clarithromycin MICs increased by six-fold on average in the G70D strains compared to the wild-type strain (Table 4.1). We also compared the experimental results with our modeling analysis: the estimate from our linear model for the azithromycin MIC of a strain that contains the RplD G70D mutation and no other resistance mutations was 0.363, which agrees well with the experimental results.

Macrolide resistance has been associated with a fitness cost in other species[30], prompting us to measure the *in vitro* growth dynamics of the RplD G70D strain. Time-course growth curves of the wild-type strain 28BI and isogenic G70D strain E9 were similar (Supplementary Figure S2.5) with overlapping estimates of doubling times: 28BI doubling time = 1.756 hours, 95% CI [1.663, 1.861] versus 28BI RplD^{G70D} (E9) doubling time = 1.787 hours, 95% CI [1.671, 1.920] (Supplementary Table S2.2). These results confirm the role of RplD G70D in mediating macrolide resistance and indicate a lack of severe associated *in vitro* fitness costs.

Table 4.1 – Macrolide MICs of laboratory strain 28BI and two isogenic derivatives confirms increased macrolide resistance conferred by RplD G70D.

Isolate	Azithromycin MIC (µg/mL)[a]	Clarithromycin MIC (µg/mL)[a]	Erythromycin MIC (µg/mL)[a]
28BI	0.094	0.25	0.38
28BI RplD^{G70D} (C5)	.25 (2.66×)	1.5 (6×)	1.5 (3.94×)
28BI RplD^{G70D} (E9)	.38 (4.04×)	1.5 (6×)	1.0 (2.63×)

[a]Fold change relative to baseline is shown in parentheses. MICs were measured once for each isogenic derivative using Etest strips placed onto GCB agar plates supplemented with 1% IsoVitaleX.

Discussion

Azithromycin resistance in *N. gonorrhoeae* is a polygenic trait involving contributions from mutations in different 50S ribosomal components, up- and down-regulation of efflux pump activity, and additional unknown factors (Supplementary Table S2.3). Genome-wide association methods offer one approach for uncovering the genotypic basis of unexplained resistance in clinical isolates, but novel causal genes associated with lower effects have been difficult to identify with traditional microbial GWAS approaches[24]. Our results indicate that extending the GWAS linear mixed model to incorporate known causal genetic variants could address some of these challenges, particularly when known genes exhibit strong penetrance and population

stratification, obfuscating signals with lower effects. After conducting conditional GWAS on azithromycin MICs, we observed a reduction in spurious results attributable to genetic linkage with known high-level resistance mutations in the 23S rRNA and an increase in power to recover secondary resistance mutations in the MtrCDE efflux pump. We also identified a resistance-associated mutation in the macrolide binding site of 50S ribosomal protein RplD as significant only after conditioning. These results are in line with studies of multi-locus methods in the human GWAS field showing increased power[31-33] and complementary methods using whole-genome elastic nets for microbial genome data[34,35].

The situations under which conditional GWAS improves the power to detect new causal genes will need to be further characterized in other bacterial species and through simulations[35]. Here, we observed both increased magnitude of effect (*rplD β* increased from 0.52 to 0.95) and increased model precision (*rplD* standard error decreased from 0.223 to 0.140) after conditioning, both of which could improve power. The success of this conditional analysis using a relatively small sample size compared to human GWAS studies may also be attributable to the degree of homoplasy, as RplD mutations have been acquired across the phylogeny (Figure 2). The degree of genetic linkage between true positives and the dominant resistance gene is likely key: if the degree of linkage is high (e.g., because a few bacterial lineages repeatedly acquire different mechanisms of resistance to the same drug), then a conditional GWAS could lead to a loss of power to detect the causal genes.

The role of RplD G70D mutations in conferring azithromycin resistance has previously been unclear, in part because of its lower effect size relative to 23S rRNA mutations. The G70D mutation was first observed in isolates from France 2013-2014[25] and in the US Centers for

Disease Control Gonococcal Isolate Surveillance Program (CDC GISP) surveillance isolates from 2000-2013[3], and a related G68D mutation was described in the GISP collection and in European isolates from 2009-2014[26]. However, these analyses reported no clear association with categorical resistance versus susceptibility. In line with this, we observed lower significance for the RplD unitig in the conditional GWAS model when the isolates were dichotomized into azithromycin susceptible versus non-susceptible (*p*-value = 3.38×10^{-09} versus 1.08×10^{-11} in the continuous case). Follow up studies in the US, Eastern China, and a historical Danish collection also reported strains with the G70D mutation[36-38], but other surveillance datasets from Canada, Switzerland, and Nanjing did not[10,39-41], indicating geography-specific circulation. As a result of this ambiguity, previous studies modeling phenotypic azithromycin resistance from genotype did not include RplD mutations[27,42].

Here, we provided confirmatory evidence that the RplD G70D mutation increases macrolide MICs several-fold and that inclusion of the mutation in resistance regression models improves model fit, in line with the GWAS analyses. While RplD G70D mutations on their own are not predicted to confer resistance levels above the clinical CLSI non-susceptibility threshold of 1.0 µg/mL, there is growing appreciation of the role that sub-breakpoint increases in resistance can play in mediating treatment failure[43]. For example, treatment failures in Japan after a 2 g azithromycin dose were associated with MICs as low as 0.5 µg/mL[44], and treatment failures in several case studies of patients treated with a 1 g azithromycin dose were associated with MICs of 0.125 to 0.25 µg/mL[45]. Low level azithromycin resistance may also serve as a stepping stone to higher level resistance, as suggested by an analysis of an outbreak of a high level azithromycin resistant *N. gonorrhoeae* lineage in the UK[46].

We also observed multiple previously undescribed mutations in the RplD macrolide binding site associated with even higher MICs than the G70D mutation. The transmission of these isolates has been relatively limited, potentially due to increased fitness costs commensurate with increased resistance. In contrast, several lines of evidence suggest that the G70D mutation carries a relatively minimal fitness cost. Time-course growth experiments indicated that the RplD G70D isogenic pair of strains have similar doubling times, and phylogenetic analyses suggest multiple acquisitions of G70D in distinct genetic backgrounds, with a lineage in New York City showing evidence of sustained transmission.

As macrolide use continues to select for increased resistance in *N. gonorrhoeae,* both the RplD G70D and rarer binding site mutations should be targets for surveillance in future whole-genome sequencing studies. The rapid increase in the prevalence of strains with mosaic *mtr* alleles conferring azithromycin reduced susceptibility underscores how quickly the molecular landscape of resistance can change[47,48] and highlights the value of early and proactive surveillance studies. Systematic genomic surveillance in turn allows for novel resistance mutations to be identified using conditional GWAS and other complementary approaches. With an increasingly refined understanding of the molecular basis of resistance, sequence-based diagnostics can then be developed by leveraging emerging point-of-care technologies such as Nanopore sequencing and CRISPR-based paper diagnostics[49,50]. The methods here can also be easily extended for other antibiotics used to treat gonococcal infections such as ceftriaxone, where resistance is of paramount concern and molecular mechanisms underlying resistance are still being uncovered[51].

In summary, by reducing genetic confounders and amplifying true signals through bacterial GWAS conditional on known effects, we identified and experimentally characterized mutations in the 50S ribosome that contribute to increased macrolide resistance in *N. gonorrhoeae*.

Methods

Genomics and GWAS. All isolates included in this study are listed in Supplementary Data 6. We conducted whole-genome sequencing assembly, resistance allele calling, phylogenetic inference, genome-wide association, and significant unitig mapping using methods from a prior GWAS[7]. Briefly, reads were downloaded using fastq-dump in SRA toolkit (version 2.8.1). We then created a recombination-corrected phylogeny by running Gubbins (version 2.3.4)[52] on an alignment of pseudogenomes generated from filtered SNPs from Pilon (version 1.16)[53] after mapping reads in BWA-MEM (version 0.7.17-r1188)[54] to the NCCP11945 reference genome (RefSeq accession: NC_011035.1). 23S rRNA mutations were called by mapping reads to a copy of the 23S rRNA locus and analyzing the frequency of variants[55]. Read mapping quality control was conducted in FastQC (version 0.11.7, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and BamQC in Qualimap (version 2.2.1)[56], and read deduplication was conducted using Picard (version 2.8.0, https://github.com/broadinstitute/picard). We also annotated assemblies with Prokka (version1.13)[57] and clustered core genes using Roary (version 3.12)[58].

All phylogenies and annotation rings were visualized in iTOL (version 5.5)[59]. As in the prior study, azithromycin MICs prior to 2005 from the CDC GISP dataset[3] were doubled to account for an MIC protocol testing change[60]. For analyses using susceptible versus non-susceptible categories as the outcome variable, isolates with adjusted azithromycin MICs of 1.0 μg/mL or lower were classified as susceptible.

We use a linear mixed model-based GWAS to control for population structure:

$$(\mathbf{1}) \quad \mathbf{Y} \sim \mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\beta + \mathbf{u} + \boldsymbol{\epsilon}$$

$$(\mathbf{2}) \quad \mathbf{u} \sim N\left(0, \sigma_g^2 \mathbf{K}\right)$$

$$(\mathbf{3}) \quad \boldsymbol{\epsilon} \sim N(0, \sigma_e^2 \mathbf{I})$$

Here, $\mathbf{Y}$ is the vector of azithromycin MICs, $\mathbf{W}$ is the covariate matrix and $\boldsymbol{\alpha}$ their fixed effects, $\mathbf{X}$ is the genetic variant / unitig under consideration and β its fixed effect, $\mathbf{u}$ is a random effect parameterized with population structure matrix $\mathbf{K}$ and additive genetic variance σ_g^2, and $\boldsymbol{\epsilon}$ is a random effect that models the non-genetic effects parameterized with variance σ_e^2 and identity matrix $\mathbf{I}$. This model is fit individually for all variants and the *p*-value for β is estimated using the likelihood-ratio test. The covariates can include isolate metadata such as country of origin or dataset as well as genetic information encoding known resistance genes.

To conduct the GWAS in Pyseer (version 1.2.0)[61], unitigs were generated from genomes assembled with SPAdes (version 3.12.0)[62] using GATB, and a population structure matrix was generated from the Gubbins phylogeny for the linear mixed model. We conducted conditional GWAS in Pyseer (version 1.2.0)[61] by including additional columns in the covariate file encoding the number of 23S rRNA mutations and including flags --covariates and --use-covariates.

We assessed genetic linkage by calculating r^2, or the squared correlation coefficient between two variants defined as $r^2 = (p_{ij} - p_i p_j)^2 / (p_i (1 - p_i)\, p_j (1 - p_j))$, where p_i is the proportion of strains with variant *i,* p_j is the proportion of strains with variant *j*, and p_{ij} is the proportion of strains with both variants[63,64]. For a given GWAS variant, we calculated r^2 between that variant and the significant unitig from the GWAS mapping to 23S rRNA C2611T. We repeated the calculation for the same variant but with the unitig mapping to 23S rRNA A2059G, and took the maximum r^2 value from the two calculations.

Azithromycin MIC regression models. Azithromycin log-transformed MICs were modeled using a panel of resistance markers[7,65] and country of origin in R (version 3.5.1) using the lm function, with and without inclusion of RplD G70D and proximal mutations:

Model 1: Log_AZI ~ Country + MtrR 39 + MtrR 45 + MtrR LOF + MtrC LOF + MtrR promoter + *mtrCDE* BAPS Cluster + 23S rRNA 2059 + 23S rRNA 2611

Model 2: Log_AZI ~ Country + MtrR 39 + MtrR 45 + MtrR LOF + MtrC LOF + MtrR promoter + *mtrCDE* BAPS Cluster + 23S rRNA 2059 + 23S rRNA 2611 + RplD G70D + RplD other 68-70 mutations

Improvement in model fit was assessed using Anova for likelihood-ratio tests for nested models in R (version 3.5.1). BAPS clusters for *mtrCDE* were called as previously described using FastBAPS (version 1.0.0) as a way to flexibly group resistance-conferring mosaic alleles for inclusion in the regression model as a categorical covariate[7,20]. Variance explained by predictors was calculated using the relaimpo R package (version 2.2.3), which assesses the change in model R2 after inclusion of a predictor. Three approaches were used to calculate this change: the “first” metric compares a model without any predictors to a model with just the predictor of interest, the “last” metric compares a model with all predictors except the one of interest to a model with all predictors, and the “lmg” method averages the change in R^2 over all possible model subsets.

