Exploring the application of whole genome sequencing to inform the control of *Neisseria gonorrhoeae*

Katherine Jane Town

Thesis submitted for the degree of

Doctor of Philosophy

UCL

2019

Declaration of Authorship

I, Katherine Jane Town, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed _____ Date _____

Abstract

Background: *Neisseria gonorrhoeae* is a sexually transmitted infection that persists globally and is of concern because of the development of antimicrobial resistance (AMR). Novel methods for investigating how *N. gonorrhoeae* spreads within sexual networks, such as phylogenetic analysis of whole genome sequencing (WGS) data could support the public health response. The aim of my PhD was to explore the application of WGS to inform the control of *N. gonorrhoeae* and associated AMR.

Methods: I conducted a systematic review to describe *N. gonorrhoeae* studies that combined molecular and epidemiological data. I created a novel WGS dataset using specimens that were broadly representative of gonorrhoea in England. I identified clusters of infection representing sexual networks and characterised these using epidemiological and phenotypic data. I explored the genetic markers of AMR and compared the sample with international WGS datasets.

Results: In the systematic review, I found that there have been few *N. gonorrhoeae* WGS studies to date. In my phylogenetic analysis, I found clusters containing men reporting sex with men (MSM) as well as men who only report heterosexual sex, indicating there may be groups of heterosexual men who would benefit from further sexual health testing and prevention messages that would normally be targeted to MSM. The clustering also indicated extensive sexual mixing between HIV-positive and HIV-negative/unknown MSM and might inform methods to identify MSM at high risk of HIV infection. *N. gonorrhoeae* with reduced susceptibility appears to have emerged separately in different sexual networks in England. Distinct lineages of *N. gonorrhoeae* in England were related to cases in Europe and the USA indicating frequent transmission between countries.

Conclusion: WGS data analysis provided new insights about the transmission of *N. gonorrhoeae* and AMR in sexual networks, which reveal the complexity of sexual mixing patterns and might inform clinical care and public health policy.

Impact Statement

I have conducted one of the largest *N. gonorrhoeae* sequencing studies using WGS to date, which provided population level insights in *N. gonorrhoeae* epidemiology and AMR. The findings from this thesis provide proof of concept for how *N. gonorrhoeae* WGS data might be combined with epidemiological and phenotypic data to improve clinical care and prevention, develop AMR surveillance programmes and design research studies.

The sexual networks investigated in this thesis were of particular relevance for public health and clinical care because they facilitated *N. gonorrhoeae* transmission in England. I demonstrated that these high-risk sexual networks commonly contained men of different sexual orientations and MSM of different HIV status. If WGS data were collected, analysed and shared in a clinically relevant time frame, clinicians might be able to use this information as part of their risk assessment to tailor the services they offer to people. For example, HIV-negative/unknown MSM found to be part of sexual networks containing MSM infected with HIV may be at increased risk of HIV infection without necessarily being aware. If this information could be shared sensitively and without deductive disclosure, it might enable clinicians to provide appropriate advice and treatment, for example proactively offering HIV pre-exposure prophylaxis (PrEP) as a prevention measure. My data might also inform mathematical modelling studies of STI transmission, which aim to predict the spread of infection and measure the impact of potential interventions, by providing estimates of population-level sexual mixing parameters.

By combining the phylogenetic analyses with AMR phenotypic data, I found that, in England, *N. gonorrhoeae* with reduced susceptibility has emerged repeatedly in different sexual networks, which suggests that these emerging strains are a result of novel mutation events or repeated introduction from external sources, rather than only by clonal expansion. Therefore, it is difficult to predict in which geographical area or population sub-group AMR will emerge and public health resources to measure, prevent and control antimicrobial resistance should cover the entire population affected by gonorrhoea, rather than focusing on specific sub-groups or locations. Additionally, one implication of external introduction is that antimicrobial stewardship activity alone is unlikely to be enough to prevent *N. gonorrhoeae* resistance in England.

The dataset created within this PhD adds novel sequence data to publicly accessible WGS repositories. Large and diverse repositories are important for genome wide association studies (GWAS), which are used to investigate genetic markers of resistance and inform the development of rapid AMR diagnostic tests. The World Health Organisation (WHO) considers the development of molecular methods to detect resistance in *N. gonorrhoeae* an essential research priority. As well as contributing new specimens to the archives, I provided further evidence that molecular tests that detect reduced susceptibility to the first-line therapy (ceftriaxone) should not focus solely on mutations in the mosaic *penA* allele, as some tests currently do, because I identified specimens with reduced susceptibility to ceftriaxone without these mutations.

Table of Contents

Publications, Conferences and Scholarships

During this PhD I have published in peer-reviewed scientific journals, presented my findings at prominent national and international conferences and been awarded scholarships to further my research by collaborating with international research groups working on similar topics.

Publications

- *Neisseria gonorrhoeae* molecular typing for understanding sexual networks and antimicrobial resistance transmission: A systematic review. Town K, Bolt H, Croxford S, Cole M, Harris S, Field N, Hughes G. *J. Infect.* 76 (2018) 507–514.[1]

- Considering the potential application of whole genome sequencing to gonorrhea prevention and control. Kirkcaldy BD, Town K, Gernert KM, Bowen VB, Torrone EA, Kersh EN, Bernstein KT. *Sex. Trans. Dis*. 2018, 45(6): e29-e32.[2]

Conferences

Oral presentations:

- Harnessing molecular technology to inform our understanding of HIV and STI epidemics. Town K. BASHH & BHIVA National Conference, Edinburgh, UK, 2018 (Invited Speaker).[3]

- The emergence and spread of antimicrobial resistant *Neisseria gonorrhoeae* in HIV-positive men who have sex with men. Town K, Field N, Furegato M, Cole M, Harris S, Hughes G. ISSTDR, Brazil, 2017.[4]

- What is the role of travel-associated sexual partnerships on the importation and spread of AMR *Neisseria gonorrhoeae* in the UK? Town K, Field N, Cole M, Harris S, Hughes G. ISSTDR, Brazil, 2017.[5]

Poster presentation:

- A transdisciplinary approach to identify and delay the spread of antimicrobial resistance *Neisseria gonorrhoeae* using epidemiology, microbiology and genomics. Town K, Field N, Harris S and Hughes G. UCL Populations and Lifelong Health Domain Symposium - The future health of the public: Towards transdisciplinary research. 2017.[6]

Scholarships

Bogue Scholarship

I was awarded funding to collaborate with the Centers for Disease Control (CDC) in the United States of America (USA), on the use of whole genome sequencing to inform the control of *N. gonorrhoeae* within the Strengthening the United States Response to Resistant Gonorrhea (SURRG) programme.[7] This research trip enhanced my understanding of WGS analyses and how these data could be used to support public health interventions to control gonorrhoea. During the trip, I was invited to present my research within the CDC Sexually Transmitted Disease branch 'Learning Hour'. I also wrote a commentary piece with colleagues in the CDC on the topic of using WGS in public health.[2] Upon returning from the CDC, I have created a *N. gonorrhoeae* WGS working group involving experts from CDC, University College London, Public Health England (PHE) and the Wellcome Sanger Institute.

School of Life and Medical Sciences Conference Scholarship

I received funding to deliver two oral presentations on my research at the World STI Congress in Brazil in 2017.

List of Tables

List of Figures

Acknowledgments

I would like to thank my supervisors Gwenda Hughes, Nigel Field and Simon Harris for their support and guidance. I would like to thank the funders of this PhD, the National Institute for Health Research Health Protection Research Unit on Sexually Transmitted Infections and Blood Borne Infections, the Wellcome Trust Sanger Institute, the UCL School of Life and Medical Sciences and the Centre for Sexual Health Research. I would also like to thank the Bogue Fellowship committee for the scholarship award that enabled me to visit and work at the Centers for Disease Control and Prevention (CDC) in the USA. At the CDC, I'd like to thank Kyle Bernstein, Ellen Kersh, Kim Gernert and Robert Kirkcaldy for their time and support. Thank you also to everyone working and living at Villa International, particularly Maureen and Camille for their warm hospitality.

I am grateful to Public Health England (PHE) for granting me access to their *Neisseria gonorrhoeae* specimens and data archive, and I would like to thank all the people working within the sexual health clinics and laboratories who participate in PHE surveillance programmes. Without their continued support, the research presented in this thesis would not be possible.

There are many colleagues, friends and family I would like to thank. At PHE, I am grateful to Michelle Cole, Rachel Pitt, Zahra Sadouki, Sharla MacTavish and Martha Valencia for their help learning how to master the art of growing, identifying and extracting DNA from *N. gonorrhoeae*. I would like to thank Leonor Sánchez-Busó for her time and patience whilst she taught me new coding languages and how to build trees. A big thank you to Lucy van Dorp for her review of this thesis. I would like to give a huge thank you to my friends and family who were always so kind and supportive (even when they'd heard it all before) – **thank you so much** – Cordelia Coltart, Sara Croxford, Tyrone Curtis, Ibidun Fakoya, Seth Francis-Graham, Meaghan Kall, David Mesher, Holly Mitchell, George Öktem, Fraser Pender, Claire Sand, Catherine Smith, Stephen Town, John Town and Shiona Dawson.

Table of Abbreviations

AGSP	Australian Gonococcal Surveillance Programme
AMR	Antimicrobial resistance
ARIBA	Antimicrobial Resistance Identification By Assembly
CDC	Centers for Disease Control and Prevention
China-GRSP	China Gonococcal Resistance Surveillance Programme
CI	Confidence interval (95% level, L = lower, U = upper)
DNA	Deoxyribonucleic acid
DS	Decreased susceptibility
EUCAST	European Committee on Antimicrobial Susceptibility Testing
Euro-GASP	European Gonococcal Antimicrobials Surveillance Programme
FE	Fisher's exact test
GASS	Gonococcal Antibiotic Surveillance in Scotland
GC	Guanine and cytosine
GGI	Gonococcal genetic island
GISP	Gonococcal Isolate Surveillance Programme
GRASP	Gonococcal Resistance to Antimicrobials Surveillance Programme
GUMCAD	Genitourinary Medicine Clinic Activity Dataset
GWAS	Genome wide association studies
Het. Men	Heterosexual men (men who only report sex with women)
HIV	Human immunodeficiency virus
HRA	Health Research Authority
MIC	Minimum inhibitory concentration
MLST	Multi-locus sequence typing
MMC	Mortimer Market Clinic
MSM	Men who report sex with men
NAATs	Nucleic acid amplification tests
Natsal	National Survey of Sexual Attitudes and Lifestyle
NG	*Neisseria gonorrhoeae*
NG-MAST	*N. gonorrhoeae* multi-antigen sequence typing
OR	Odds ratio (cOR = crude odds ratio, aOR = adjusted odds ratio)
P	Probability value
PHE	Public Health England
PID	Pelvic inflammatory disease
POCT	Point of care test
PPV	Positive predictive value
PrEP	Pre-exposure prophylaxis
RAxML	Randomized Axelerated Maximum Likelihood
REC	NHS Research and Ethics Committee
REGG	Research Ethics and Governance Group
RNA	Ribonucleic acid
RS	Reduced susceptibility
SNP	Single nucleotide polymorphism
STI	Sexually transmitted infection
TBE	Transfer bootstrap expectation
TOC	Test of cure
UK	United Kingdom
USA	United States of America
WGS	Whole genome sequencing
WHO	World Health Organisation

Glossary of Key Terms

Antimicrobial resistance	The ability of a microorganism to grow in the presence of a concentration of an antimicrobial that would otherwise have killed or prevent growth of the organism
Cluster	A group of specimens with similar molecular data
Decreased or reduced susceptibility	An MIC value that demonstrates a microorganism is less sensitive to an antimicrobial but not resistant
Gene	A distinct sequence of nucleotides that code for a protein
Genome	All the genetic material within an organism
Genetic recombination	The exchange of similar or identical nucleotide sequences
HIV-positive people	People who are living with HIV
HIV-negative/unknown people	People who were known to be HIV-negative at the time of their gonorrhoea diagnosis or their HIV status was unknown
Minimum inhibitory concentration	The concentration of antimicrobial at which the bacterium can no longer grow
Molecular epidemiology	The use of pathogen molecular data in the study of the distribution, dynamics, and determinants of health and disease in human populations
Multi-locus sequence typing	Molecular typing technique used for *N. gonorrhoeae*: the nucleotide sequence of the internal fragments of seven or more chromosomal housekeeping genes are identified and used to categorise samples into strain types
***Neisseria gonorrhoeae* multi-antigen sequence typing**	Molecular typing technique used for *N. gonorrhoeae:* the nucleotide sequence of the porin PorB gene (*porB*) and the subunit B of the transferrin-binding protein gene (*tbpB*) are identified and used to categorise specimens into strain types
PhD study sample	All *N. gonorrhoeae* specimens that were sequenced and used for analysis in the PhD
Phylogenetic tree	A diagram that shows the inferred relatedness and evolutionary relationships between specimens
Seroadaptive behaviours	Behaviours where individuals engage in sexual practices to reduce the risk of transmission or acquisition of HIV
Serosorting	Condomless anal intercourse with partners according to their known or presumed HIV status
Sexual network	A pattern of connections between people as defined by their sexual relationships
Single nucleotide polymorphism	One nucleotide difference between sequences being compared
Specimen	Sample of *N. gonorrhoeae*
Strain	A genetic variant or subtype of a microorganism
Whole genome sequencing	The process of identifying all the nucleotides in the DNA of an organism

1 Introduction to *Neisseria gonorrhoeae* and molecular epidemiology

1.1 Chapter summary

In this chapter, I describe the biology, genetics and epidemiology of *Neisseria gonorrhoeae*, including the public health strategies used to prevent and control the infection. I discuss the prevalence of antimicrobial resistance in *N. gonorrhoeae*, highlight some of the functional mechanisms underlying resistance phenotypes and describe the consequences for treatment regimens. I summarise the principles and tools used within molecular epidemiology and phylogenetics to understand the transmission and spread of *N. gonorrhoeae*. Finally, I present the key aims and research questions of this thesis.

1.2 Aetiology

N. gonorrhoeae is a gram-negative bacterium that causes the sexually transmitted infection (STI) gonorrhoea. Humans are the only natural host of *N. gonorrhoeae* in whom it infects mucosal membranes such as the urethra or endocervix. Half of genital infections in women are asymptomatic, whereas the majority (>90%) of genital infections in men are symptomatic.[8,9] Symptoms of genital infections include vaginal or urethral discharge and dysuria.[10] Extra-genital infections occur at the rectum, pharynx or eye and are often asymptomatic. If infection is left untreated, people can experience painful and debilitating sequelae, including pelvic inflammatory disease (PID) in women and epididymitis in men, both of which can cause infertility. Disseminated gonococcal infection is rare but can cause skin lesions and arthritis.[11,12]

1.3 Genome

The genome of *N. gonorrhoeae* typically contains around 2.2 million nucleic acid base pairs in a circular chromosome, which codes for about 2000 core genes shared amongst all strains and 500-600 accessory genes that can vary widely.[13,14] *N. gonorrhoeae* is highly competent at homologous recombination and horizontal gene transfer, including the transfer of plasmids (extra-chromosomal circular deoxyribonucleic acid (DNA)) through conjugation.[15,16] Transfer of DNA into the chromosome is facilitated by a high number (~1,900) of DNA uptake sequences, which are found throughout the *N. gonorrhoeae* genome, particularly around genes

involved in DNA repair, recombination and replication.[13,17-21] *N. gonorrhoeae* can incorporate DNA into the genome from other *N. gonorrhoeae* organisms and from other *Neisseria* species, including *Neisseria meningitidis.*[22] Pharyngeal infection is thought to be a primary location for the exchange of genetic material in the host because many different organisms, including other *Neisseria* species, are commensals of this niche.[23,24]

1.4 Epidemiology

1.4.1 Number of diagnoses

In England, the number of gonorrhoea diagnoses has fluctuated since recording began in the early 20th Century, with the highest number of diagnoses recorded in 1975 (58,956 cases).[25,26] Despite a decline in prevalence in England during the late 1980s and early 1990s to less than 10,000 new cases, likely attributable to health promotion campaigns related to human immunodeficiency virus (HIV) prevention, there has been a resurgence in the number of new diagnoses over the last decade.[27,28] Between 2008 and 2017, the number of new diagnoses rose annually from 14,985 in 2008 to 44,676 in 2017 (Figure 1.1). In 2016, there was a decline in cases, which was likely due to increased testing and shorter treatment turnaround times in men reporting sex with men (MSM) particularly in London.[29]

Despite the increase in gonorrhoea diagnoses, *N. gonorrhoeae* is still considered a rare infection in the United Kingdom (UK), restricted to people with recognised risk factors, such as multiple sexual partners and condomless sex. In the sexually active British general population aged 16 to 44 years, the prevalence was estimated to be <0.1% (data from the National Survey of Sexual Attitudes and Lifestyle 3 - Natsal-3, a nationally representative survey conducted between September 2010 and August 2012).[30] Even in specialist sexual health clinics, in young heterosexuals (aged <25 years), the prevalence was estimated to be <1% across England in 2015.[31]

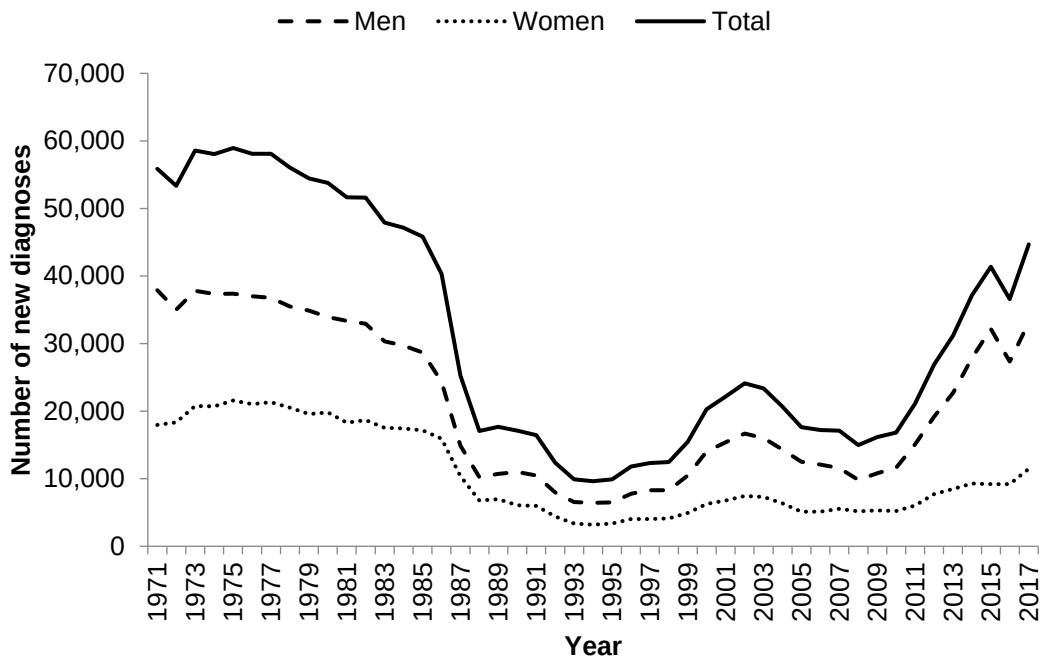


Figure 1.1 Number of new diagnoses of *N. gonorrhoeae* by gender in England: 1971-2017
Combined data from national surveillance of specialist sexual health services (1971-2017) and non-specialist services (2012-2017) in England. Gender is provided as per legend at the top with "total" providing the number of new diagnoses including those with unknown or unreported gender.[26]

1.4.2 Transmission and sexual networks

Transmission of *N. gonorrhoeae* generally requires sexual contact between individuals. As such, sexual partnerships can be considered as linked transmission chains, forming a network through which infection can spread. STI transmission through sexual networks depend on several different elements including the biology of the pathogen, the clinical manifestation of the infection, the sexual behaviour of the individual and the social and cultural contexts that influence these behaviours.[32-36] Characterising a sexual network may allow identification of individuals at most risk of infection, so that they can be targeted for health promotion, testing and treatment. In addition, information about infected individuals in the network can be used to determine specific risk factors associated with STI transmission, thereby increasing our understanding of the epidemiology, which can help tailor public health interventions that prevent and control infection and improve monitoring.[37,38] For example, by constructing a sexual network and investigating the behaviours of those in the network, De *et al*. identified that the people who attended a particular bar were at highest risk of *N. gonorrhoeae* infection. Therefore, targeting patrons of the bar

with sexual health promotional material was considered a potential approach to help control infection spread.[37]

To assist in constructing a sexual network, information on sexual contacts can be obtained from partner notification i.e. where sexual partners of people infected with a STI are notified that that they have been exposed and are offered treatment and support services.[39] Notification can either be via the patient who was initially diagnosed (patient referral) or by the health professional either in the first instance (provider referral) or at a later date if the patient has been unable to contact the partner (contract or conditional referral).[40,41] Partner notification is a routine and essential procedure for the control and prevention of STIs and HIV[39,41,42] but is challenging, often resulting in incomplete data on sexual contacts within a sexual network.[40,43-45]

Mathematical models can be used to predict the spread of *N. gonorrhoeae* and measure the impact of potential interventions.[46-50] In these models, the reproduction number (R0) is a core concept and is defined as the number of secondary cases arising from an index case. For a given pathogen and population, R0 can be estimated by multiplying the transmission probability (β), rate of contact (c) and duration of infection (D) ($R_0=\beta cD$).[51]

Data from epidemiological studies on the sexual mixing of different sub-groups of the population are used to parameterise the rate of contact (c). These sexual mixing patterns can either be assortative or disassortative.[33,49,52] Assortative sexual mixing describes people preferentially choosing partners with similar characteristics to themselves (e.g. assortative mixing by age might involve teenagers choosing other teenagers as sexual partners). Whereas disassortative sexual mixing describes people preferentially choosing partners with dissimilar characteristics to them (e.g. disassortative mixing by ethnicity might involve a person of black Caribbean ethnicity choosing someone of white ethnicity as their sexual partner). The sexual mixing pattern of infected individuals affects how the infection spread in the population.[32,52,53] Assortative sexual mixing will tend to concentrate infection in discrete population sub-groups, whereas disassortative sexual mixing will foster wider dissemination of infection across the population.[49] Therefore, understanding sexual mixing patterns can help predict and explain the distribution of infection in the population.

The other parameters to calculate the reproduction number (β and D) vary depending on the type of sexual act and anatomical site of infection. For example, in a recent systematic review, Chow *et al*. found that pharyngeal gonorrhoea had a shorter duration of infection (D) (114-138 days) than rectal gonorrhoea (346 days).[54] Similarly, asymptomatic infections are important to consider in gonorrhoea transmission modelling because they are less likely to be diagnosed and treated, leading to longer duration of infection.[55-57] Recent models have suggested that condomless oral sex or even kissing may be important factors in the perpetuation of *N. gonorrhoeae* at a population level, particularly in MSM.[58,59]

1.4.3 Characteristics of people most at risk of *N. gonorrhoeae* infection

Risk factors associated with *N. gonorrhoeae* infection can be determined from studies or public health surveillance using classical epidemiological techniques to analyse data on time, person and place.[30,60,61] *N. gonorrhoeae* is unequally distributed in the population.[60] Specific population groups are at a higher risk of infection because of sexual behaviours that facilitate transmission.[53,62] *N. gonorrhoeae* persists in these sexual networks because infection prevalence is high, and because network members have relatively large numbers of sexual partners, faster rates of partner change and assortative sexual mixing patterns.[63-65] In England, there are two key populations at greatest risk of infection:

i. MSM

The rise of *N. gonorrhoeae* in the last decade has primarily occurred among MSM.[28] MSM report higher numbers of sexual partners than men who have sex exclusively with women and are more likely to engage in sexual behaviour that puts them at risk of STIs, including condomless sex with multiple partners.[65,66] The sharpest rise in cases has occurred amongst MSM living with HIV.[67] Over a third of MSM living with HIV were estimated to have had a second *N. gonorrhoeae* infection within one year compared to one fifth of HIV-negative/status unknown (hereafter referred to as HIV-negative/unknown) MSM diagnosed only once.[67] The increase in *N. gonorrhoeae* cases may be linked to changes in sexual behaviour including more condomless sex with multiple partners and the adoption of HIV seroadaptive behaviours.[67-75] Seroadaptive behaviours involve individuals engaging in sexual practices to reduce the risk of transmission or acquisition of HIV, such as condomless anal intercourse with partners according to their known or presumed HIV status (serosorting) or only

practicing insertive anal sex with potentially discordant partners because the risk of HIV transmission is lower for the insertive partner (seropositioning).[76] Recent studies found that nearly half of MSM reported condomless anal intercourse and more than a quarter practiced serosorting.[68,74] *N. gonorrhoeae* infection has also been shown to be associated with greater risk of HIV acquisition in several studies.[77]

ii. People of black Caribbean and black other ethnicity

Compared to other ethnic groups, people of black Caribbean or black other ethnicities (self-defined) are at higher risk of infection and reinfection.[25,78,79] The reasons for the unequal distribution of *N. gonorrhoeae* by ethnic group are multifactorial. Socioeconomic deprivation likely contributes to high rates of infection in these ethnic groups, as do differences in the access and provision of health care services.[25,26] Additionally, epidemiological studies have shown some evidence of sexual risk behaviours such as concurrent partnerships and condomless sex being more common in these ethnic groups. However, these do not fully explain the differences in STI risk.[80-83]

1.5 Prevention and control

1.5.1 Testing

The recent rise in *N. gonorrhoeae* diagnoses is in part due to improvements in STI detection through widespread use of nucleic acid amplification tests (NAATs).[28] NAATs are molecular tests that detect pathogen-specific genetic material. NAATs are faster, cheaper and more reliable in terms of sensitivity and specificity than traditional culture or microbiological techniques and are superior for detecting asymptomatic infections.[11,84] National clinical guidelines recommend NAATs for detecting genital and extra-genital infections; extra-genital testing is routinely recommended for MSM attending sexual health clinics and for women who are sexual contacts of people infected by gonorrhoea and is guided by sexual risk assessment and symptoms for everyone else.[11,85] However, NAATs are not infallible given that cross reaction with non-gonococcal *Neisseria* species can lead to a false positive result, particularly if the test is at the pharynx, where colonisation by non-gonococcal *Neisseria* species is common.[84,86] False positive test results are also more likely to occur in populations with a low prevalence of gonorrhoea, resulting in low positive predictive values (PPV).[31,87,88] Consequently, it is recommended that a

supplementary validation test designed to detect a different nucleic acid target is performed.[84] Specimens for culture should be taken from patients with signs and symptoms of gonorrhoea and/or a positive NAAT result to enable antimicrobial susceptibility testing.[84] Microscopy is also used to visualise *N. gonorrhoeae* directly from a clinical specimen, which may allow more rapid treatment, but has lower sensitivity than NAATs, particularly for specimens from women and asymptomatic infections, and therefore is not recommended as an independent diagnostic technique.[11]

1.5.2 Treatment

In 1940, sulphonamides were the first antimicrobials used to treat *N. gonorrhoeae*[89,90] followed by penicillin in 1943.[91] Treatment guidelines changed in the 1980s to ciprofloxacin as the prevalence of penicillin AMR increased.[89] Since the turn of the century, the recommended treatment for gonorrhoea has changed three times from ciprofloxacin to cefixime in 2005, to ceftriaxone and azithromycin dual therapy in 2011 and to ceftriaxone monotherapy in 2018 (this most recent change was under consultation at the time of writing).[10,11] These changes have been made in an attempt to control the emergence and spread of AMR (see section 1.6).

Treatment is given to people diagnosed with *N. gonorrhoeae* by culture and/or NAATs or if there is evidence through Gram stain microscopy of diplococci. For those presenting with symptoms suggestive of gonorrhoea, or sexual contacts of those diagnosed, treatment is often given empirically with the instruction to avoid sexual contact for at least one week to reduce the risk of onward transmission before laboratory tests confirm infection. Test of cure (TOC) is recommended for everyone diagnosed with gonorrhoea, with priority follow-up for people with persistent symptoms, pharyngeal infection, treatment using an antimicrobial other than the first line therapy and for people who acquired their infection in the Asia-Pacific region if the antimicrobial susceptibility profile of the infection is unknown. TOC aims to detect treatment failures and reinfections and to inform continued partner notification and health promotion efforts.[11]

1.5.3 Interventions

Strategies used to control and prevent the transmission of *N. gonorrhoeae* typically involve promoting behavioural change (e.g. condom use) and frequent testing and treatment for high risk populations, such as MSM. Vaccine development has been difficult due to poor quality animal models for experimentation and weak evidence for

inducible natural immunity in humans.[92-95] Although vaccine targets have been identified, a successful vaccine has yet to be produced.[94,95] Another option in development is to use vaccines that have been developed for other *Neisseria* species. In theory, this might be an effective strategy because of the large proportion (80-90%) of shared genetic sequence between *N. meningitidis* and *N. gonorrhoeae*.[96-98] A retrospective case-control study of sexual health clinic attendees in New Zealand found that those vaccinated with the outer membrane vesicle meningococcal B vaccine were less likely than those who had not been vaccinated to be infected with *N. gonorrhoeae.*[98]

1.6 Antimicrobial resistance

1.6.1 Mechanisms

Antimicrobial resistance (AMR) in *N. gonorrhoeae* is caused by chromosomal mutations or plasmid-mediated genes that provide phenotypic mechanisms that prevent antimicrobials reaching their target or interrupt their function and enable the bacteria to remain viable.[99] Antimicrobial susceptibility is commonly measured by the minimum inhibitory concentration (MIC), which defines the lowest concentration of antimicrobial able to visibly inhibit growth. Therefore, the higher the MIC, the less susceptible the bacteria are to the antimicrobial. For most antimicrobials, there is a standard MIC threshold that defines clinically relevant resistance,[100] which is related to the dosage used to treat infection. In *N. gonorrhoeae*, functional mechanisms of AMR include:[89,99]

- reduced influx of antimicrobials into the bacterium, for example, some mutations in the *porB1b* gene lead to changes in the outer membrane porin and restrict the influx of antimicrobials such as penicillin or cephalosporins (cefixime and ceftriaxone),

- increased efflux of antimicrobials out of the bacterium, for example, mutations in the *mtrR* gene lead to overexpression of the MtrCDE efflux pump, which leads to efflux of antimicrobials such as penicillin or cephalosporins (cefixime and ceftriaxone),

- production of enzymes that inhibit the action of the antimicrobial, for example, penicillinase encoded by the TEM-1 or TEM-135 plasmids inactivate penicillin,

- structural changes to the target of the antimicrobial, for example, mutations in the 23S ribosomal ribonucleic acid (rRNA) target lead to reduced binding of azithromycin to the 50S ribosome.

AMR has likely increased in frequency in the *N. gonorrhoeae* population because of the selection pressure from antimicrobial prescribing patterns[99] and the extensive capacity of the *N. gonorrhoeae* to mutate, recombine and acquire genetic determinants of AMR, most of which seemingly do not affect the fitness of the organism, so are retained in the pathogen population.[89]

1.6.2 Prevalence

N. gonorrhoeae AMR is a major concern globally, as it now affects all classes of antimicrobials used for treatment. Previous first-line recommended therapies in the UK and elsewhere included penicillin, ciprofloxacin and cefixime, but these were replaced as first-line treatment, because the prevalence of AMR breached the World Health Organisation (WHO) recommended threshold (≥5% of infections resistant to the first line therapy) (Figure 1.2).[10,11,28,101,102] In March 2017, *N. gonorrhoeae* was included as a high priority on the WHO list of pathogens that require the urgent development of new antibiotics to remain treatable.[103]

Antimicrobial susceptibility surveillance programmes are used to monitor the emergence and spread of *N. gonorrhoeae* AMR. In England and Wales, the Gonococcal Resistance to Antimicrobials Surveillance Programme (GRASP) was initiated in 2000[104] and includes a sentinel programme that tests a sample of *N. gonorrhoeae* diagnosed during a three-month period for susceptibility to seven antimicrobials, and combines these data with patient demographic, behavioural and clinical data. GRASP has twice provided data, in conjunction with the Gonococcal Antibiotic Surveillance in Scotland (GASS),[105] to directly influence changes in treatment guidelines in the UK. Other surveillance systems include the WHO Gonococcal Antimicrobial Surveillance Programme (GASP),[106] which includes data from over 50 countries worldwide,[107] the European Gonococcal Antimicrobials Surveillance Programme (Euro-GASP), which collects data from over 20 countries and includes a subset of GRASP specimens from England,[108] the Gonococcal Isolate Surveillance Programme (GISP)[109] in the United States of America (USA), the Australian Gonococcal Surveillance Programme (AGSP)[110] and the China Gonococcal Resistance Surveillance Programme (China-GRSP).[111]

1.6.3 Risk groups

The prevalence of AMR *N. gonorrhoeae* varies by patient demographics and sexual behaviours. In England and the USA, surveillance data show that AMR prevalence appears to be higher amongst MSM compared to heterosexual populations.[28,109] However, in other European surveillance datasets (Euro-GASP) AMR prevalence appears to be highest in heterosexual men,[112] but missing sexual orientation data for over 40% of people included in Euro-GASP survey suggest these trends should be interpreted with caution.

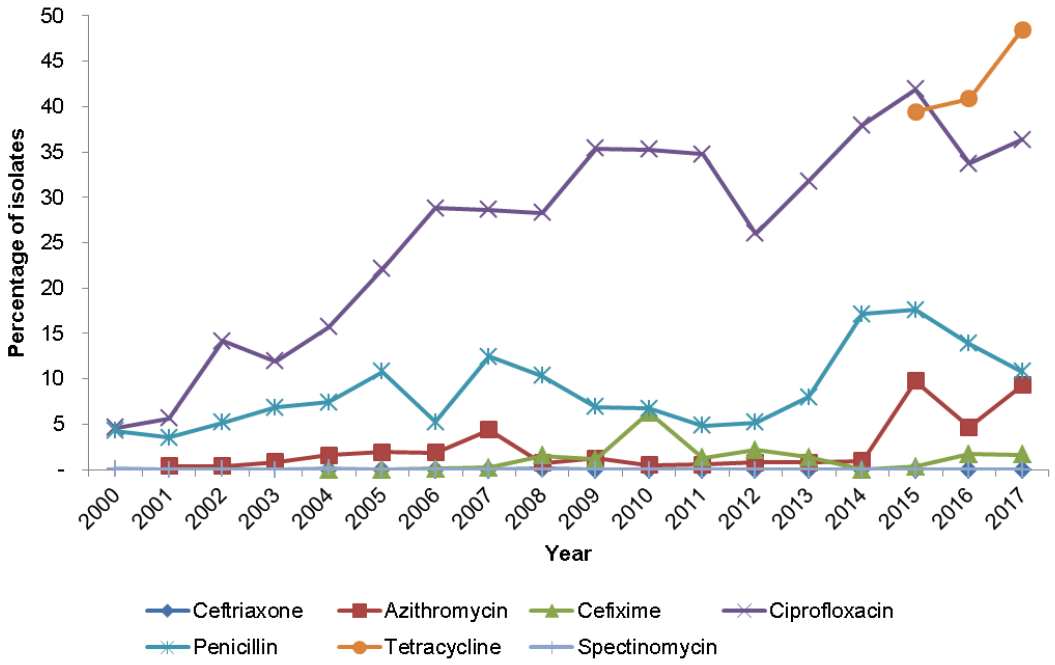


Figure 1.2 *N. gonorrhoeae* AMR identified in England: 2000 to 2017
Figure reproduced from the GRASP annual report[28]

1.7 Molecular epidemiology

The accumulation of genetic changes over time in lineages of *N. gonorrhoeae* enables molecular typing data to be used to investigate the distribution of different strains in circulation in the human population. In the field of molecular epidemiology, specimens (each being a sample of a pathogen) can be grouped according to similarities in molecular markers.[113] Specimens with similar genetic profiles are more likely to be related and the degree of similarity provides a proxy for time since their last common ancestor. Therefore, it follows that people infected with these strains are more likely to be closely linked in a sexual network. When combined with epidemiological data on time, person and place, these data can be used to describe

characteristics associated with sexual networks, and test hypotheses about how transmission occurs.

1.7.1 Phylogenetic analysis

Phylogenetic analyses based on single nucleotide polymorphisms (SNPs) between specimens provide a reconstruction of the relatedness of specimen sequences. In a phylogenetic tree, each sequence is represented as a tip node, and tip nodes are connected by internal nodes. The lengths of the edges (or branches) between nodes represents a measure of genetic distance. The internal nodes represent the points of genetic divergence between specimens. A variety of methods, including distance-based, maximum likelihood and maximum parsimony are used to build phylogenetic trees that reconstruct the position of each specimen relative to every other.[114,115]

Time-calibrated phylogenies can be used to estimate when genetic divergence occurred by estimating the time to most recent common ancestor between specimens. The rate at which genetic differences accumulate between specimens, known as the mutation rate,[116] and transmission chains with some estimate of directionality can be calculated from time-calibrated phylogenies.[117] The mutation rate in *N. gonorrhoeae* is estimated to be around three SNPs per year,[118-121] which is similar to the rates estimated in other important AMR pathogens such as *Streptococcus pneumoniae* and *Staphylococcus aureus*.[122] Knowing the approximate mutation rate is important in the context of phylogenetic and transmission network reconstructions. If the expected rate of genetic change for a given pathogen or genome section is slow or the rate of change is irregular (also known as “not clocklike”), genetic similarity may be less accurate as a marker of being part of the same transmission chain. Conversely, rapid changes in the genome might lead to isolates from closely linked people appearing distant on a phylogenetic tree and lead to the inappropriate inference that people are not linked in a network.[113]

Recombination events, where homologous DNA from an extraneous source are incorporated via horizontal transformation into a pathogen genome are an important driver of genetic diversity for organisms such as *N. gonorrhoeae*, and require careful consideration when constructing phylogenetic trees. These events tend to distort phylogenetic relatedness between two isolates and generally introduce more variation than would be expected from SNP mutation alone.[123] For example, the genomes of *N. gonorrhoeae* isolated from two people that were directly linked by a transmission event might appear very different if recombination events occurred

between transmission and diagnosis leading to the wrong conclusions that the specimens were not linked by a direct transmission event. In bacteria that frequently recombine such as *N. gonorrhoeae,* it is recommended that phylogenetic analyses are undertaken with caution and take into consideration the extent to which recombination has occurred.[123-125]

1.7.2 Molecular typing techniques

The typing methods used for *N. gonorrhoeae* molecular epidemiology have advanced in recent years, enabling more precise characterisation and comparisons of specimens. Initially, phenotypic methods were used to group specimens according to the bacterium's growth requirements or protein expression (autotyping and serotyping). Then, Sanger sequencing[126] and polymerase chain reaction methods were used to amplify specific DNA fragments. The molecular weights of these fragments were compared using gel electrophoresis (pulse field gel electrophoresis). These fragments could be specific genes (e.g. *opa* typing) or repeat regions of the genome (multiple-locus variable-number tandem repeat analysis). Other gel-based techniques cut the entire genome into fragments and compared the band patterns on gels (restriction fragment length polymorphism).

Further advances in DNA sequencing improved the amount and quality of genetic data available to allow specimen comparisons. Analysis of sequence data allowed characterisation of the nucleic acids that make up sections of a genome and facilitated the development of sequence typing schemes, based on the presence or absence of a distinct set of taxonomically relevant marker genes. The dominant typing schemes in *N. gonorrhoeae* are: *N. gonorrhoeae* multi-antigen sequence typing (NG-MAST) and multi-locus sequence typing (MLST). To discriminate between isolates, NG-MAST uses sequencing of an internal fragment of the porin PorB gene (*porB*) and subunit B of the transferrin-binding protein gene (*tbpB*),[127] both areas of the genome that mutate rapidly. MLST uses internal fragments of seven or more *N. gonorrhoeae* chromosomal housekeeping genes, which mutate less frequently than the NG-MAST loci.[128,129] The sequence information generated from NG-MAST or MLST is used to group specimens into sequence types. NG-MAST has been used within national AMR surveillance systems in the UK[31,105] to describe the antimicrobial susceptibility profile of common sequence types and identify clusters of related specimens. NG-MAST type may be useful to predict AMR

where no cultured isolate exists[130] and can help distinguish isolates that belong to an outbreak.[44]

Whole genome sequencing (WGS) is the latest sequencing-based typing technique to be used in molecular epidemiological studies of *N. gonorrhoeae*. This method generates the whole genome of a specimen and can be used to make comparisons at much higher resolution. WGS allows the detection of finer-scale differences than NG-MAST or MLST, as these methods only use a small part of the genome and may therefore misclassify isolates through over- or underestimating relatedness. The basis of the sequencing method remains the same as that developed by Sanger *et al* in 1977[126] but advances in automation and microchemistry have made the process faster and cheaper, enabling the whole genome and many specimens to be sequenced simultaneously. High-throughput WGS technologies, such as Illumina dye sequencing, have four key steps (Figure 1.3):

1) Library Preparation: DNA from the specimen is broken up into DNA fragments (the 'library') and tags are added to each fragment to track sequences that are close together in the genome

2) Cluster Amplification: the DNA fragments are added to a flow cell that contains short strands of DNA which are bound to the surface, the DNA fragments hybridise to the relevant complementary strand, and the DNA is amplified to create thousands of copies of each fragment

3) Sequencing: fluorescently tagged nucleotides are added to the flow cell along with other sequencing reagents. As each nucleotide is added and incorporated into the DNA sequence, a specialised camera takes an image of the cell and records the emission wavelength produced, the intensity of this wavelength indicates which nucleotide has been added. The sequence identified is called a 'read'. This step happens simultaneously across all the fragments creating millions of reads

4) Data analysis: reads with similar sequences cluster together. These clusters are mapped back to the reference genome and the consensus sequence for the specimen is determined

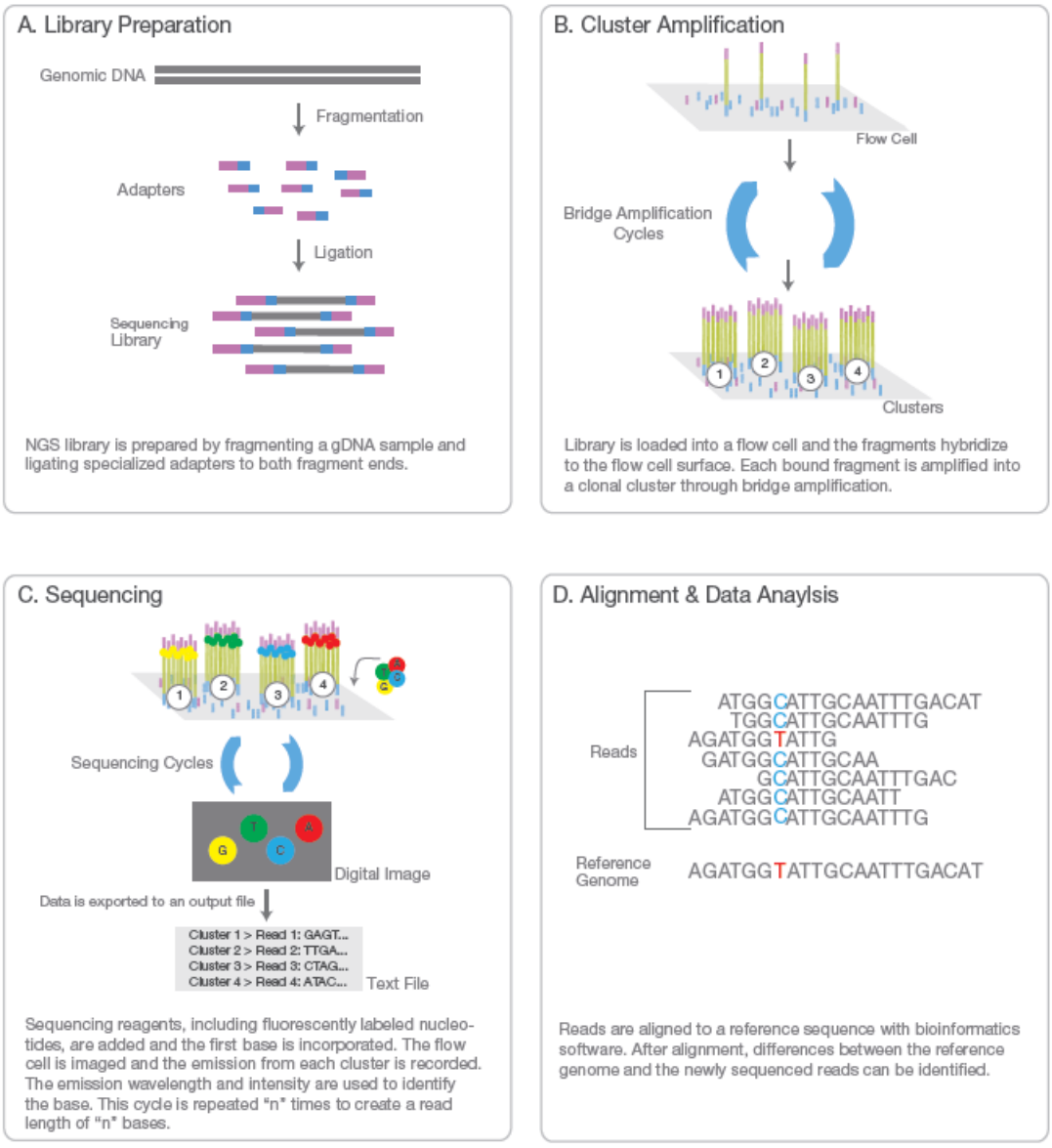


Figure 1.3 Illumina whole genome sequencing methodology
Adapted from "An introduction to next generation sequencing by Illumina technologies".[131]

1.7.3 Molecular epidemiology using whole genome sequencing

The number of published *N. gonorrhoeae* molecular epidemiology studies using WGS methods has expanded considerably over the last five years, as the speed of WGS increases and costs decline. Most studies have focused on exploring the genetic determinants of AMR or investigated the emergence and spread of AMR in the gonococcal population.[15,120,132-155] These studies support the development of point of care tests (POCT) that guide patient treatment based on the predicted antimicrobial susceptibility of the infection, although no POCT is currently being used routinely in clinical settings.[149,156-160]

Studies that combine WGS data with patient demographic and behavioural information provide insights into the spread of *N. gonorrhoeae* and AMR through different sexual networks.[118,119,121,140,161-168] For example, in 2014, Grad *et al.*[121] investigated the spread of cefixime resistant *N. gonorrhoeae* across the USA by conducting phylogenetic analyses combined with data on time, location and patient sexual orientation. The researchers found that cefixime resistance appeared on the west coast and spread eastwards primarily within MSM networks with a small number of introductions into heterosexual networks. Most of the WGS studies to date have had restrictive sampling criteria and sparse epidemiological information available for the study patients. There have been no WGS studies in England including specimens from multiple clinics and people from different demographic and sexual behaviour groups. In Chapter 2, I present a systematic review in which I identify and describe studies to date that combine molecular typing data and patient epidemiological information to investigate the spread of *N. gonorrhoeae* in sexual networks.

1.8 Thesis rationale, aim and research questions

1.8.1 Rationale and aim

The spread of *N. gonorrhoeae* is of global health concern due to the persistent increase in diagnoses and development of AMR. Novel methods for investigating how *N. gonorrhoeae* and resistant infections spread within sexual networks could support the public health response and improve control. WGS is a relatively new typing technique that holds enormous potential to provide greater insight into the epidemiology of gonorrhoea and AMR and the structure of sexual networks. These sexual networks are of particular relevance for public health and clinical care because they facilitate *N. gonorrhoeae* transmission in England. The aim of my PhD thesis was to explore the application of WGS to inform the control of *N. gonorrhoeae* and associated AMR. My work had a number of important strengths that allowed me to build on and improve upon previous WGS studies by selecting a large, broadly representative sample of *N. gonorrhoeae* from England that included all main risk populations and spanned four years, thereby improving generalisability of the findings from phylogenetic analyses.

1.8.2 Research questions

The following research questions underpinned the research presented in this thesis:

- How have *N. gonorrhoeae* molecular epidemiology studies to date been used to inform understanding of gonorrhoea epidemiology and improve infection control? (Chapter 2)
- What is the genetic variation of *N. gonorrhoeae* in England within and between different geographical areas and how does this vary over time? (Chapter 5)
- How can WGS data combined with patient epidemiological data be used to identify and describe gonorrhoea transmission networks? (Chapter 5)
- What is the genomic and phenotypic variation in *N. gonorrhoeae* antimicrobial susceptibility in England? (Chapter 6)
- What are the characteristics of sexual networks in which *N. gonorrhoeae* with reduced susceptibility to antimicrobials disseminates? (Chapter 6)
- What is the relationship between travel-associated sexual partnerships and infection with *N. gonorrhoeae* exhibiting reduced susceptibility to antimicrobials in England? (Chapter 7)
- How genetically similar is *N. gonorrhoeae* in England to those circulating globally? (Chapter 7)
- Which patient groups are more likely to be infected with globally circulating *N. gonorrhoeae*? (Chapter 7)

Other chapters in this thesis cover the methodology used for data collection (Chapter 3), a description of the epidemiological and phenotypic characteristics of the specimens used in the PhD along with an assessment of data quality for the WGS data I created (Chapter 4), and a discussion of the overall findings of the PhD research and conclusions on how WGS data can be used to improve the public health response and control of *N. gonorrhoeae* (Chapter 8).

2 Systematic review of *N. gonorrhoeae* molecular typing methods to understand sexual networks and antimicrobial resistance transmission

2.1 Introduction

In this chapter, I present the systematic review conducted at the start of the PhD, which identified and described studies that combined patient and molecular typing data to investigate the spread of *N. gonorrhoeae* in sexual networks. Although there have been previous reviews focusing on the use of different molecular typing techniques,[89,129,169] there has been no systematic review of how these data were used to understand sexual networks for *N. gonorrhoeae* transmission and the value for public health interventions that aim to control the spread of infection and AMR. The aim of the review presented in this chapter was to identify and summarise studies that have linked patient demographic, behavioural or clinical information to molecular data from sequence-based DNA typing of *N. gonorrhoeae*. I have focused on sequence-based DNA typing techniques that were recommended for *N. gonorrhoeae* molecular epidemiology:[129] MLST, NG-MAST and WGS.

The objectives of the review were to describe:

- the extent to which *N. gonorrhoeae* molecular typing studies have linked molecular typing to epidemiological data
- the patterns of *N. gonorrhoeae* transmission in sexual networks, as identified through the linked molecular and epidemiological data
- the public health actions proposed or resulting from the combined molecular and epidemiological data that support control measures to reduce the transmission of *N. gonorrhoeae*

The review informed the development of the PhD study and subsequent analyses presented in later chapters of the thesis, and has been published in the *Journal of Infection*.[1]

2.2 Methods

2.2.1 Protocol

The protocol for this systematic review was pre-specified and published on the international prospective public register of systematic reviews, PROSPERO.[170]

2.2.2 Search strategy

Five databases (Web of Science,[171] Scopus,[172] MEDLINE,[173] EMBASE,[174] and the Cochrane library[175]) were systematically searched using the search terms detailed in Table 2.1. These databases cover the biomedical literature most likely to contain studies relevant to the study objectives. The search terms were identified by selecting subject headings related to *N. gonorrhoeae* and the typing methods of interest (NG-MAST, MLST, and WGS). Other specific terminology related to these methods such as "molecular epidemiology" and "genotype" were also included. These search terms were used in a subject heading search and a keyword search. All terms were searched individually and then combined with results from the *N. gonorrhoeae* search. The results of each database search were exported to Endnote[176] where duplicates were discarded.

Abstract books for the following conferences were searched for relevant studies: World STI & HIV Congress (2013 & 2015), International Union against STIs (IUSTI) Europe (2014 & 2015), IUSTI Asia-Pacific (2014), British Association of Sexual Health and HIV (BASHH) (2014 & 2015), STD Prevention Conference (2014), European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) (2014 & 2015), Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC) (2014 & 2015) and the Federation of European Microbiological Societies (FEMS) (2013 & 2015).

Table 2.1 Systematic review search terms

Search concept	Subject heading	Keyword group
Neisseria gonorrhoeae	"neisseria gonorrhoeae"	"neisseria gonorrhoeae" or "n gonorrhoeae" or "gonococcus neisseria" or "micrococcus gonorrhoeae" or "neisseria gonococcus" or "neisseria gonorrhoea" or "neisseria gonorrhoea"
	gonorrhea	gonorrhea or "gonococcal infection" or "gonococcal infections" or gonococcosis or "gonococcus infection" or "gonorrhea epidemiology" or "gonorrhea treatment" or gonorrhoea or "neisseria gonorrhoeae infection"
	"gonococcal arthritis"	"gonococcal arthritis" or "gonococcal septic arthritis" or "gonorrheal arthritis"

	“gonococcal conjunctivitis”	“gonococcal conjunctivitis” or “gonococcal opthalmia” or “gonorrheal conjunctivitis” or “gonorrheal opthalmia” or “neisseria gonorrhoeae conjunctivitis”
	“gonococcal urethritis”	“gonococcal urethritis” or “gonococal urethritis” or “gonococcic urethritis” or “gonococcus urethritis” or “gonorrheal urethritis” or “gonorrhoic urethritis” or “specific urethritis” or “urethritis gonorrhoica”
		"gonococcal strain"
		"gonococcal strains"
Neisseria gonorrhoeae multi-antigen sequence typing	*no subject headings*	“neisseria gonorrhoeae multiantigen sequence typing”
		“neisseria gonorrhoeae multi antigen sequence typing”
		“neisseria gonorrhoea multi antigen sequence typing”
		“neisseria gonorrhoea multiantigen sequence typing”
		“gonorrhoea multiantigen sequence typing”
		“gonorrhea multiantigen sequence typing”
		“NG-MAST”
		“NGMAST”
		“neisseria gonorrhoeae MAST”
		“gonorrhoea MAST”
		“gonorrhea MAST”
		“multiantigen sequence typing”
		“multi antigen sequence typing”
		“multi antigen sequence type”
		“multiantigen sequence”
		“multi antigen sequence”
Multilocus sequence typing	“multilocus sequence typing”	“multilocus sequence typing” or “multilocus sequence analysis”
		"multi locus sequence typing”
		“multilocus sequence data”
		“multi locus sequence data”
		“multilocus sequencing”
		“multi locus sequencing”
		“multilocus typing”
		“multi locus typing”
		“multilocus type”
		“multilocus typing”
Genome sequencing	"Genome"	“genome” or “genome components” or “genome, protozoan” or “protozoan genome”
	"bacterial genome"	"bacteria genome" or "bacterium genome" or "genome, bacterial"
	"microbial genome"	"microbial genome" or "prokaryote genome" or "prokaryotic genome"
	"genome analysis"	“genome analysis” or “genome organisation” or “genome organization”
	"gene library"	“gene library” or “genome library” or “genomic library” or “library, gene”
	"metagenome"	"metagenome" or "metagenomes" or "metagenomic" or "metagenomics"
	"genomics"	"genomic" or “genomics” or “biomics”
	“genome DNA”	“genome DNA” or “DNA, genome” or “DNA, genomic”
	"phylogenomics"	"phylogenomics" or "phylogenomic"

	“gene mapping”	“gene mapping” or “gene map” or “genetic map” or “genetic mapping” or “genome mapping” or “genomic mapping” or “map, gene” or “mapping, gene” or “mapping, genetic” or “mapping, genome”
		"whole genome sequencing" or "whole genome sequence" or "whole genome sequences" or "whole genomic sequence" or "whole genomic sequencing" or "genome sequencing" or "genome sequence" or "genome sequences" or "genomic sequencing" or "genomic sequence" or "genomic sequences"
Molecular typing	"molecular typing"	"molecular typing"
		"genetic typing" or "gene typing" or "genome typing" or "genomic typing" or "genomes typing"
Phylogenetics	"molecular phylogeny"	"molecular phylogeny" or "genetic relationship" or "relationship, genetic"
	"phylogeny"	"genealogical relationship" or "phylogenesis" or "phylogenesis model" or "phylogenesis analysis" or "phylogenetic analysis” or "phylogenetic relationship" or “phylogenetics" or "phylogeny model" or "phylogeny relationship" or "relationship, genealogical" or "relationship, phylogenetic” or "phylogenetic” or "phylogeny"
	"phylogenetic tree"	"phylogenetic tree” or "dendrogram" or "evolution tree" or "evolutionary tree" or “phylogeny tree"
Molecular epidemiology	"molecular epidemiology"	"molecular epidemiology” or "epidemiology, molecular"
Genotype	"genotype"	"genotype" or "genotyping"

2.2.3 Study inclusion and exclusion criteria

After deduplication, each remaining study was independently considered by myself and another reviewer against the eligibility criteria. Studies were included if at least one of the sequence-based DNA typing methods of interest were used (NG-MAST, MLST, WGS) and the typing data were linked to patient-level epidemiological data. I considered epidemiological data to include demographic data about the patients (e.g. gender, age, and ethnicity), the patient’s sexual behaviours (e.g. sexual orientation, sex work, condom use, number of partners, sex abroad) or characteristics about the patient’s infection (e.g. symptoms, site of infection, concurrent STIs, HIV status). All types of studies with primary data published before April 2017 were included.

Studies were excluded if the articles were not in English, if the typing data were only linked to the geographical location and/or date of the isolates and not to patient-level epidemiological data, if typing data were only used to investigate the pathobiology of *N. gonorrhoeae*, or to develop methods for typing of *N. gonorrhoeae.* Review articles were excluded.

The title and abstract of articles were initially reviewed. Then, if considered eligible, the full text of articles was assessed. In cases of disagreement between reviewers, eligibility was resolved through discussion. Where no full text was available online, study authors were contacted by email to request a copy of the paper. Authors of conference abstracts were also contacted for posters or presentation slide sets.

2.2.4 Data extraction and synthesis

Independent data extraction was undertaken by both reviewers using a standardised form I developed. Common themes were identified from the data extracted and are presented using a narrative approach. Risk of bias was assessed by considering (1) selection bias affecting the interpretation and generalisability of results based on the specimens with typing data available and the wider population that the study aimed to represent, (2) missing data bias, and whether data were missing at random, and (3) potential reporting bias, particularly for reported behavioural variables. Disagreements were resolved through discussion between the two reviewers.

2.3 Results

2.3.1 Article screening

In total, 4,759 studies were identified; of which 2,101 were unique (Figure 2.1). Most studies were excluded following title and abstract review (94%; 1,982/2,101). Of the remaining 119 studies, 50 met the study eligibility criteria (42%; 50/119) during the full text review. Studies were most commonly excluded at this stage because no patient-level epidemiological data were reported (58%; 69/119). One study could not be reviewed as the full text could not be accessed. Nine conference abstracts were identified, four of which have been subsequently published. Authors for the remaining five were contacted for the presented poster or slide set but none replied (Figure 2.2). Therefore, overall, 49 studies were included in the review. A summary of all the included studies is presented at the end of this chapter (Table 2.3) and a more detailed table with study aims and conclusions is presented in the Appendix (Table 10.1).

2.3.2 Molecular typing methods

Nearly all the studies used NG-MAST (82%; 40/49).[44,127,130,177-212] Of the remaining studies, two used MLST (4%; 2/49)[213,214] and seven used WGS (14%; 7/49).[118,119,121,135,141,144,162] Each year from 2004, three to five studies using NG-MAST were published. This declined in 2015 and 2016 to one per year. I observed a

corresponding rise in the number of publications using WGS during these years (Figure 2.3). The WGS studies also determined the NG-MAST and MLST strain types *in silico* and compared the findings of each typing method. The NG-MAST and MLST strain types broadly correlated with the clustering of isolates in the WGS phylogenetic analyses. However, the WGS data enabled further differentiation of these groups and showed that the other methods sometimes misclassified clusters of specimens. For example, using WGS data, Demczuk *et al*.[135] identified three genetically distinct clusters of specimens that had the same NG-MAST strain type, and Didelot *et al*.[119] showed that specimens with an identical NG-MAST type found at the same time and location, were genetically distinct using WGS.

2.3.3 Study sampling strategies

A third of studies were conducted using specimens from the UK (37%; 18/49).[44,118,119,127,130,162,177,179,185,186,189,190,194,198,202,212,215] The remaining studies originated from across the globe, including from China,[183,200,211] Taiwan,[182,184,210] Australia,[196,208,214] Canada,[135,201,204] and the USA.[121,141,178] There were no studies from Africa or South America. The majority of specimens were selected from people attending specialist STI clinics or were specimens sent routinely to national reference laboratories. A third of studies (35%; 17/49) only selected resistant specimens or specimens of a particular strain type previously identified. A third of studies selected specimens from consecutive patients (35%; 17/49).[118,130,177,178,183,186-190,192,193,200,203,210,214,215] Study sample sizes ranged from six to 3,326, and included a variety of different patient groups, such as MSM and sex workers. Over a third of studies covered up to one-year (37%; 18/49),[121,127,162,177,178,183,187,188,192,196,199-203,212,214,215] but some extended to several years with varying sample sizes each year (range: less than six months up to nine years). Several studies included non-consecutive years. The earliest isolates used were collected in 1989.[135]

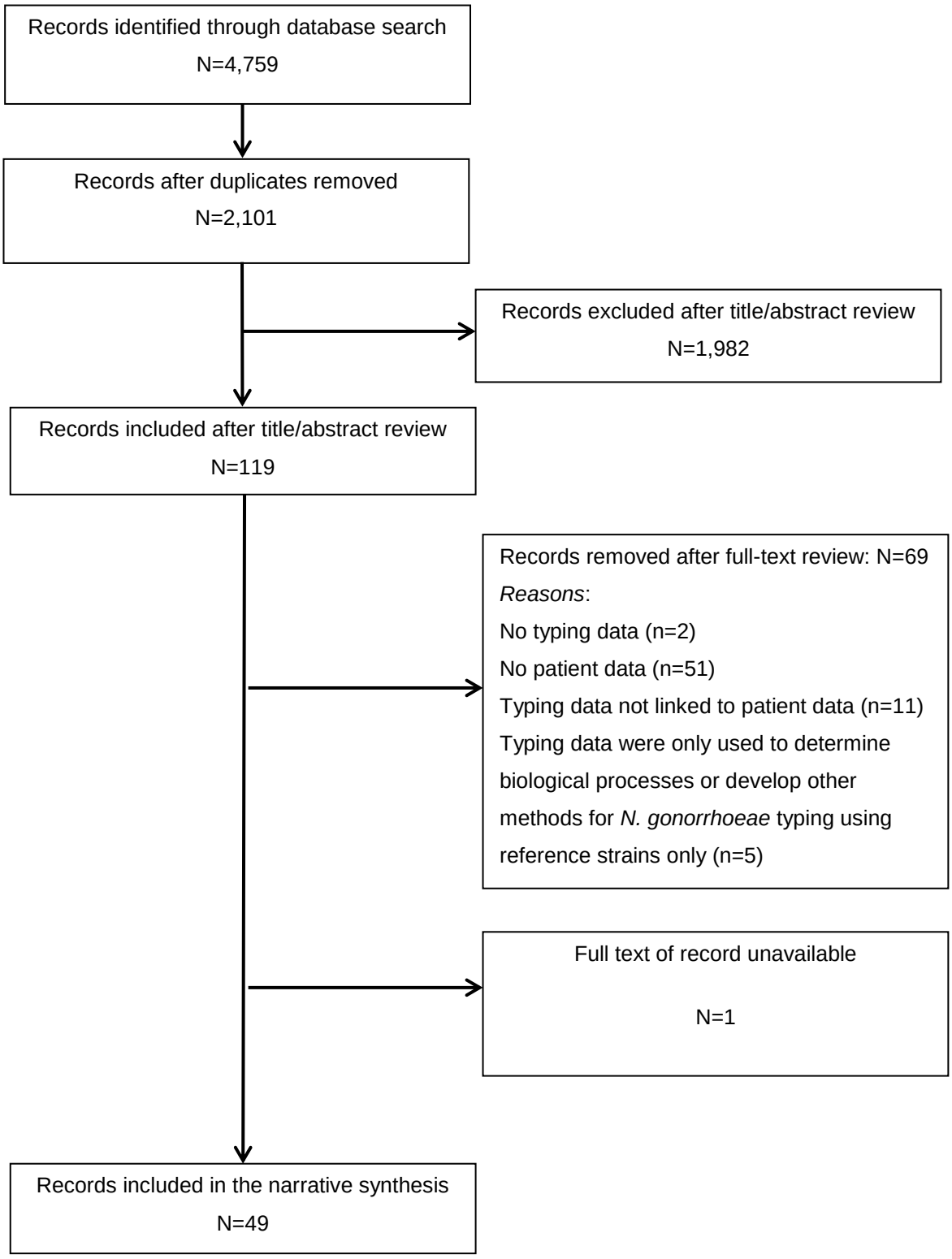


Figure 2.1 PRISMA flowchart indicating the systematic selection of journal articles for inclusion in this review

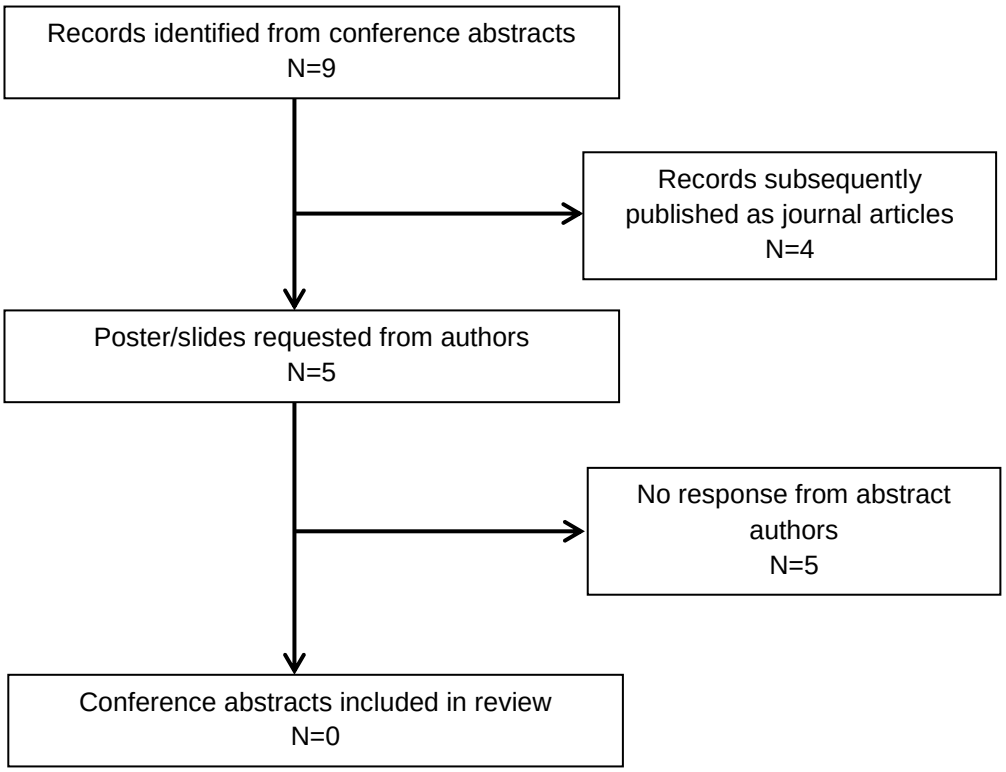


Figure 2.2 PRISMA flowchart indicating the systematic selection of conference abstracts for inclusion in this review

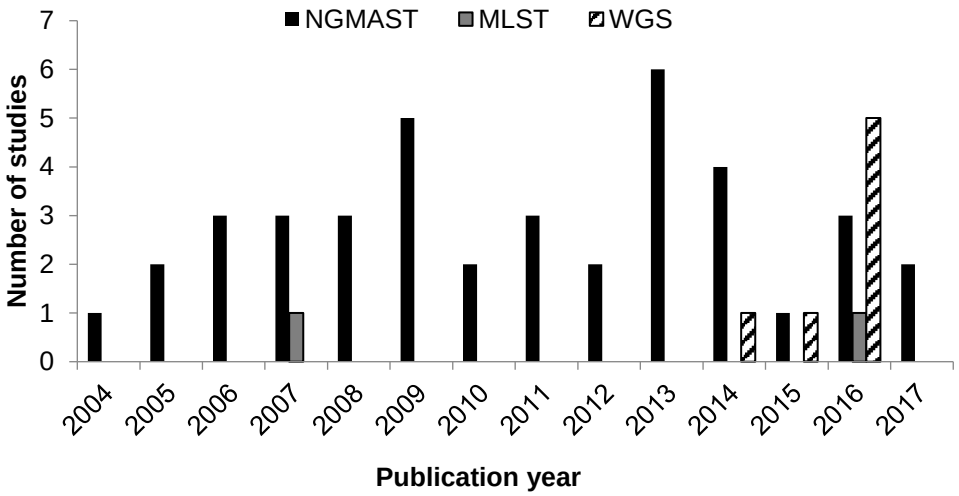


Figure 2.3 Number of studies by typing method over time
Total number of studies = 49

2.3.4 Risk of bias

Most studies applied specific selection criteria for specimens, which were mainly related to the setting, location and period for data collection. Some studies also placed restriction criteria on the antimicrobial susceptibility profile of the specimens, typically to include specimens that were resistant. These restrictions may limit the generalisability of the results. The majority of studies included ≥70% of the specimens eligible under the predetermined selection criteria for the study. Where reported, the main reasons for not including all eligible specimens were that the specimen could not be retrieved from the archive or the typing method failed. Where described, missing epidemiological data within each study was minimal. Most studies did not define how patient-level variables were collected; it was therefore often difficult to assess reporting bias. A summary of the risk of bias assessment for all included studies is presented in the Appendix (Table 10.2).