Diversity of RplD macrolide binding site mutations. We ran BLASTn (version 2.6.0)[66] on the *de novo* assemblies using a query *rplD* sequence from FA1090 (RefSeq accession: NC_002946.2). *rplD* sequences were aligned using MAFFT (version 7.450)[67]. Binding site mutations were identified after *in silico* translation of nucleotide alignments in Geneious Prime (version 2019.2.1, https://www.geneious.com). Subsequent analyses identifying prevalence, geometric mean azithromycin MIC, and MIC distribution differences were conducted in Python (version 3.6.5) using the Biopython package (version 1.69)[68] and R (version 3.5.1).

Experimental validation. We cultured *N. gonorrhoeae* on GCB agar (Difco) plates supplemented with 1% Kellogg's supplements (GCBK) at 37°C in a 5% CO2 incubator[69]. We conducted antimicrobial susceptibility testing using Etests (bioMérieux) placed onto GCB agar plates supplemented with 1% IsoVitaleX (Becton Dickinson). We selected laboratory strain 28BI for construction of isogenic strains and measured its MIC for azithromycin, clarithromycin, and erythromycin[20]. *rplD* encoding the G70D mutation was PCR amplified from RplD G70D isolate GCGS1043[3] using primers rplD_FWD_DUS (5' CATGCCGTCTGAACAAGACCCGGGTCGCG 3') (containing a DUS tag to enhance transformation[70]) and rplD_REV (5' TTCAGAAACGACAGGCGCC 3'). The resulting ~1 kb amplicon was spot transformed[69] into 28BI. We selected for transformants by plating onto GCBK plates with clarithromycin 0.4 µg/mL and erythromycin 0.4 µg/mL. We confirmed via Sanger sequencing that transformants had acquired the RplD G70D mutation and selected one transformant from each selection condition (strain C5 for clarithromycin and strain E9 for erythromycin) for further characterization. We confirmed that for all macrolides used for selection, no spontaneous resistant mutants were observed after conducting control transformations in the absence of GCGS1043 PCR product. We did not construct strains with the mutation complemented.

Growth assays. We streaked 28BI and 28BI RplD^{G70D} (E9) onto GCBK plates and grew them overnight for 16 hours at 37°C in a 5% CO2 atmosphere. We prepared 1 L of fresh Graver Wade (GW) media[71] and re-suspended overnight cultures into 1 mL of GW. After normalizing cultures to OD 0.1, we diluted cultures 1:10^5 and inoculated central wells of a 24-well plate with 1.5 mL GW and cells in triplicate. Edge wells were filled with 1.5 mL water. After growth for 1 hour to acclimate to media conditions, we sampled CFUs every 2 hours for a total of 12 hours. For each timepoint, we aspirated using a P1000 micropipette to dissolve clumps and then plated

serial dilutions onto a GCBK plate. We counted CFUs the following day and used GraphPad Prism (version 8.2.0 for Windows, GraphPad Software) to graph the data and estimate exponential phase growth rates following removal of lag phase data points and log-transformation of CFUs / mL.

Data availability

In Supplementary Data S2.6, we have included accession numbers (via publicly hosted database NCBI SRA) for accessing all raw sequence data used for *N. gonorrhoeae* analyses. Intermediate outputs from the genomics pipeline (e.g., *de novo* assemblies) may also be available from the authors upon request. An interactive and downloadable version of the phylogeny and annotation rings used in Figure 4.2 and Supplementary Figure S2.1 is hosted at https://itol.embl.de/tree/12810322453513515973382 46. Source data are available at https://github.com/gradlab/rplD-conditional-GWAS (doi:10.5281/zenodo.4042334)[72].

Code availability

Code to reproduce the analyses and figures is available at https://github.com/gradlab/rplD-conditional-GWAS (doi:10.5281/zenodo.4042334)[72] or from the authors upon request.

References

1 Unemo, M. & Shafer, W. M. Antimicrobial resistance in Neisseria gonorrhoeae in the 21st century: past, evolution, and future. *Clin Microbiol Rev* **27**, 587-613, doi:10.1128/CMR.00010-14 (2014).

2 Wi, T. *et al.* Antimicrobial resistance in Neisseria gonorrhoeae: Global surveillance and a call for international collaborative action. *PLoS Med* **14**, e1002344, doi:10.1371/journal.pmed.1002344 (2017).

3 Grad, Y. H. *et al.* Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United States, 2000-2013. *J Infect Dis* **214**, 1579-1587, doi:10.1093/infdis/jiw420 (2016).

4 Fifer, H., Saunders, J., Soni, S., Sadiq, S. T. & Fitzgerald, M. *British Association for Sexual Health and HIV national guideline for the management of infection with Neisseria gonorrhoeae*, <https://www.bashhguidelines.org/media/1208/gc-2019.pdf> (2019).

5 Allan-Blitz, L.-T. *et al.* A Cost Analysis of Gyrase A Testing and Targeted Ciprofloxacin Therapy Versus Recommended 2-Drug Therapy for Neisseria gonorrhoeae Infection. *Sex. Transm. Dis.* **45**, 87-91, doi:10.1097/OLQ.0000000000000698 (2018).

6 Tuite, A. R. *et al.* Impact of Rapid Susceptibility Testing and Antibiotic Selection Strategy on the Emergence and Spread of Antibiotic Resistance in Gonorrhea. *J. Infect. Dis.* **216**, 1141-1149, doi:10.1093/infdis/jix450 (2017).

7 Ma, K. C. *et al.* Adaptation to the cervical environment is associated with increased antibiotic susceptibility in Neisseria gonorrhoeae. *Nat Commun* **11**, 4126, doi:10.1038/s41467-020-17980-1 (2020).

8 Hicks, A. L. *et al.* Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. *PLoS Comput Biol* **15**, e1007349, doi:10.1371/journal.pcbi.1007349 (2019).

9 Gaynor, M. & Mankin, A. S. Macrolide antibiotics: binding site, mechanism of action, resistance. *Curr Top Med Chem* **3**, 949-961, doi:10.2174/1568026033452159 (2003).

10 Ng, L. K., Martin, I., Liu, G. & Bryden, L. Mutation in 23S rRNA associated with macrolide resistance in Neisseria gonorrhoeae. *Antimicrob Agents Chemother* **46**, 3020-3025, doi:10.1128/aac.46.9.3020-3025.2002 (2002).

11 Zhang, J. & van der Veen, S. Neisseria gonorrhoeae 23S rRNA A2059G mutation is the only determinant necessary for high-level azithromycin resistance and improves in vivo biological fitness. *J. Antimicrob. Chemother.* **74**, 407-415, doi:10.1093/jac/dky438 (2019).

12 Rouquette, C., Harmon, J. B. & Shafer, W. M. Induction of the mtrCDE-encoded efflux pump system of Neisseria gonorrhoeae requires MtrA, an AraC-like protein. *Mol Microbiol* **33**, 651-658, doi:10.1046/j.1365-2958.1999.01517.x (1999).

13 Shafer, W. M. *et al.* in *National Institute of Allergy and Infectious Diseases, NIH* 55-63 (Humana Press, 2008).

14 Veal, W. L., Nicholas, R. A. & Shafer, W. M. Overexpression of the MtrC-MtrD-MtrE efflux pump due to an mtrR mutation is required for chromosomally mediated penicillin resistance in Neisseria gonorrhoeae. *J Bacteriol* **184**, 5619-5624, doi:10.1128/jb.184.20.5619-5624.2002 (2002).

15 Warner, D. M., Folster, J. P., Shafer, W. M. & Jerse, A. E. Regulation of the MtrC-MtrD-MtrE efflux-pump system modulates the in vivo fitness of Neisseria gonorrhoeae. *J Infect Dis* **196**, 1804-1812, doi:10.1086/522964 (2007).

16 Warner, D. M., Shafer, W. M. & Jerse, A. E. Clinically relevant mutations that cause derepression of the Neisseria gonorrhoeae MtrC-MtrD-MtrE Efflux pump system confer different levels of antimicrobial resistance and in vivo fitness. *Mol Microbiol* **70**, 462-478, doi:10.1111/j.1365-2958.2008.06424.x (2008).

17 Zalucki, Y. M., Dhulipala, V. & Shafer, W. M. Dueling regulatory properties of a transcriptional activator (MtrA) and repressor (MtrR) that control efflux pump gene expression in Neisseria gonorrhoeae. *MBio* **3**, e00446-00412, doi:10.1128/mBio.00446-12 (2012).

18 Zarantonelli, L., Borthagaray, G., Lee, E.-H. & Shafer, W. M. Decreased Azithromycin Susceptibility ofNeisseria gonorrhoeae Due to mtrRMutations. *Antimicrob. Agents Chemother.* **43**, 2468-2472 (1999).

19 Rouquette-Loughlin, C. E. *et al.* Mechanistic Basis for Decreased Antimicrobial Susceptibility in a Clinical Isolate of Neisseria gonorrhoeae Possessing a Mosaic-Like mtr Efflux Pump Locus. *MBio* **9**, doi:10.1128/mBio.02281-18 (2018).

20 Wadsworth, C. B., Arnold, B. J., Sater, M. R. A. & Grad, Y. H. Azithromycin Resistance through Interspecific Acquisition of an Epistasis-Dependent Efflux Pump Component and Transcriptional Regulator in Neisseria gonorrhoeae. *MBio* **9**, doi:10.1128/mBio.01419-18 (2018).

21 Cousin, S. L., Jr., Whittington, W. L. H. & Roberts, M. C. Acquired macrolide resistance genes and the 1 bp deletion in the mtrR promoter in Neisseria gonorrhoeae. *J. Antimicrob. Chemother.* **51**, 131-133 (2003).

22 Lees, J. A. *et al.* Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat Commun* **7**, 12797, doi:10.1038/ncomms12797 (2016).

23 Diner, E. J. & Hayes, C. S. Recombineering reveals a diverse collection of ribosomal proteins L4 and L22 that confer resistance to macrolide antibiotics. *J Mol Biol* **386**, 300-315, doi:10.1016/j.jmb.2008.12.064 (2009).

24 Wheeler, N. E. *et al.* Contrasting approaches to genome-wide association studies impact the detection of resistance mechanisms in Staphylococcus aureus. *bioRxiv*, doi:10.1101/758144 (2019).

25 Belkacem, A. *et al.* Molecular epidemiology and mechanisms of resistance of azithromycin-resistant Neisseria gonorrhoeae isolated in France during 2013-14. *J Antimicrob Chemother* **71**, 2471-2478, doi:10.1093/jac/dkw182 (2016).

26 Jacobsson, S. *et al.* WGS analysis and molecular resistance mechanisms of azithromycin-resistant (MIC >2 mg/L) Neisseria gonorrhoeae isolates in Europe from 2009 to 2014. *J Antimicrob Chemother* **71**, 3109-3116, doi:10.1093/jac/dkw279 (2016).

27 Demczuk, W. *et al.* Equations to predict antimicrobial minimum inhibitory concentrations in Neisseria gonorrhoeae using molecular antimicrobial resistance determinants. *Antimicrob. Agents Chemother.*, doi:10.1128/AAC.02005-19 (2019).

28 Mortimer, T. D. *et al.* The distribution and spread of susceptible and resistant Neisseria gonorrhoeae across demographic groups in a major metropolitan center. *Clin Infect Dis*, doi:10.1093/cid/ciaa1229 (2020).

29 Yahara, K. *et al.* Genomic surveillance of Neisseria gonorrhoeae to investigate the distribution and evolution of antimicrobial-resistance determinants and lineages. *Microb Genom* **4**, doi:10.1099/mgen.0.000205 (2018).

30 Zeitouni, S., Collin, O., Andraud, M., Ermel, G. & Kempf, I. Fitness of macrolide resistant Campylobacter coli and Campylobacter jejuni. *Microb. Drug Resist.* **18**, 101-108, doi:10.1089/mdr.2011.0188 (2012).

31 Ma, L., Han, S., Yang, J. & Da, Y. Multi-locus test conditional on confirmed effects leads to increased power in genome-wide association studies. *PLoS One* **5**, e15006, doi:10.1371/journal.pone.0015006 (2010).

32 Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* **44**, 825-830, doi:10.1038/ng.2314 (2012).

33 Bennett, D., Morris, D. & Seoighe, C. Controlling for off-target genetic effects using polygenic scores improves the power of genome-wide association studies. *bioRxiv* (2020).

34 Lees, J. A. T. M., T; Galardini, Marco; Wheeler, Nicole E; Corander, Jukka. Improved inference and prediction of bacterial genotype-phenotype associations using pangenome-spanning regressions. *bioRxiv*, doi:10.1101/852426 (2019).

35 Saber, M. & Shapiro, B. Benchmarking bacterial genome-wide association study methods using simulated genomes and phenotypes. *Microbial Genomics*, doi:doi:10.1099/mgen.0.000337 (2020).

36 Zheng, Z. *et al.* Antimicrobial Resistance And Molecular Characteristics Among Neisseria gonorrhoeae Clinical Isolates In A Chinese Tertiary Hospital. *Infect Drug Resist* **12**, 3301-3309, doi:10.2147/IDR.S221109 (2019).

37 Thomas, J. C. *et al.* Evidence of Recent Genomic Evolution in Gonococcal Strains with Decreased Susceptibility to Cephalosporins or Azithromycin in the United States, 2014-2016. *J. Infect. Dis.*, doi:10.1093/infdis/jiz079 (2019).

38 Golparian, D. *et al.* Genomic evolution of Neisseria gonorrhoeae since the preantibiotic era (1928-2013): antimicrobial use/misuse selects for resistance and drives evolution. *BMC Genomics* **21**, 116, doi:10.1186/s12864-020-6511-6 (2020).

39 Demczuk, W. *et al.* Genomic Epidemiology and Molecular Resistance Mechanisms of Azithromycin-Resistant Neisseria gonorrhoeae in Canada from 1997 to 2014. *J Clin Microbiol* **54**, 1304-1313, doi:10.1128/JCM.03195-15 (2016).

40 Endimiani, A. *et al.* Characterization of Neisseria gonorrhoeae isolates detected in Switzerland (1998-2012): emergence of multidrug-resistant clones less susceptible to cephalosporins. *BMC Infect Dis* **14**, 106, doi:10.1186/1471-2334-14-106 (2014).

41 Wan, C. *et al.* Increasing resistance to azithromycin ofNeisseria gonorrhoeaein eastern Chinese cities: mechanisms and genetic diversity of resistant Nanjing isolates. *Antimicrob. Agents Chemother.*, doi:10.1128/AAC.02499-17 (2018).

42 Harrison, O. B. *et al.* Genomic analyses of Neisseria gonorrhoeae reveal an association of the gonococcal genetic island with antimicrobial resistance. *J Infect* **73**, 578-587, doi:10.1016/j.jinf.2016.08.010 (2016).

43 Colangeli, R. *et al.* Bacterial Factors That Predict Relapse after Tuberculosis Therapy. *N Engl J Med* **379**, 823-833, doi:10.1056/NEJMoa1715849 (2018).

44 Yasuda, M., Ito, S., Hatazaki, K. & Deguchi, T. Remarkable increase of Neisseria gonorrhoeae with decreased susceptibility of azithromycin and increase in the failure of azithromycin therapy in male gonococcal urethritis in Sendai in 2015. *J Infect Chemother* **22**, 841-843, doi:10.1016/j.jiac.2016.07.012 (2016).

45 Tapsall, J. W. *et al.* Failure of azithromycin therapy in gonorrhea and discorrelation with laboratory test parameters. *Sex. Transm. Dis.* **25**, 505-508, doi:10.1097/00007435-199811000-00002 (1998).

46 Fifer, H. *et al.* Sustained transmission of high-level azithromycin-resistant Neisseria gonorrhoeae in England: an observational study. *Lancet Infect. Dis.*, doi:10.1016/S1473-3099(18)30122-1 (2018).

47 Williamson, D. A. *et al.* Bridging of Neisseria gonorrhoeae lineages across sexual networks in the HIV pre-exposure prophylaxis era. *Nat Commun* **10**, 3988, doi:10.1038/s41467-019-12053-4 (2019).

48 Gernert, K. M. S., Sandra; Schmerer, Matthew W; Thomas IV, Jesse C; Pham, Cau D; St Cyr, Sancta; Schlanger, Karen; Weinstock, Hillard; Shafer, William M; Raphael, Brian H; Kersh, Ellen N; Hun, Sopheay; Hua, Chi; Ruiz, Ryan; Soge, Olusegun O; Dominguez, Catherine; Patel, Ami; Loomis, Jillian; Leavitt, John; Zhang, Jenny; Baldwin, Tamara; Wang, Chun; Moore, Christina; Whelen, Christian; O'Brien, Pamela; Harvey, Alesia. Azithromycin susceptibility of Neisseria gonorrhoeae in the USA in 2017: a genomic analysis of surveillance data. *The Lancet Microbe* **1**, e154-e164, doi:10.1016/S2666-5247(20)30059-8 (Aug 01, 2020).

49 Břinda, K. *et al.* Rapid heuristic inference of antibiotic resistance and susceptibility by genomic neighbor typing. *bioRxiv*, doi:10.1101/403204 (2019).

50 Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438-442, doi:10.1126/science.aam9321 (2017).

51 Palace, S. G. *et al.* RNA polymerase mutations cause cephalosporin resistance in clinical Neisseria gonorrhoeae isolates. *Elife* **9**, doi:10.7554/eLife.51407 (2020).

52 Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**, e15, doi:10.1093/nar/gku1196 (2015).

53 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).

54 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint* **1303.3997** (2013).

55 Johnson, S. R., Grad, Y., Abrams, A. J., Pettus, K. & Trees, D. L. Use of whole-genome sequencing data to analyze 23S rRNA-mediated azithromycin resistance. *Int J Antimicrob Agents* **49**, 252-254, doi:10.1016/j.ijantimicag.2016.10.023 (2017).

56 Garcia-Alcalde, F. *et al.* Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**, 2678-2679, doi:10.1093/bioinformatics/bts503 (2012).

57 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).

58 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691-3693, doi:10.1093/bioinformatics/btv421 (2015).

59 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-W259, doi:10.1093/nar/gkz239 (2019).

60 Kersh, E. N. *et al.* Rationale for a Neisseria gonorrhoeae Susceptible Only Interpretive Breakpoint for Azithromycin. *Clin Infect Dis*, doi:10.1093/cid/ciz292 (2019).

61 Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**, 4310-4312, doi:10.1093/bioinformatics/bty539 (2018).

62 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

63 Arnold, B. *et al.* Fine-scale haplotype structure reveals strong signatures of positive selection in a recombining bacterial pathogen. *Mol. Biol. Evol.*, doi:10.1093/molbev/msz225 (2019).

64 Hill, W. G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**, 226-231, doi:10.1007/BF01245622 (1968).

65 Demczuk, W. *et al.* Equations To Predict Antimicrobial MICs in Neisseria gonorrhoeae Using Molecular Antimicrobial Resistance Determinants. *Antimicrob Agents Chemother* **64**, doi:10.1128/AAC.02005-19 (2020).

66 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

67 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

68 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423, doi:10.1093/bioinformatics/btp163 (2009).

69 Dillard, J. P. Genetic Manipulation of Neisseria gonorrhoeae. *Curr Protoc Microbiol* **Chapter 4**, Unit4A 2, doi:10.1002/9780471729259.mc04a02s23 (2011).

70 Ambur, O. H., Frye, S. A. & Tønjum, T. New functional identity for the DNA uptake sequence in transformation and its presence in transcriptional terminators. *J. Bacteriol.* **189**, 2077-2085, doi:10.1128/JB.01408-06 (2007).

71 Wade, J. J. & Graver, M. A. A fully defined, clear and protein-free liquid medium permitting dense growth of Neisseria gonorrhoeae from very low inocula. *FEMS Microbiol. Lett.* **273**, 35-37, doi:10.1111/j.1574-6968.2007.00776.x (2007).

72 Ma, K. C. gradlab/rplD-conditional-gwas 1.0.0 (Version 1.0.0). *Zenodo*, doi:http://doi.org/10.5281/zenodo.4042334 (2020).

Chapter 5: Concluding remarks

Concluding remarks

Here, we have used genetic association studies and genomic epidemiology to investigate causal determinants of antibiotic MIC variation in gonococci. In Chapter 2, we demonstrated that truncations in the *mtrC* component of the MtrCDE efflux pump are the primary mechanism for acquired antibiotic susceptibility to multiple antibiotics. Our results suggest that efflux pump loss may be adaptive for cervical colonization in pathogenic *Neisseria*, but further experimental work in appropriate *in vitro* or animal models will be necessary to fully characterize this fitness benefit and to understand the molecular mechanisms underlying it. In addition to deepening our understanding of how niche and drug selective pressures interact, these results also raise the possibility of targeting the MtrCDE pump with small molecules, phages, or other therapeutic agents to revert antibiotic resistant strains back to sensitivity[1].

In Chapter 3, we identified strains with rare mutations in MtrD conferring azithromycin non-susceptibility, highlighting another genetic pathway clinical isolates have taken towards resistance. To enhance the power of GWAS to detect these and other secondary resistance mutations, we extended the GWAS LMM to condition on known causal genetic variants in Chapter 4; the key idea is to explicitly account for the effect of known resistance genes, which typically exhibit strong penetrance and population stratification and obfuscate signals with lower effects. We observed a reduction in spurious results attributable to genetic linkage and an increase in power to recover resistance mutations in ribosomal proteins. Our findings highlight the ability of conditional microbial GWAS to identify causal variants for polygenic bacterial phenotypes, similar to studies using multi-locus methods in the human GWAS field[2, 3]. In support of the utility of GWAS for diagnostics development, adding the variants discovered in

Chapters 2 and 4 (e.g., *mtrC* LOF and *rplD* mutations) to our linear regression models significantly decreased the amount of unexplained variability in MICs.

The results of the studies presented here raise several further avenues for investigation. The linear regression models of MICs including the GWAS variants still exhibit unexplained variance, particularly for azithromycin and ceftriaxone. The simplest explanation is that this missing heritability is an artifact due to laboratory noise or error; it is commonly accepted that MICs can vary by up to one doubling dilution depending on testing conditions and methods[4], and sample mix-ups can occasionally occur leading to mismatches between genome sequences and phenotypic metadata. While meta-analysis datasets such as the one we have used here are powerful due to the number of strains and the genetic diversity present, careful curation, adjustment for variable phenotyping methods, and the development of methods to detect outlier strains will further facilitate working with such cohorts. Another explanation for the missing heritability is that the linear models we use to model MICs fail to capture interactions and non-linear effects across resistance genes[5, 6]. While including these dimensions will likely increase the proportion of variability explained and, at least in some ways, better reflect reality compared to a strictly linear model, careful experimental testing will be needed to validate these additional model features. The final explanation is that there are indeed more causal genes to be detected, and our current GWAS methods are simply not robust or powerful enough to uncover them.

There are several scenarios in which the GWAS framework that we employ could fail. The first is that the causal mutations are simply too rare to be detected. This was the case with the mutations we investigated in Chapter 3 (see also work by Palace et al.[7]), where only a handful of strains had acquired these resistance mutations at the time of the study, requiring us to rely

on data from structural biology studies to guide our genomic investigations. These rare variants are difficult to uncover via GWAS and also have the potential to cause false positive associations[8], but are also some of the resistance mutations we are most interested in monitoring because they have not yet disseminated through a population. One promising direction for improving microbial GWAS here is to adopt rare-variant testing approaches from human GWAS studies[9]. The simplest method is to conduct a "burden" test, which collapses variants in some window or gene together as one statistical unit to be tested; further adjustments can be made depending on whether mutations are synonymous or nonsynonymous, whether amino acid changes are biochemically similar or not, and so on. In Chapter 4, we identified a prevalent RplD macrolide resistance variant using GWAS and through targeted gene analysis, we also identified rarer mutations adjacent to this variant that were also associated with increased resistance. By grouping all of these variants into a single unit, we showed that the association of this "grouped variant" with resistance was more significant than for any of the individual variants, demonstrating the power of the burden test approach. One drawback is that rare variant testing methods have generally relied on mapping-based variant calling instead of *de novo* assemblies, and merging these two strategies – e.g., conducting burden tests on pangenomes to account for indels and the accessory genome – remains an open technical challenge.

Another limitation to our current GWAS methodologies is that variants are generally tested independently of one another; in Chapter 2, regression was run independently for each variant, and in our conditional GWAS in Chapter 4, we only conditioned on a small subset of known resistance variants. Epistatic interactions between co-varying sites across the genome are theorized to be a substantial contributor to antibiotic resistance[10]. The challenge for methods designed to detect epistatic interactions is the considerable space of possible combinations of

variants that could be tested, and as always, care must be taken to adjust for population structure and other genetic confounders (as in Chapter 4). Machine learning methods such as regularized regression and neural nets also are promising, but will need to be critically assessed for their ability to identify causal markers if the goal is to understand the biology underlying resistance, and not just to optimize for predictive performance[6].

In these studies, we have primarily analyzed variant associations to each antibiotic independently. To improve the power of GWAS for complex traits, human genetic association studies sometimes jointly analyze multiple correlated and related phenotypes[11]. Similar approaches could be extended to microbial association studies: multiple phenotype analyses could identify markers associated with resistance to broad classes of drugs, such as the beta-lactams or macrolides, or identify mechanisms involved generally in multi-drug resistance or susceptibility. These approaches may also provide insight into enabling or compensatory mutations that can play a role in facilitating the emergence of antibiotic resistance broadly in *N. gonorrhoeae*[12]*.* As the characterized genomic diversity of gonococcus continues to expand, improved GWAS methods will enable the identification of emerging resistance pathways and drive forward both our basic understanding of gonococcal biology and the development of next-generation diagnostics.