2.3.5 Patient-level epidemiological data

The most commonly reported and linked patient-level epidemiological data were gender (reported 100%; 49/49, linked 90%; 44/49),[44,118,119,127,130,162,177-181,183,185-190,192-194,197,200-206,208,210-212,214,215] and sexual orientation (reported 82%; 40/49,[44,119,121,127,130,141,144,162,177,178,180-192,194-199,201,202,204-207,209,212,215] linked 71%; 35/49).[44,119,121,127,130,141,162,177,178,180-192,194,195,197,199,201,202,204-206,210,212,215] Other patient-level epidemiological data reported but not always linked to typing data included age, sexual partner history (such as the number of sexual partners), location of patient residence or clinic, site of infection, travel-associated sexual partnerships, STI history, patient HIV status, patient ethnicity, infection symptoms and sexual behaviour (Figure 2.4).

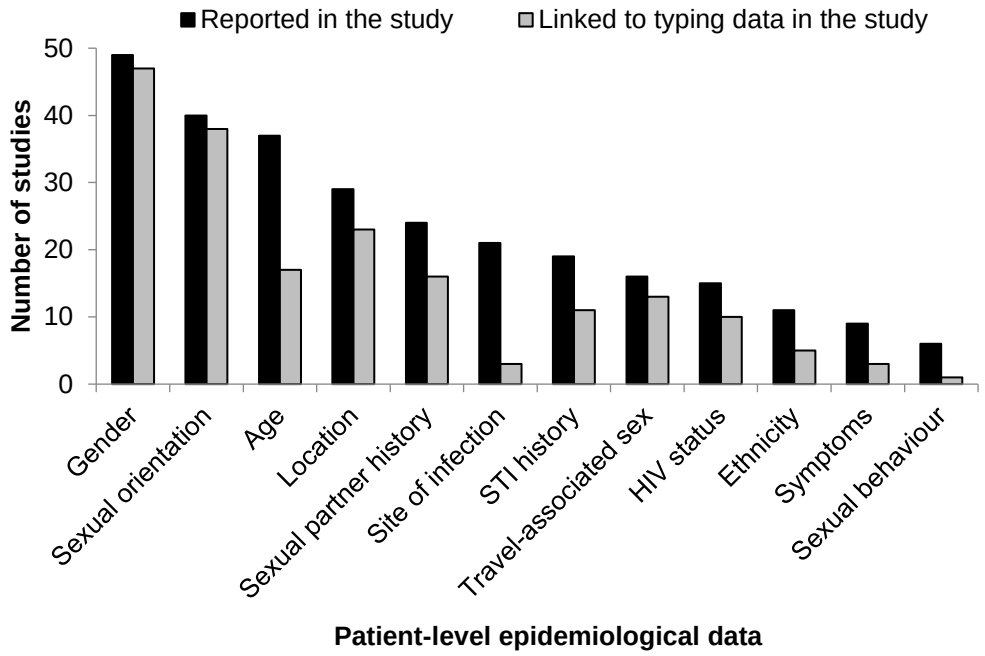


Figure 2.4 Reporting and linkage of patient-level epidemiological data to typing data
Total number of studies = 49

2.3.6 Summary of study findings

2.3.6.1 Identifying clusters of infection and sexual networks

In many studies, molecular typing data were used to identify specimens that were genetically similar and classify them into clusters. These studies inferred that specimens with the same NG-MAST or MSLT type, or few SNP differences between whole genome sequences, were closely related and that the people infected were more likely to be part of the same sexual network. Most studies defined a cluster by the presence of two or more specimens with identical NG-MAST or MLST types.[127,187,189,193,198,206] Most studies identified a large number of different sequence types. Many of these sequence types had not been identified before ('novel' sequence type) and/or were represented by one specimen ('unique' sequence types).[187,191-193] For example, Chisholm *et al.* (2013)[187] found 406 different NG-MAST types amongst 1,066 specimens. Over half of these were novel (216/406) and/or were unique (281/406). Other studies with smaller samples also identified a large number of sequence types. For example, Florindo *et al.* (2010)[191] found 104 different

NG-MAST types in 236 isolates, more than half were novel (60/104) and/or were unique (69/104).

2.3.6.2 Describing sexual networks using typing and patient-level epidemiological data

Patient-level epidemiological data were used to characterise and compare clusters. For example, Horn *et al.* (2014)[193] found specimens classified as NG-MAST genogroup 25 were predominantly from women, whereas specimens classified as NG-MAST genogroup 1407 were predominantly from men. Usually, only the most common sequence types making up the largest clusters in the study sample were described like this.[130,177,178,184,188-191,193-195,198,202,204,207,210] For sequence types associated with one group of people, such as MSM, authors inferred that this represented a discrete sexual network. Examples of studies comparing patient-level epidemiological data of different clusters are described below.

(i) Sexual orientation:

In total, 38 studies presented data on clusters by sexual orientation. The majority (87%; 33/38) found that sequence types from MSM and heterosexual populations clustered separately. For example, Cole *et al*. (2013)[189] demonstrated that NG-MAST types 147, 4, 1634 and 64 were more likely to be from MSM than heterosexual people, and Florindo *et al*. (2010)[191] found that NG-MAST type 783 and 1318 were associated with specimens from heterosexual people. However, sequence type clusters that were associated with a particular sexual orientation also contained a small number of specimens from other populations. For example, Wong *et al*. (2008)[210] found that NG-MAST type 547 was associated with MSM but four of the 16 isolates were from heterosexual men.

There was also evidence of the same sequence type being associated with different sexual orientations in other studies. For example, using specimens collected in 2003, Abu-Rajab *et al.* (2009)[177] found that the NG-MAST type 210 was associated with heterosexuals in Glasgow, Scotland, and a year later in 2004, the same sequence type was found to be associated with MSM in London, England, by Choudhury *et al*. (2006).[188]

(ii) Sub-groups within MSM and heterosexual populations:

For half of the studies that included sexual orientation (45%; 17/38), further characterisation of the networks was possible from the additional patient-level epidemiological data that were available. For example, Choudhury *et al.* (2006)[188] found that specimens from heterosexual people had different NG-MAST types depending on the ethnicity and age of the person, suggesting separate sexual networks within the heterosexual population. They did not find any significant differences between the characteristics of MSM and different sequence types. However, Bernstein *et al.* (2013)[178] found that different NG-MAST types were associated with specific characteristics of MSM. For example, NG-MAST type 2992 specimens were more likely to be found in MSM who reported oral sex only, and other sequence types were found in MSM reporting three or more sexual partners compared to MSM who reported fewer sexual partners.

(iii) HIV status:

One fifth of studies (20%; 10/49) used HIV status of the patient to investigate and describe clusters of gonococcal infection.[178,182,184,188-191,206,207,210] Fernando *et al.* (2010)[190] found NG-MAST sequence types that were unique to people living with HIV and concluded that this sequence type could be used to identify HIV-negative people at higher risk of HIV infection if they were infected with this sequence type, as this is evidence that these people are having condomless sex in a sexual network including people living with HIV. However, Bernstein *et al.* (2013)[178] and Cheng *et al.* (2016)[184] did not find an association between specific sequence types and HIV status in their samples. Using WGS data, Didelot *et al*.[119] found that *N. gonorrhoeae* was more likely to be transmitted between two people living with HIV than between two serodiscordant individuals.

(iv) Travel-associated sexual partnerships:

The role of travel-associated sexual partnerships on *N. gonorrhoeae* transmission and acquisition was investigated using molecular typing data in 13 studies.[118,119,127,177,183,185,188,189,192,198,202,208,212] In some studies, the assumption made was that unique sequence types were not likely to have been acquired locally, whereas common sequence types that formed clusters

were more likely to be locally acquired.[188,198,202,212] Fernando *et al.* (2009)[190] found that people infected with *N. gonorrhoeae* with a unique sequence type were more likely to have reported recent sexual contacts from outside the local area. Choudhury *et al.* (2006)[188] also found that people who reported sex outside of the UK were more likely to be infected with *N. gonorrhoeae* with a unique sequence type. Similarly, Martin *et al.* (2005)[198] found that the common sequences types were more likely to be from people who did not report sex abroad.

2.3.6.3 Describing AMR within sequence types

In 83% (41/49) of studies, phenotypic AMR data were combined with molecular typing and patient-level epidemiological data to describe the distribution and spread of resistant *N. gonorrhoeae* infection in sexual networks or across populations. *N. gonorrhoeae* susceptible to antimicrobials was found to be more genetically diverse than non-susceptible *N. gonorrhoeae*, which tended to be more clonal.[180,199,213] For example, *N. gonorrhoeae* with decreased susceptibility (DS) to cefixime was usually found in specimens with NG-MAST type 1407.[180,181,185,191-195] However, azithromycin resistance tended to be identified sporadically rather than associated with a particular clone,[141,206] although high-level azithromycin resistance (usually defined as specimens with an MIC ≥256 mg/L) was found in clusters of genetically similar *N. gonorrhoeae*.[162,186,200,211]

The combination of molecular, epidemiological and phenotypic data provided greater insight into the spread of antimicrobial resistant *N. gonorrhoeae* in sexual networks and across populations.[121,130,135,178,184,186,187,198,207] For example, using WGS data, Grad *et al.* (2014)[121] reported that *N. gonorrhoeae* with DS to cefixime first appeared on the west coast of the USA and then spread eastwards primarily within MSM networks, but with a small number of introductions into heterosexual networks. Authors often compared their findings to other studies in order to make an assessment of whether there were sexual networks operating between cities or countries and how this contributed to the spread of resistant infection. For example, Chisholm *et al.* (2009)[186] speculated that high-level azithromycin resistant *N. gonorrhoeae* identified in Liverpool, England has spread from Scotland because the NG-MAST type (ST649), AMR phenotype and sexual orientation of the infected patients matched those of the cases previously reported by Palmer *et al.* (2008)[130]

2.3.7 Public health application of molecular typing data

Similar suggestions about how molecular typing data along with epidemiological data could be used in public health were made across many of studies (Table 2.2). These suggestions included using molecular typing data to confirm outbreaks of new strains, including antimicrobial resistant infections.[44,127,162,196,212] Other suggestions included using the molecular typing data to assess the effectiveness of partner notification in a population. Monfort *et al*. (2009)[199] suggested that contact tracing had not been effective in the population they studied because of the high number of specimens that did not cluster with any other specimen in the study sample – suggesting many undiagnosed infections that might have been identified through partner notification. Other studies suggested molecular typing data could help complement or confirm information provided through contact tracing.[44,118,119,121,162,177,179,188,189,202,209,215] Use of the molecular data in the clinical setting to inform clinical management of the infected patients was suggested by over one third of studies (39%; 19/49).[44,127,141,144,162,177,178,180-184,187,189,190,202,206,210,214]

Table 2.2 Summary of the suggested applications of molecular typing data within public health

Table adapted from the published article[170]

Topic	Application of use of molecular typing data
Outbreak investigation	• the genetic similarity of *N. gonorrhoeae* infections was used to confirm whether cases were related and part of an outbreak[44,162,196] • WGS data were used within mathematical models to estimate the number of undetected cases in an outbreak[119]
Connections between people in sexual networks	• genetic similarities or differences between *N. gonorrhoeae* was suggested to be of use for sexual partner tracing and could complement partner notification data[118,119,121,177,179,188,189,202,209,215] • WGS could be used to identify likely transmission events between people and describe the extent of sexual networks and speed of transmission between people[118,119]
Treatment	• identifying sequence types associated with antimicrobial resistance was suggested to be used to guide patient treatment[141,144,177,189,206,214]
Targeted health promotion/behavioural interventions aimed at controlling gonorrhoea	• identifying sexual networks could help quantify mixing between different populations, such as MSM and heterosexuals, which could help determine the effectiveness of delivering public health interventions to one group or the whole population[188,189] • services provided to patients could be adapted based on the sequence type of the infecting *N. gonorrhoeae*, for example, if the sequence type was associated with people living with HIV, it would be recommended that HIV-negative/unknown people infected with this *N. gonorrhoeae* sequence type are tested for HIV and provided with other appropriate health promotion information[118,177,178,190]

2.4 Discussion

This systematic review identified 49 studies that used molecular typing data linked to patient-level epidemiological data to describe sexual networks infected by *N. gonorrhoeae* and AMR infection. NG-MAST was the most commonly used typing method. Most studies identified a large number of sequence types, and many were novel and/or unique. Clusters of specimens with the same sequence type were used to infer sexual networks in the populations sampled and these sexual networks were described using the patient-level epidemiological data available. Sexual orientation and gender were the most commonly used patient-level epidemiological variables. A variety of different sampling strategies were used, such as restricting the inclusion criteria to specimens with a specific AMR phenotype or from a specific group of people, such as MSM. The public health applications of molecular typing data were hypothesised but not tested in the majority of studies included in this review.

2.4.1 Review strengths and limitations

The strength of the review was the systematic methods I used to identify and assess the relevant literature. The use of precise search terms applied to large medical literature databases reduced the risk of missing studies. Use of two independent reviewers to select studies and extract data minimised the risk of observer bias. A limitation of the review was that only English language studies were included.

2.4.2 Understanding sexual networks from molecular typing data

The included studies demonstrated that molecular data combined with patient-level epidemiological data can provide information about sexual networks. The typing data were particularly useful for interpreting the interaction of multiple different patient-level epidemiological variables in a sexual network. For example, Choudhury *et al.* (2006)[188] identified discrete sexual networks within heterosexual populations based on ethnicity and age using different strain types associated with these sub-groups. However, most studies found that specimens with the same sequence types were from people with heterogeneous characteristics, even if most specimens were from a broadly similar population. For example, within most MSM-associated clusters there were small numbers of specimens from heterosexual people. This might reflect data coding errors or reporting bias associated with sensitive sexual behaviour, i.e. these may be men who identify as heterosexual but have sex with men, and may bridge between sexual networks.

Mixed clusters might also occur due to misclassification of clustering when using NG-MAST and MLST, which could lead to incorrect inferences about relatedness within networks. WGS provides more genomic data for comparison and is likely to improve the accuracy of clustering. Only seven studies using WGS to assess sexual networks were identified in this review.[118,119,121,135,141,144,162] These studies found that phylogenetic analysis provided greater resolution to the grouping of specimens than NG-MAST or MLST.

Although most studies only analysed the most common sequence type clusters for associations with patient-level epidemiological data to describe sexual networks, many of these sequence type clusters contained fewer than 10 specimens. The small number of specimens within clusters was due to the large number of sequence types identified in nearly every study, many of which were unique. This heterogeneity of sequence types could limit the interpretation by inaccurately separating people into different sexual networks.

2.4.3 Sampling strategies

It is important to consider the study sampling strategy when interpreting the results of each study and its potential application public health interventions. There is a risk of low internal and external validity if restrictive sampling criteria or convenience samples are used without consideration of the specimens and/or population groups that are excluded. To improve the validity and utility of molecular typing data, the sampling strategy should aim to reduce the chance of selection bias, for example by including consecutive specimens across a representative sample of the population. However, it is important to note that the aim of the studies included in this review was not always to identify sexual networks.

2.4.4 Implications for my PhD

My systematic review has demonstrated the importance of including data that allows appropriate comparisons between different patient groups and AMR phenotypes in molecular epidemiological studies. By selecting one particular set of specimens, such as resistant infections, the interpretation of the findings is limited. To address this limitation, in my PhD, I selected consecutive isolates from five geographically dispersed clinics, so that specimens in my study were more likely to be representative of the wider *N. gonorrhoeae* population in England in terms of antimicrobial sensitivities and risk groups covered.

Table 2.3 Summary of studies included in the systematic review

Ordered by molecular typing method and year of publication. *1=sexual orientation or gender, 2=age, 3=location, 4=site of infection, 5=HIV status, 6=travel, 7=ethnicity, 8=symptoms, 9=other sexual behaviour, 10=AMR, 11 = STI history/co-infection

Reference	Study year(s)	Location	Sample size	Sample selection criteria for typing	Variables linked between molecular & patient data*
Studies using NG-MAST					
Martin (2004)[127]	2000	UK	464	Selected diverse strains; sexual contact pairs, alternate sampling of surveillance programme isolates, all resistant isolates from surveillance sample	1, 3, 6, 10
Martin (2005)[198]	2000-03	UK	192	Ciprofloxacin resistant isolates only	1, 6, 9, 10, 11
Palmer (2005)[212]	2002	UK	106	Isolates with RS or resistance to ciprofloxacin	1, 3, 6, 10
Palmer (2006)[202]	2003	UK	56	Only ciprofloxacin-resistant serogroup WI isolates were typed and were from a previous study	1, 6, 9
Lundback (2006)[197]	1999, 2002-04	Sweden	14	Azithromycin resistant isolates only	9, 10
Choudhury (2006)[188]	2004	UK	2,045	Any isolate from people attending STI clinics selected, these are part of the national sentinel surveillance programme	1, 2, 5, 6, 7, 8
Bilek (2007)[179]	1995-2000	UK	231	Known sexual contacts	9
Unemo (2007)[209]	2002-03	Sweden	47	Isolates selected based on serotype IB-10 only	1, 10
Risley (2007)[215]	2004	UK	1,882	Any isolate from people attending STI clinics selected, these are part of the national sentinel surveillance programme	1, 2, 3, 7
Palmer (2008)[130]	2004-07	UK	3,326	Any isolate from people attending STI clinics	1, 2, 3, 10
Wong (2008)[210]	2006-07	Taiwan	149	Consecutive diagnoses made in the clinic	1, 5, 10
Starnino (2008)[207]	2003-05	Italy	164	50% of isolates from consecutive sample from sentinel surveillance which had sexual orientation data	1, 3
Abu-Rajab (2009)[177]	2003-04	UK	170	Consecutive isolates	1, 2, 6, 9, 11
Starnino (2009)[206]	2007-08	Italy	22	Only azithromycin resistant isolates from men attending STI clinic selected	1, 10

Reference	Study year(s)	Location	Sample size	Sample selection criteria for typing	Variables linked between molecular & patient data*
Fernando (2009)[190]	2004-06	UK	370	All people with a culture-confirmed diagnosis attending the Edinburgh STI clinic	1, 3, 5, 9
Chisholm (2009)[186]	2006-07	UK	75	Purposive sampling of azithromycin resistant and azithromycin sensitive isolates from sentinel surveillance programme plus in areas with high-level azithromycin resistance	10
Monfort (2009)[199]	2006	France	93	Purposive selection of isolates with different AMR profiles, from different locations, of different genders and infected at different anatomical sites. Additional rectal samples from MSM to describe clusters within specific population	10
Starnino (2010)[205]	2003-05 & 2007-08	Italy	137	Ciprofloxacin resistant isolates from people that provided sexual orientation data	1
Florindo (2010)[191]	2004-09	Portugal	236	Over 100 local laboratories were asked to send NG isolates to national reference laboratory: 25 participated	1, 2, 10
Chisholm (2011)[185]	2005-09	UK	96	Only typed cefixime DS isolates identified within national sentinel surveillance programme	1, 6, 10
Yuan (2011)[211]	2008-09	China	17	Isolates exhibiting azithromycin resistance	10
Ota (2011)[201]	2006	Canada	104	Only ciprofloxacin resistant isolates typed	1, 10
Carannante (2012)[180]	2003-12	Italy	120	Samples from people who provided sexual orientation data from previous study (Starnino 2008), resistant isolates only	1
Hjelmevoll (2012)[192]	2009	Norway	126	Consecutive isolates	1, 10
Cole (2013)[189]	2005-06	UK	475	Consecutive isolates received by laboratories	1, 10
Bernstein (2013)[178]	2009	USA	212	Molecular typing data described for MSM only, sample from national surveillance programme GISP	1, 7, 9, 10
Chen (2013)[182]	2006-12	Taiwan	47	Only isolates identified as ST4378 from surveillance programme	1, 10
Ison (2013)[194]	2007-11	UK	534	Only cefixime DS isolates typed from sentinel surveillance programme	1, 10

Reference	Study year(s)	Location	Sample size	Sample selection criteria for typing	Variables linked between molecular & patient data*
Chisholm (2013)[185]	2010	Europe	1,066	Consecutive isolates from 21 European countries participating in sentinel surveillance programme	1, 3, 10
Singh (2013)[204]	2007-11	Canada	238	AMR specimens submitted to reference laboratory	1, 10
Horn (2014)[193]	2010-11	Germany	213	Consecutive isolates sent to the selected laboratories	1, 2, 3, 10
Carannante (2014)[181]	2003-12	Italy	81	Multi-drug resistant isolates identified by the network of participating laboratories	1, 10
Jeverica (2014)[195]	2006-12	Slovenia	194	Any available gonococcal isolate in Slovenia	1, 10
Cheng (2016)[184]	2006-13	Taiwan	1,090	Symptomatic patients at Taipei sexual health clinic	1, 10
Stevens (2015)[208]	2011-13	Australia	6	High-level azithromycin resistant isolates only	6
Chen (2016)[183]	2012-13	China	920	Selected isolates from clinics participating in the national sentinel surveillance programme for NG AMR	10
Foster (2016)[44]	2010-13	UK	284	Isolates available from outbreak and non-outbreak region plus isolates from sentinel surveillance in the outbreak area from previous years	1, 2
Ni (2016)[200]	2011-12	China	118	All NG cases with culture from one hospital	10
Lahra (2017)[196]	2016	Australia	28	Isolates exhibiting azithromycin resistance from national antimicrobial resistance surveillance programme	10
Serra-Pladevall (2017)[203]	2013	Spain	111	All isolates from heterosexuals attending STI clinic, random selection of MSM attending clinic	1, 10
Studies using MLST					
Perez-Losada (2007)[213]	2000-01	Israel	48	Random selection of resistant and susceptible quinolone isolates from previous study	9, 10
Trembizki (2016)[214]	2012	Australia	2,218	Consecutive culture isolates received by laboratories	1, 3, 10

Reference	Study year(s)	Location	Sample size	Sample selection criteria for typing	Variables linked between molecular & patient data*
Studies using WGS					
Grad (2014)[121]	2009-10	USA	236	Cefixime resistant plus sensitive isolates matched according to location, date and sexual orientation. Isolates from national sentinel surveillance programme	1, 3, 10
Demczuk (2015)[135]	1989 & 2013	Canada	169	Convenience sample with aim to include isolates that were across time, ceftriaxone DS, of different NG-MAST strain types and locations	1, 10
Chisholm (2016)[162]	2014-15	England	16	Eight cases defined as outbreak cases (high-level azithromycin resistant NG in Leeds) (seven sequenced), eight control strains chosen to be from similar area with varying azithromycin resistant profiles	3, 10
De Silva (2016)	2011-15	England	1,061	Consecutive culture positive NG samples from attendees at sexual health clinic and primary care. Additional convenience sample of previously published WGS data for comparison.	6
Didelot (2016)[119]	1995-2000 & 2004	England	237	Sheffield: isolates from one sexual health clinic identified as NG-MAST strain type 12. London: isolates from across city sexual health clinics identified as NG-MAST strain type 225	1, 2, 3, 5, 9
Grad (2016)[141]	2000-13	USA	1,102	Purposive selection of isolates based on antimicrobial susceptibility profile focused predominantly on isolates with RS to cefixime, ceftriaxone, azithromycin and quinolones; isolates from national sentinel surveillance programme	1, 3, 10
Jaccobson (2016)[144]	2009-14	Europe	75	Purposive selection of azithromycin resistant NG from 21 European countries participating in sentinel surveillance programme	3, 10

3 Data sources, sampling and collection methodology

3.1 Chapter summary

In this chapter I present the sources, sampling strategy and collection methodology for the epidemiological, antimicrobial susceptibility and WGS data of the *N. gonorrhoeae* specimens used in the PhD study.

3.2 Data sources

The *N. gonorrhoeae* specimens used in this PhD were selected from the archive of the Gonococcal Resistance to Antimicrobials Surveillance Programme (GRASP) held at Public Health England (PHE).[216] GRASP is a sentinel programme that collects consecutive specimens from all culture-positive *N. gonorrhoeae* diagnoses made in 26 specialist STI clinics in England and Wales over a three-month period (July-September) annually. The GRASP methodology prioritises rectal specimens over all other specimen types. The specimens are tested by agar dilution methods in the national reference laboratory at PHE for antimicrobial susceptibility to seven antimicrobials that were previously or are currently used for the treatment of gonorrhoea (penicillin, tetracycline, ciprofloxacin, cefixime, ceftriaxone, azithromycin and spectinomycin). Epidemiological data collected through the Genitourinary Medicine Clinic Activity Dataset (GUMCAD)[61] are linked to the antimicrobial susceptibility data of each specimen. GUMCAD contains longitudinal patient, clinical and behavioural data on all STI diagnoses made and services provided by specialist STI clinics and some non-specialist STI services, such as general practitioners in England. These data are supplemented by more detailed clinical and behavioural data directly from participating GRASP clinics.[104]

3.3 Sampling strategy

To investigate the genetic variation of *N. gonorrhoeae* over time, and across different geographic locations and sexual networks in England, I chose specimens that were broadly representative of *N. gonorrhoeae* circulating in England over a four-year period (2013-2016). To achieve this, I selected all specimens in the GRASP archive from five sexual health clinics that provided specimens covering the key population groups affected. The most recent years of GRASP data were selected in order to investigate prevailing trends and patterns. There was limited available information on

population variation of the *N. gonorrhoeae* genome in England, which prevented meaningful sample size calculations (discussed further in Chapter 8 Section 8.6.1).

GRASP provides a broadly representative sample of people and sexual networks at risk of gonococcal infection in England, but slightly oversamples MSM and people from London (see section 4.2).[217,218] Additionally, of necessity for antimicrobial susceptibility testing, GRASP only includes specimens of *N. gonorrhoeae* from cultures, and consequently may over-sample symptomatic or genital infections which are more likely to be successfully cultured than asymptomatic or pharyngeal infections.[219] Furthermore, not all diagnoses made in participating clinics are included because culture may not be attempted in all cases. Overall, just less than 50% of gonorrhoea diagnoses in England are cultured.[219]

3.4 Producing whole genome sequence data

The process of producing WGS data involved (i) retrieving the organisms from the archive, (ii) extracting the DNA from each specimen, and (iii) sequencing the DNA. I completed the retrieval and extraction using standard operating procedures developed by PHE and described below. Sequencing was performed by staff at the Wellcome Sanger Institute[220] and a consensus whole genome sequence for each sample was created using a computer programming script previously developed in-house at the Sanger institute.

3.4.1 Retrieving *N. gonorrhoeae* and DNA extraction

The specimens were previously confirmed to be *N. gonorrhoeae* and stored in Microbank beads at -80°C, as part of standard GRASP laboratory procedures.[104] For the PhD, one Microbank bead from each specimen archive was streaked across a GCVIT agar plate (non-selective agar (Difco BBL GC II Agar Base (Product Code 212171)) and 1% IsoVitalex) using the streak-plate technique to isolate single colonies.[221,222] Each specimen plate was incubated at 36°C with 5% CO_2 for 20 hours. If there was growth on the plate after this time, I proceeded to identify whether the organism was *N. gonorrhoeae*.[223] If there was no growth or the specimen was not identified as *N. gonorrhoeae*, retrieval was repeated once more during another run.

3.4.1.1 Identification

Three tests were used to identify pure colonies of *N. gonorrhoeae* for sub-culturing;

(1) Visual inspection of the agar plate for contaminants:

N. gonorrhoeae was visually identified by the presence of consistent translucent growth on the plate in the areas where the plate has been streaked (identifiable by the slight indent on the agar). Contaminants were often large opaque-orange circles in areas outside of the pattern of streaks. If there was translucent growth, I proceeded to the next stage of identification.

(2) Oxidase test:

N. gonorrhoeae is an oxidase positive bacterium meaning the cells produce cytochrome oxidase, which catalyses the transport of electrons from donor compounds to electron acceptors. The test reagent *N, N, N', N'*-tetramethyl-p-phenylenediamine dihydrochloride acts as an artificial electron acceptor for the enzyme oxidase. In the presence of *N. gonorrhoeae,* the oxidised reagent forms the coloured compound indophenol blue.

To create the reagent, 10µl of *N, N, N', N'*-tetramethyl-p-phenylenediamine dihydrochloride was added to 5ml sterile distilled water and the container immediately covered in aluminium foil to prevent light speeding up the process of oxidation. The test reagent was used immediately after production, as the oxidation process will continue despite being shielded from light.

Using a plastic pipette, one drop of reagent was added to filter paper. Using a disposable 1µl sterile plastic loop for each specimen, I picked one colony that had been visually identified as *N. gonorrhoeae* and rubbed this into the reagent on the filter paper. A positive reaction turned the spot blue/deep purple within 10 seconds and confirmed that the bacteria selected were oxidase positive. These positive specimens were selected for the next stage of *N. gonorrhoeae* identification.

(3) Gram stain microscopy:

Putative *N. gonorrhoeae* specimens from the first two stages of *N. gonorrhoeae* identification were selected for Gram stain microscopy. *N. gonorrhoeae* is a Gram-negative bacterium. Gram-negative bacteria do not retain the first dye (crystal violet

and iodine) applied during the Gram staining procedure because they do not have sufficient peptidoglycan in their cell wall.[224,225] Instead, *N. gonorrhoeae* has a thin peptidoglycan layer covered by an outer membrane, which cannot retain the first dye but is coloured by the counter stain (safranin). Additionally, they are round, often in pairs (diplococci), as opposed to rod shaped (bacilli) bacteria.

On a glass slide, individual drops of sterile distilled water were added. I picked one colony per specimen using a 1µl sterile plastic loop and glided the loop through one droplet of water to suspend the bacteria. The slide was placed on a heating block at 70°C to dry for five minutes. Once dry, the slide was flooded with crystal violet for 60 seconds and then washed with tap water. Then the slide was flooded with Gram's iodine for 60 seconds and washed with tap water. After this, the slide was washed with Gram's differentiator without leaving the solution on the slide for more than three seconds, then washed with tap water. Finally, the slide was flooded with safranin for 60 seconds, then washed with tap water.

Before viewing using a light microscope, the slide was thoroughly dried using blotting paper. A drop of immersion oil was added to each specimen on the slide and the slide was viewed under x 100 oil immersion lens. *N. gonorrhoeae* was identified if the cells were pink and round.

3.4.1.2 Sub-culture

Specimens that were successfully identified as *N. gonorrhoeae* were sub-cultured to attain pure growth. A single colony for each specimen was picked using a 1µl plastic sterile loop and streaked using the streak-plate technique on GCVIT and incubated at 36°C and 5% CO_2 for 20 hours.[221,222]

3.4.1.3 Cell lysis

N. gonorrhoeae grown on the sub-culture plates were visually inspected for contaminants. If contaminant-free, 1µl loop of bacterium was harvested from the plate and transferred to 250ml of lysis buffer (ATL buffer 226µl, RNase A 4µl, Proteinase K 20µl)[226] in a 96 well plate. Plates were sealed and incubated on a shaker for 30 minutes at 56°C. This procedure inactivated the bacteria and lysed the cell walls to allow access and extraction of the DNA from each cell.

3.4.1.4 DNA extraction

DNA was extracted from the lysed cell mix into 100µl eluate using the automated QIAsymphony DNA mini kit technology.[226] I used the QIAsymphony Tissue_HC_200_V7_DSP protocol, which was designed to extract DNA from ≥200µL of pre-lysed bacterial suspension and produced an elution of 100µl of DNA. The DNA plate was stored in an -80°C freezer until ready for transportation to the Wellcome Sanger Institute for WGS. The extracted DNA was transported on dry ice by courier.

3.4.2 Whole Genome Sequencing

At the Wellcome Sanger Institute, WGS was conducted using the Illumina HiSeq X Ten system (Figure 1.3). [131,227] For each specimen, the output from the Illumina sequencing process were short raw DNA sequences, known as reads. I used Sanger in-house computer programming scripts to align the raw reads to a reference genome used frequently in *N. gonorrhoeae* phylogenetics (FA1090) to create a consensus whole genome sequence for each specimen.[228] The script used the Burrows-Wheeler Aligner Maximal Exact Match (BWA-MEM) algorithm[229] with the option to flag duplicate shorter reads that match as secondary for removal (option –M). The Sequence Alignment/Map (SAM) file output was converted into a Binary Alignment/Map (BAM) file using SAMTools[230] to reduce the size of the file for faster computer processing. The Genome Analysis Toolkit (GATK)[231] was used to realign indels, which helps the process of identifying SNPs. SAMTools mpileup was used to identify the variant nucleotides identified in each read and the haploid option of Binary Call Format (BCF) tools from SAMTools filtered this information to select the variant nucleotides based on the following conditions: the minimum base call quality was ≥50 (quality of the base was previously determined using the Phred score system in SAMTools); the minimum mapping quality score by BWA-MEM was 20; at least eight reads have the same variant and at least three of these are from each strand direction (forward and back); that the specific variant called is the same in ≥80% of the reads used. The consensus sequence for each specimen was compiled into one multiple fasta file and used for the analyses presented in the next chapters.

3.5 Information governance and ethical approval

Information governance advice and ethical approval were sought from the PHE Research Ethics and Governance Group (REGG). The Head of this group (Dr. Elizabeth Coates) consulted with the Health Research Authority (HRA), who administers the NHS Research and Ethics Committee (REC) service, regarding whether NHS REC approval or PHE REGG approval was required. The HRA recommended PHE REGG approval, as there would be no breach of patient confidentiality during the research. The PHE REGG approved the research for the PhD. Key points related to the information governance procedures and patient confidentiality are outlined below.

3.5.1 Existing approvals for GRASP

PHE has permission to process patient confidential data obtained by GRASP under Regulation 3 of the Health Service (Control of Patient Information) Regulations 2002.[232]

3.5.2 Patient confidentiality

GRASP holds pseudo-anonymised patient data. Pseudo-anonymised data held on any individual person cannot be used independently to identify that person. However, pseudo-anonymised data can theoretically be linked to the original data source with patient identifying information, in this case the confidential patient clinical record held by the clinic. I did not have access to the clinical record held by the clinic and therefore could not identify any patient.

3.5.3 Materials transfer agreement

A materials transfer agreement was drawn up and signed by PHE and the Wellcome Sanger Institute. This stipulated that the sequencing data created would be stored and partially analysed on Sanger Institute servers. When published, the sequence data will be made publically available, free of charge, on the European Genome-phenome Archive,[233] for other researchers to use. Data will be held by the Sanger Institute until data analysis is complete. Any residual DNA material will be held for five years by the Sanger Institute as per standard protocol.

4 Epidemiological characteristics, antimicrobial susceptibility profile and genome sequence quality of the PhD study sample

4.1 Chapter summary

In this chapter, I compare the epidemiological characteristics of specimens successfully sequenced and used for phylogenetic analyses in the subsequent chapters (hereafter referred to as the PhD study sample) with those of the original sample selected for sequencing and all gonorrhoea diagnoses in England. I also compare the phenotypic antimicrobial susceptibility profile of the PhD study sample with all GRASP specimens in the study period and I present data assessing the quality of the whole genome sequences produced.

4.2 Epidemiological characteristics

4.2.1 Comparison of the PhD study sample with the original sample selected for sequencing

Using the sampling strategy outlined in Chapter 3, 1,407 *N. gonorrhoeae* specimens were eligible for inclusion in this study. Of these, 1,277 (91%) were sequenced and 130 (9%) were not because the bacteria were no longer viable for DNA extraction. A significantly higher percentage of specimens that were not sequenced were from 2016 (not sequenced - 52% (68/130), sequenced – 20% (251/1,277), P<0.001) or the West London clinic (not sequenced - 33.1% (43/130), sequenced – 15% (186/1,277), P<0.001) (Table 4.1).

Overall, the 1,277 specimens in the PhD study sample came from 1,256 people. Of these people, 1,235 had one and 21 had two diagnoses. Of the people with two diagnoses, one person was re-infected two months after the first diagnosis, 16 were re-infected between 10 and 14 months after and four were re-infected between 22 and 34 months after.

4.2.2 Comparison of the PhD study sample with all gonorrhoea diagnoses in England

4.2.2.1 Year and geographical location

The PhD study sample contained a similar number of specimens each year between 2013 and 2015. However, between 2015 and 2016 the number of specimens declined by 23% from 326 to 251 (Table 4.1), primarily associated with a decline in diagnoses in MSM at the Mortimer Market Centre (MMC) (116 in 2015 to 53 in 2016) (Table 4.2). The decline reflected the overall annual trend in gonorrhoea diagnoses in MSM between 2015 and 2016 across England (22,419 to 17,584).[28] However, there was still a significantly lower percentage of specimens from 2016 in the PhD study sample compared to the percentage of gonorrhoea diagnoses in England in 2016 (PhD study sample – 20% (251/1,277), England – 25% (36,583/146,369), P<0.001) (Table 4.3). There was also a significantly lower percentage of specimens from London in the PhD study sample compared to the national data (PhD study sample – 45% (572/1,277), England – 50% (72,809/146,369), P<0.001) (Table 4.3).

4.2.2.2 Sexual orientation

The percentage of specimens in the PhD study sample that were from MSM was 60% (766/1,277), which was significantly higher than the percentage of diagnoses made in MSM nationally (50% (72,660/146,369), P<0.001). The percentage of specimens in the PhD study sample that were from women was significantly lower than the percentage of diagnoses in women in England during the study period (PhD study sample – 16% (206/1,277), England – 25% (36,178/146,369), P<0.001) (Table 4.3).

The distribution of specimens by sexual orientation in the PhD study sample varied by clinic: specimens from the two clinics in London (West London and MMC) were predominantly from MSM (74% (138/186) and 91% (353/386), respectively), whereas specimens from the Birmingham and Liverpool clinics were predominantly from women and men who only report heterosexual behaviours (hereafter referred to as heterosexual men) (66% (266/404) and 64% (89/138) respectively) (Table 4.2). Birmingham provided the largest number of specimens, making up 32% (404/1,277) of the total sample, and MMC was the second largest clinic, contributing 30% (386/1,277) of the total.

4.2.2.3 Age

The percentage of specimens in the PhD study sample that were from older people (≥35 years) was significantly higher than the percentage of diagnoses made in this age group nationally (PhD study sample – 30% (390/1,277), England – 25% (37,197/146,369), P<0.001). The percentage of specimens in the PhD study sample that were from people ≤24 years was significantly lower than the percentage of diagnoses made in people ≤24 years nationally (PhD study sample – 30% (384/1,277), England – 38% (55,029/146,369), P<0.001) (Table 4.3).

The distribution of specimens by age group in the PhD study sample varied by clinic (Table 4.2). The age distribution of people from clinics with more MSM, such as MMC, tended to be older (MMC ≥25 years: 86% (331/386)), compared with clinics with predominantly heterosexual populations, such as Liverpool (≤24 years: 49% (68/138)).

4.2.2.4 Ethnicity and country of birth

The majority (64%; 824/1,277) of people in the PhD study sample were white, although overall the PhD study sample contained a significantly lower percentage of white people compared to all diagnoses made in England (PhD study sample – 64% (824/1,277), England – 70% (104,028/146,369), P<0.001) (Table 4.3). Black Caribbean people were the second largest ethnic group, making up 10% (132/1,277) of the PhD study sample, which was a significantly higher percentage compared to all diagnoses made in England (5% (8,280/146,369), P<0.001). Two thirds of black Caribbean people in the PhD study sample were from the Birmingham clinic (66%; 88/132). The PhD study sample contained a higher percentage of specimens from people of Asian ethnicity compared to all diagnoses in England (PhD study sample – 6% (74/1,277), England – 4% (5,750/146,369), P=0.026).

The percentage of specimens from people born in the UK was significantly lower than the national percentage, 61% (782,1,277) vs. 65% (96,189/146,369) (P<0.001) (Table 4.3). Within the PhD study sample, the distribution of people born in the UK varied by clinic ranging from 33% (128/386) in MMC to 85% (118/138) in Liverpool (Table 4.2).

4.2.2.5 Site of infection and clinical presentation

Data on symptoms at presentation and site of infection were only available for the PhD study sample and not for all gonorrhoea diagnoses in England thereby precluding comparisons across datasets. Within the PhD study, the majority (62%; 803/1,277) of people presented with symptoms, probably because culture is more successful from symptomatic infections.[219] The majority of specimens were from genital infections (69%; 881/1,277).

4.2.2.6 Prior and concurrent STIs

One fifth (20%; 262/1,277) of people in the PhD study sample had been diagnosed with an STI in the year prior to their gonorrhoea diagnosis, which is same as the percentage reported nationally (Table 4.3). However, in the PhD study sample, the percentage ranged from 14% (56/404) in the Birmingham clinic to 38% (141/386) in MMC (Table 4.2).

Overall, 18% (226/1,277) of people in the PhD study sample were known to be people living with HIV (hereafter referred to as HIV-positive people),[234] which was significantly higher than the national percentage of people living with HIV and diagnosed with gonorrhoea (11% (16,171/146,369), P<0.001) (Table 4.3). However, in the PhD study sample, the percentage ranged from 5% (8/163) in the Bristol clinic to 39% (149/386) in MMC (Table 4.2). Nearly all the HIV-positive people in the PhD study sample were MSM (97%, 220/226).

4.2.2.7 Sexual behaviour

Data on sexual behaviour (number of partners in the UK and travel-associated sexual partnerships) were only available for people in the PhD study sample and not the wider population of people diagnosed with gonorrhoea. One third of cases in the PhD study sample were missing sexual behavioural data (33%; 422/1,277) and most of the missing data were from MMC (Table 4.2) (I attempted to rectify this by requesting the missing data from MMC but it was not possible to access these data). Where complete, half of people had reported two or more sexual partners in the UK three months prior to their diagnosis (49%, 471/855). For records with information on travel-associated sexual partnerships, 7% (88/855) had a sexual partnership whilst travelling abroad.

Table 4.1 Epidemiological characteristics of eligible specimens stratified by sequencing success

	Sequenced		Not sequenced		Two sample proportions z-test p-value*
	N	%	N	%	
Total	**1,277**	**100**	**130**	**100**	
Year					
2013	326	25.5	22	16.9	**0.030**
2014	333	26.1	9	6.9	**<0.001**
2015	367	28.7	31	23.8	0.238
2016	251	19.7	68	52.3	**<0.001**
Clinic					
Bristol	163	12.8	7	5.4	**0.014**
Liverpool	138	10.8	8	6.2	0.098
West London	186	14.6	43	33.1	**<0.001**
Birmingham	404	31.6	43	33.1	0.737
MMC	386	30.2	29	30.0	0.059
Gender & sexual orientation					
MSM	766	60.0	67	51.5	0.059
Heterosexual men	304	23.8	34	26.1	0.551
Women	206	16.1	29	22.3	0.072
Missing	1	0.1	0	0.0	N/A
Age group (years)					
≤24	384	30.1	30	23.1	0.096
25-34	503	39.4	61	46.9	0.095
≥35	390	30.5	39	30.0	0.899
Ethnicity					
White	824	64.5	76	58.5	0.170
Black Caribbean	132	10.3	18	13.8	0.217
Black African	47	3.7	6	4.6	0.594
Black Other	10	0.8	1	0.8	N/A
Asian	74	5.8	10	7.7	0.384
Other	32	2.5	3	2.3	N/A
Mixed	105	8.2	8	6.2	0408
Missing	53	0.4	8	6.2	0.285
Country of birth					
UK	782	61.2	73	56.1	0.258
Not UK	407	31.9	42	32.3	0.919
Missing	88	6.9	15	11.5	0.053
Symptomatic infection					
No	338	26.5	28	21.5	0.222
Yes	803	62.9	88	67.7	0.278
Missing	136	10.6	14	10.8	0.966
Anatomical specimen site					
Genital	881	69.0	91	70.0	0.813
Rectal	299	23.4	27	20.8	0.496
Throat	80	6.3	12	9.2	0.193
Other	6	0.5	0	0.0	N/A
Missing	11	2.7	0	0.0	N/A
Diagnoses with a new STI (excluding HIV) in the past year					
No/Unknown	979	76.7	108	83.1	0.097
Yes	298	23.3	22	16.9	0.097

Table continued on the next page

	Sequenced		Not sequenced		Two sample proportions z-test p-value*
	N	%	N	%	
HIV status					
Negative/ Unknown	1,051	82.3	109	83.8	0.659
Positive	226	17.7	21	16.2	0.659
Number of UK partners three months prior to diagnosis					
0	47	3.7	5	3.8	0.924
1	337	26.4	43	33.1	0.102
≥2	471	36.9	57	43.8	0.118
Missing	422	33.0	25	19.2	**0.001**
Travel-associated sexual partnerships in the three months prior to diagnosis					
No	767	60.1	97	74.6	**0.001**
Yes	88	6.9	8	6.2	0.750
Missing	422	33.0	25	19.2	**0.001**

*Test to see if the two percentages on each row are significantly different (null hypothesis is that they are the same)[235]

Table 4.2 Epidemiological characteristics of the PhD study sample by clinic
W. London = West London, B.ham = Birmingham, MMC = Mortimer Market Centre

	Bristol		Liverpool		W. London		B.ham		MMC	
	N	%	N	%	N	%	N	%	N	%
Total	**163**	**100**	**138**	**100**	**186**	**100**	**404**	**100**	**386**	**100**
Year										
2013	46	28.2	35	25.4	46	24.7	92	22.8	107	27.7
2014	39	23.9	24	17.4	63	33.9	97	24.0	110	28.5
2015	48	29.4	55	39.9	44	23.7	104	25.7	116	30.1
2016	30	18.4	24	17.4	33	17.7	111	27.5	53	13.7
Gender & sexual orientation										
MSM	88	54.0	49	35.5	138	74.2	138	34.2	353	91.5
Heterosexual men	42	25.8	61	44.2	30	16.1	150	37.1	21	5.4
Women	33	20.2	28	20.3	17	9.1	116	28.7	12	3.1
Missing	0	0.0	0	0.0	1	0.5	0	0.0	0	0.0
Age group (years)										
≤24	60	36.8	68	49.3	43	23.1	158	39.1	55	14.2
25-34	62	38.0	43	31.2	86	46.2	154	38.1	158	40.9
≥35	41	25.2	27	19.6	57	30.6	92	22.8	173	44.8
Ethnicity										
White	133	81.6	120	87.0	120	64.5	176	43.6	275	71.2
Black Caribbean	9	5.5	0	0.0	14	7.5	88	21.8	21	5.4
Black African	2	1.2	3	2.2	9	4.8	14	3.5	19	4.9
Black Other	0	0.0	3	2.2	2	1.1	2	0.5	3	0.8
Asian	3	1.8	2	1.2	16	8.6	32	7.9	21	5.2
Other	2	1.2	2	1.2	8	4.3	7	1.7	13	3.2
Mixed	7	4.3	8	4,9	12	6.5	48	11.9	30	7.4
Missing	7	4.3	0	0.0	5	2.7	37	9.2	4	1.0
Country of birth										
UK	134	82.2	118	85.5	98	52.7	304	75.2	128	33.2
Not UK	29	17.8	20	14.5	83	44.6	67	16.6	208	53.9
Missing	0	0.0	0	0.0	5	2.7	33	8.2	50	13.0

Table continued on the next page

	Bristol		Liverpool		W. London		B.ham		MMC	
	N	%	N	%	N	%	N	%	N	%
Symptomatic infection										
No	28	17.2	19	13.8	58	31.2	123	30.4	110	28.5
Yes	121	74.2	104	75.4	107	57.5	244	60.4	227	58.8
Missing	14	8.6	15	10.9	21	11.3	37	9.2	49	12.7
Anatomical specimen site										
Genital	117	71.8	103	74.6	99	53.2	336	83.2	226	58.5
Rectal	31	19.0	24	17.4	62	33.3	47	11.6	135	35.0
Throat	10	6.1	11	8.0	24	12.9	11	2.7	24	6.2
Other	1	0.6	0	0.0	0	0.0	5	1.2	0	0.0
Missing	4	2.5	0	0.0	1	0.5	5	1.2	1	0.2
Diagnoses with a new STI (excluding HIV) in the past year										
No/Unknown	131	80.4	113	81.9	145	78.0	348	86.1	242	62.7
Yes	32	19.6	25	18.1	41	22.0	56	13.9	144	37.7
HIV status										
Negative/Unknown	155	95.1	127	92.0	152	81.7	380	94.1	237	61.4
Positive	8	4.9	11	8.0	34	18.3	24	5.9	149	38.6
Number of UK partners three months prior to diagnosis										
0	11	6.7	4	2.9	10	5.4	22	5.4	0	0.0
1	42	25.8	52	37.7	57	30.6	184	45.5	2	0.5
≥2	96	58.9	81	58.7	114	61.3	176	43.6	4	1.0
Missing	14	8.6	1	0.7	5	2.7	22	5.4	380	98.4
Travel-associated sexual partnerships in the three months prior to diagnosis										
No	127	77.9	123	89.1	157	84.4	354	87.6	6	1.6
Yes	22	13.5	14	10.1	24	12.9	28	6.9	0	0.0
Missing	14	8.6	1	0.7	5	2.7	22	5.4	380	98.4

Table 4.3 Epidemiological characteristics of the PhD study sample compared to all gonorrhoea diagnoses in England during the study period (2013-2016)

	PhD study sample		England*		Two sample proportions z-test p-value**
	n	%	n	%	
Total	1277	100	**146,369**	**100**	
Year					
2013	326	25.5	31,213	21.3	**<0.001**
2014	333	26.1	37,178	25.4	0.580
2015	367	28.7	41,396	28.3	0.718
2016	251	19.7	36,582	25.0	**<0.001**
Geographical location					
London	572	44.8	72,809	49.7	**<0.001**
Outside London	705	55.2	73,560	50.2	**<0.001**
Gender & sexual orientation					
MSM	766	60.0	72,660	49.6	**<0.001**
Heterosexual men	304	23.8	34,330	23.5	0.768
Women	206	16.1	36,178	24.7	**<0.001**
Missing	1	<0.1	3,201	2.2	N/A
Age group (years)					
≤24	384	30.1	55,029	37.6	**<0.001**
25-34	503	39.4	54,143	37.0	0.077
≥35	390	30.5	37,197	25.4	**<0.001**

Table continued on the next page

	PhD study sample		England*		Two sample proportions z-test p-value**
	n	%	n	%	
Total	1277	100	**146,369**	**100**	
Ethnicity					
White	824	64.5	104,028	71.1	**<0.001**
Black Caribbean	132	10.3	8,280	5.7	**<0.001**
Black African	47	3.7	5,858	4.0	0.559
Black Other	10	0.8	3,238	2.2	**<0.001**
Asian	74	5.8	5,750	3.9	**0.026**
Other	32	2.5	4,747	3.2	0.138
Mixed	105	8.2	8,614	5.9	**<0.001**
Missing	53	4.2	5,815	4.0	0.747
Country of birth					
UK	782	61.2	96,189	65.7	**<0.001**
Not UK	407	31.9	38,334	26.2	**<0.001**
Missing	88	6.8	11,846	8.1	0.117
Diagnoses with a new STI (excluding HIV) in the past year					
No/Unknown	1,015	79.5	117,493	80.3	0.481
Yes	262	20.5	28,876	19.7	0.481
HIV status					
Negative/Unknown	1,051	82.3	130,198	89.0	**<0.001**
Positive	226	17.7	16,171	11.0	**<0.001**

*Data covering the whole year, not just the GRASP sampling period (July-September) ** Test to see if the two percentages on each row are significantly different (null hypothesis is that they are the same)[235]

4.3 Phenotypic antimicrobial susceptibility profile

4.3.1 Antimicrobial susceptibility of PhD study sample compared to GRASP

The PhD study sample covers 20% of GRASP specimens over the study period. For all five antimicrobials tested, the antimicrobial minimum inhibitory concentration (MIC) threshold used to define resistance of *N. gonorrhoeae* specimens are presented in Table 4.4. The MIC distribution of specimens from the PhD study sample were similar to those of all GRASP specimens (Table 4.5). In the PhD study sample, the modal MIC for ceftriaxone was 0.004 mg/L and no specimens were classed as resistant (>0.125mg/L) (Table 4.5). The modal MIC for azithromycin was 0.25mg/L, 4% (46/1,267) of specimens were resistant and two were highly resistant (≥256mg/L). The modal MIC for cefixime was 0.015mg/L and <1% (8/1,267) of specimens were resistant (>0.125mg/L). The MIC distribution for ciprofloxacin between the years 2013-2015 was bimodal with a peak at the highly sensitive MIC category 0.03 mg/L and a second peak at 8.00mg/L, which is above the resistance threshold (>0.06mg/L). In 2016, the routine susceptibility testing for ciprofloxacin was changed to breakpoint plates instead of the full MIC spectrum. Overall, between 2013 and 2016, 35% (449/1,267) of specimens were resistant to ciprofloxacin in the

PhD study sample. The modal MIC for penicillin was 0.25mg/L and 17% (211/1,267) of specimens were resistant.

Table 4.4 Antimicrobial minimum inhibitory concentration threshold used to define resistance of *N. gonorrhoeae* specimens

	Minimum inhibitory concentration thresholds (mg/L)
Antimicrobial	**Resistance***
Azithromycin	>0.5
Ceftriaxone	>0.125
Cefixime	>0.125
Ciprofloxacin	>0.06
Penicillin	>1 or β-lactamase positive

*As defined by EUCAST[100]

Table 4.5 Phenotypic antimicrobial susceptibility profile of PhD study sample and all GRASP specimens during the study period (2013-2016)

MIC	PhD study sample		GRASP		Chi² test P value**	PhD study sample as a % of GRASP
	N*	%	N	%		
Total	**1,267**	**100.0**	**6,184**	**100.0**		**20.5**
Ceftriaxone (MIC resistance threshold >0.125 mg/L)						
<0.002	207	16.3	938	15.2		22.1
0.004	409	32.3	1,999	32.3		20.5
0.008	374	29.5	1,760	28.5		21.3
0.015	150	11.8	833	13.5	0.569	18.0
0.03	121	9.6	601	9.7		20.1
0.06	7	0.6	52	0.8		13.5
0.125	0	0.0	1	<0.1		0.0
Azithromycin (MIC resistance threshold >0.5 mg/L)						
<0.03	127	10.0	598	9.7		21.2
0.06	186	14.7	924	14.9		20.1
0.125	382	30.1	1,873	30.3		20.4
0.25	402	31.7	1,913	30.9		21.0
0.50	125	9.9	622	10.1		20.1
1.00	39	3.1	192	3.1	0.847	20.3
2.00	1	0.1	31	0.5		3.2
4.00	3	0.2	15	0.2		20.0
8.00	1	0.1	3	0.0		33.3
16.0	0	0.0	3	0.0		0.0
≥256	2	0.2	10	0.2		20.0
Cefixime (MIC resistance threshold >0.125 mg/L)						
0.002	37	2.9	200	3.2		18.5
0.004	62	4.9	305	4.9		20.3
0.008	331	26.1	1,691	27.3		19.6
0.015	484	38.2	2,187	35.4		22.1
0.03	174	13.7	882	14.3	0.795	19.7
0.06	144	11.4	729	11.8		19.8
0.125	28	2.2	140	2.3		20.0
0.25	8	0.6	46	0.7		17.4
0.50	0	0.0	4	0.1		0.0

Table continued on the next page

MIC	PhD study sample		GRASP		Chi² test P value**	PhD study sample as a % of GRASP
	N*	%	N	%		
Total	**1,267**	**100.0**	**6,184**	**100.0**		**20.5**
Ciprofloxacin (2013-2015) (MIC resistance threshold >0.06 mg/L)						
0.03	632	49.9	3,009	48.7	0.952	21.0
0.06	12	0.9	65	1.1		18.5
0.125	6	0.5	28	0.5		21.4
0.25	2	0.2	14	0.2		14.3
0.50	3	0.2	28	0.5		10.7
1.00	18	1.4	84	1.4		21.4
2.00	13	1.0	78	1.3		16.7
4.00	68	5.4	329	5.3		20.7
8.00	145	11.4	698	11.3		20.8
16.0	90	7.1	427	6.9		21.1
32.0	37	2.9	140	2.3		26.4
Ciprofloxacin (2016 breakpoint plates) (MIC resistance threshold >0.06 mg/L)						
≤0.06	164	12.9	851	13.8	0.302	19.3
>0.06 & <0.50	5	0.4	13	0.2		38.5
≥0.50	73	5.8	420	6.8		17.4
Penicillin (MIC resistance threshold >1.0 mg/L)						
0.06	92	7.3	386	6.2	0.066	23.8
0.125	298	23.5	1,400	22.6		21.3
0.25	399	31.5	2,112	34.2		18.9
0.50	186	14.7	1,009	16.3		18.4
1.0	82	6.5	413	6.7		19.9
2.0	34	2.7	168	2.7		20.2
4.0	44	3.5	156	2.5		28.2
8.0	133	10.5	540	8.7		24.6

*Denominator is less than PhD study sample as 10 samples from 2016 failed routine phenotypic testing so no AMR data is available **P value from Chi²-test comparing distribution of MICs in the PhD study sample to the GRASP sample

4.4 Whole genome sequence data quality

The following measures were used to assess the quality of the WGS data for each specimen included in the phylogenetic analyses: Phred score (a measure of sequencing quality), species identification, length of contigs (the DNA fragments assembled from the raw sequence data), sequence length, sequence guanine and cytosine (GC) content, percentage of the reference genome covered by reads and the number of SNPs identified. These summary measures were calculated during the sequencing pipeline process at the Wellcome Sanger Institute.[236]

4.4.1 Phred score

The Phred score measures the quality of the nucleotides called during the sequencing process.[236] A higher Phred score indicates a lower probability of error.[237] A score of 10 equals an error probability of 1 in 10, a score of 20 equals 1 in 100 and a score of 30 equals 1 in 1000 etc. As a rule of thumb, scores over 30 are

considered extremely confident base calls. The mean Phred score for the PhD study sample specimens was 37 (median: 37, minimum: 35, maximum: 38).

4.4.2 Species identification

The species identity of the specimens was confirmed using Kraken.[238] Kraken assigns sequencing reads to a taxonomic classification by comparing kmers in the raw sequence reads to a large, public database of genomes.[239] For all specimens, the majority of reads were assigned to *N. gonorrhoeae*.

4.4.3 Contigs

During the sequencing pipeline process at Sanger, the raw sequencing reads were assembled into a genome assembly made up of contiguous sequence fragments (contigs).[236] A useful measure of the contiguity (i.e. how well reads were joined into longer fragments) of the assembly is the N50 value. Given a set of contigs sorted by decreasing length, the N50 value is the size of the contig at 50% of the total assembly length. The mean N50 across the PhD study sample was 93,070 nucleotides (median: 85,526, minimum: 40,600, maximum: 208,556).

4.4.4 Assembly length

The mean assembly length was 2,165,750 nucleotides (median: 2,165,330, minimum: 2,093,067, maximum: 4,729,643). There were two assemblies that were almost double the average length. Both these specimens also contained relatively large percentage of unclassified sequence in the species identification analysis (20% and 40%), although the majority of reads matched *N. gonorrhoeae*. This suggests these two samples contained low level contamination from an unknown source. Removing these outliers, the mean sequence length was 2,162,966 (minimum: 2,093,067, maximum: 2,259,256). The length of the reference *N. gonorrhoeae* genome used in this PhD (FA1090) is 2,153,922 nucleotides.

4.4.5 Assembly GC content

The percentage of guanine (G) and cytosine (C) nucleotides was also measured to assess the sequence data quality. The mean percentage GC content for each assembly in the PhD study sample was 52% (median: 52%, minimum: 43%, maximum: 53%). The specimen assembly with the lowest GC content (43%) was the outlier described above with a long assembly length. The GC content of the reference *N. gonorrhoeae* genome used in this PhD (FA1090) is 53%.

4.4.6 Mapping

The raw sequencing reads of each specimen were used to create a full sequence alignment of the genome sequence by mapping to a reference genome (FA1090) (method described in Chapter 3). These alignments were used in the phylogenetic analyses. The mean percentage of the reference genome that was successfully mapped was 93% (median: 93%, minimum: 91%, maximum: 94%). The mean number of SNPs relative to the reference identified per sequence was 4,731 (median: 4,723, minimum: 3,138, maximum: 5,362).

4.5 Conclusion

The PhD study sample was broadly representative in terms of the epidemiological characteristics and antimicrobial susceptibility profile of *N. gonorrhoeae* circulating in England. Although there were several significant differences between the characteristics of the specimens in PhD study sample compared to the national data, the absolute difference was usually small. Therefore, these findings indicate that the study sample is likely to provide reliable insights into the molecular epidemiology of *N. gonorrhoeae* that are generalisable at a national level. Nevertheless, there are some caveats that need to be considered.

The PhD study sample contained a higher percentage of specimens from MSM than was seen nationally, likely due to inclusion of two clinics from London that serve large MSM populations, and the overrepresentation of MSM in GRASP.[217] Likewise, the larger percentage drop in cases in the PhD study sample between 2015 and 2016 than seen nationally was likely due to inclusion of the two London clinics and also a higher culture failure rate for 2016. The national decline in gonorrhoea diagnoses was possible due to increased testing and shorter treatment turnaround times in MSM, particularly in London.[29]

An expected, but important, finding from these descriptive analyses was that the distribution of patient characteristics varied across clinics, which had to be considered when interpreting results in later chapters. While the percentage of missing epidemiological data was low (<10%) for most variables and clinics, sexual behaviour data from one clinic (MMC) were not available (number of sexual partners and travel-associated sexual partnerships). Therefore, any results from analyses using these data items will only reflect data from a subset of the clinics investigated.

Overall the sequence data quality was high. There were a small number of specimens with a high percentage of contaminant DNA, as evidenced by large assemblies with abnormal GC content. Despite these issues the affected specimens were kept in the PhD study sample, as there was enough *N. gonorrhoeae* DNA to make a consensus *N. gonorrhoeae* genome.

5 Identification and description of *N. gonorrhoeae* transmission in sexual networks using combined WGS and epidemiological data

5.1 Introduction

Classical epidemiological methods, including surveillance of diagnoses and comparisons by person, time and place, are used to identify population groups at greater risk of *N. gonorrhoeae* infection and their associated sexual behaviours. However, in the absence of partner notification data, observational epidemiological studies are limited in their ability to determine whether infected individuals with the same characteristics are part of the same transmission network. The central hypothesis considered in this chapter is that studies incorporating molecular data can provide additional information about STI transmission within and between sexual networks by grouping infections according to their genetic similarity.

In this chapter, I aimed to answer the following research questions:

- What is the genetic variation of *N. gonorrhoeae* in England within and between different geographical areas and how does this vary over time?
- How can WGS data combined with patient epidemiological data be used to identify and describe gonorrhoea sexual transmission networks?

I used WGS data and phylogenetic analyses to identify clusters of genetically related specimens of *N. gonorrhoeae*, representing different sexual networks and I characterised the sexual networks using epidemiological data.

5.2 Methods

Data collection is described in Chapters 3 and 4. To describe the genetic variation of the PhD study sample by time and place and to identify sexual networks, I used a systematic approach employing methods of increasing complexity (Figure 5.1). I began with visualisation of an annotated phylogenetic tree and then used the number of SNPs between specimens to measure differences in the distribution of epidemiological variables in phylogenetic clusters.

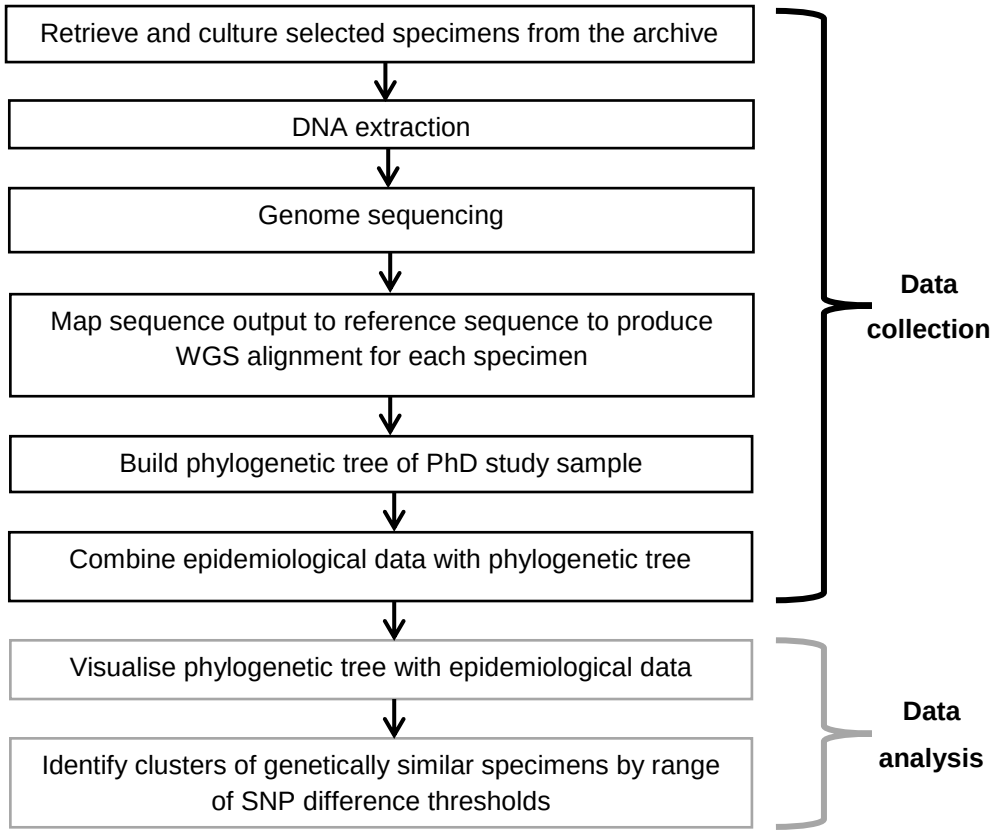


Figure 5.1 Pipeline of WGS data collection and analysis

5.2.1 Creating the phylogenetic tree

A phylogenetic tree with genetic recombination events removed was created using the programme Gubbins (Version 2.4.0)[117] with the default settings (five iterations and a minimum number of three base substitutions to identify a recombination) and the tree building option Randomized Axelerated Maximum Likelihood (RAxML) (Version 8.2.8).[240] Prior to this, the *opa* and *pil* genes, phages[241] and the Gonococcal Genetic Island (GGI)[143] were manually removed from the alignment. These add noise to the phylogenetic inferences made during the tree building exercise because they recombine intragenomically (*opa* and *pil* genes) or are mobile elements (phages and GGI)[15] which move between bacteria at a different rate to the mutation rate of the pathogen.