References

1. Chan BK, Sistrom M, Wertz JE, Kortright KE, Narayan D, Turner PE. Phage selection restores antibiotic sensitivity in MDR Pseudomonas aeruginosa. *Sci Rep* **6**, 26717 (2016).

2. Ma L, Han S, Yang J, Da Y. Multi-locus test conditional on confirmed effects leads to increased power in genome-wide association studies. *PLoS One* **5**, e15006 (2010).

3. Segura V*, et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* **44**, 825-830 (2012).

4. EUCAST. Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. *Clin Microbiol Infect* **9**, ix-xv (2003).

5. Nguyen M*, et al.* Using Machine Learning To Predict Antimicrobial MICs and Associated Genomic Features for Nontyphoidal Salmonella. *J Clin Microbiol* **57**, (2019).

6. Lees JA, Tien Mai T, Galardini M, Wheeler NE, Corander J. Improved inference and prediction of bacterial genotype-phenotype associations using pangenome-spanning regressions. *bioRxiv*, (2019).

7. Palace SG*, et al.* RNA polymerase mutations cause cephalosporin resistance in clinical Neisseria gonorrhoeae isolates. *Elife* **9**, (2020).

8. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* **8**, e1000294 (2010).

9. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**, 5-23 (2014).

10. Pensar J*, et al.* Genome-wide epistasis and co-selection study using mutual information. *bioRxiv*, (2019).

11. Deng X, Wang B, Fisher V, Peloso G, Cupples A, Liu CT. Genome-wide association study for multiple phenotype analysis. *BMC Proc* **12**, 55 (2018).

12. Rubin DHF, Ma KC, A WK, Hullahalli K, Waldor MK, Grad YH. Variation in supplemental carbon dioxide requirements defines lineage-specific antibiotic resistance acquisition in Neisseria gonorrhoeae. *bioRxiv*, 2022.2002.2024.481660 (2022).

Appendix 1: Supplementary materials accompanying Chapter 2

Supplementary Tables

Supplementary Figures

Supplementary Figure S1.1 – Diagnostic Q-Q plots of expected versus observed p-values for GWAS on a) azithromycin, b) ceftriaxone, and c) ciprofloxacin.

Supplementary Figure S1.2 – Alignment of nucleotide sequences for strains with representative LOF mutations observed in a) *mtrC*, b) *mtrA*, and c) *farA* in the global dataset.

Supplementary Figure S1.3 – Phylogenetic distribution of gonococcal *mtrC*, *mtrA*, and *farA* LOF alleles with patient site of infection (n=2742) in global dataset.

Supplementary Figure S1.4 – MIC distributions for isolates with indicated resistance determinants for azithromycin, ceftriaxone, and ciprofloxacin respectively, stratified by *mtrC* genotypic status.

Supplementary Figure S1.5 – Gonococcal *mtrC* and *farA* LOF mutations are associated with sexual behavior and site of infection in the validation dataset.

Supplementary Figure S1.6 – Phylogenetic distribution of gonococcal *mtrC* LOF alleles with patient site of infection (n=2186) in the validation dataset.

Supplementary References

Supplementary Table S1.1 – Datasets included in global meta-analysis collection.

Publication	Study summary	Timespan	Included number
Mortimer et al., 2020[1]	Transmission and AMR surveillance in New York City, USA	2011-2015	888
Sánchez-Busó et al., 2019[2]	Worldwide phylogeography and evolution of gonococcus	1979-2012	408
Yahara et al., 2018[3]	AMR surveillance in Kyoto and Osaka, Japan	1996-2015	260
Ryan et al., 2018[4]	AMR surveillance in Ireland	2012-2016	39
Harris et al., 2018[5]	Genomic survey across European Euro-GASP participant countries (n=20)	2013	1048
Fifer et al., 2018[6]	High-level azithromycin resistance outbreak in UK	2004-2017	50
Lee et al., 2018[7]	Genomic epidemiology in New Zealand	2014-2015	397
Kwong et al., 2017[8]	Transmission among MSM in Melbourne, Australia	2005-2014	94
Eyre et al., 2017[9] and De Silva et al., 2016[10]	Transmission in Brighton, UK	2004-2011	231
Grad et al., 2016[11] and 2014[12]	AMR surveillance across CDC GISP clinics, USA	2000-2013	1100
Demczuk et al., 2016[13]	Azithromycin resistance surveillance in Canada	1991-2014	199
Demczuk et al., 2015[14]	Cephalosporin decreased susceptibility surveillance in Canada	1989-2013	114
Ezewudo et al., 2015[15]	Population structure and AMR surveillance	1982-2011	54

All isolates that passed genomics quality control filters (see methods) with associated azithromycin, or ceftriaxone, or ciprofloxacin metadata were included (n=4852 in total). Euro-GASP = European Gonococcal Antimicrobial Surveillance Program; CDC GISP = Centers for Disease Control and Prevention Gonococcal Isolate Surveillance Project.

Supplementary Table S1.2 – Restoration of the *mtrC* coding frame in a clinical isolate by transformation increases MICs to azithromycin, ceftriaxone, and ciprofloxacin.

Strain	Azithromycin	Ceftriaxone	Ciprofloxacin
NY0195	0.064, 0.094, 0.094	0.016, 0.023, 0.023	6, 6, 4
NY0195 *mtrC*(in-frame)	0.5, 0.75, 0.75	0.064, 0.064, 0.064	≥ 32, ≥ 32, ≥ 32
	p = 0.0187	p = 0.00289	p = 0.000624

MIC Etests were conducted in triplicate with all results reported. Statistical significance for MIC differences between parental strain and transformant strain was assessed by a two-sample t-test after setting the value of "≥ 32" to 32.

Supplementary Table S1.3 – Prevalence of predicted LOF mutations in efflux pump genes for both global and Australian datasets.

Dataset	Gene	N LOF	N Total	Percentage
Global	*farA*	332	4838	6.86%
Global	*farB*	2	4850	0.04%
Global	*norM*	2	4852	0.04%
Global	*macA*	1	4847	0.02%
Global	*macB*	13	4845	0.27%
Global	*mtrR*	386	4845	7.97%
Global	*mtrA*	268 (362)	4842	5.45% (7.48%)
Global	*mtrC*	185	4847	3.82%
Global	*mtrD*	10	4815	0.21%
Global	*mtrE*	0	4849	0.00%
Australia	*farA*	225	2180	10.32%
Australia	*farB*	1	2186	0.05%
Australia	*norM*	0	2186	0.00%
Australia	*macA*	0	2186	0.00%
Australia	*macB*	0	2186	0.00%
Australia	*mtrA*	4 (85)	2186	0.18% (3.89%)
Australia	*mtrR*	253	2183	11.59%
Australia	*mtrC*	35	2186	1.60%
Australia	*mtrD*	2	2185	0.09%
Australia	*mtrE*	0	2186	0.00%

Counts for *mtrA* LOF mutations in the absence of other epistatic *mtrCDE* overexpression mutations are listed first, with counts for total number of *mtrA* LOF mutations regardless of *mtrCDE* overexpression status listed in parentheses.

Supplementary Table S1.4 – Association between sexual behavior and *mtrC* LOF in global dataset. *mtrC* LOF is significantly associated with sexual behavior (p = 0.04021, Two-Sided Fisher's Exact Test for Count Data).

mtrC	**WSM**	**MSW**	**MSMW**	**MSM**
Intact	30	598	123	1158
LOF	3	28	4	31

Supplementary Table S1.5 – Association between site of infection and mtrC LOF in global dataset. *mtrC* LOF is significantly associated with site of infection ($p = 6.487 \times 10^{-5}$, Two-Sided Fisher's Exact Test for Count Data).

mtrC	**Cervix**	**Pharynx**	**Rectum**	**Urethra**
Intact	113	103	242	2167
LOF	16	3	4	82

Supplementary Table S1.6 – Association between site of infection and *mtrC* LOF in validation dataset. *mtrC* LOF is significantly associated with site of infection (p = 0.02561, Two-Sided Fisher's Exact Test for Count Data).

mtrC	**Cervix**	**Pharynx**	**Rectum**	**Urethra**
Intact	218	383	625	867
LOF	9	3	7	15

Supplementary Table S1.7 – Association between sexual behavior and *mtrC* LOF in validation dataset. *mtrC* LOF is significantly associated with sexual behavior (p = 0.01803, Two-Sided Fisher's Exact Test for Count Data).

mtrC	**WSM**	**MSW**	**MSM**
Intact	268	243	1424
LOF	10	4	17

Supplementary Table S1.8 – *mtrA* LOF occurs most often in genetic backgrounds without *mtrR* or *mtr* promoter mutations in global dataset.

***mtrCDE* upregulation background**	***mtrA***	**n**
Yes	Intact	4113
No	Intact	355
Yes	LOF	89
No	LOF	268
Unknown	Intact	52
Unknown	LOF	5

Supplementary Table S1.9 – Association between site of infection and *mtrA* LOF in global dataset. *mtrA* LOF is significantly associated with site of infection ($p = 1.626 \times 10^{-12}$, Two-Sided Fisher's Exact Test for Count Data).

mtrA	**Cervix**	**Pharynx**	**Rectum**	**Urethra**
Intact	104	102	242	2187
LOF	25	4	4	61

Supplementary Table S1.10 – Association between sexual behavior and mtrA LOF in global dataset. *mtrA* LOF is significantly associated with sexual behavior ($p = 1.805 \times 10^{-11}$, Two-Sided Fisher's Exact Test for Count Data).

mtrA	**WSM**	**MSW**	**MSMW**	**MSM**
Intact	27	598	127	1181
LOF	6	28	0	7

Supplementary Table S1.11 – *mtrA* LOF occurs most often in genetic backgrounds with *mtrR* or *mtr* promoter mutations in validation dataset.

***mtrCDE* upregulation background**	***mtrA***	**n**
Yes	Intact	2012
No	Intact	87
Yes	LOF	81
No	LOF	4
Unknown	Intact	2

Supplementary Table S1.12 – *mtrC* LOF occurs most often in genetic backgrounds with *mtrR* or *mtr* promoter mutations in global dataset.

***mtrCDE* upregulation background**	***mtrC***	**n**
Yes	Intact	4028
No	Intact	617
Yes	LOF	174
No	LOF	6
Unknown	Intact	52
Unknown	LOF	5

Supplementary Table S1.13 – *mtrC* LOF occurs most often in genetic backgrounds with *mtrR* or *mtr* promoter mutations in validation dataset.

***mtrCDE* upregulation background**	***mtrC***	**n**
Yes	Intact	2060
No	Intact	89
Yes	LOF	33
No	LOF	2
Unknown	Intact	2

Supplementary Table S1.14 – Association between site of infection and *farA* LOF in global dataset. *mtrA* LOF is significantly associated with site of infection ($p = 1.775 \times 10^{-12}$, Two-Sided Fisher's Exact Test for Count Data).

farA	**Cervix**	**Pharynx**	**Rectum**	**Urethra**
Intact	96	103	234	2129
LOF	33	3	12	117

Supplementary Table S1.15 – Association between sexual behavior and *farA* LOF in global dataset. *mtrA* LOF is significantly associated with sexual behavior ($p = 5.055 \times 10^{-10}$, Two-Sided Fisher's Exact Test for Count Data).

farA	**WSM**	**MSW**	**MSMW**	**MSM**
Intact	25	585	125	1166
LOF	8	40	2	22

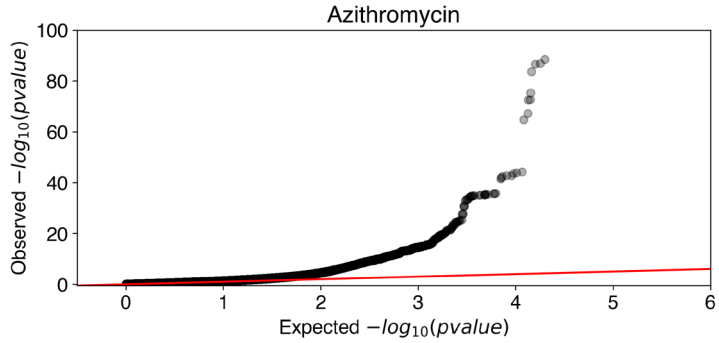
Supplementary Table S1.16 – Association between site of infection and farA LOF in validation dataset. *mtrA* LOF is significantly associated with site of infection ($p < 2.2 \times 10^{-16}$, Two-Sided Fisher's Exact Test for Count Data).

farA	**Cervix**	**Pharynx**	**Rectum**	**Urethra**
Intact	160	354	609	782
LOF	67	31	21	98

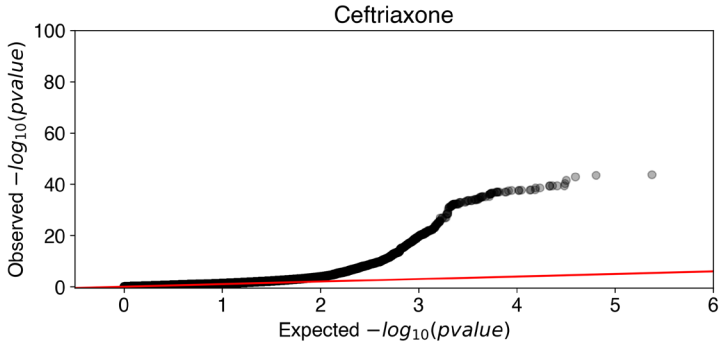
Supplementary Table S1.17 – Association between sexual behavior and farA LOF in validation dataset. *mtrA* LOF is significantly associated with sexual behavior ($p < 2.2 \times 10^{-16}$, Two-Sided Fisher's Exact Test for Count Data).

farA	**WSM**	**MSW**	**MSM**
Intact	197	186	1386
LOF	81	60	51

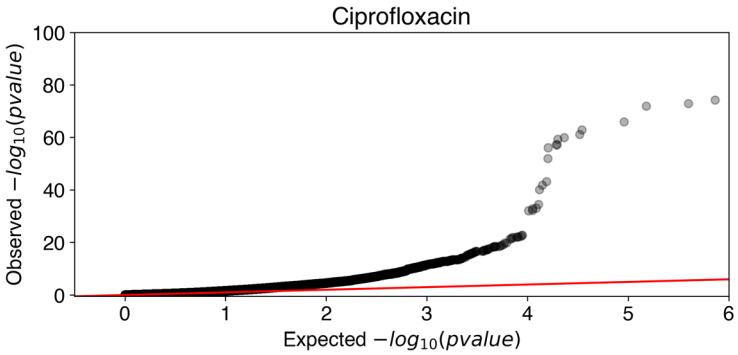
a)



b)

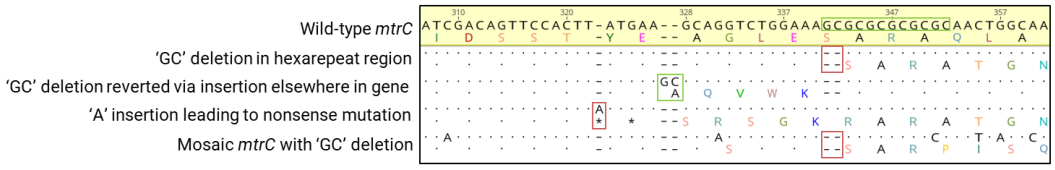


c)



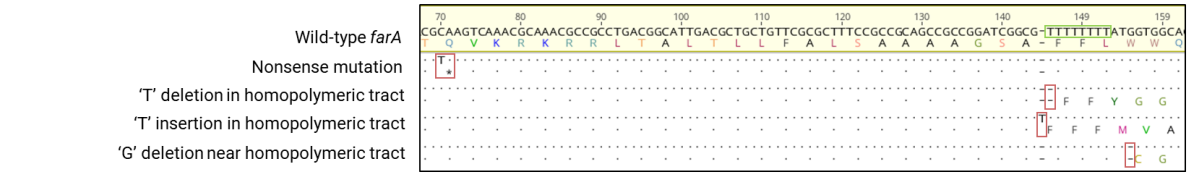
Supplementary Figure S1.1 – Diagnostic Q-Q plots of expected versus observed p-values for GWAS on a) azithromycin, b) ceftriaxone, and c) ciprofloxacin. In the absence of confounders such as population structure, *p*-values are distributed uniformly and would be expected to lie along the y=x line (in red) before diverging at higher -$\log_{10}$(*p*-values) due to true causal variants[16]. Q-Q plots for all three antibiotics appear to be well-behaved, indicating that the steps we have taken to control for population structure (i.e., using a linear mixed model parameterized by the recombination-corrected phylogeny) were adequate. Highly significant markers corresponding to diverging variants at higher -$\log_{10}$(*p*-values) were confirmed to map to known causal variants for all three antibiotics (see Supplementary Data S1.2).

a) Mutational diversity in *mtrC*

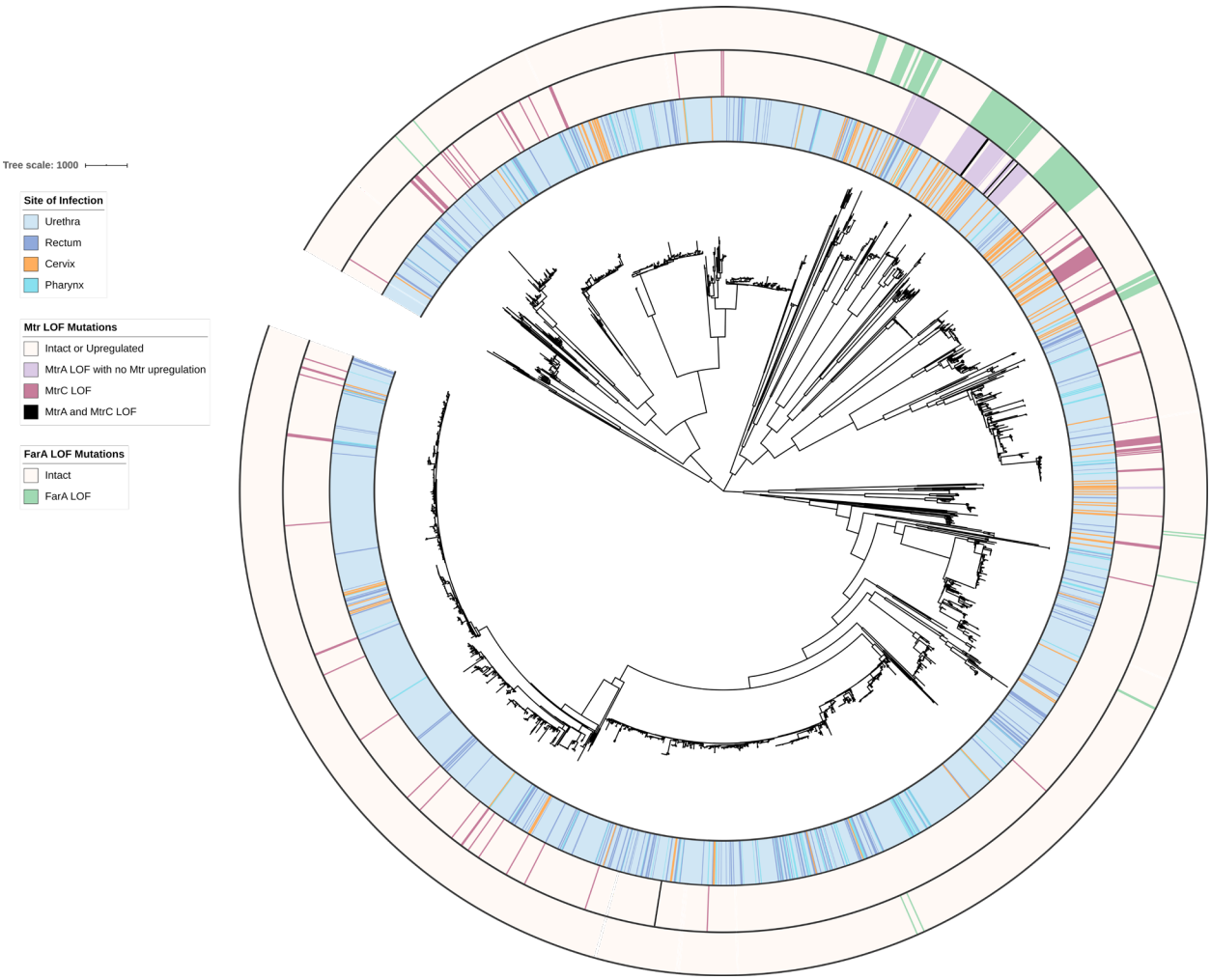


b) Mutational diversity in *mtrA*

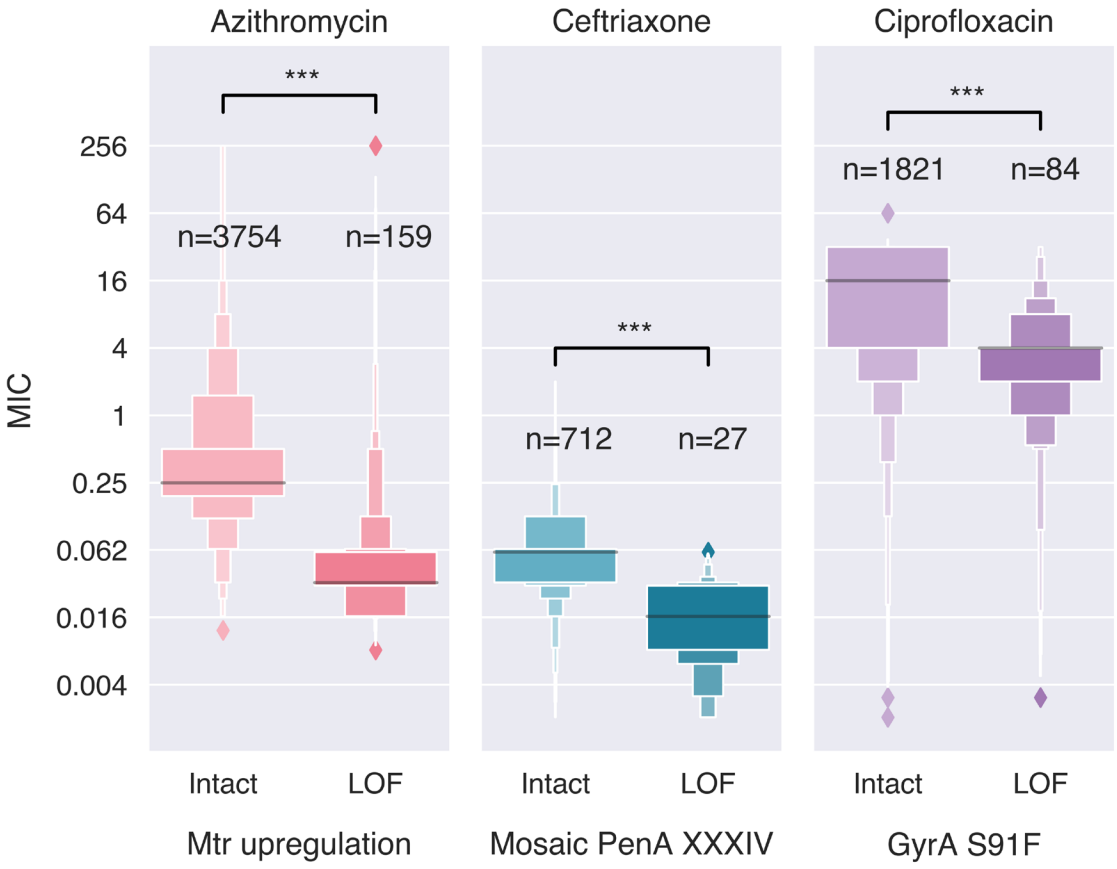
c) Mutational diversity in *farA*



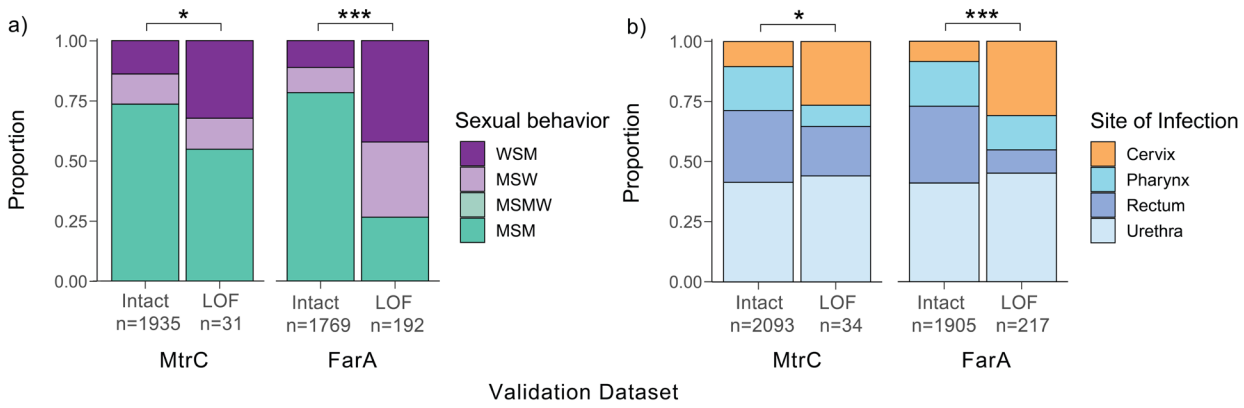
Supplementary Figure S1.2 – Alignment of nucleotide sequences for strains with representative LOF mutations observed in a) *mtrC*, b) *mtrA*, and c) *farA* in the global dataset. The wild-type reference sequences (FA1090 for *mtrC* and *farA*, FA19 for *mtrA*) are shown at the top of the alignment highlighted in yellow. Nucleotide sequences were depicted in black with the corresponding amino acid translations directly under. Dots in LOF sequences represent exact match to the wild-type reference sequence. For *mtrC*, the hexarepeat tract was boxed in the reference genome in green, and mutations leading to LOFs were boxed in red. For *mtrA*, the 11-bp deletion leading to *mtrA* LOF was boxed in red. For *farA*, the repeat tract of Ts was boxed in green, and mutations leading to LOFs were boxed in red. All alignments were visualized in Geneious Prime (see methods).