Gubbins identified possible recombination events by assessing the SNP density on each branch. Parts of the alignment considered to be due to recombination were removed and the tree was redrawn with the remaining data. The process was repeated until the tree no longer changed or it reached the number of iteration phases specified (default 5). RAxML used a heuristic approach to find the tree with

the maximum likelihood of producing the data given the model. RAxML was chosen over other maximum likelihood methods because it is fast and has been shown to give higher likelihoods and more accurate trees.[114]

The input data were the mapped alignment of all the sequences in the PhD study sample plus the reference sequence (FA1090). The output files used for analysis were the phylogenetic tree and a summary file of recombination events per branch of the tree. The output phylogenetic tree was midpoint rooted (meaning the root of the tree was placed half-way between the two most distant specimens) using Figtree (Version 1.4.3)[242] and to aid visualisation the branches were ladderised i.e. branches were rotated so that they were ordered by increasing clade size at each node. The summary file of recombination events produced by Gubbins was used to assess the extent of recombination in the PhD study sample. For each branch the number of SNPs from recombination events (known as r) was divided by the number SNPs from point mutations (knowns as m) to create the *r/m* value. The mean *r/m* across all branches was calculated.

Statistical support for relationships in the phylogenetic tree was assessed using the programme Booster.[243] The phylogenetic tree was reconstructed using Gubbins and RAxML with the bootstrap option to create 100 trees and input into Booster along with the tree used for analysis (the reference tree). Booster calculated a value known as the transfer bootstrap expectation (TBE), which quantifies the presence of each branch at a particular position in the bootstrap trees. A value of one indicates that the branch is present in all bootstrap trees and a value of zero indicates that the bootstrap trees are random.

5.2.2 Visualisation with the epidemiological data

To visualise the phylogenetic tree with the epidemiological data I used Phandango.[244] The input data were the phylogenetic tree and a metadata file including the identifier data, which was identical to the tip label on the tree. Phandango enabled interactive visualisation of the tree with the epidemiological data using different colours for the variable values.

5.2.3 Identifying *N. gonorrhoeae* clusters

I used the number of SNPs between pairs of specimens to group specimens as clusters, which represent sexual networks with *N. gonorrhoeae* transmission. The number of SNPs between each pair of specimens was estimated from the branch

length distance, which was extracted from the phylogenetic tree using SeaView (Version 4.7).[245] Branch lengths are an estimate of the total number of substitutions that have occurred between a pair of specimens. To translate this into an estimated SNP difference, the branch length was multiplied by the number of nucleotides in the sequence alignment used to create the phylogenetic tree.

The pairwise SNP difference data were input with epidemiological metadata into the programme MicrobeTrace (Version 0.1.10).[246] As there is no universally agreed number of SNPs for defining sexual networks in *N. gonorrhoeae* transmission, I used a range of values to define the threshold (5, 10 and 30 SNP differences). Specimens were included if they were similar (according to the SNP threshold) to at least one other specimen. Consequently, the SNP difference range across all specimens in a cluster may be larger than the SNP threshold defined. As a result, more complex relationships between the specimens can be identified. Clusters were described by year, clinic, gender and sexual orientation, age, ethnicity and HIV status. Additionally, I investigated whether any of the HIV negative/unknown people included in the PhD study sample were subsequently diagnosed with HIV using data from GUMCAD up to the end of 2016, and I described the characteristics of the clusters these people were found in.

5.2.4 Statistical methods of comparison

Univariate analysis of odds ratios was used to assess differences in the epidemiological characteristics of specimens grouped according to data extracted from the phylogenetic analyses. The Chi^2 test was used to assess if the odds of the outcome (indicated in the text and table for each analysis) was significantly different between the baseline and other categories of the explanatory variables.[235] Significance was determined by whether the confidence interval of the odds ratio crossed 1.0 and if the P value was <0.05. Following the univariate analysis, multivariable logistic regression models using odds ratios were developed in a forward step-wise approach including only the variables identified as significantly associated with the outcome in the univariate models to control for possible confounding between variables. The likelihood ratio test was used to determine if the explanatory variables should remain in the multivariable model. Explanatory variables were kept in the model if the Chi^2 test approximate P value was <0.05.[235]

5.3 Results

5.3.1 Genetic variation across the whole PhD study sample

5.3.1.1 Describing the phylogenetic tree

Statistical support for the phylogenetic tree was high: 79% (1,014/1,276) of nodes had a TBE value greater than 70%. In the PhD study sample more SNPs on branches were due to recombination events (r) than point mutations (m) (mean r/m value: 1.13). After removing the recombination events, two distinct high-level lineages were visible in the phylogenetic tree: Lineage A contained 838 specimens, and lineage B contained 439 specimens. Both lineages contained specimens from all four years (Figure 5.2, Table 5.1).

In the adjusted analysis, compared to lineage B, lineage A was significantly more likely to contain specimens from the Liverpool clinic than the Bristol clinic (base: Bristol clinic, Liverpool clinic adjusted odds ratio (aOR) 2.20, 95% confidence interval (CI) 1.27-3.80, P value (P)=0.005), specimens from people aged ≥35 years compared to people aged ≤24 years (base: ≤24 years, ≥35 years aOR 1.92, CI 1.31-2.82, P=0.001) and specimens from people of Asian ethnicity compared to people of white ethnicity (base: white ethnicity, Asian ethnicity aOR 2.71, CI 1.42-5.17, P=0.002).

Lineage A was significantly less likely to contain specimens from the Birmingham clinic compared to the Bristol clinic (base: Bristol clinic, Birmingham clinic aOR 0.49, CI 0.32-0.77, P=0.002) or from women or heterosexual men compared to MSM (base: MSM, women aOR 0.13, CI 0.09-0.20, P<0.001; heterosexual men aOR 0.27, CI 0.19-0.40, P<0.001).

Other variables that were significantly associated with lineage A compared to lineage B in the univariate model but that did not remain significant in the multivariable model after adjusting for clinic, sexual orientation, age and ethnicity were: HIV status (base: HIV-negative/unknown, HIV-positive crude odds ratio (cOR) 2.75, CI 1.91-3.96, P<0.001), diagnosed with a new STI (excluding HIV) in the past year (base: no, yes cOR 1.75, CI 1.75-1.31, P<0.001), country of birth (base: UK, not UK cOR 1.95, CI 1.49-2.56, P<0.001) and reported recent travel-associated sexual partnership (base: no, yes cOR 1.96, CI 1.20-3.21, P=0.006).

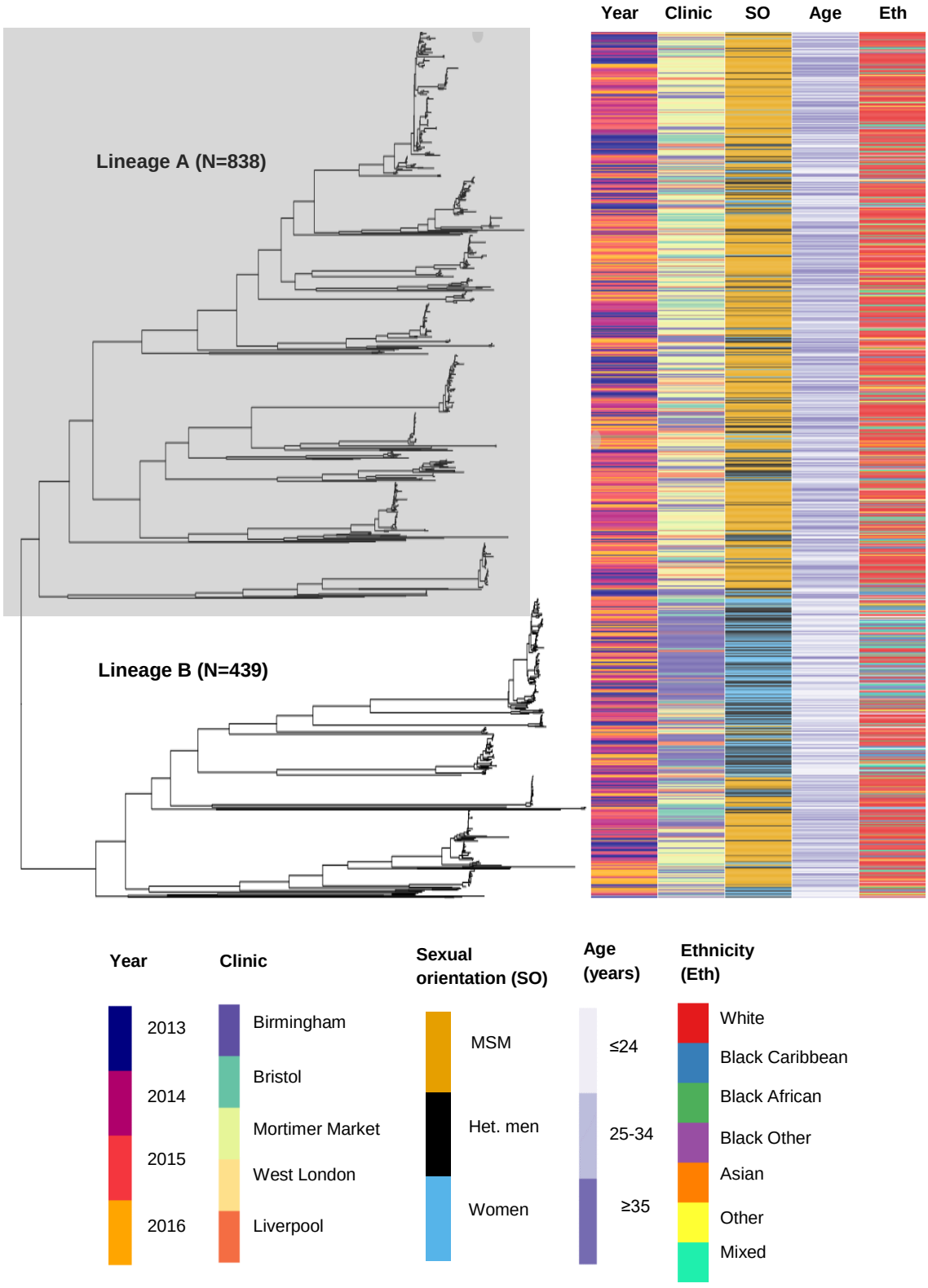


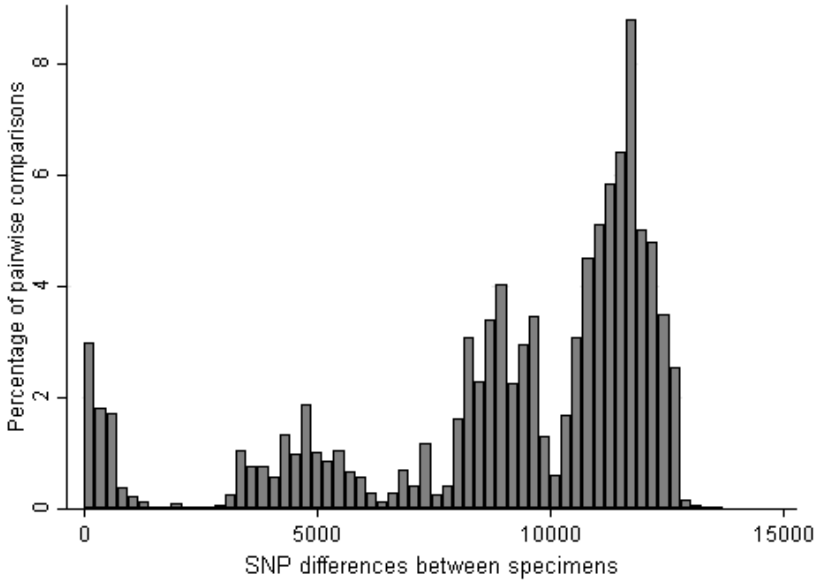
Figure 5.2 Maximum likelihood phylogeny with recombination events removed of all *N. gonorrhoeae* specimens annotated with lineage (A in shaded box), year, clinic, sexual orientation, age and ethnicity

Table 5.1 Univariate and multivariable analysis comparing the epidemiological characteristics of cases of *N. gonorrhoeae* between the two lineages in the phylogeny

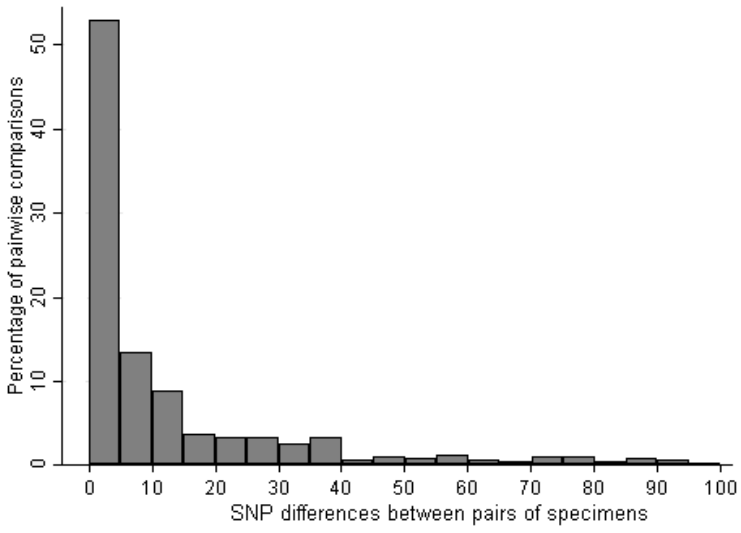
	Lineage (n)		Outcome: specimen in lineage A							
			Univariate				Multivariable			
	B	A	OR	LCI	UCI	P	OR	LCI	UCI	P
Total	439	838								
Year										
2013	106	220	1	-	-	-				
2014	123	210	0.82	0.60	1.13	0.234				
2015	107	260	1.17	0.85	1.62	0.339				
2016	103	148	**0.69**	**0.49**	**0.98**	**0.035**				
Clinic										
Bristol	60	103	1	-	-	-	1	-	-	-
Liverpool	38	100	1.53	0.94	2.51	0.088	**2.20**	**1.27**	**3.80**	**0.005**
West London	38	148	**2.27**	**1.39**	**3.69**	**<0.001**	1.63	0.95	2.81	0.077
Birmingham	232	172	**0.43**	**0.29**	**0.63**	**<0.001**	**0.49**	**0.32**	**0.77**	**0.002**
MMC	71	315	**2.58**	**1.70**	**3.93**	**<0.001**	1.16	0.72	1.86	0.545
Sexual orientation and gender										
MSM	136	630	1	-	-	-	1	-	-	-
Het. M	154	150	**0.21**	**0.15**	**0.29**	**<0.001**	**0.27**	**0.19**	**0.40**	**<0.001**
Women	149	57	**0.08**	**0.05**	**0.12**	**<0.001**	**0.13**	**0.09**	**0.20**	**<0.001**
Age (years)										
≤24	196	188	1	-	-	-	1	-	-	-
25-34	161	342	**2.21**	**1.67**	**2.93**	**<0.001**	1.30	0.93	1.82	0.126
≥35	82	308	**3.92**	**2.81**	**5.46**	**<0.001**	**1.92**	**1.31**	**2.82**	**0.001**
Ethnicity										
White	238	586	1	-	-	-	1	-	-	-
Black Caribbean	81	51	**0.26**	**0.17**	**0.38**	**<0.001**	0.77	0.48	1.24	0.280
Black African	20	27	**0.55**	**0.30**	**1.00**	**0.046**	1.06	0.53	2.13	0.864
Black Other	4	6	0.61	0.17	2.18	0.442	0.47	0.11	2.10	0.326
Asian	17	57	1.36	0.78	2.39	0.280	**2.71**	**1.42**	**5.17**	**0.002**
Other	8	24	1.22	0.54	2.75	0.634	1.14	0.45	2.87	0.781
Mixed	43	62	**0.59**	**0.39**	**0.89**	**0.011**	1.07	0.65	1.79	0.782
Country of birth										
UK	309	473	1	-	-	-				
Not UK	102	305	**1.95**	**1.49**	**2.56**	**<0.001**				
Symptomatic infection										
No	119	219	1	-	-	-				
Yes	277	526	1.03	0.79	1.35	0.818				
Diagnosed with a new STI (excluding HIV) in the past year										
No/Unknown	363	615	1	-	-	-				
Yes	75	223	**1.75**	**1.31**	**2.35**	**<0.001**				
HIV status										
Negative/ Unknown	398	653	1	-	-	-				
Positive	41	185	**2.75**	**1.91**	**3.96**	**<0.001**				
Number of UK partners three months prior to diagnosis										
0	20	27	1	-	-	-				
1	162	175	0.80	0.43	1.48	0.478				
≥2	167	304	1.35	0.73	2.48	0.335				
Travel- associated sexual partnerships in the three months prior to diagnosis										
No	325	442	1	-	-	-				
Yes	24	64	**1.96**	**1.20**	**3.21**	**0.006**				

5.3.1.2 SNP differences between specimen pairs across the whole PhD study sample

The largest difference between specimen pairs was 13,706 SNPs (Figure 5.3a). Half of the specimens (49%; 630/1,277) had a neighbouring specimen that was ≤5 SNPs different (Figure 5.3b).



(A)



(B)

Figure 5.3 Pairwise SNP differences comparison between specimens

(a) Estimated SNP difference between every specimen pairwise comparison. In total, there are 816,003 pairwise comparisons in the PhD study sample of 1,277 specimens. (b) Estimated SNP difference between every specimen and the specimen closest to it (where SNP difference is <100 SNPs (N=1,149))

5.3.2 Identifying and characterising *N. gonorrhoeae* clusters

Clusters of *N. gonorrhoeae*, which represent transmission networks, were defined by the number of SNP differences between specimens. As expected, the size of clusters increased with the threshold of SNP differences used to define clusters (Figure 5.4 and Figure 5.5). Most clusters contained only two specimens (using a five SNP threshold: 63% (135/213), using a 30 SNP threshold: 43% (76/175)) (Figure 5.4). Overall, half of specimens clustered when using a five SNP threshold and three-quarters of specimens clustered with a 30 SNP threshold.

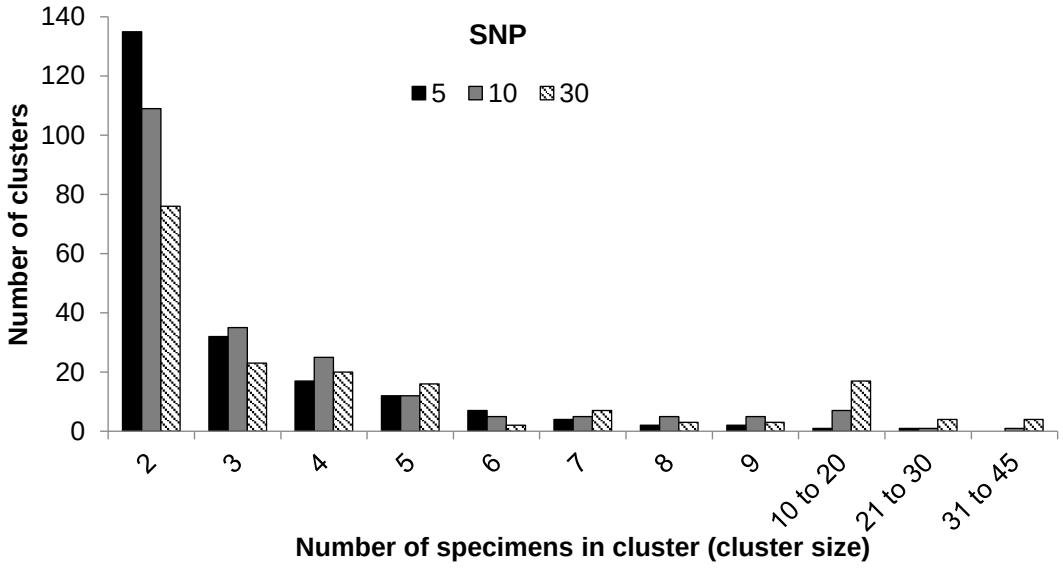
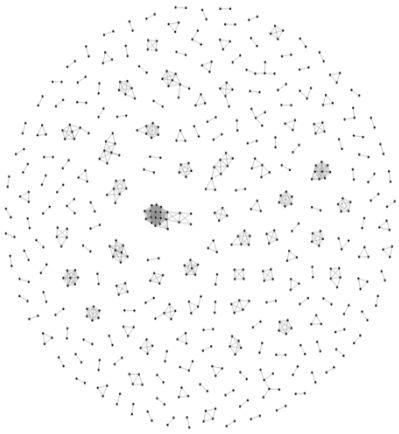
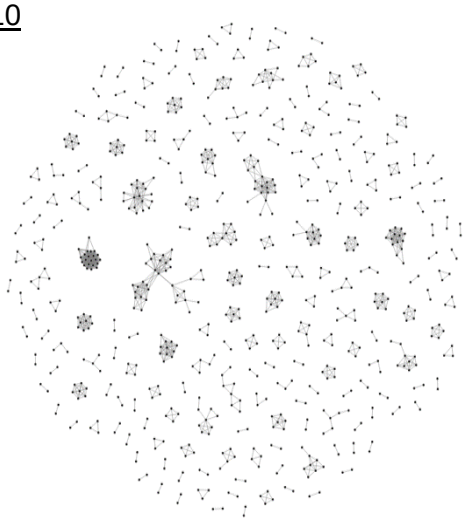


Figure 5.4 Frequency and size of *N. gonorrhoeae* clusters by using SNP difference thresholds

SNP threshold: 5
N=213, R=2 to 21



SNP threshold: 10
N=210, R=2 to 34



SNP threshold: 30
N=175, R=2 to 45

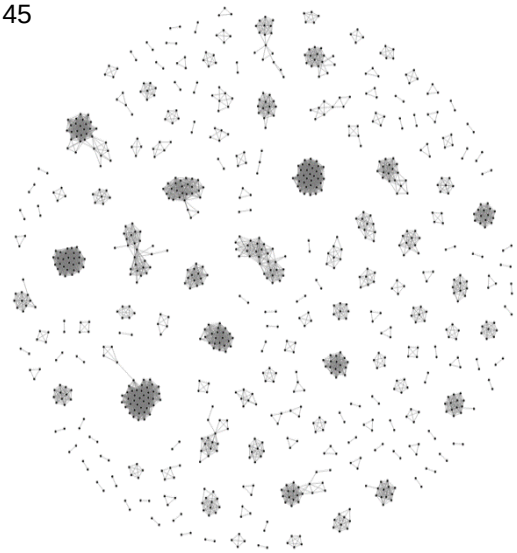


Figure 5.5 Clusters of genetically similar *N. gonorrhoeae* as defined by SNP thresholds from five to 30 SNPs
(N=number of clusters, R=range of cluster size)

5.3.2.1 Description of clusters

For the following sections, I focused the analysis on clusters defined by a five SNP threshold to describe the sexual networks with specimens most similar to each other.

5.3.2.1.1 By year, clinic, ethnicity and age

There were five clusters with specimens from three or more different years (Figure 5.6). The clusters that persisted over multiple years contained specimens from multiple clinics and a diverse mix of people from different sexual orientation, age and ethnic groups. The persistent clusters were found in lineage A and B of the phylogenetic tree.

Around half of the specimens from each clinic, ethnicity and age sub-group clustered with at least one other specimen in the PhD study sample (Table 5.2). Although single clinic/ethnicity/age group clusters were common, the majority of clusters contained a mix of specimens from different groups (Table 5.3).

Table 5.2 Distribution of specimens that cluster using a five SNP threshold by clinic, ethnicity and age

		Specimens that cluster		
	Total (N)	**n**	**Row %**	**Col %**
Clinic				
Total	**1,277**	**630**	**49.3**	**100**
Bristol	163	90	55.2	14.3
Liverpool	138	64	46.4	10.2
West London	186	80	43.0	12.7
Birmingham	404	205	50.7	32.5
MMC	386	191	49.5	30.3
Ethnicity				
Total	**1,277**	**630**	**49.3**	**100**
White	824	416	50.5	66.0
Black Caribbean	132	63	47.7	10.0
Black African	47	19	40.4	3.0
Black other	10	4	10.0	0.2
Asian	74	31	41.9	4.9
Mixed	105	54	51.4	8.6
Other	32	17	53.1	2.7
Missing	53	26	49.1	4.1
Age (years)				
Total	**1,277**	**630**	**49.3**	**100**
≤24	384	192	50.0	30.5
25 to 34	503	248	49.3	39.4
≥35	390	190	48.7	30.2

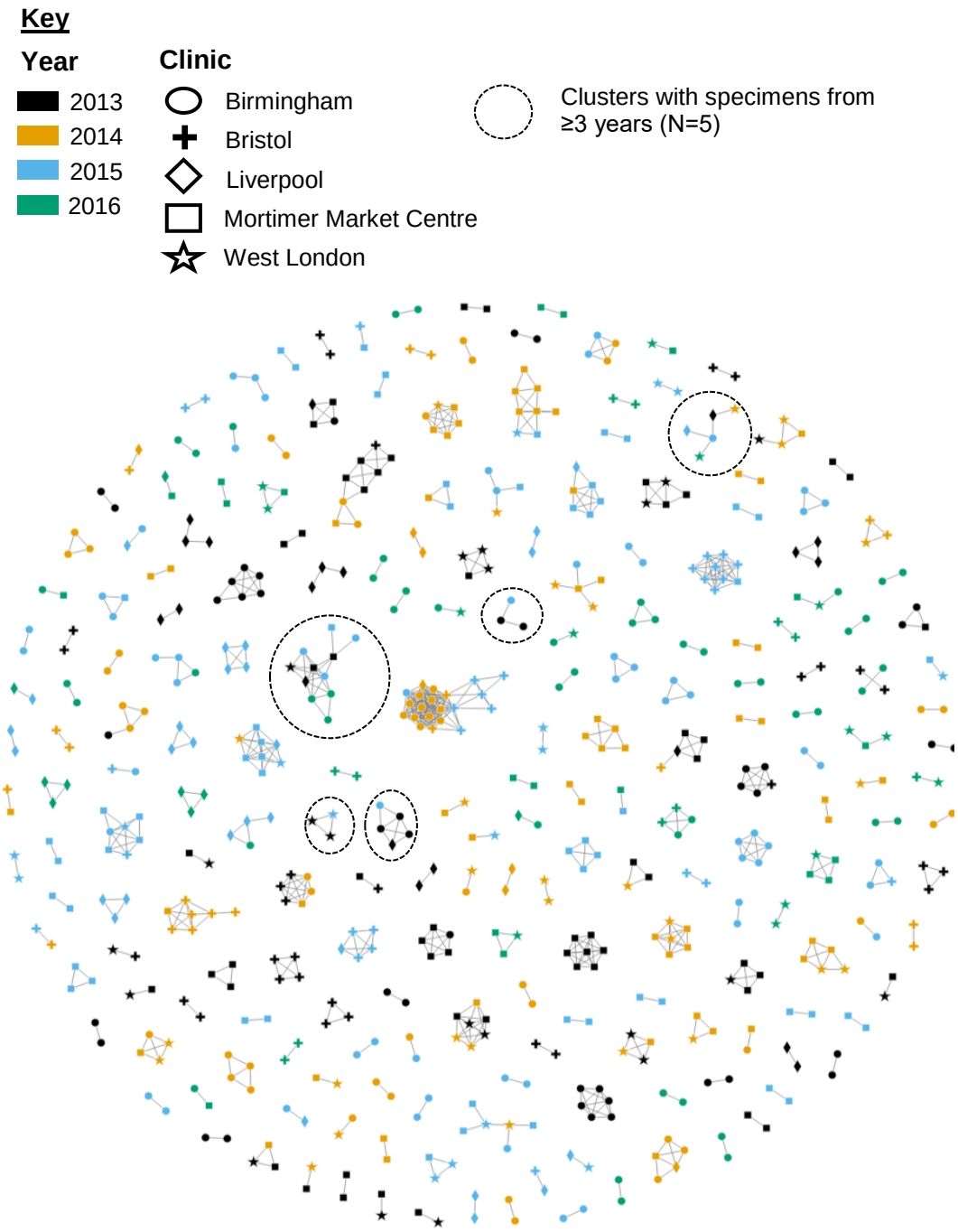


Figure 5.6 Clusters of genetically similar *N. gonorrhoeae* as defined by a five SNP threshold described by year (colours) and clinic location (shapes)

Table 5.3 Description of clusters defined by a five SNP threshold by clinic, ethnicity and age

	Clusters with at least one specimen from variable		Clusters only containing specimens from variable		
	N	Col %	n	Col %	Row %
Clinic					
Total	**213**	**100**	**135**	**100**	**63.4**
Bristol	40	18.8	20	14.8	50.0
Liverpool	35	16.4	17	12.6	48.6
West London	53	24.9	6	4.4	11.3
Birmingham	91	42.7	58	43.0	63.7
MMC	91	42.7	34	25.2	37.4
Ethnicity					
Total	**213**	**100**	**94**	**100**	**44.1**
White	181	85.0	80	85.1	44.2
Black Caribbean	47	22.1	7	7.4	14.9
Black African	19	8.9	0	0.0	0.0
Black other	4	1.9	0	0.0	0.0
Asian	25	11.7	4	4.3	16.0
Mixed	44	20.7	1	1.1	2.3
Other	17	8.0	0	0.0	0.0
Missing	24	11.3	2	2.1	8.3
Age (years)					
Total	**213**	**100**	**79**	**79**	**37.1**
≤24	122	57.3	33	41.8	27.0
25 to 34	144	67.6	27	34.2	18.8
≥35	107	50.2	19	24.1	17.8

5.3.2.1.2 Sexual orientation

Over half of the MSM specimens (54%; 414/766) and two-fifths of the specimens from women (44.2%; 91/206) or heterosexual men (41%; 125/304) were clustered. Half of the clusters only included MSM (51%; 108/213) and a quarter only specimens from women and heterosexual men (23%; 49/213) (Table 5.4). Only 2% of clusters contained specimens from all three sexual orientation groups (2%; 5/213). There were 30 clusters (14%; 30/213) that only contained specimens from heterosexual men and MSM, considerably higher than the percentage containing specimens from only women and MSM (1%; 3/213). The largest clusters identified (total specimens in cluster: N=21 and N=11) contained specimens from both heterosexual men and MSM (Table 5.5, Figure 5.7). Both clusters persisted over multiple years.

I investigated whether the characteristics of heterosexual men clustering only with women or only with MSM might differ but I found no evidence for this using the variables available (Table 5.6). A significantly higher percentage of MSM clustering with heterosexual men and MSM compared to MSM only were from the Birmingham clinic (clustered with heterosexual men and MSM: 40% (31/80); clustered with MSM:

14% (45/323); P<0.001) (Table 5.7). A higher percentage of MSM clustering with heterosexual men and MSM were born in the UK compared to those only clustered with MSM (clustered with heterosexual men and MSM: 77% (62/80); clustered with MSM only: 56% (180/323), P=0.001).

Table 5.4 Distribution of sexual orientation groups across clusters identified using a five SNP threshold

Cluster description	Clusters N (col %)	Number of specimens in cluster		
		Women N (col %)	Het. men N (col %)	MSM N (col %)
Total	**213 (100)**	**91 (100)**	**125 (100)**	**414 (100)**
Only women	9 (4.2)	20 (22.0)	-	-
Only het. men	9 (4.2)	-	18 (14.4)	-
Only MSM	108 (50.7)	-	-	323 (78.0)
Only women & het. men	49 (23.0)	60 (65.9)	63 (50.4)	-
Only women & MSM	3 (1.4)	3 (3.3)	-	3 (0.7)
Only het. men & MSM	30 (14.1)	-	36 (28.8)	80 (19.3)
Women, het. men & MSM	5 (2.3)	8 (8.8)	8 (6.4)	8 (1.9)

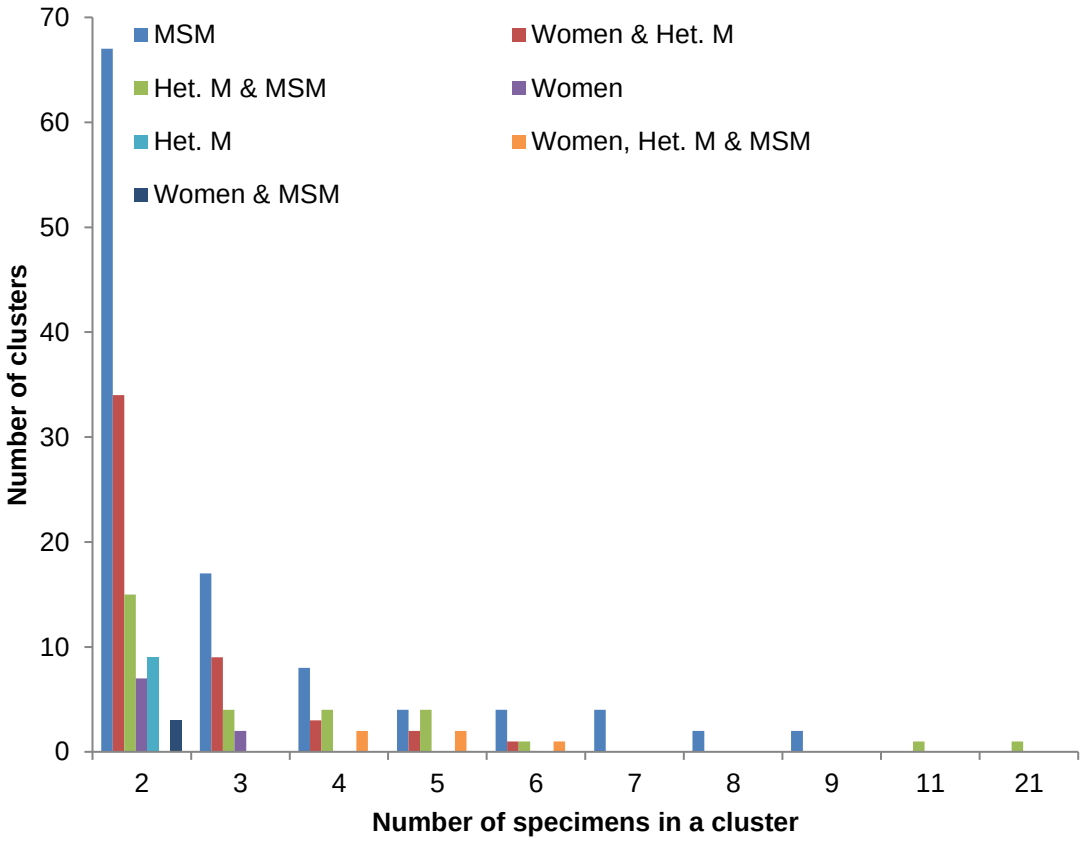


Figure 5.7 Size and frequency of clusters by sexual orientation

Table 5.5 Description of specimens in the two largest clusters identified in the PhD study sample

	Cluster N=21		Cluster N=11	
	N	%	N	%
Total	**21**	**100**	**11**	**100**
Year				
2013	0	0.0	4	36.4
2014	14	66.7	0	0.0
2015	7	33.3	4	36.4
2016	0	0.0	3	27.3
Clinic				
Bristol	8	38.0	0	0.0
Liverpool	1	4.8	1	9.1
West London	0	0.0	1	9.1
Birmingham	11	52.4	6	54.5
MMC	1	4.8	3	27.3
Sexual orientation				
Heterosexual men	3	14.3	3	27.3
MSM	18	85.7	8	72.7
Age (years)				
≤24	2	9.5	2	18.2
25-34	9	42.	5	45.4
≥35	10	47.6	4	36.2
Ethnicity				
White	16	76.2	8	72.7
Black Caribbean	3	14.3	1	9.1
Black African	1	4.8	0	0.0
Black Other	0	0.0	0	0.0
Asian	0	0.0	1	9.1
Other	0	0.0	0	0.0
Mixed	1	4.8	1	9.1
Country of birth				
UK	18	85.7	7	63.6
Not UK	3	14.3	4	36.4
Number of sexual partners in the UK in the three months prior to diagnosis				
0	6	28.6	4	36.4
1	9	42.9	4	36.4
≥2	4	19.0	0	0.0
Missing	2	9.5	3	27.3
Symptoms				
No	5	23.8	3	27.3
Yes	12	57.1	7	63.6
Missing	4	19.0	1	9.1
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis				
No/Unknown	16	76.2	8	72.7
Yes	5	23.8	3	27.3
HIV status				
Negative/Unknown	13	61.9	9	81.8
Positive	8	38.1	2	18.2
Travel-associated sexual partnership in the three months prior to diagnosis				
No	18	85.7	7	63.6
Yes	1	4.8	1	9.0
Missing	2	9.5	3	27.3

Table 5.6 Comparison of epidemiological characteristics of specimens from heterosexual men that clustered with specimens from women only or specimens from MSM only using a five SNP threshold to define clusters

	Clustered only with specimens from women		Clustered only with specimens from MSM		P Value**
	N	%*	N	%*	
Total	**63**	**100.0**	**36**	**100.0**	
Year					
2013	17	27.0	6	16.7	0.578
2014	16	25.4	8	22.2	
2015	17	27.0	13	36.1	
2016	13	20.6	9	25.0	
Clinic					
Bristol	8	12.7	8	22.2	0.210 (FE)
Liverpool	12	19.0	6	16.7	
West London	2	3.2	4	11.1	
Birmingham	38	60.3	15	41.7	
MMC	3	4.8	3	8.3	
Age (years)					
≤24	29	46.0	8	22.2	0.06
25-34	23	36.5	18	50.0	
≥35	11	17.5	10	27.8	
Ethnicity					
White	28	46.7	17	51.5	0.252 (FE)
Black Caribbean	17	28.3	3	9.1	
Black African	2	3.3	3	9.1	
Black Other	1	1.7	0	0	
Asian	5	8.3	5	15.2	
Other	2	3.3	1	3.0	
Mixed	5	8.3	4	12.1	
Country of birth					
UK	46	73.0	24	66.7	0.516
Not UK	14	22.2	10	27.8	
Symptoms					
No	10	16.1	6	17.6	0.849
Yes	52	83.9	28	82.4	
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis					
No/Unknown	57	90.5	30	83.3	0.345 (FE)
Yes	6	9.5	6	16.7	
HIV status					
Negative/Unknown	63	100.0	36	100.0	N/A
Positive	0	0.0	0	0.0	
Number of sexual partners in the UK in the three months prior to diagnosis					
0	4	7.0	1	3.2	0.618 (FE)
1	24	42.1	11	35.5	
≥2	29	50.9	19	61.3	
Travel-associated sexual partnership in the three months prior to diagnosis					
No	50	87.7	30	96.8	0.251 (FE)
Yes	7	12.3	1	3.2	

*Percentages may not add up to 100 because of missing data **Chi^2 test was used except where expected cell size was <5, then the Fisher's exact test was used (indicated by FE)[235]

Table 5.7 Comparison of epidemiological characteristics of specimens from MSM that clustered with specimens from heterosexual men and MSM or only clustered with specimens from MSM using a five SNP threshold to define clusters

	Clustered with specimens from heterosexual men and MSM		Clustered only with specimens from MSM		P Value**
	N	%*	N	%*	
Total	**80**	**100**	**323**	**100**	
Year					
2013	24	30.0	89	27.6	0.445
2014	27	33.8	89	27.6	
2015	22	27.5	99	30.7	
2016	7	8.8	46	14.2	
Clinic					
Bristol	15	18.8	41	12.7	**<0.001**
Liverpool	7	8.8	16	5.0	
West London	11	13.8	60	18.6	
Birmingham	31	38.8	45	13.9	
MMC	16	20.0	161	49.8	
Age (years)					
≤24	13	16.3	67	20.7	0.659
25-34	36	45.0	140	43.3	
≥35	31	38.8	116	35.9	
Ethnicity					
White	69	87.3	240	75.7	0.149 (FE)
Black Caribbean	5	6.3	13	4.1	
Black African	0	0.0	8	2.5	
Black Other	0	0.0	2	0.6	
Asian	1	1.3	14	4.4	
Other	0	0.0	12	3.8	
Mixed	4	5.1	28	8.8	
Country of birth					
UK	62	77.5	180	55.7	**0.004**
Not UK	17	21.3	116	35.9	
Number of sexual partners in the UK in the three months prior to diagnosis					
0	3	4.9	3	1.9	0.201 (FE)
1	21	34.4	42	27.3	
≥2	37	60.7	109	70.8	
Symptoms					
No	18	25.7	97	34.6	0.155
Yes	52	74.3	183	65.4	
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis					
No/Unknown	60	75.0	208	64.6	0.077
Yes	20	25.0	114	35.4	
HIV status					
Negative/Unknown	59	73.8	222	68.7	0.382
Positive	21	26.3	101	31.3	
Travel-associated sexual partnership in the three months prior to diagnosis					
No	54	88.5	141	91.6	0.490
Yes	7	11.5	13	8.4	

*Chi^2 test was used except where expected cell size was <5, then the Fisher's exact test was used, indicated by (FE)[235]

5.3.2.1.3 HIV status

Most clusters were comprised of HIV-negative/unknown people only (68%; 146/213) but 3% (7/213) comprised of HIV-positive people only (Figure 5.8). Of the clusters including both HIV-positive and HIV-negative/unknown people (28%; 60/213), 85% (51/60) included only MSM, 13% (8/60) included heterosexual men and MSM and one cluster contained two specimens from heterosexual men only (Figure 5.9). Within the clusters containing MSM and heterosexual men, all the heterosexual men were HIV-negative/unknown (100%; 36/36) and 26% (21/80) of the MSM were HIV-positive. The largest clusters identified in the PhD study sample (total specimens: N=11 and N=21) were comprised of specimens from HIV-positive people and HIV-negative/unknown people, and as described in the previous section (5.3.2.1.2), contained specimens from MSM and heterosexual men (Table 5.5).

Of the HIV-negative/unknown MSM that clustered with other MSM, 42% (124/292) clustered with HIV-positive MSM (Table 5.8a). I investigated the characteristics of these men and in univariate analysis found that HIV-negative/unknown MSM aged between 25-34 years were more likely to cluster with HIV-positive MSM than younger HIV-negative/unknown MSM (base: ≤24 years, cOR 2.06, CI 1.3-4.11, P=0.04), but this association did not remain in the multivariable model. In the adjusted analysis, HIV-negative/unknown MSM from the West London, Birmingham or MMC clinics were more likely to cluster with HIV-positive MSM compared to HIV-negative/unknown MSM attending the Bristol clinic (base: Bristol clinic, West London: aOR 7.22, CI 2.65-19.68, P<0.001, Birmingham: aOR 4.26, 1.47-12.33, P=0.008, MMC: aOR 3.89, CI 1.63-9.27, P=0.002). Clustering of HIV-negative/unknown MSM with HIV-positive MSM was less likely to occur in 2016 compared to 2013 (base: 2013, 2016: aOR 0.22, CI 0.08-0.06, P=0.003) (Table 5.8b).

5.3.2.1.4 Follow-up of HIV-negative/unknown people

Between 2013 and 2016, 33 people included in the PhD study sample were diagnosed with HIV after they were infected with gonorrhoea (3% (33/1,051) of all HIV-negative/unknown people). All were MSM (6% (33/546) of all HIV-negative/unknown MSM) and the majority were diagnosed within one year (75%; 25/33) (although the follow-up time was not equal for all people in the study as data were only available until the end of 2016). Half of the people diagnosed with HIV after they were infected with gonorrhoea clustered with others in the PhD study

sample (52%; 17/33). Most clustered with other MSM only (65%; 11/17) and half were in clusters containing a mix of HIV-positive and HIV-negative/unknown people (47%; 8/17).

There were three people diagnosed with HIV after they were infected with gonorrhoea who had two gonorrhoea diagnoses in the PhD study sample. Two of these people were diagnosed with HIV after the first gonorrhoea infection but before the second. Both were clustered in mixed HIV status clusters during the first gonorrhoea episode. The third person was diagnosed with HIV after the second gonorrhoea infection. This third person was clustered with HIV-negative/unknown MSM only during the first infection, and clustered with HIV-positive and HIV-negative/unknown people during the second infection.

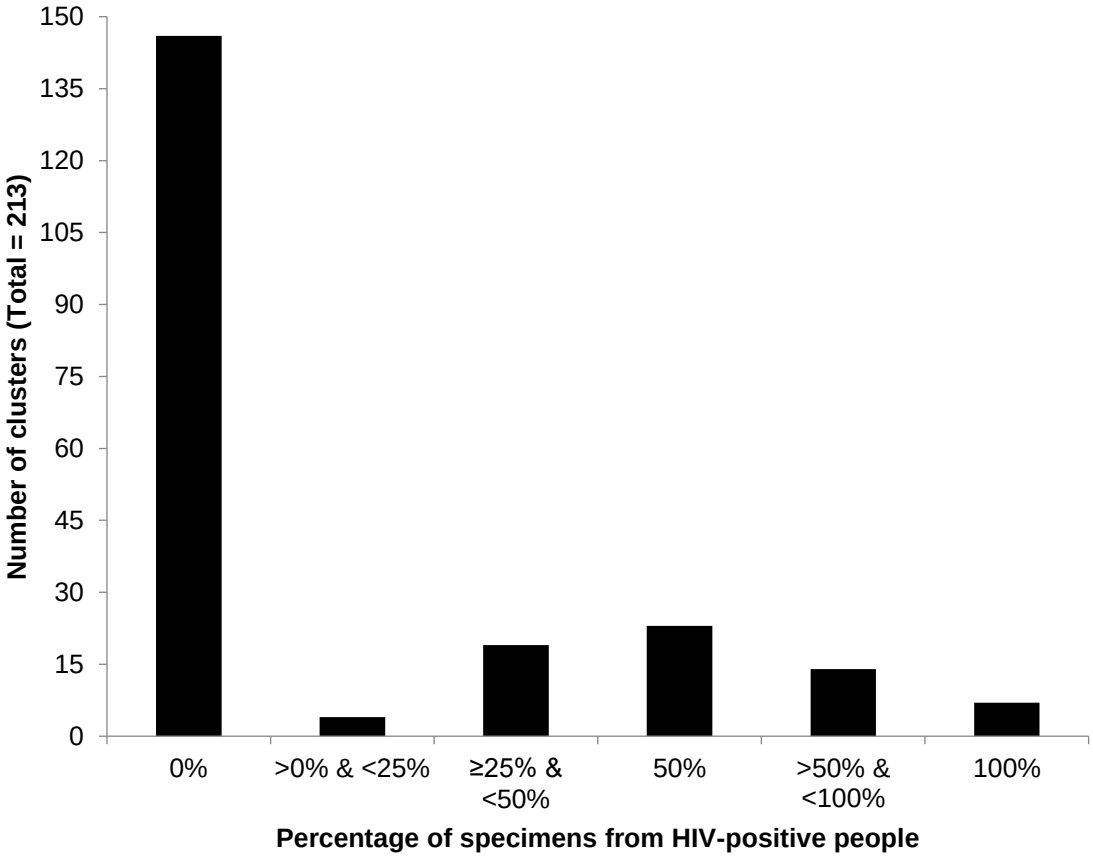


Figure 5.8 Number of clusters by percentage of specimens from HIV-positive people in each cluster using a five SNP threshold

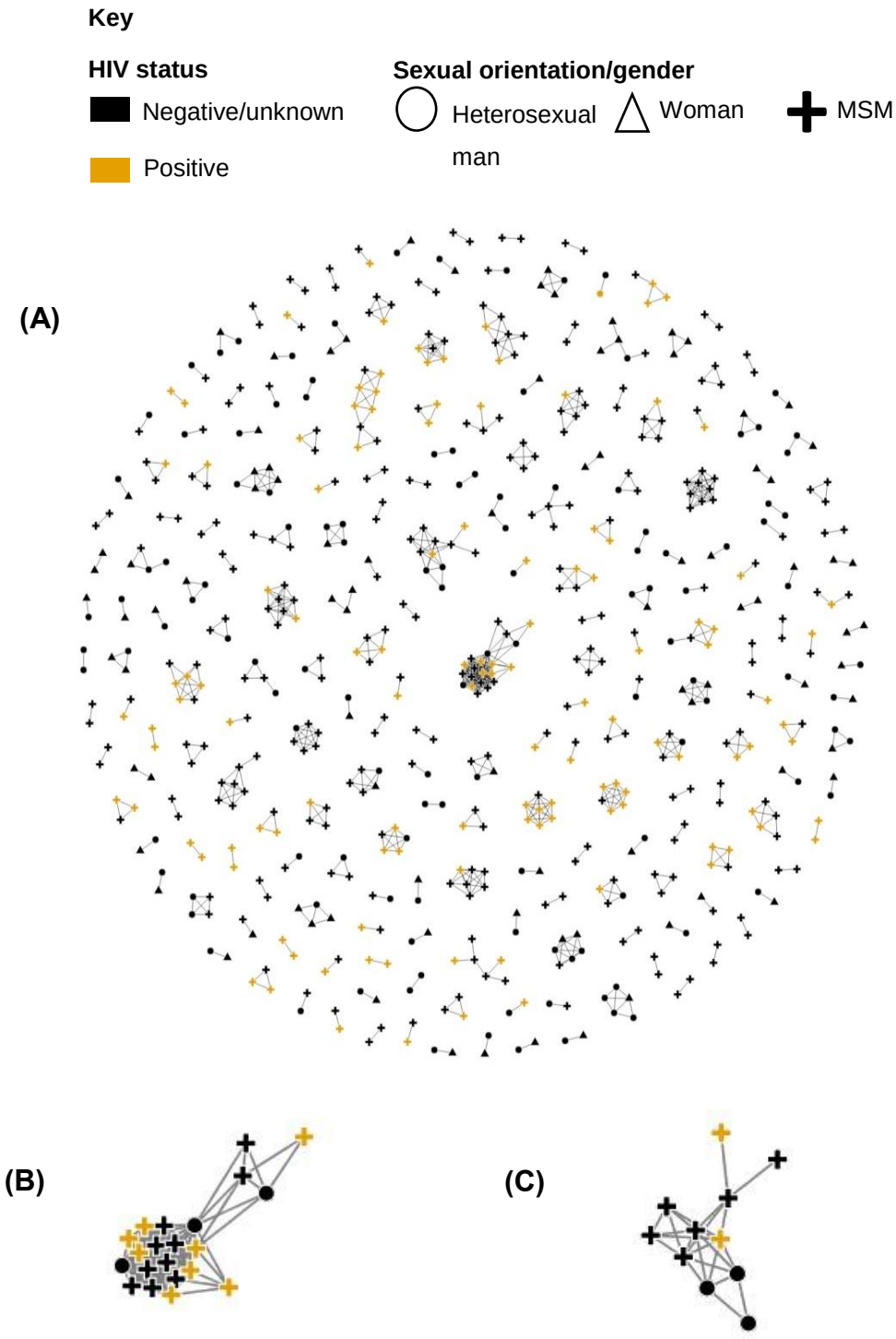


Figure 5.9 Clusters of genetically similar *N. gonorrhoeae* using a five SNP threshold with sexual orientation/gender (shape) and HIV status (colour) of the person

(A) all clusters (total clusters=213), (B) largest cluster (total specimens=21), (C) second largest cluster (total specimens=11)

Table 5.8 Univariate and multivariable analysis comparing the epidemiological characteristics of HIV-negative/unknown MSM clustering with (i) HIV-negative/unknown MSM only and (ii) HIV-positive MSM, using five SNP threshold to define clusters
(A) Univariate analysis

cOR = crude odds ratio, LCI = lower 95% confidence interval, UCI = upper 95% confidence interval

		Cluster with specimens from HIV-negative/ unknown MSM only		Cluster with specimens from HIV-positive MSM		Outcome: cluster with specimens from HIV-positive MSM			
		n	Row %	n	Row %	cOR	LCI	UCI	P
Total	**222**	**120**	**54.1**	**102**	**45.9**				
Year									
2013	56	29	51.8	27	48.2	1	-	-	-
2014	60	26	43.3	34	56.7	1.40	0.67	2.94	0.364
2015	68	36	52.9	32	47.1	0.95	0.47	1.94	0.898
2016	38	29	76.3	9	23.7	**0.33**	**0.13**	**0.86**	**0.017**
Clinic									
Bristol	38	29	76.3	9	23.7	1	-	-	-
Liverpool	12	10	83.3	2	16.7	0.64	0.12	3.58	0.613
West London	45	17	37.8	28	62.2	**5.31**	**1.86**	**15.1**	**0.001**
Birmingham	36	20	55.6	16	44.4	2.58	0.92	7.20	0.061
MMC	91	44	48.4	47	51.6	**3.44**	**1.42**	**8.35**	**0.004**
Age (years)									
≤24	55	37	67.3	18	32.7	1	-	-	-
25-34	104	52	50.0	52	50.0	**2.06**	**1.03**	**4.11**	**0.04**
≥35	63	31	49.2	32	50.8	2.12	0.99	4.56	0.05
Ethnicity									
White	168	92	54.8	76	45.2	1	-	-	-
Black Caribbean	4	1	25.0	3	75.0	3.63	0.36	36.2	0.239
Black African	5	1	20.0	4	80.0	4.84	0.52	45.2	0.126
Black Other	1	1	100	0	0	-	-	-	-
Asian	10	3	30.0	7	70.0	2.82	0.70	11.4	0.128
Other	8	5	62.5	3	37.5	0.73	0.17	3.15	0.668
Mixed	21	14	66.7	7	33.3	0.61	0.23	1.58	0.301
Country of birth									
UK	136	77	56.6	59	43.4	1	-	-	-
Not UK	76	38	50.0	38	50.0	1.31	0.74	2.30	0.355
Symptomatic infection									
No	67	37	55.2	30	44.8	1	-	-	-
Yes	126	65	51.6	61	48.4	1.16	0.64	2.10	0.631
Diagnosed with an STI (excluding HIV) in the previous year									
No/Unknown	161	90	55.9	71	44.1	1	-	-	-
Yes	61	30	49.2	31	50.8	1.31	0.72	2.37	0.371
Number of sexual partners in the three months prior to diagnosis									
0	3	1	33.3	2	66.7	1	-	-	-
1	32	19	59.4	13	40.6	0.34	0.03	4.45	0.390
≥2	89	51	57.3	38	42.7	0.37	0.03	4.36	0.413
Travel-associated sexual partnership in the three months prior to diagnosis									
No	113	63	55.8	50	44.2	1	-	-	-
Yes	11	8	72.7	3	27.3	0.47	0.12	1.90	0.279

(B) Multivariable analysis

aOR = adjusted odds ratio, LCI = lower 95% confidence interval, UCI = upper 95% confidence interval

	Outcome: cluster with specimens from HIV-positive MSM			
	OR	**LCI**	**UCI**	**P value**
Year				
2013	1	-	-	-
2014	1.29	0.59	2.82	0.523
2015	0.94	0.44	2.00	0.867
2016	**0.22**	**0.08**	**0.60**	**0.003**
Clinic				
Bristol	1	-	-	-
Liverpool	0.69	0.13	3.78	0.667
West London	**7.22**	**2.65**	**19.7**	**0.000**
Birmingham	**4.26**	**1.47**	**12.3**	**0.008**
MMC	**3.89**	**1.63**	**9.27**	**0.002**

5.4 Discussion and conclusions

In this chapter I investigated the transmission of *N. gonorrhoeae* lineages within and between defined population groups. The analysis was unique in using sequence data for a large number of specimens collected over several years from different clinics in England. I observed substantial genetic heterogeneity in the *N. gonorrhoeae* population, with pronounced variation observed by time, person and place. However, clusters of genetically similar specimens representing sexual networks with potentially rapid gonorrhoea transmission were also identified, including clusters that spanned large geographical distances.

Through cluster analysis the WGS data revealed a population of men who reported solely heterosexual behaviours but who were closely linked to sexual networks transmitting gonorrhoea between MSM. Heterosexually-identifying MSM are likely to be at high risk of STIs and HIV but are unlikely to benefit from the comprehensive sexual health testing and prevention messages routinely targeted at gay and bisexual-identifying MSM.[247-252] While the cluster analysis supported the likelihood of HIV serosorting behaviours occurring among MSM, most clusters had mixed HIV status highlighting a group of MSM who were not infected but could be at risk of HIV transmission. These findings may be useful in the clinical setting in refining individual patient risk assessments – potentially allowing clinicians to demonstrate to patients who might be unaware that they are linked to high-risk networks. This might also

inform clinical and patient-based decisions about pre-exposure prophylaxis (PrEP) and other treatment or prevention messages.

At a public health level the WGS data combined with epidemiological information provides further insight into the sexual mixing of different sub-groups of the population, which could help inform the development and targeting of *N. gonorrhoeae* and HIV prevention and control measures, as discussed in the following sections.

5.4.1 The genetic variation of *N. gonorrhoeae* in England across time and place

I identified several genetically similar clusters that persisted across three or more years. These were found in multiple clinics and may represent endemic, rather than imported, *N. gonorrhoeae* strains. These persistent strains were primarily found in the larger of the two high-level lineages (A rather than B). Investigating why some strains persist while others may not might further understanding of why, despite control efforts, *N. gonorrhoeae* continues to spread in some populations but not others. These persistent strains might have characteristics that facilitate the persistence of the pathogen, such as causing fewer symptoms, which might delay detection and treatment resulting in an increased duration of infection and likelihood of transmission. Alternatively, these strains may be associated with transmission within specific sexual networks, perhaps where there is a high rate of sexual partner change such that *N. gonorrhoeae* is spread before detected, or that a significant proportion of the people in those sexual networks do not attend STI testing services.

There was considerable evidence of rapid geographic dissemination of *N. gonorrhoeae.* The large number of clusters containing specimens from multiple clinics indicates that sexual networks in England are spread over large distances although it may be that only a few individuals facilitate geographical dissemination.

5.4.2 The epidemiological characteristics of gonorrhoea sexual transmission networks

5.4.2.1 Sexual mixing by age and ethnicity

Disassortative sexual mixing by age was demonstrated by the evidence that most clusters contained specimens from different age groups, which is similar to the results of other epidemiological studies investigating sexual mixing patterns by age.[63,253-255] The ability to detect differences between clusters by ethnicity was

constrained because most specimens in the PhD study sample were from people of white ethnicity. However, some clusters comprised specimens from at least two different ethnic groups, suggesting disassortative sexual mixing by ethnicity occurs, as shown in a recent study of sexual health clinic attendees in England.[80]

5.4.2.2 Sexual networks containing heterosexual men and MSM

Clusters usually consisted of people of the same sexual orientation. However, the genomic data provide good evidence of men that only report heterosexual behaviour being closely linked to sexual networks where *N. gonorrhoeae* is transmitted through sex between men. Mixed clusters of men reporting heterosexual behaviour and MSM were considerably more common than clusters of MSM and women. Care is required in the interpretation because the WGS data do not provide evidence of direct transmission (or the direction of transmission) even where the specimens are identical genetically and it is possible that the heterosexual men have intervening female partners who have sex with MSM. The epidemiological characteristics of the heterosexual men who clustered only with MSM did not differ from those of heterosexual men who clustered only with women for any of the characteristics with data available. However, this could be explored further if more behavioural data were available. Other studies have found that men reporting heterosexual behaviours but in sexual networks with MSM were more likely to be of black African ethnicity.[256] Sexual orientation data reported to GUMCAD are supposed to be based on self-reported behaviour of the individual.[61,257] Although miscoding of sexual orientation by clinic staff may bias interpretation and lead to overestimation of the number of apparently heterosexual men who are part of MSM networks transmitting *N. gonorrhoeae*, this is likely to affect the minority of cases.

A mismatch between self-identified sexual orientation and reported sexual behaviour has previously been described in in-depth confidential qualitative interviews as part of a hepatitis B outbreak investigation[258] and in HIV phylogenetic studies.[256,259] A British population-based survey has shown that more than a quarter of men reporting sex with men identified as heterosexual.[260] Where there is discrepancy between identity and male same-sex behaviour, public health and clinical practice interventions and messaging that target gay and bisexual-identifying MSM only may miss and even disenfranchise heterosexual-identifying MSM.[248,260] In the longer term, if WGS data are available in a clinically relevant timeframe and analysed with WGS data from other specimens to provide appropriate context, this could provide

additional information to clinicians about sexual risk profiles and support patient management, including testing for other STIs, such as hepatitis B[261], and HIV prevention in the form of PrEP and behavioural risk reduction advice. These actions would need to be implemented in an appropriately sensitive manner with regard for confidentiality and the potential for deductive disclosure[248] but would be applicable to anyone found to be in a close sexual network with MSM.

5.4.2.3 Sexual networks of MSM by HIV status

There was evidence of possible sero-sorting behaviours amongst MSM, as some clusters of infection only contained people of the same HIV status. However, a larger number of clusters comprised both HIV-positive and HIV-negative/unknown MSM, which is perhaps not surprising given the inherent problems with serosorting as a strategy (i.e. undiagnosed infections and misreporting of status). Some of the HIV-negative/unknown MSM in these mixed status networks were subsequently diagnosed with HIV. These findings imply groups of MSM without HIV who are participating in sexual networks transmitting gonorrhoea who might therefore be particularly in need of HIV prevention management, such as PrEP, as well as other STI tests, including hepatitis C.[262]

Using the data available, I investigated whether these MSM were in any way measurably different but the epidemiological characteristics of people who were HIV-negative/unknown who clustered with HIV-positive MSM could not be distinguished from those that did not cluster. However, the sexual behaviour data available in this study was limited and more detailed information might reveal important differences between these groups. The increase in availability and use of anti-retroviral therapies in both HIV-positive and HIV-negative MSM may explain the substantial number of mixed HIV status clusters and this may have changed over time as treatment as prevention (TasP) and PrEP became more readily available and/or widely used. Sexual behaviour and specifically condom use has likely been affected by the increase in HIV-positive MSM with an undetectable viral load[263] and PrEP use in HIV-negative MSM.[264] With the reduced risk of HIV transmission, HIV sero-adaptive behaviours may have declined or become more complex.[265] For example, some MSM may make decisions on their sexual activities based on their own viral load or PrEP use or that of their prospective partners.[75,266] A recent study of *N. gonorrhoeae* WGS data from MSM in Brighton also found clusters of mixed HIV status more frequently than clusters including only HIV-positive people.[167]

5.4.3 Methodological considerations for defining clusters

The methods used in this study rely on the underlying premise of molecular epidemiology that specimens with similar genomes, i.e. fewer SNP differences between them, are more likely to share a common ancestor than specimens with dissimilar genomes, and that the degree of similarity reflects the time since divergence from a common ancestor. It follows that the people infected with *N. gonorrhoeae* from the same genetic cluster are likely to be part of a sexual network, although it is not possible to know whether sexual transmission has been direct or indirect without detailed case interviews, nor is it easy to establish the direction of transmission, although time of diagnosis can provide clues to this. More advanced phylogenetic analyses that incorporate time into the model can estimate the temporal relationship between specimens more precisely. These methods require a strong positive correlation between specimen isolation dates and the genetic distance from the root of the phylogenetic tree to the tip (known as a good temporal signal). However, a good temporal signal is difficult to achieve if the sampling period is short or if only a small percentage of infections in the transmission network are included or in the presence of horizontally acquired genetic variation, for example by homologous recombination.[267,268] Therefore, as the sampling period for the PhD study was only three months for each year, these advanced phylogenetic analyses were not completed.

One of the interesting challenges I faced in this analysis was that there is no standardised method to define a genetic cluster, nor is it clear how well a given cluster represents a sexual network. To address this problem, I used a range of SNP thresholds from five to 30. Five SNPs difference suggests nearly identical specimens and that people are within a close sexual network and transmission timeframe. Naturally, as the SNP threshold increased, a higher percentage of the specimens clustered. Most clusters identified through SNP differences consisted of only two specimens, even when the SNP threshold was relaxed. The number of small clusters is related to the sampling strategy, which includes an unknown but probably small proportion of all infections in a given sexual network.

A study of 1,218 *N. gonorrhoeae* specimens from 21 countries in Europe found that when the SNP threshold was increased to ≥30 SNPs, any two specimens being compared could have come from any of the countries in the study.[142] Another study of 1,407 specimens from consecutive gonorrhoea cases in one clinic in England

collected over four years calculated the range of SNPs expected between two specimens within the same sexual network.[118] This was achieved by modelling data on *N. gonorrhoeae* mutation rates, genomic diversity between known sexual contacts and the date of infection. The authors predicted specimens collected simultaneously and up to nine SNPs apart might be part of the same sexual network, and that there might be up to 14 SNPs between specimens collected four years apart. Given these two studies, it is likely that a 30 SNP threshold is too large to identify recent transmission clusters and the five SNP threshold used in this chapter was probably a reasonable compromise to allow me to identify people within close sexual networks.

Another consideration for *N. gonorrhoeae* in defining clusters by SNP differences is the high frequency of DNA recombination. Recombination causes large sections of DNA to be moved within the genome in a short space of time. If they are not removed from analyses, SNPs within the inserted sections would be counted in the SNP difference cluster analysis. In this event, specimens with a recent common ancestor may be incorrectly separated and grouped with other specimens.[165] In my PhD study sample, I found the amount of genetic diversity introduced by recombination events compared to point mutation was similar to previous studies.[119,121] Although efforts were taken to identify and remove recombination events, it is likely that some remained and this may have introduced errors in the clustering.

5.4.4 Public health application

Although the sexual mixing patterns described using the WGS data have been identified before in epidemiological studies, the WGS data has helped to quantify the frequency and density of these networks, which improves our understanding of the relative importance of different sexual networks in the spread of *N. gonorrhoeae*. For example, the third most common sexual network type by sexual orientation consisted of men who only reported heterosexual behaviour and MSM, and the largest clusters only comprised people in these two groups. This finding highlights the complexity of the sexual network structure and that the relationships between men reporting heterosexual behaviour and MSM are important for shaping the *N. gonorrhoeae* population. This complexity should be considered when parameterising sexual mixing patterns in mathematical models of *N. gonorrhoeae* transmission and when developing interventions to prevent and control infection. The WGS data could also

be used to determine the sexual networks an individual belongs to, which might inform individual level clinical risk assessment and management considerations.

5.4.5 Chapter summary

To answer the research questions posed in this chapter, I have (i) described and compared the genetic variation of *N. gonorrhoeae* by time and location, and (ii) identified and described gonorrhoea transmission networks by using a combination of epidemiological and WGS data. In summary:

- Clusters of genetically similar *N. gonorrhoeae* indicate that some strains persisted over multiple years
- There was evidence of rapid geographic dissemination of *N. gonorrhoeae*
- WGS data identified a considerable number of sexual networks containing both men who report only heterosexual behaviours and MSM
- Analysis of WGS data suggested extensive sexual mixing between HIV-positive and HIV-negative/unknown MSM

The implications of these findings are:

- Further biological and epidemiological investigation into persistent strains may help inform our understanding of how to control *N. gonorrhoeae*
- By providing information about the sexual networks a patient belongs to and likely sexual health risks, WGS data could influence clinical decision-making, which could support more tailored and relevant interventions, such as HIV PrEP
- Sexual mixing patterns of men who only report only heterosexual behaviour and MSM should be included in mathematical models that use sexual mixing parameters to explore the transmission dynamics of *N. gonorrhoeae* and AMR

6 Phenotypic and genotypic characteristics of *N. gonorrhoeae* antimicrobial resistance in sexual networks

6.1 Introduction

To understand the spread of antimicrobial resistant (AMR) *N. gonorrhoeae* and which patient groups are at risk of these infections, surveillance programmes and research studies often link phenotypic microbiological data on *N. gonorrhoeae* susceptibility with data on the epidemiological characteristics of cases. However, these analyses are limited because specimens with identical phenotypes may be genetically unrelated and genetically related specimens may have different phenotypes, which makes it impossible to know whether AMR spread is due to clonal transmission or separate introductions. Combining phenotypic with genomic data can help resolve these uncertainties and identify specific genetic markers of resistance.