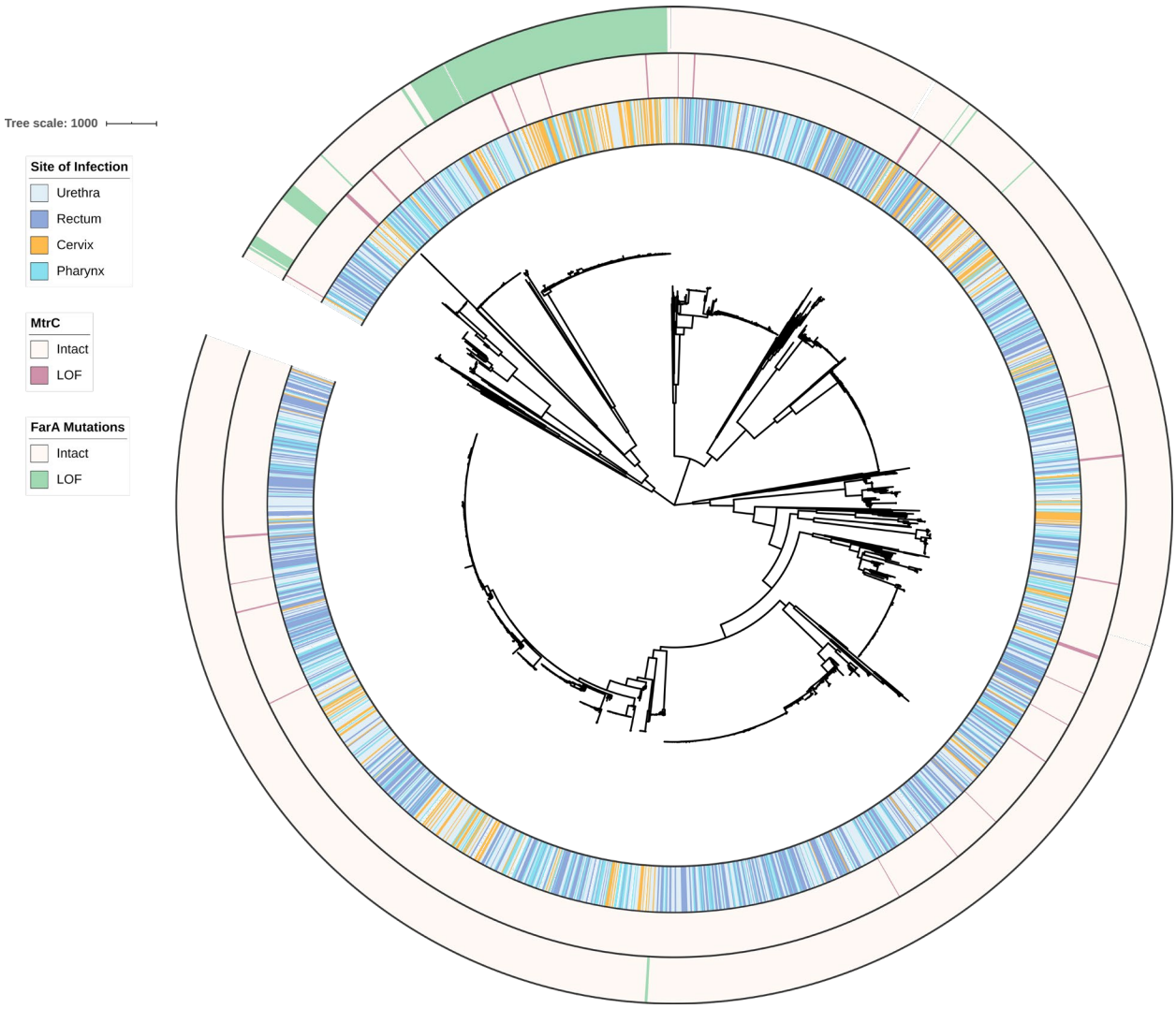
Supplementary Figure S1.3 – Phylogenetic distribution of gonococcal *mtrC*, *mtrA*, and *farA* LOF alleles with patient site of infection (n=2742) in global dataset. The recombination-corrected maximum likelihood phylogeny based on 36347 SNPs is shown annotated with rings (from innermost to outermost) for site of infection, *mtrCDE* regulon LOF mutations, and *farA* LOF mutations. Branch length represents total number of substitutions after removal of predicted recombination.



Supplementary Figure S1.4 – MIC distributions for isolates with indicated resistance determinants for azithromycin, ceftriaxone, and ciprofloxacin respectively, stratified by *mtrC* genotypic status. Statistical significance between *mtrC* intact and LOF MIC distributions was assessed by Mann-Whitney U Test: * p<0.05, ** p<0.01, and *** p<0.001. Enhanced box plots (also known as letter-value plots[17]) were drawn using default settings in the Python Seaborn visualization package (version 0.9.0). Interpretation differs from standard box-and-whisker plots: the central half of the MIC distribution is within the widest box; the central three-quarters of the distribution is within the widest and second widest box; the central three-eighths of the distribution is within the widest, second widest, and third widest box; and so on. The proportion of outliers was set to the default 0.007, and the number of boxes depicted is selected proportionally with respect to the total number of samples.



Supplementary Figure S1.5 – Gonococcal *mtrC* and *farA* LOF mutations are associated with sexual behavior and site of infection in the validation dataset. a) Sex partner information in patients infected with isolates with either intact or LOF alleles of *mtrC* (left) or *farA* (right). MSMW patients were labelled as MSM in the validation dataset[18]. b) Site of infection in patients infected with isolates with either intact or LOF alleles of *mtrC* (left) or *farA* (right) datasets. Statistical significance between intact versus LOF patient metadata distributions was assessed by Fisher's exact test: * $p<0.05$, ** $p<0.01$, and *** $p<0.001$. Exact *p*-values from left to right for analyses conducted in a) were 0.0180 and $< 2.20\times10^{-16}$, and for b) were 0.0256 and $< 2.20\times10^{-16}$. WSM = women who have sex with men, MSW = men who have sex with women, MSMW = men who have sex with men and women, MSM = men who have sex with men.



Supplementary Figure S1.6 – Phylogenetic distribution of gonococcal *mtrC* LOF alleles with patient site of infection (n=2186) in the validation dataset. The recombination-corrected maximum likelihood phylogeny based on 26669 SNPs is shown annotated with rings (from innermost to outermost) for site of infection, *mtrC* LOF mutations, and *farA* LOF mutations. Branch length represents total number of substitutions after removal of predicted recombination.

Supplementary References

1 Mortimer, T. D. *et al.* The distribution and spread of susceptible and resistant *Neisseria gonorrhoeae* across demographic groups in a major metropolitan center. Preprint at https://www.medrxiv.org/content/10.1101/2020.04.30.20086413v1 (2020).

2 Sanchez-Buso, L. *et al.* The impact of antimicrobials on gonococcal evolution. *Nat Microbiol*, doi:10.1038/s41564-019-0501-y (2019).

3 Yahara, K. *et al.* Genomic surveillance of Neisseria gonorrhoeae to investigate the distribution and evolution of antimicrobial-resistance determinants and lineages. *Microb Genom* **4**, doi:10.1099/mgen.0.000205 (2018).

4 Ryan, L. *et al.* Antimicrobial resistance and molecular epidemiology using whole-genome sequencing of Neisseria gonorrhoeae in Ireland, 2014-2016: focus on extended-spectrum cephalosporins and azithromycin. *Eur J Clin Microbiol Infect Dis*, doi:10.1007/s10096-018-3296-5 (2018).

5 Harris, S. R. *et al.* Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a genomic survey. *Lancet Infect Dis* **18**, 758-768, doi:10.1016/S1473-3099(18)30225-1 (2018).

6 Fifer, H. *et al.* Sustained transmission of high-level azithromycin-resistant Neisseria gonorrhoeae in England: an observational study. *Lancet Infect Dis* **18**, 573-581, doi:10.1016/S1473-3099(18)30122-1 (2018).

7 Lee, R. S. *et al.* Genomic epidemiology and antimicrobial resistance of Neisseria gonorrhoeae in New Zealand. *J Antimicrob Chemother* **73**, 353-364, doi:10.1093/jac/dkx405 (2018).

8 Kwong, J. C. *et al.* 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, doi:10.1136/sextrans-2017-053287 (2018).

9 Eyre, D. W. *et al.* WGS to predict antibiotic MICs for Neisseria gonorrhoeae. *J Antimicrob Chemother* **72**, 1937-1947, doi:10.1093/jac/dkx067 (2017).

10 De Silva, D. *et al.* Whole-genome sequencing to determine transmission of Neisseria gonorrhoeae: an observational study. *Lancet Infect Dis* **16**, 1295-1303, doi:10.1016/S1473-3099(16)30157-8 (2016).

11 Grad, Y. H. *et al.* Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United States, 2000-2013. *J Infect Dis* **214**, 1579-1587, doi:10.1093/infdis/jiw420 (2016).

12 Grad, Y. H. *et al.* Genomic epidemiology of Neisseria gonorrhoeae with reduced susceptibility to cefixime in the USA: a retrospective observational study. *Lancet Infect Dis* **14**, 220-226, doi:10.1016/S1473-3099(13)70693-5 (2014).

13 Demczuk, W. *et al.* Genomic Epidemiology and Molecular Resistance Mechanisms of Azithromycin-Resistant Neisseria gonorrhoeae in Canada from 1997 to 2014. *J Clin Microbiol* **54**, 1304-1313, doi:10.1128/JCM.03195-15 (2016).

14 Demczuk, W. *et al.* Whole-genome phylogenomic heterogeneity of Neisseria gonorrhoeae isolates with decreased cephalosporin susceptibility collected in Canada between 1989 and 2013. *J Clin Microbiol* **53**, 191-200, doi:10.1128/JCM.02589-14 (2015).

15 Ezewudo, M. N. *et al.* Population structure of Neisseria gonorrhoeae based on whole genome data and its relationship with antibiotic resistance. *PeerJ* **3**, e806, doi:10.7717/peerj.806 (2015).

16 Power, R. A., Parkhill, J. & de Oliveira, T. Microbial genome-wide association studies: lessons from human GWAS. *Nat Rev Genet* **18**, 41-50, doi:10.1038/nrg.2016.132 (2017).

17 Hofmann, H., Wickham, H. & Kafadar, K. Letter-Value Plots: Boxplots for Large Data. *J Comput Graph Stat* **26**, 469-477, doi:10.1080/10618600.2017.1305277 (2017).

18 Williamson, D. A. *et al.* Bridging of Neisseria gonorrhoeae lineages across sexual networks in the HIV pre-exposure prophylaxis era. *Nat Commun* **10**, 3988, doi:10.1038/s41467-019-12053-4 (2019).

Appendix 2: Supplementary materials accompanying Chapter 4

Supplementary Tables

Supplementary Figures

Supplementary References

Supplementary Table S2.1 – Prevalence of RplD macrolide binding site mutations varies across datasets.