In this chapter I aimed to answer the following research questions:

- What is the genomic and phenotypic variation in *N. gonorrhoeae* antimicrobial susceptibility in England?
- What are the epidemiological characteristics of sexual transmission networks (defined by genetic clustering) that more frequently transmit *N. gonorrhoeae* with reduced susceptibility to antimicrobials?

To do this, I used the WGS data to identify genetic markers of antimicrobial resistance and compare these markers with the phenotypic antimicrobial susceptibility data. I focused on mutations in the *penA* allele, particularly the mosaic *penAXXXIV* allele, which plays an important role in antimicrobial resistance to the current and last line of therapy, ceftriaxone. I also compared the phenotypic, genotypic and epidemiological characteristics of genetically distinct *N. gonorrhoeae* clusters. The results of these analyses will inform understanding about how *N. gonorrhoeae* with reduced susceptibility (RS) to antimicrobials are distributed in sexual networks in England.

6.2 Methods

6.2.1 Association between *N. gonorrhoeae* phenotypic antimicrobial susceptibility and genetic lineages

To investigate whether specimens in the two distinct lineages (A and B) of *N. gonorrhoeae* identified in Chapter 5 had different phenotypic antimicrobial susceptibility profiles, I conducted univariate and multivariable logistic regression (using the same methods reported in Chapter 5) with the outcome that the specimen was found in lineage A and the explanatory variable the antimicrobial susceptibility profile of the specimen. Reduced susceptibility thresholds presented in Table 6.1 were used for azithromycin, ceftriaxone and cefixime because there were insufficient specimens in the PhD study sample exhibiting phenotypic resistance as defined by international standards.[100] Standard resistance thresholds were used for ciprofloxacin and penicillin. A separate multivariable model was created for each of the five antimicrobials assessed and each adjusted for the patient variables found to be significantly associated with lineage in Chapter 5 (clinic, age, sexual orientation and ethnicity).

Table 6.1 Antimicrobial minimum inhibitory concentration threshold used to define reduced susceptibility and resistance of *N. gonorrhoeae* specimens

	MIC thresholds (mg/L)	
Antimicrobial	**Reduced susceptibility**	**Resistance***
Azithromycin	≥0.25	>0.5
Ceftriaxone	≥0.015	>0.125
Cefixime	≥0.03	>0.125
Ciprofloxacin	N/A	>0.06
Penicillin	≥1	>1 or β-lactamase positive

* as defined by EUCAST[100]

6.2.2 Identifying genetic markers associated with resistance to antimicrobials

Known genetic markers of antimicrobial resistance were identified using ARIBA (Antimicrobial Resistance Identification By Assembly).[269] ARIBA was selected because it is computationally fast and performs well. The input data were the raw sequence reads files for each specimen and a database of known genetic marker sequences.[168] The phenotypic MIC values for five antimicrobials (azithromycin, ceftriaxone, cefixime, ciprofloxacin and penicillin) were compared graphically to the genetic markers using ARIBA. For ease of visualisation the graphs were restricted to combinations of markers with at least five specimens. For specimens from 2016, data were only available on whether the specimen was susceptible or resistant to

ciprofloxacin, so for MIC analyses all resistant specimens were allocated an MIC of 1.0 mg/L and all sensitive specimens were allocated an MIC of 0.03 mg/L.

6.2.3 Describing the distribution of *penA* alleles

Phandango[244] was used to visualise the distribution of *penA* alleles and phenotypic ceftriaxone and cefixime MIC values across the phylogenetic tree produced in Chapter 5. The epidemiological characteristics of specimens found in large clusters (>10 specimens) with elevated MICs to ceftriaxone and cefixime and the same *penA* allele were compared. The Chi^2 test was used to assess if the differences in epidemiological characteristics of cases between clusters was significant using a P value <0.5.[235] The NG-MAST sequence types for these specimens were extracted *in silico* using the online platform for genome sequencing analysis Pathogenwatch v2.6.0.[270] Pathogenwatch compares the raw sequence reads for the genes used in NG-MAST with the online databases of sequence types.

6.2.4 Describing the phenotypic, genotypic and epidemiological characteristics of clusters

The phenotypic, genotypic and epidemiological characteristics of specimens in clusters identified in Chapter 5 using the five SNP threshold were described and compared. The epidemiological characteristics of specimens in each cluster were aggregated to describe the overall epidemiological characteristics of each cluster, known from here as the 'cluster type'. Associations between the cluster type and maximum modal MIC of the cluster were assessed using univariate and multivariable logistic regression following the same methods reported in Chapter 5. The modal MIC threshold for the outcome was the reduced susceptibility threshold presented in Table 6.1. No analyses were completed for ciprofloxacin because data were only available on whether the specimen was susceptible or resistant to ciprofloxacin for specimens from 2016, so it was not possible to calculate a modal ciprofloxacin MIC for each cluster.

6.3 Results

6.3.1 Association between *N. gonorrhoeae* phenotypic antimicrobial susceptibility and genetic lineages

Specimens with reduced susceptibility to ceftriaxone, cefixime, azithromycin, ciprofloxacin and penicillin were dispersed across the phylogenetic tree (Figure 6.1). However, there were sections of the tree containing more specimens with reduced susceptibility to antimicrobials. Specimens in lineage A were more likely to have higher MICs than specimens in lineage B (Table 6.2).

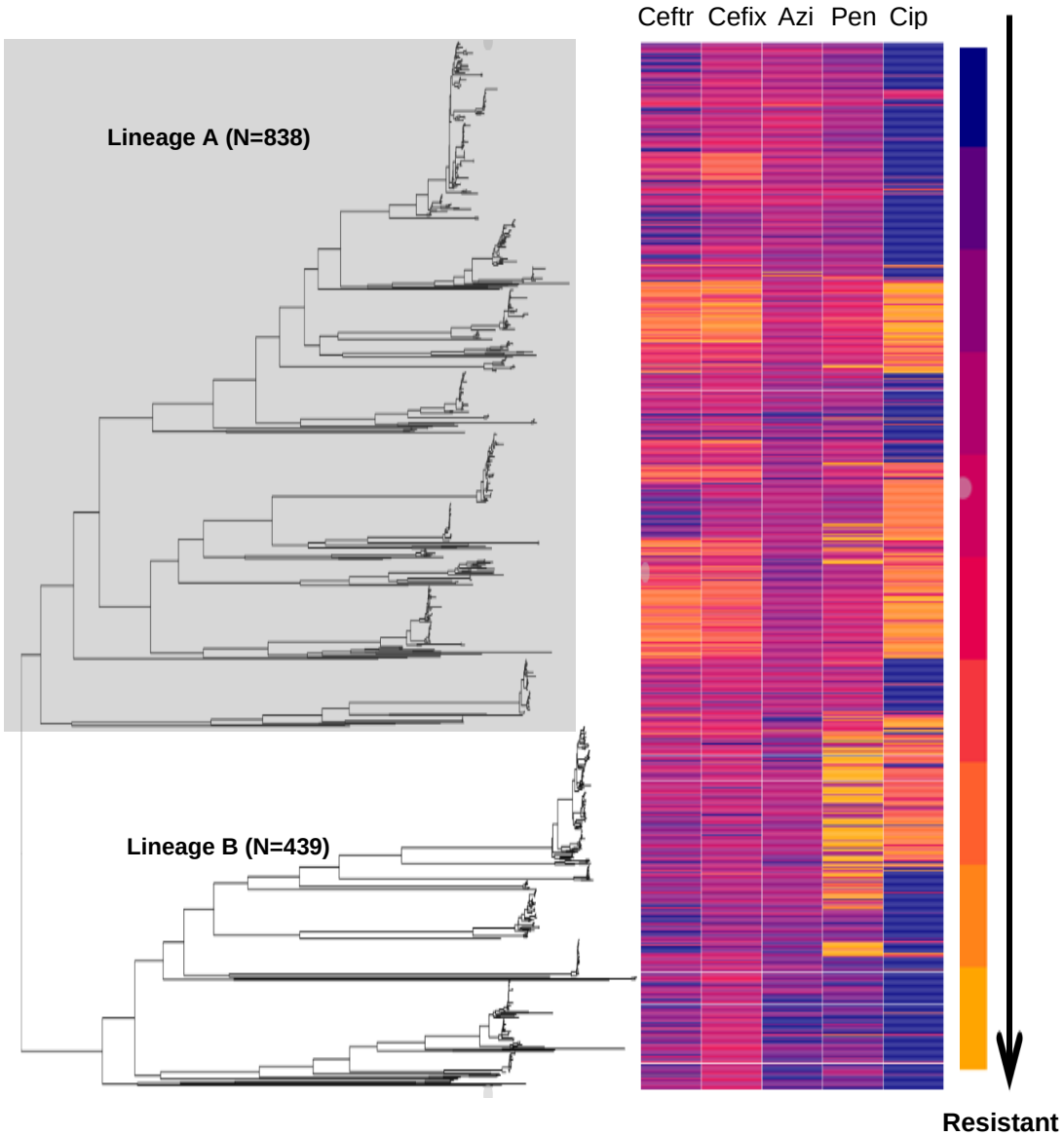


Figure 6.1 Distribution of antimicrobial susceptibility of *N. gonorrhoeae* specimens across the phylogeny
Ceftr= ceftriaxone, Cefix = cefixime, Azi = azithromycin, Pen = penicillin, Cip = ciprofloxacin

Table 6.2 Association between the antimicrobial susceptibility of *N. gonorrhoeae* specimens and presence in lineage A of the phylogeny
aOR = adjusted odds ratio, LCI = lower 95% confidence interval, UCI = upper 95% confidence interval. Each model is adjusted for clinic, age, sexual orientation and ethnicity. The total number of specimens differs to that reported in Chapter 5 because nine specimens did not have any MIC data.

	Specimen is in lineage B (n)	Specimen is in lineage A (n)	aOR	LCI	UCI	P value
Reduced susceptibility to ceftriaxone (MIC ≥0.015 mg/L)						
No	418	572	1.00	-	-	-
Yes	15	263	**16.4**	**8.9**	**30.0**	**<0.001**
Reduced susceptibility to cefixime (MIC ≥0.03 mg/L)						
No	370	544	1.00	-	-	-
Yes	63	291	**4.33**	**2.97**	**6.32**	**<0.001**
Reduced susceptibility to azithromycin (MIC ≥0.25 mg/L)						
No	367	328	1.00	-	-	-
Yes	66	507	**7.55**	**5.36**	**10.60**	**<0.001**
Resistance to penicillin (MIC >1 mg/L and/or β-lactamase positive)						
No	378	671	1.00	-	-	-
Yes	55	164	1.38	0.94	2.01	0.100
Resistance to ciprofloxacin (Mic >0.06 mg/L)						
No	408	400	1.00	-	-	-
Yes	25	435	**17.1**	**10.7**	**27.4**	**<0.001**

6.3.2 Genetic markers of antimicrobial resistance

In the PhD study sample, 32 known mutations were identified in eight genes associated with resistance to azithromycin, ceftriaxone, cefixime, ciprofloxacin or penicillin in previous studies (Table 6.3).[129,269] The genetic markers associated with resistance were compared to the phenotypic MIC for each specimen (Figure 6.2, Figure 6.3, Figure 6.4, Figure 6.5, Figure 6.6). For all antimicrobials, there were specimens with the same combination of genotypic markers of resistance but differing phenotypic MICs.

6.3.2.1 Ceftriaxone and cefixime

Specimens with the highest ceftriaxone and cefixime MICs contained a combination of mutations in the *penA*, *porB1b* and *pon* genes (Figure 6.2 and Figure 6.3). All cefixime resistant (>0.125 mg/L) specimens contained three mutations found in mosaic *penA* alleles (I312M, V316T and N512Y).[89,137,271] For other specimens, there were a range of ceftriaxone and cefixime MICs across all mutation combinations. There were also specimens with elevated MICs (≥0.015 mg/L) that had none of the mutations investigated in this study.

6.3.2.2 Azithromycin, ciprofloxacin and penicillin

A range of MICs for different combinations of genetic markers was seen for the other antimicrobials. Two specimens with high-level azithromycin resistance (≥256 mg/L) contained a mutation in the 23S rDNA (A2045G) (Figure 6.4). These were the only specimens with this mutation. Most specimens phenotypically resistant to ciprofloxacin had a known genetic marker of reduced susceptibility (90%; 416/460) (Figure 6.5). However, there were five specimens phenotypically sensitive to ciprofloxacin that contained genetic mutations associated with resistance (*gyrA* S91F and/or *gyrA* D95G). One third (38%; 80/211) of specimens that were phenotypically penicillin resistant (>1 mg/L) did not have any of the genetic markers of resistance investigated (Figure 6.6).

Table 6.3 Genetic markers of decreased susceptibility or resistance identified in the PhD sample

Mutated gene and resulting amino acid change (wild type, amino acid position, variant)	**Number of specimens with mutation**	**% of total (N=1,277)**
***mtrR* A39T**	377	29.5
***mtrR* G45D**	134	10.5
***mtrR* promoter 195C**	3	0.2
***penA* I312M**	89	7.0
***penA* N512Y**	107	8.4
***penA* V316T**	89	7.0
***penA* A502T**	94	7.4
***penA* A502V**	30	2.3
***penA* G542S**	1	0.1
***penA* G543S**	32	2.5
***penA* G545S**	88	6.9
***penA* P551L**	1	0.1
***penA* P552L**	168	13.2
***penA* P552S**	30	2.3
***ponA* L421P**	433	33.9
***porB1b* A121D**	245	19.2
***porB1b* A121N**	69	5.4
***porB1b* D121N**	7	0.5
***porB1b* G120K**	327	25.6
23S 2045A/G	2	0.2
23S 2045G	2	0.2
23S 2597C/T	6	0.5
23S 2597T	3	0.2
***gyrA* D95G**	202	15.8
***gyrA* D95N**	7	0.5
***gyrA* S91F**	422	33.0
***norM* promoter 231T**	1	0.1
***parC* D86N**	88	6.9
***parC* E91K**	2	0.2
***parC* S87I**	10	0.8
***parC* S87R**	190	14.9
***parC* S88P**	2	0.2

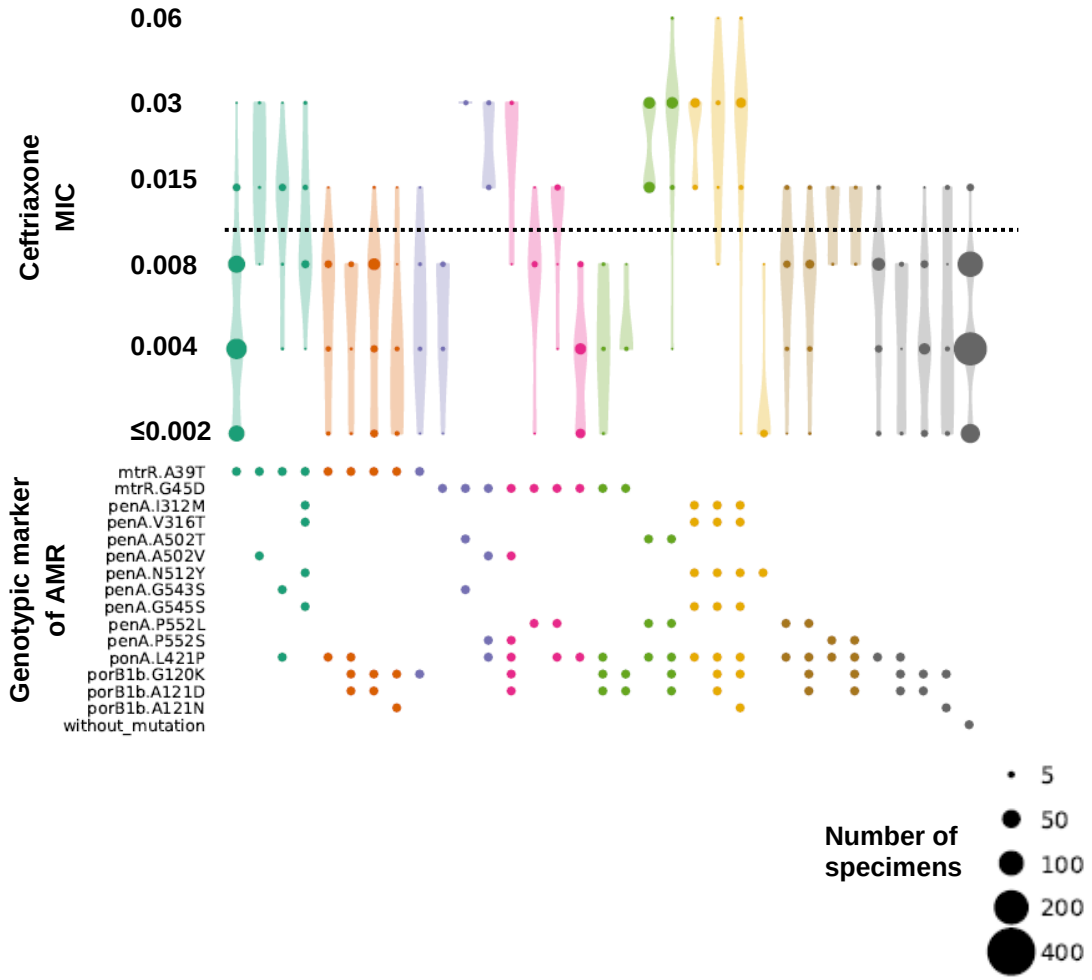


Figure 6.2 Genetic markers of antimicrobial resistance and MIC for ceftriaxone
Only marker combinations with ≥5 specimens included, dotted line indicates reduced susceptibility threshold (≥0.015 mg/L), the colours are a visual aide to distinguish every four groups of genotypic markers and are not representative of any genotypic types

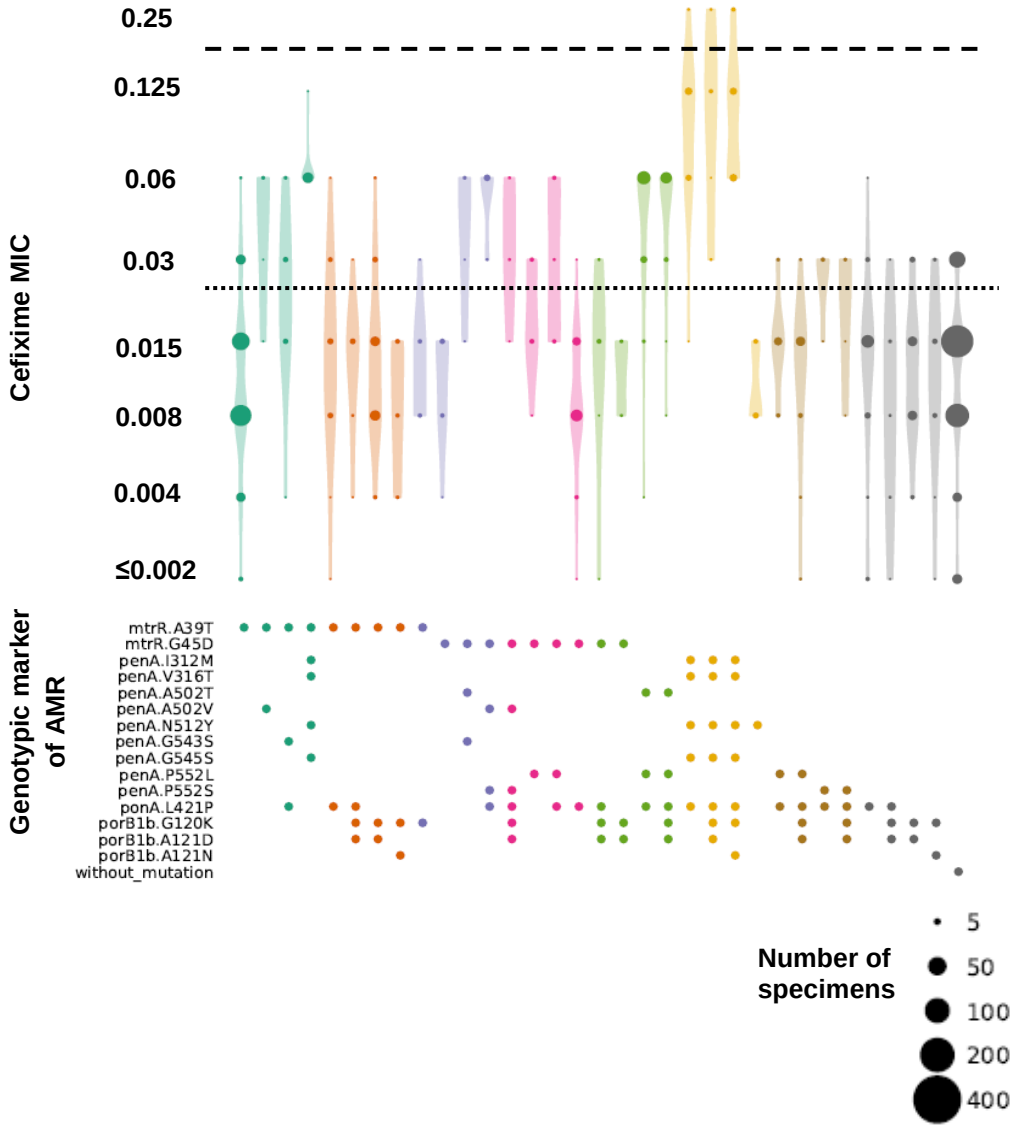


Figure 6.3 Genetic markers of antimicrobial resistance and MIC for cefixime
Only marker combinations with ≥5 specimens included, dotted line indicates reduced susceptibility threshold (≥0.015 mg/L), dashed line indicates resistance threshold (>0.125 mg/L), the colours are a visual aide to distinguish every four groups of genotypic markers and are not representative of any genotypic types

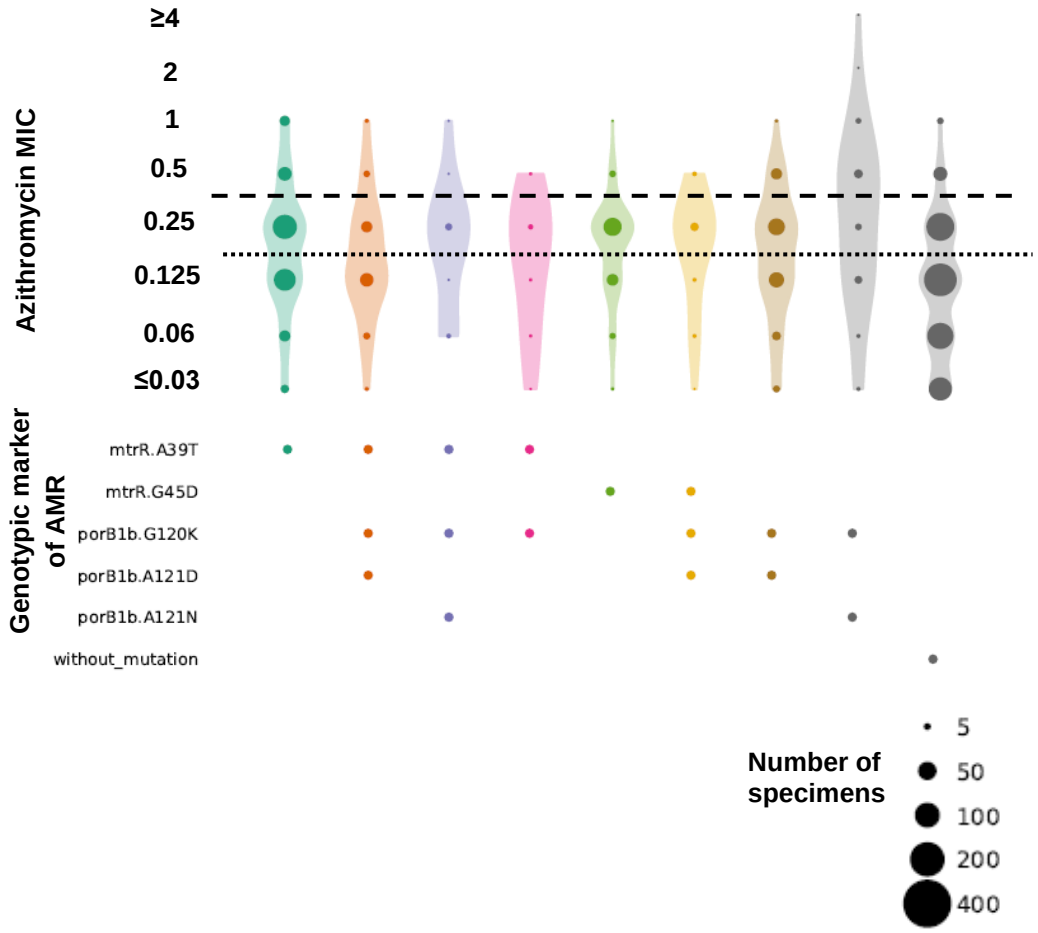


Figure 6.4 Genetic markers of antimicrobial resistance and MIC for azithromycin

Only marker combinations with ≥5 specimens included, dotted line indicates reduced susceptibility threshold (≥0.25 mg/L), dashed line indicates resistance threshold (>0.5 mg/L), the colours are a visual aide to distinguish every group of genotypic markers

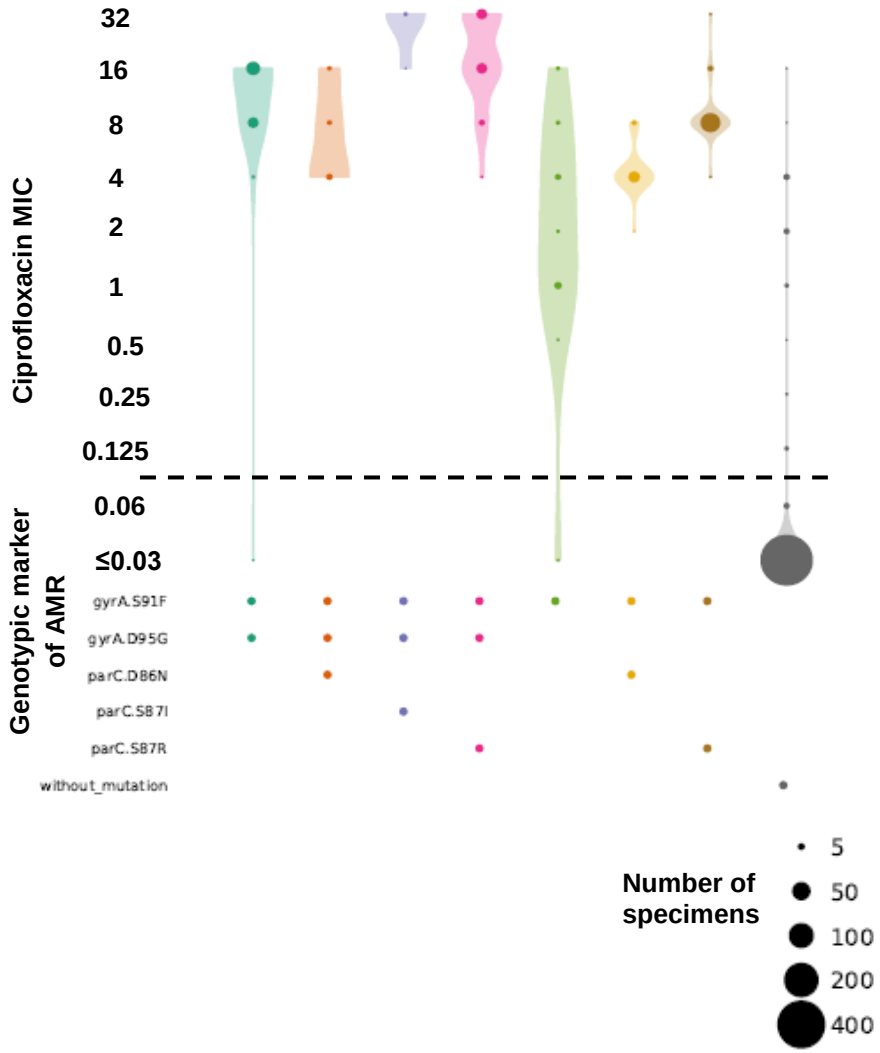


Figure 6.5 Genetic markers of antimicrobial resistance and MIC for ciprofloxacin

Only marker combinations with ≥5 specimens included, dashed line indicates resistance threshold (>0.06 mg/L), the colours are a visual aide to distinguish every group of genotypic markers, for specimens from 2016, only data on whether the specimen was susceptible or resistant to ciprofloxacin were available, so for MIC analyses all resistant specimens were allocated an MIC of 1.0 mg/L and all sensitive specimens were allocated an MIC of ≤0.03 mg/L.

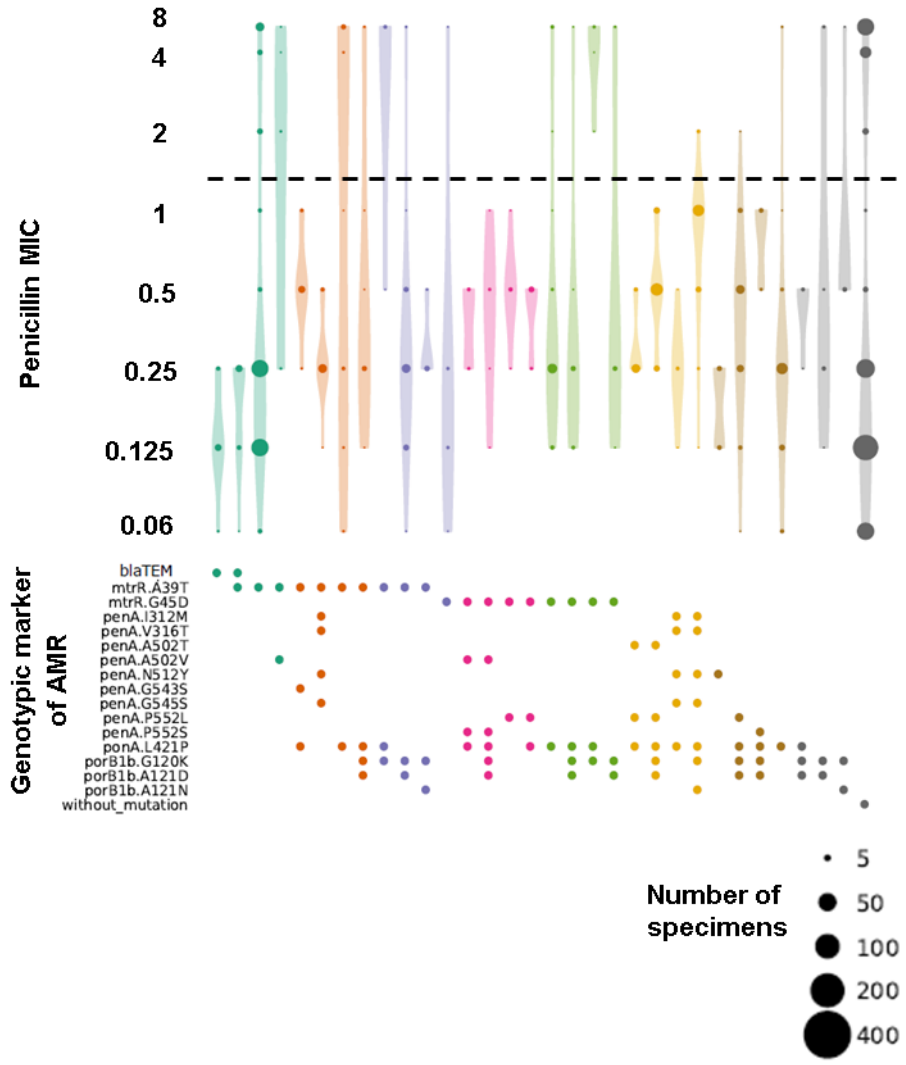


Figure 6.6 Genetic markers of antimicrobial resistance and MIC for penicillin
Only marker combinations with ≥5 specimens included, dashed line indicates resistance threshold (>1.0 mg/L), the colours are a visual aide to distinguish every four groups of genotypic markers

6.3.3 Distribution of *penA* alleles across the phylogeny

A range of *penA* alleles were found in the PhD study sample (Table 6.4). The largest type was the non-mosaic *penAII* allele (36%; 465/1,277). Three types of mosaic *penA* alleles were found, the largest being *penAXXXIV* (7%; 86/1,277). The other two types were found in a small number of specimens: three were *penAX* and one was *penAXXXVIIII.* There were multiple occurrences of specimens with the same *penA* alleles in distinct parts of the phylogenetic tree (Figure 6.7). However, most specimens with the same *penA* allele were found in one of the two main lineages. The larger, distinct clusters with elevated ceftriaxone and cefixime MICs contained the mosaic *penAXXXIV* allele and the non-mosaic *penA*.44.001 allele and are described in more detail using the epidemiological data in the following sections.

Table 6.4 *penA* allele types found in the PhD study sample

***pen*A allele type**	**N**	**% (N=1,277)**	**Mosaic**	**Lineage A (n)**	**Lineage B (n)**
1	164	12.8	No	1	163
2	465	36.4	No	454	11
4	1	0.1	No	1	0
5	24	1.9	No	24	0
9	83	6.5	No	83	0
10	3	0.2	Yes	3	0
12	13	1.0	No	13	0
13	17	1.3	No	17	0
14	130	10.2	No	13	117
15	10	0.8	No	0	10
18	9	0.7	No	9	0
19	28	2.2	No	27	1
21	4	0.3	No	4	0
22	116	9.1	No	0	116
34	**86**	**6.7**	**Yes**	**86**	**0**
39	1	0.1	Yes	1	0
41	7	0.5	No	7	0
43	9	0.7	No	9	0
44	**84**	**6.6**	**No**	**84**	**0**
45	18	1.4	No	0	18
49	1	0.1	No	1	0
52	1	0.1	No	1	0
Wild Type	3	0.2	No	0	3

6.3.3.1.1 Description of specimens with penA alleles associated with reduced susceptibility to cefixime and ceftriaxone

There were five places on the tree where the mosaic *penAXXXIV* allele was found (Figure 6.7). The two largest groups (N=26 and N=57) were genetically distinct from each other indicating separate mutation events. There were multiple NG-MAST sequence types within these two groups. The smaller group contained four different

NG-MAST types, most (81%; 21/26) were NG-MAST 4244. There was also one specimen with a novel NG-MAST type. The larger of the two mosaic *penAXXXIV* allele groups contained 22 NG-MAST types; the largest was NG-MAST 1407 (25%; 14/57) and four specimens had a novel NG-MAST type.