Dataset	Region and Timespan	Prevalence of RplD macrolide binding site mutations
Demczuk et al., 2015[1]	Canada 1989-2013	0.0439
Demczuk et al., 2016[2]	Canada 1982-2011	0.0804
Eyre et al., 2017[3]	Brighton, UK 2004-2011	0.0346
Ezewudo et al., 2015[4]	Global 1982-2011	0
Fifer et al., 2018[5]	UK 2004-2017	0.0200
Grad et al., 2016[6] and 2014[7]	US 2000-2013	0.0582
Harris et al., 2018[8]	Europe 2013	0.0258
Kwong et al., 2017[9]	Melbourne, Australia 2005-2014	0.0532
Lee et al., 2018[10]	New Zealand 2014-2015	0.0050
Mortimer et al., 2020[11]	**New York City 2011-2015**	**0.1014**
Ryan et al., 2018[12]	Ireland 2012-2016	0.0513
Sánchez-Busó et al., 2019[13]	Global 1979-2012	0.0212
Yahara et al., 2018[14]	**Kyoto and Osaka, Japan 1996-2015**	**0.1346**

Datasets with prevalence over 10% are bolded.

Supplementary Table S2.2 – No substantial *in vitro* growth difference for the RplD G70D mutation in the 28BI laboratory strain background.

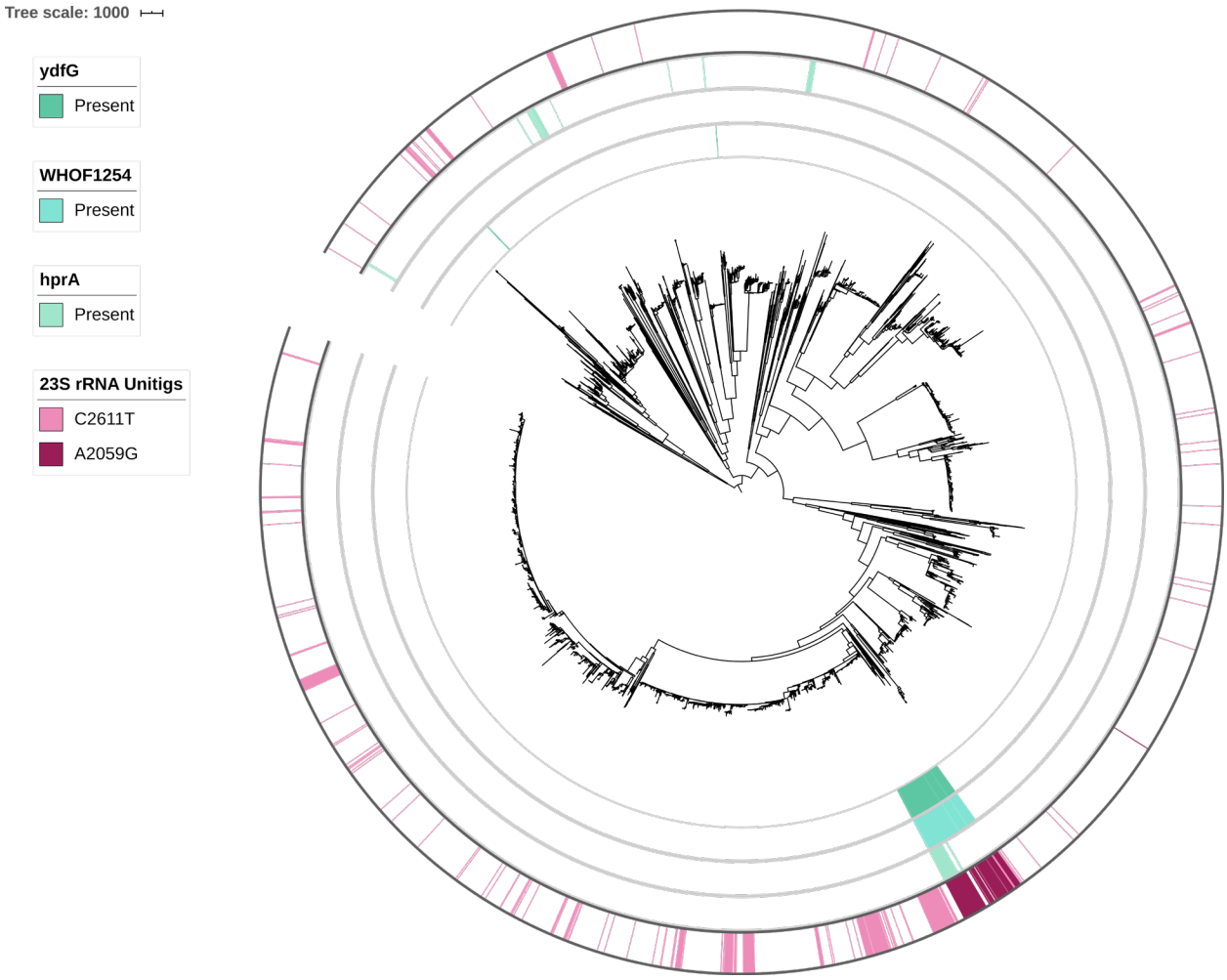
	28BI	28BI RplD^{G70D} (E9)
Best-fit values		
logY0	3.285	3.822
k	0.3947	0.3879
Doubling Time	1.756	1.787
95% CI (profile likelihood)		
logY0	3.113 to 3.458	3.612 to 4.032
k	0.3725 to 0.4169	0.3610 to 0.4149
Doubling Time	1.663 to 1.861	1.671 to 1.920
Goodness of Fit		
Degrees of Freedom	16	16
R squared	0.9889	0.9831
Sum of Squares	0.3673	0.5431
Sy.x	0.1515	0.1842
Number of points		
# of X values	18	18
# Y values analyzed	18	18

Estimation of exponential phase growth parameters using GraphPad Prism following removal of lag phase data points and log-transformation of CFUs / mL; see Supplementary Figure S2.5 (bottom) for estimated best fit lines.

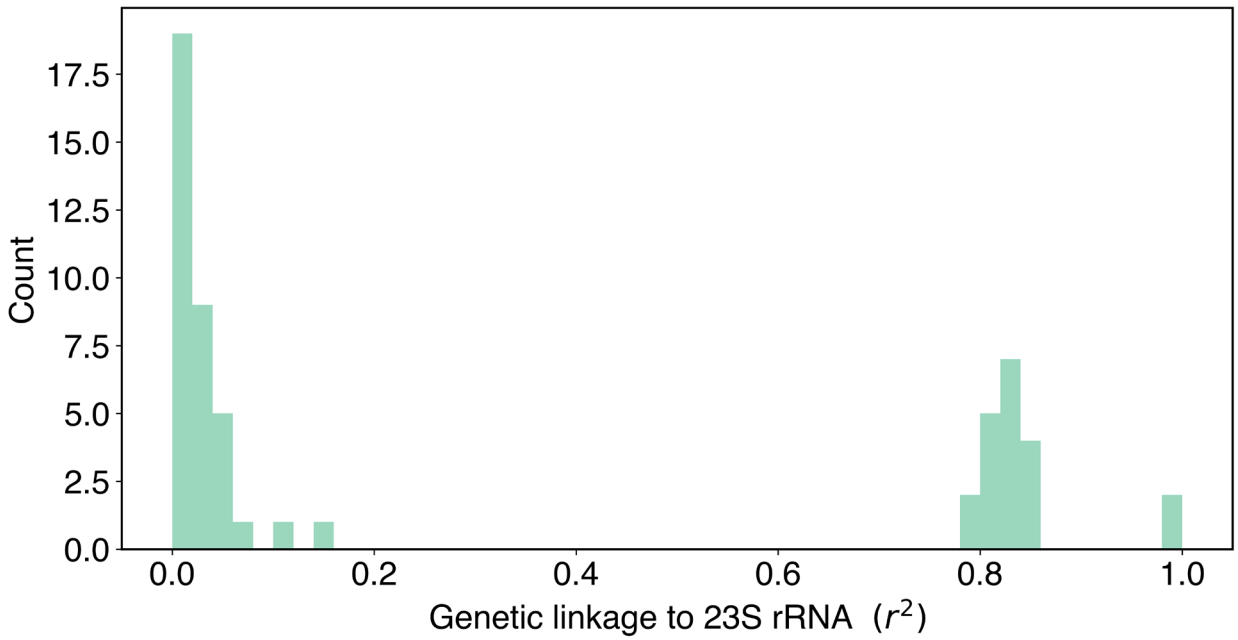
Supplementary Table S2.3 – Variance in azithromycin MICs explained by known resistance and susceptibility genes.

Mutation	Function	R^2 (lmg)	R^2 (first)	R^2 (last)
MtrR A39T	Pump upregulation	0.008703907	4.08E-05	0.014769
MtrR G45D	Pump upregulation	0.008911109	2.41E-03	0.015492
MtrR promoter	Pump upregulation	0.041388494	4.17E-02	0.037885
MtrR LOF	Pump upregulation	0.004493833	2.61E-04	0.0107
Mosaic Mtr	Pump upregulation and structural changes	0.076295392	4.20E-02	0.104465
23S rRNA A2059G	Ribosomal binding site modification	0.171732297	1.80E-01	0.167698
23S rRNA C611T	Ribosomal binding site modification	0.286807641	2.82E-01	0.290899
MtrC LOF	Pump loss-of-function (increased susceptibility)	0.056270124	4.85E-02	0.065024
RplD binding site mutations	Ribosomal binding site modification	0.010389559	1.62E-02	0.003228

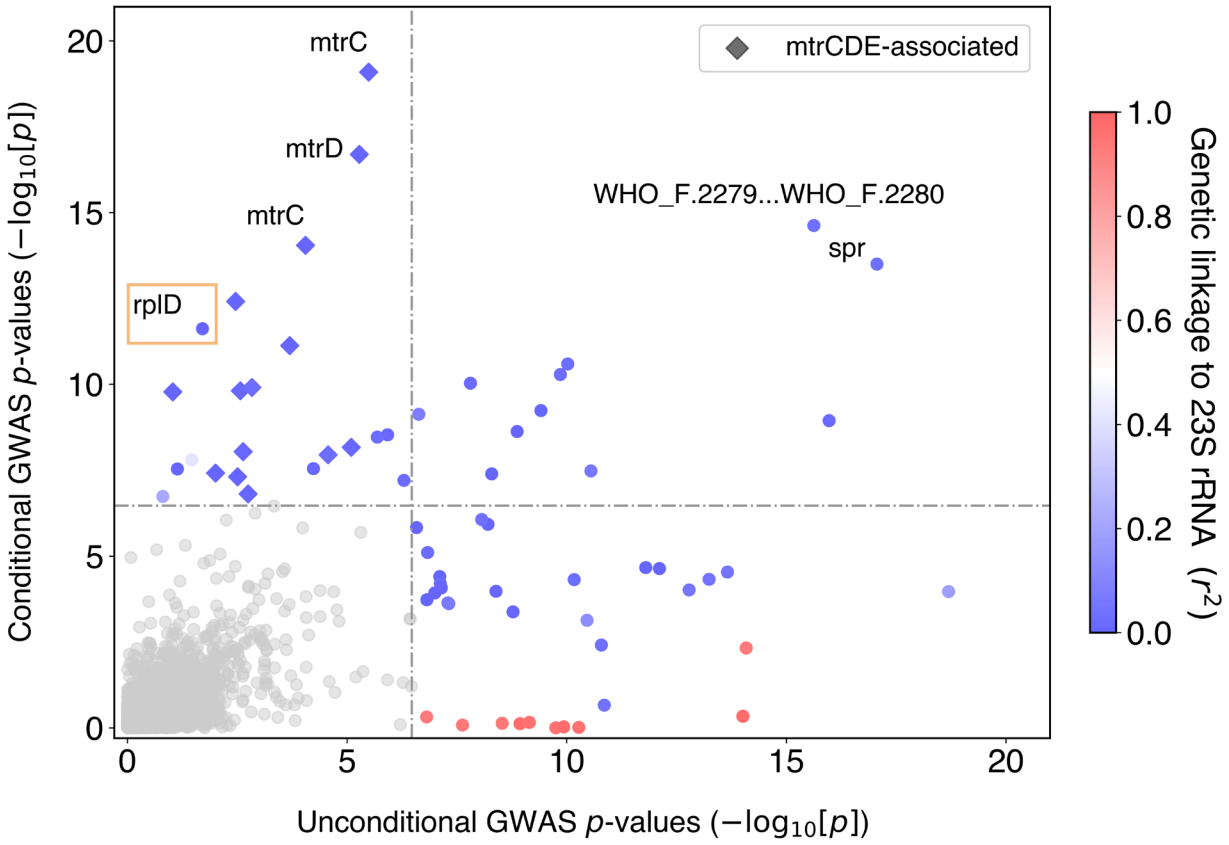
Variance explained by predictors was calculated using the relaimpo R package (version 2.2.3). The overall proportion of variance explained by the model was 66.5%. The variance explained by a predictor was determined by the change in model R^2 after inclusion of that predictor. Three approaches were used to calculate this change: the “first” metric compares a model without any predictors to a model with just the predictor of interest, the “last” metric compares a model with all predictors except the one of interest to a model with all predictors, and the “lmg” method averages the change in R^2 over all possible model subsets.



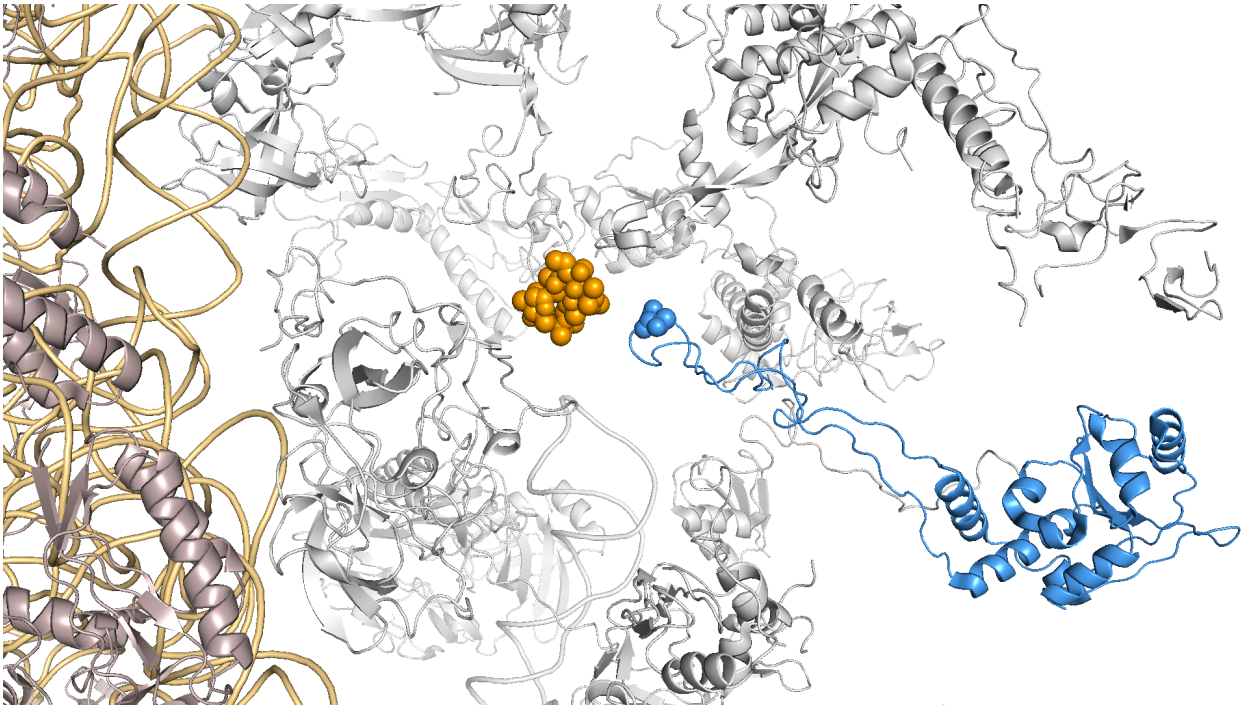
Supplementary Figure S2.1 – High genetic linkage between significant azithromycin MIC-associated variants in the GWAS. The recombination-corrected phylogeny from Figure 4.2 was annotated with the presence and absence of significant variants from the GWAS corresponding to 23S rRNA, *hprA*, WHO_F.1254, and *ydfG* (outermost to innermost). Branch length represents total number of substitutions after removal of predicted recombination.



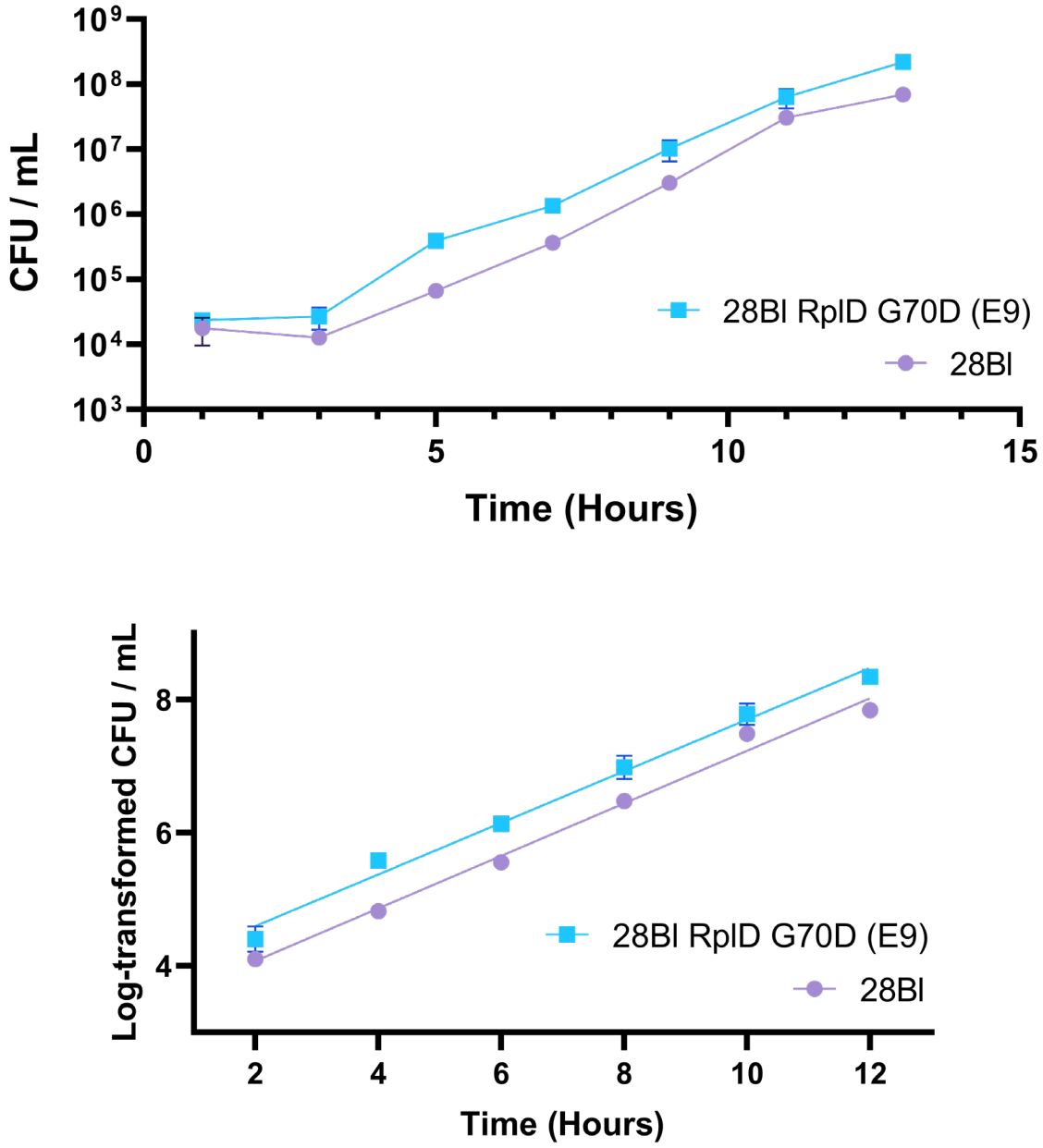
Supplementary Figure S2.2 – Distribution of r^2 values between significant variants (p-value < 2.97×10^{-7}) and 23S rRNA-associated unitigs in the single-locus GWAS. Significant variants with high linkage to 23S rRNA are likely to be spurious associations. See methods for details on calculation of r^2.



Supplementary Figure S2.3 – GWAS conditional on 23S rRNA mutations compared to unconditional GWAS results recovers similar results as in Figure 4.1 but does not control for dataset-specific confounders (*spr* and the intergenic region between genes WHO_F.2279 and 2280). As in Figure 4.1, genetic linkage measured by r^2 to 23S rRNA mutations A2059G and C2611T for significant variants is colored as indicated on the right. Variants associated with previously experimentally verified resistance mechanisms in the *mtrR* and *mtrCDE* promoters and coding regions are denoted in the legend. Bonferroni thresholds, calculated using the number of unique patterns, for both GWASes are depicted using a dashed line at 3.38×10^{-7}. Plot axes are limited to highlight variants associated with lower-level resistance; as a result, the highly significant 23S rRNA substitutions and *mtrC* indel mutations are not shown.



Supplementary Figure S2.4 – RplD G70 is part of the azithromycin binding pocket in the 50S ribosome from *Thermus thermophilus* (PDB ID: 4v7y). *N. gonorrhoeae* RplD is relatively similar to its *T. thermophilus* homolog (28.4% identical, 49.1% similarity using a BLOSUM62 matrix over 218 amino acids with 20 insertions/deletions). PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) was used to depict azithromycin in orange and RplD in blue (with the G70 amino acid highlighted as blue spheres) and to hide the 23S rRNA for clarity.



Supplementary Figure S2.5 – Growth curve experiments for RplD G70D isogenic strains show no *in vitro* fitness cost. Data are presented as mean values +/- SD calculated across the three technical replicates. Top – Calculated CFUs / mL from the full growth experiment for the two strains graphed on a logarithmic axis. Bottom – estimation of exponential phase best fit lines using GraphPad Prism following removal of lag phase data points and log-transformation of CFUs / mL; see Supplementary Table S2.2 for estimated parameters.

Supplementary References

1 Demczuk, W. *et al.* Whole-genome phylogenomic heterogeneity of Neisseria gonorrhoeae isolates with decreased cephalosporin susceptibility collected in Canada between 1989 and 2013. *J Clin Microbiol* **53**, 191-200, doi:10.1128/JCM.02589-14 (2015).

2 Demczuk, W. *et al.* Genomic Epidemiology and Molecular Resistance Mechanisms of Azithromycin-Resistant Neisseria gonorrhoeae in Canada from 1997 to 2014. *J Clin Microbiol* **54**, 1304-1313, doi:10.1128/JCM.03195-15 (2016).

3 Eyre, D. W. *et al.* WGS to predict antibiotic MICs for Neisseria gonorrhoeae. *J Antimicrob Chemother* **72**, 1937-1947, doi:10.1093/jac/dkx067 (2017).

4 Ezewudo, M. N. *et al.* Population structure of Neisseria gonorrhoeae based on whole genome data and its relationship with antibiotic resistance. *PeerJ* **3**, e806, doi:10.7717/peerj.806 (2015).

5 Fifer, H. *et al.* Sustained transmission of high-level azithromycin-resistant Neisseria gonorrhoeae in England: an observational study. *Lancet Infect Dis* **18**, 573-581, doi:10.1016/S1473-3099(18)30122-1 (2018).

6 Grad, Y. H. *et al.* Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United States, 2000-2013. *J Infect Dis* **214**, 1579-1587, doi:10.1093/infdis/jiw420 (2016).

7 Grad, Y. H. *et al.* Genomic epidemiology of Neisseria gonorrhoeae with reduced susceptibility to cefixime in the USA: a retrospective observational study. *Lancet Infect Dis* **14**, 220-226, doi:10.1016/S1473-3099(13)70693-5 (2014).

8 Harris, S. R. *et al.* Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a genomic survey. *Lancet Infect Dis* **18**, 758-768, doi:10.1016/S1473-3099(18)30225-1 (2018).

9 Kwong, J. C. *et al.* 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, doi:10.1136/sextrans-2017-053287 (2018).

10 Lee, R. S. *et al.* Genomic epidemiology and antimicrobial resistance of Neisseria gonorrhoeae in New Zealand. *J Antimicrob Chemother* **73**, 353-364, doi:10.1093/jac/dkx405 (2018).

11 Mortimer, T. D. *et al.* The distribution and spread of susceptible and resistant *Neisseria gonorrhoeae* across demographic groups in a major metropolitan center. *medRxiv*, 2020.2004.2030.20086413, doi:10.1101/2020.04.30.20086413 (2020).

12 Ryan, L. *et al.* Antimicrobial resistance and molecular epidemiology using whole-genome sequencing of Neisseria gonorrhoeae in Ireland, 2014-2016: focus on extended-spectrum cephalosporins and azithromycin. *Eur J Clin Microbiol Infect Dis*, doi:10.1007/s10096-018-3296-5 (2018).

13 Sanchez-Buso, L. *et al.* The impact of antimicrobials on gonococcal evolution. *Nat Microbiol*, doi:10.1038/s41564-019-0501-y (2019).

14 Yahara, K. *et al.* Genomic surveillance of Neisseria gonorrhoeae to investigate the distribution and evolution of antimicrobial-resistance determinants and lineages. *Microb Genom* **4**, doi:10.1099/mgen.0.000205 (2018).