Specimens in the two groups differed significantly by year, clinic, sexual orientation and HIV status (Table 6.5) with those in the smaller group predominantly from MMC in 2014 and 2015 (65%; 17/26), and those in the larger group distributed across all clinics and years (albeit progressively fewer each year). The smaller group mainly contained specimens from MSM (96%; 25/26) whereas the larger group contained specimens from women (21%; 13/57), heterosexual men (37%; 21/57) and MSM (42%; 24/57). The smaller group had a higher percentage of people with HIV (34%; 9/26 vs 7%; 4/57).

The other large group of specimens with reduced susceptibility to ceftriaxone and cefixime consisted of 84 specimens with the non-mosaic *penA*44.001 allele (Figure 6.7). The *penA*44.001 allele contained mutations A502T and P552L, which cause the same amino acid changes linked to ceftriaxone and cefixime reduced susceptibility in previous studies.[89] Most of the specimens in this *penA*44.001 group were from MSM (82%; 69/84). The *penA*44.001 allele persisted over all four years of the study and was found in all five clinics (Table 6.6). The characteristics of specimens with the *penA*44.001 allele were more similar to the characteristics of specimens in the smallest *penAXXXIV* group compared with the larger *penAXXXIV* group.

Key: No mutation Mutation

MIC: lighter colour = higher MIC

★ = groups of specimens with reduced susceptibility to ceftriaxone and cefixime: two groups contain the *penAXXXIV* allele, one group contains the *pen*A44.001 allele

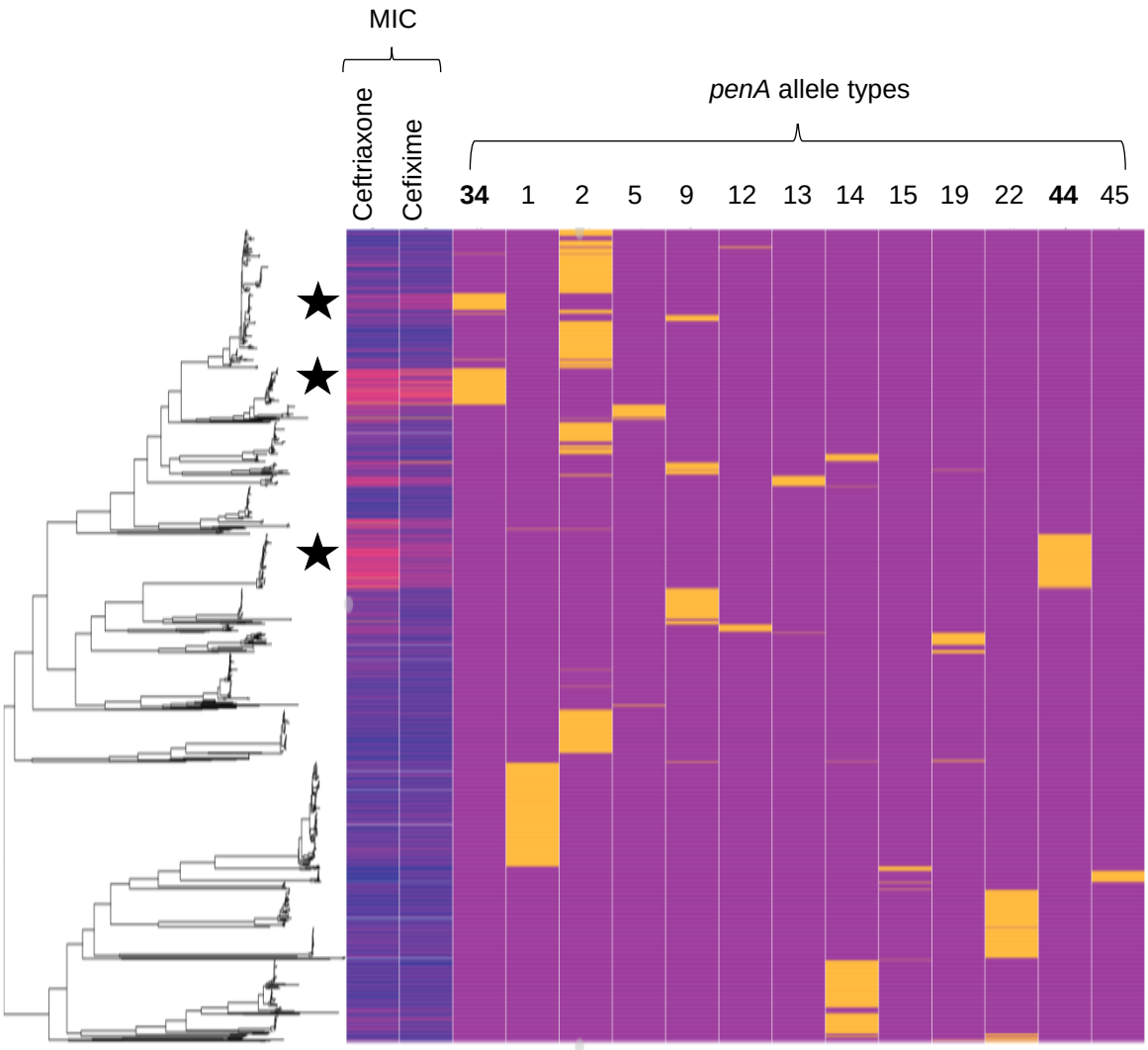


Figure 6.7 Maximum likelihood phylogenetic tree with recombination removed and ceftriaxone MIC, cefixime MIC and *penA* alleles highlighted
Only allele types with ≥10 samples have been included for ease of visualisation

Table 6.5 Epidemiological characteristics of specimens in the two largest *penAXXXIV* clusters

	Cluster N=57		Cluster N=26		Chi² test P value*
	n	%	n	%	
Year					
2013	26	45.6	0	0.0	**<0.001 (FE)**
2014	14	24.6	15	57.7	
2015	10	17.5	10	38.5	
2016	7	12.3	1	3.8	
Gender & sexual orientation					
MSM	24	42.1	25	96.2	**<0.001 (FE)**
Heterosexual men	21	36.8	0	0.0	
Women	12	21.1	1	3.8	
Clinic					
Bristol	18	31.6	0	0.0	**<0.001 (FE)**
Liverpool	11	19.3	1	3.8	
West London	7	12.3	4	15.4	
Birmingham	9	15.8	4	15.4	
MMC	12	21.1	17	65.4	
Age group (years)					
≤24	23	40.4	5	19.2	0.081
25-34	20	35.1	9	34.6	
≥35	14	24.6	12	46.2	
Ethnicity					
White	37	64.9	22	84.6	0.408 (FE)
Black Caribbean	6	10.5	0	0.0	
Black Other	2	3.5	0	0.0	
Asian	4	7.0	1	3.8	
Other	2	3.5	2	7.7	
Mixed	3	5.3	1	3.8	
Missing	3	5.3	0	0.0	
Country of birth					
UK	30	52.6	10	38.5	0.465 (FE)
Not UK	26	45.6	14	53.8	
Missing	1	1.7	2	7.7	
Symptomatic Infection					
No	13	22.8	15	57.7	**0.004**
Yes	40	70.2	11	42.3	
Missing	3	5.3	0	0.0	
Diagnosed with a new STI (excluding HIV) in the past year					
No/Unknown	51	89.5	17	65.4	**0.013 (FE)**
Yes	6	10.5	9	34.6	
HIV Status					
Negative/Unknown	53	93.0	17	65.4	**0.003 (FE)**
Positive	4	7.0	9	34.6	
Number of sexual partners in the UK 3 months prior to diagnosis					
0	6	10.5	0	0.0	0.675 (FE)
1	16	28.1	3	11.5	
≥2	23	40.3	6	23.1	
Missing	12	21.0	17	65.4	
Travel-associated sexual partnership					
No	34	59.6	9	34.6	0.178 (FE)
Yes	11	19.3	0	0.0	
Missing	12	21.0	17	65.4	

* Not including missing, FE = Fisher's exact test was used instead of Chi²

Table 6.6 Epidemiological characteristics of cases in *penA*44.001 group compared to two *penAXXXIV* groups

	***penA*44.001**		***penAXXXIV* comparison P value***	
	Total (N=84)	**%**	**Small group (N=26)**	**Large group (N=57)**
Year				
2013	39	46.4	**<0.001 (FE)**	0.952
2014	23	27.4		
2015	12	14.3		
2016	10	11.9		
Gender & sexual orientation				
MSM	69	82.1	0.116 (FE)	**<0.001**
Heterosexual men	11	13.1		
Women	4	4.8		
Clinic				
Bristol	10	11.9	0.079 (FE)	**0.023**
Liverpool	11	13.1		
West London	15	17.9		
Birmingham	17	20.2		
MMC	31	36.9		
Age group (years)				
≤24	16	19.0	0.508	**0.021**
25-34	39	46.4		
≥35	29	34.5		
Ethnicity				
White	60	71.4	0.836 (FE)	0.867 (FE)
Black Caribbean	5	5.9		
Black African	1	1.2		
Black Other	1	1.2		
Asian	6	7.1		
Other	4	4.8		
Mixed	6	7.1		
Missing	1	1.2		
Country of birth				
UK	56	66.7	**0. 030 (FE)**	0.107 (FE)
Not UK	26	31.0		
Missing	2	2.4		
Symptomatic infection				
No	19	22.6	**0.003**	0.917
Yes	56	66.6		
Missing	9	10.7		
Diagnosed with a new STI (excluding HIV) in the past year				
No/Unknown	60	71.4	0.557	**0.01**
Yes	24	28.6		
HIV Status				
Negative/Unknown	61	72.6	0.478	**0.003**
Positive	23	27.4		
Number of sexual partners in the UK 3 months prior to diagnosis				
0	1	1.2	0.915	0.082
1	18	21.4		
≥2	34	40.5		
Missing	31	36.9		
Travel-associated sexual partnership				
No	51	96.2	1.00 (FE)	**0.003**
Yes	2	3.8		
Missing	31	36.9		

* Not including missing, FE = Fisher's exact test was used instead of Chi^2

6.3.4 Phenotypic and genotypic characteristics of specimens in clusters

In each cluster defined by a five SNP threshold (N=213, Chapter 5), specimens had the same genetic markers of resistance as every other specimen in the same cluster, except six clusters that contained specimens with different *porB1b* genes and one cluster that contained specimens with different *mtrR* genes. The three largest clusters defined by the five SNP threshold had the following non-mosaic *penA* alleles: *penAXIV* (N=21)*, penAVIIII* (N=11) and *penAXXII* (N=9)*.* Less than 10% of all clusters defined by a five SNP threshold contained a mix of specimens with more than two MIC dilutions between them (Table 6.7).

Table 6.7 Clusters (defined by five SNP threshold) with MIC distributions spanning >2 MIC dilutions

Antimicrobial	Number of clusters with specimen MIC distribution >2 dilutions	% of all clusters (N=213)
Azithromycin	9	4.2
Ceftriaxone	8	3.8
Cefixime	12	5.6
Penicillin	16	7.5

6.3.4.1 Association between the phenotypic and epidemiological characteristics of clusters

6.3.4.1.1 Ceftriaxone

Clusters comprising HIV-negative/unknown people were more likely to have reduced susceptibility to ceftriaxone compared to clusters with specimens from clusters with both HIV-positive and HIV-negative/unknown people (base: mixed HIV status clusters; clusters with only HIV-negative/unknown people crude odds ratio (cOR) 2.6, 95% CI 1.11-5.99, P=0.022) (Table 6.8). There were no other cluster types significantly associated with reduced susceptibility to ceftriaxone.

6.3.4.1.2 Cefixime

Clusters comprising only people aged 25-34 years were more likely to have reduced susceptibility to cefixime compared to clusters with specimens from people in other age groups (base: mixed age clusters; clusters comprising people aged 25-34 years only cOR 3.08, 95%CI 1.3-7.4, P=0.008) (Table 6.9).

6.3.4.1.3 Azithromycin

Clusters that comprised women and heterosexual men only and clusters that only contained specimens from women were less likely to have reduced susceptibility to

azithromycin compared to clusters with specimens from MSM only (base: MSM only cluster; women and heterosexual men cluster cOR: 0.27, 95% CI 0.13-0.58, P<0.001; women only cluster cOR 0.06, 95% CI 0.006-0.56, P<0.001) (Table 6.10).

6.3.4.1.4 Penicillin

Clusters from Birmingham were less likely to have reduced susceptibility to penicillin compared to clusters with specimens from multiple clinics (cOR: 0.38, 95% CI 0.14-0.97), P=0.033, Table 6.11).

Table 6.8 Epidemiological characteristics of clusters (defined by the five SNP threshold) with a modal ceftriaxone MIC ≥0.015 mg/L compared to clusters with a modal ceftriaxone MIC <0.015 mg/L
cOR = crude odds ratio, LCI = lower 95% confidence interval, UCI = upper 95% confidence interval

Cluster Type	Cluster total (N)	Cluster modal MIC ≥0.015 mg/L (n)	%	cOR	LCI	UCI	P Value
Clinic							
Multiple clinics	78	20	25.6	1.00	-	-	-
Birmingham	58	10	17.2	0.60	0.26	1.42	0.244
Bristol	20	10	50.0	1.33	0.44	3.98	0.607
Liverpool	17	5	29.4	1.29	0.40	4.18	0.666
Mortimer Market	34	8	23.5	0.80	0.30	2.15	0.665
West London	6	2	33.3	0.62	0.07	5.74	0.672
Gender & sexual orientation							
MSM	108	24	22.2	1.00	-	-	-
Women & Het. Men	49	10	20.4	0.90	0.39	2.06	0.799
Het. Men & MSM	30	7	26.7	1.06	0.40	2.79	0.898
Women	9	1	11.1	0.44	0.05	3.73	0.437
Het. Men	9	3	33.3	1.75	0.40	7.60	0.449
Women, Het. Men & MSM	5	1	20.0	0.87	0.09	8.28	0.907
Women & MSM	3	1	66.7	1.75	0.15	20.4	0.651
Age (years)							
Mixed	26	4	26.1	1.00	-	-	-
≤24	33	7	21.2	0.65	0.23	1.83	0.408
25-34	27	8	29.6	1.52	0.60	3.86	0.370
≥35	19	5	26.3	1.29	0.43	3.90	0.648
Ethnicity							
White	80	25	31.3	1.00	-	-	-
Black Caribbean	7	1	14.3	0.50	0.06	4.49	0.528
Asian	4	2	50.0	3.00	0.39	23.3	0.270
Mixed	120	27	22.5	0.75	0.38	1.48	0.404
HIV status							
Mixed	64	12	18.8	1.00	-	-	-
Only HIV negative/unknown	141	42	29.8	**2.58**	**1.11**	**5.99**	**0.022**
Only HIV positive	8	1	12.5	1.00	0.11	9.37	1.00

Table 6.9 Epidemiological characteristics of clusters (defined by the five SNP threshold) with a modal cefixime MIC ≥0.03 mg/L compared to clusters with a modal cefixime MIC <0.03 mg/L
(A) Univariate analysis

cOR = crude odds ratio, LCI = lower 95% confidence interval, UCI = upper 95% confidence interval

Cluster Type	Cluster total (N)	Cluster modal MIC ≥0.03 mg/L (n)	%	cOR	LCI	UCI	P value
Clinic							
Multiple	78	28	35.9	1.00	-	-	-
Birmingham	58	18	31.0	0.80	0.39	1.66	0.555
Bristol	20	8	40.0	1.19	0.43	3.28	0.736
Liverpool	17	5	29.4	0.74	0.24	2.35	0.613
Mortimer Market	34	12	35.3	0.97	0.42	2.27	0.951
West London	6	3	50.0	1.79	0.33	9.59	0.493
Gender & sexual orientation							
MSM only	108	38	35.2	1.00	-	-	-
Women & Het. Men	49	16	32.6	0.89	0.43	1.83	0.758
Het. Men & MSM	30	10	33.3	0.92	0.39	2.17	0.851
Women	9	3	33.3	0.92	0.22	3.92	0.911
Het. Men	9	3	33.3	0.92	0.22	3.92	0.911
Women, Het. M & MSM	5	1	20.0	0.46	0.05	4.33	0.487
Women & MSM	3	3	100.0	-			
Age (years)							
Mixed	134	43	32.1	1.00	-	-	-
≤24	33	8	24.2	0.68	0.28	1.63	0.382
25-34	27	16	59.3	**3.08**	**1.29**	**7.36**	**0.008**
≥35	19	7	36.8	1.23	0.45	3.37	0.680
Ethnicity							
White	80	28	35.0	1.00	-	-	-
Black Caribbean	7	3	42.9	1.39	0.29	6.74	0.679
Asian	4	2	50.0	1.86	0.24	14.1	0.544
Mixed	120	41	34.2	0.96	0.53	1.75	0.904
HIV status							
Mixed	64	17	26.6	1.00	-	-	-
Only HIV negative/unknown	141	56	39.7	1.82	0.95	3.51	0.069
Only HIV positive	8	1	12.5	0.39	0.04	3.54	0.390

Table 6.10 Epidemiological characteristics of clusters (defined by the five SNP threshold) with a modal azithromycin MIC ≥0.25 mg/L compared to clusters with a modal azithromycin MIC <0.25 mg/L

cOR = crude odds ratio, LCI = lower 95% confidence interval, UCI = upper 95% confidence interval

Note: when clinic and sexual orientation were included in the same model, only the association with sexual orientation remained and the cOR value was the same

Cluster Type	Clusters total (N)	Cluster modal MIC ≥0.25 mg/L (n)	%	cOR	LCI	UCI	P value
Clinic							
Multiple clinics	78	53	67.9	1.00	-	-	-
Birmingham	58	22	37.9	**0.29**	**0.14**	**0.61**	**<0.001**
Bristol	20	13	65.0	0.88	0.31	2.48	0.803
Liverpool	17	8	47.1	0.42	0.14	1.24	0.105
Mortimer Market	34	18	52.9	0.53	0.23	1.23	0.131
West London	6	5	83.3	2.36	0.26	21.7	0.435
Gender & sexual orientation							
MSM only	108	73	67.6	1.00	-	-	-
Women & Het. Men	49	18	36.7	**0.27**	**0.13**	**0.58**	**<0.001**
Het. Men & MSM	30	17	56.7	0.63	0.27	1.44	0.268
Women only	9	5	55.6	**0.06**	**0.006**	**0.56**	**<0.001**
Het. Men only	9	1	11.1	0.60	0.15	2.39	0.464
Women, Het. Men & MSM	5	3	60.0	0.72	0.11	4.54	0.725
Women & MSM	3	2	66.7	0.96	0.08	11.1	0.973
Age (years)							
Mixed	26	15	57.7	1.00	-	-	-
≤24	33	16	48.5	0.68	0.31	1.46	0.315
25-34	27	18	66.7	1.43	0.60	3.44	0.415
≥35	19	7	36.8	0.42	0.15	1.15	0.080
Ethnicity							
White	80	48	60.0	1.00	-	-	-
Black Caribbean	7	2	28.6	0.27	0.05	1.51	0.109
Asian	4	2	50.0	0.67	0.09	5.05	0.693
Mixed	120	66	55.0	0.81	0.46	1.45	0.485
HIV status							
Mixed	64	36	56.3	1.00	-	-	-
Only HIV negative/unknown	141	77	54.6	0.94	0.52	1.70	0.827
Only HIV positive	8	6	75.0	2.33	0.43	12.8	0.314

Table 6.11 Epidemiological characteristics of clusters (defined by the five SNP threshold) with a modal penicillin MIC ≥1.0 mg/L compared to clusters with a modal penicillin MIC <1.0 mg/L
cOR = crude odds ratio, LCI = lower 95% confidence interval, UCI = upper 95% confidence interval

Cluster Type	Clusters total (N)	Cluster modal MIC ≥1.0 mg/L (n)	%	cOR	LCI	UCI	P Value
Clinic							
Multiple clinics	78	23	29.5	1.00	-	-	-
Birmingham	58	8	13.8	**0.38**	**0.14**	**0.97**	**0.033**
Bristol	20	6	30.0	0.68	0.20	2.28	0.528
Liverpool	17	6	35.3	0.83	0.24	2.87	0.775
Mortimer Market	34	9	26.5	0.70	0.26	1.87	0.479
West London	6	3	50.0	1.36	0.23	8.06	0.736
Gender & sexual orientation							
MSM	108	25	23.1	1.00	-	-	-
Women & Het. Men	49	8	16.3	0.65	0.27	1.57	0.333
Het. Men & MSM	30	4	13.3	0.51	0.16	1.62	0.245
Women only	9	1	2.2	0.41	0.05	3.53	0.406
Het Men. only	9	3	6.7	1.66	0.38	7.19	0.493
Women, Het. Men & MSM	5	3	60.0	4.98	0.76	32.7	0.063
Women & MSM	3	1	33.3	1.66	0.14	19.3	0.620
Age (years)							
Mixed	134	36	26.9	1.00	-	-	-
≤24	33	5	15.2	0.33	0.09	1.18	0.074
25-34	27	8	29.6	0.95	0.35	2.57	0.918
≥35	19	6	31.6	1.19	0.39	3.57	0.760
Ethnicity							
White	80	20	25.0	1.00	-	-	-
Black Caribbean	7	2	28.6	1.73	0.30	9.95	0.532
Asian	4	3	75.0	4.33	0.54	34.7	0.131
Mixed	120	30	25.0	1.20	0.59	2.44	0.618
HIV status							
Mixed	64	19	29.7	1.00	-	-	-
Only HIV negative/unknown	141	33	23.4	0.72	0.37	1.41	0.339
Only HIV positive	8	3	37.5	1.42	0.30	6.64	0.653

6.4 Discussion and conclusions

In this chapter I analysed WGS, epidemiological and phenotypic data to investigate the distribution of *N. gonorrhoeae* with reduced susceptibility to antimicrobials and the genotypic markers associated with these phenotypes. I found that *N. gonorrhoeae* with reduced susceptibility has emerged separately in different sexual networks in England, as was evidenced by three distinct clusters with different mutations of the *penA* allele, different epidemiological characteristics and reduced susceptibility to cephalosporins (ceftriaxone and cefixime). Therefore, it is difficult to predict in which geographical area or population sub-group AMR will emerge and

public health resources to measure, prevent and control antimicrobial resistance should cover the entire population affected by gonorrhoea, rather than focusing on specific sub-groups or locations. These data also suggest that, in England, *N. gonorrhoeae* with reduced susceptibility has emerged through novel mutation events and/or repeated introduction from external sources, as well as by clonal expansion.

6.4.1 Genomic variation of *N. gonorrhoeae* in England: two distinct lineages with differing antimicrobial susceptibility profiles

This is the largest and most representative study on genomic variation in antimicrobial susceptibility in *N. gonorrhoeae* in England to date. I observed two distinct lineages with different antimicrobial susceptibility profiles. The larger lineage contained specimens with elevated MICs to all five antimicrobials, consistent with findings from recent global and European studies.[142,168] The authors of these studies hypothesised that differing susceptibility profiles of the two lineages were associated with different sexual orientation networks, although both studies were limited by a lack of complete data on sexual orientation to support this. My study had sexual orientation data for 99% of cases and my findings strongly support this hypothesis: The lineage with reduced susceptibility to antimicrobials was more likely to be associated with MSM and the more susceptible lineage was more likely to be associated with heterosexuals. One explanation put forward by Sanchez-Buso *et al*. is that MSM experience more diagnosed bacterial STIs and greater exposure to antimicrobials, thereby increasing selection pressures for antimicrobial resistance,[99,168,272] and this hypothesis is supported by mathematical models.[273,274] It is also possible that resistant strains might persist in the absence of selective pressure because the biological fitness of the organism is not always affected or, via compensatory mutations that mitigate deleterious effect.[209,275]

6.4.2 Distribution of *penA* alleles in different sexual networks

Within the less susceptible lineage there were distinct groups of specimens with different genetic markers of resistance. The largest groups contained specimens with reduced susceptibility to cephalosporins, and different *penA* alleles: two groups contained the mosaic *penAXXXIV* and one group contained the non-mosaic *penA*.44.001. Both of *penAXXXIV* and *penA*.44.001 alleles contain mutations that were found to be associated with elevated MICs in previous studies.[89,99,271,276-279] Specimens from the two *penAXXXIV* allele groups were genetically distinct from each other and the epidemiological characteristics of these two groups provided

further evidence that the strains were circulating in different sexual networks (the smaller *penAXXXIV* group consisted of specimens predominantly from MSM in the MMC clinic, whereas specimens in the larger *penAXXXIV* group consisted of specimens from both MSM and heterosexuals from all five study clinics). The larger group contained specimens with the NG-MAST type 1407, which is a strain that has been circulating globally and in England for over a decade.[280] Therefore, the larger *penAXXXIV* group may represent clonal spread of a previously identified endemic strain of *N. gonorrhoeae,* whereas, the smaller *penAXXXIV* group represents a new strain emerging in a different sexual network.

Identifying that the dissemination of *N. gonorrhoeae* with reduced susceptibility to antimicrobials is associated with the emergence of new strains as well as the spread of existing clonal strains highlights a need for further research into how and why new strains with this phenotype emerge. Investigating the biological properties of new strains may uncover novel information about the mechanisms of antimicrobial resistance that could be used in the development of new treatments.[281]

Only through analysis of WGS data combined with epidemiological and phenotypic data is it possible to show that reduced susceptibility to cephalosporins has emerged repeatedly in separate sexual networks in England. Analyses using only epidemiological and phenotypic data would have grouped all *penAXXXIV* samples from MSM together and if sequencing had been restricted to the *penA* gene rather than the whole genome, the distinct clusters of *penAXXXIV* would not have been identified.

6.4.3 Using WGS to identify and describe sexual networks containing *N. gonorrhoeae* with reduced susceptibility to antimicrobials

There were few differences in the epidemiological characteristics of the sexual networks where *N. gonorrhoeae* with higher MICs were circulating, demonstrating that *N. gonorrhoeae* with reduced susceptibility are ubiquitous. However, despite this being one of the largest *N. gonorrhoeae* sequencing studies undertaken to date, my analyses were limited by relatively small numbers of specimens from some population sub-groups, such as people of Asian ethnicity, and smaller numbers of specimens with high MICs.

6.4.4 Public health application

6.4.4.1 Surveillance

The key finding from this chapter is that *N. gonorrhoeae* with reduced susceptibility to antimicrobials has emerged separately in different sexual networks across England, as evidenced by the distinct clusters of *N. gonorrhoeae* with reduced susceptibility to antimicrobials and different epidemiological profiles. Therefore, it is difficult to predict in which geographical area or population sub-group AMR will emerge and public health resources to measure, prevent and control antimicrobial resistance should cover the entire population affected by gonorrhoea, rather than focusing on specific sub-groups or locations. This supports continuation of the sampling approach of GRASP in England and Wales, which is geographically diverse and is not restricted to people from specific risk groups.

6.4.4.2 Molecular tests for antimicrobial resistance

The development of rapid molecular tests for genetic markers of resistance that are highly predictive of a resistant phenotype could lead to more effective use of antimicrobials for therapy. Tests that rapidly detect markers of ciprofloxacin and cephalosporin resistance are already in development.[156,158,282] However, as found in this study and elsewhere, the association between the genotype and phenotype is much stronger for ciprofloxacin resistance than for cephalosporins.[89,141,283] Most rapid tests for cephalosporin resistance focus on detecting mutations in the mosaic *penA* allele. However, reduced susceptibility to cephalosporins can also be caused by mutations in the non-mosaic *penA* allele, such as *penA*44.001, as found in my study and elsewhere.[183]

Molecular tests that focus on the presence or absence of one mutation without considering the additive effect of multiple mutations may not be sufficient for detecting resistance and likely treatment failure. Epistasis is known to occur in *N. gonorrhoeae*, meaning that phenotypic resistance is dependent on the complex interactions of multiple mutant genes.[89,99,276] Within the PhD study sample, the antimicrobial susceptibility of specimens with the same genetic markers of resistance varied substantially, and most specimens with these markers were sensitive. However, the presence of one mutation that is part of the complex of mutations required for resistance is indicative of the potential for that strain to become resistant. Therefore, it would be prudent of clinicians to prioritise patients infected with these strains for a test of cure or possibly use alternative antimicrobials

unaffected by the marker of resistance. In any event, mathematical modelling studies have shown that the implementation of molecular tests should only occur if they are highly sensitive; otherwise, they risk accelerating the spread of resistance.[273]

A better understanding of the mechanisms and genomic markers of cephalosporin resistance is needed and will be achieved through a combination of microbiological and genomic studies, including genome wide association studies (GWAS). WGS cannot, however, replace phenotypic testing for antimicrobial susceptibility because it can only detect known mutations associated with resistance; novel mutations associated with resistance are constantly developing in *N. gonorrhoeae*.

6.4.5 Chapter summary

In answering the questions posed at the start of this chapter, I have (i) described and compared the genotypic and phenotypic variation in *N. gonorrhoeae* antimicrobial susceptibility, and (ii) characterised the sexual networks transmitting *N. gonorrhoeae* that has reduced susceptibility to antimicrobials. In summary:

- two distinct lineages of *N. gonorrhoeae* have circulated in England and have different antimicrobial susceptibility profiles; the lineage associated with MSM was generally less susceptible to antimicrobials
- there were genetically distinct strains of *N. gonorrhoeae* with the same antimicrobial susceptibility profiles but different epidemiological profiles, indicating that reduced susceptibility to antimicrobials has probably emerged separately in different sexual networks in England
- epidemiological and phenotypic data alone cannot adequately discriminate between *N. gonorrhoeae* strains circulating in England; routine use of WGS in surveillance could be used to avoid incorrect inferences about the relatedness of strains grouped by epidemiological and/or phenotypic characteristics

7 Epidemiological, phenotypic and genomic insights into travel-associated *N. gonorrhoeae*

7.1 Introduction

Resistance to the current first line therapy, ceftriaxone, has been reported in many countries, most frequently in East and South East Asia probably due to poor antimicrobial stewardship.[99,183,284] In England, recent cases of antimicrobial resistant *N. gonorrhoeae* that resulted in treatment failure occurred in people who acquired infection whilst travelling in this area.[139,164] Ceftriaxone and cefixime resistance have been linked to genotypic mutations in the *penA* gene, particularly the mosaic *penAXXXIV* allele, which has been reported in *N. gonorrhoeae* found in several continents, including North America and Europe.[118,121,135,141,142] Travel-associated sexual partnerships are therefore a potentially important risk factor for acquiring a resistant strain and treatment failure. However, robust epidemiological information on travel-associated sexual exposure is difficult to collect. Studies are also complicated by the potential for exposure through sex with a partner (UK or non-UK resident) who was infected abroad. Phylogenetic analyses that include global collections may provide additional information about whether an infection is likely to have been acquired abroad or domestically, and the extent to which travel might influence the overall *N. gonorrhoeae* pathogen population structure.

In this chapter I aimed to answer the following research questions:

- What is the relationship between travel-associated sexual partnerships and infection with *N. gonorrhoeae* exhibiting reduced susceptibility to antimicrobials in England*?*
- How genetically similar are *N. gonorrhoeae* in England to those circulating globally?
- Which patient groups are more likely to be infected with globally circulating *N. gonorrhoeae*?

To answer these questions, I used available epidemiological and phenotypic data in England to investigate the association between travel-associated sexual partnerships and infection with *N. gonorrhoeae* with reduced susceptibility to each of five antimicrobials. I subsequently combined the PhD study sample data with other

publically available sequence datasets from Europe and the USA. I investigated whether *N. gonorrhoeae* from England with specific epidemiological characteristics, reduced susceptibility to antimicrobials or with mutations in the mosaic *penAXXXIV* allele were associated with *N. gonorrhoeae* found in other countries.

7.2 Methods

7.2.1 Epidemiological and phenotypic characteristics of cases with recent travel-associated sexual partnerships

Univariate and multivariable logistic regression was used as described in Chapter 5 to model the association between recent travel-associated sexual partnerships (outcome) and the following explanatory variables: year, age group, ethnicity, country of birth, STI in the previous year (excluding HIV), HIV status, number of sexual partners in the three months prior to diagnosis, reduced susceptibility or resistance (see Table 6.1) to ceftriaxone, cefixime, azithromycin, ciprofloxacin or penicillin. All models were stratified by sexual orientation (MSM and heterosexuals) because *N. gonorrhoeae* strains with differing antimicrobial susceptibility profiles circulate in these two population sub-groups.[141,280] The dataset included all diagnoses reported in GRASP between 2013 and 2016 with complete information on travel-associated sexual partnerships, gender, sexual orientation and antimicrobial susceptibility data (see Chapter 3).

7.2.2 Creating the phylogenetic trees

Published WGS and associated metadata for *N. gonorrhoeae* specimens from two international studies[141,142] were compared to the PhD study sample. For each international dataset, the consensus sequences were combined with the PhD study consensus sequences and the mobile and repetitive elements removed (Chapter 5). Phylogenetic trees with genetic recombination events removed were created using the programme Gubbins (Version 2.4.0)[124] with the default settings (five iterations and the minimum number of base substitutions required to identify a recombination was three) and the tree building option FastTree (Version 2.1.4). Due to the large number of sequences in the PhD study sample combined with the international dataset, RAxML (the tree building programme used in previous Chapters 5 and 6) was unable to complete the analysis. Therefore, I used FastTree, which also uses a heuristic approach to find the tree with the maximum likelihood of producing the data given the model. FastTree has been shown to produce similar phylogenetic trees as RAxML methods but can analyse larger datasets within one day.[114,285] To compare

and validate the two methods, I created phylogenetic trees with recombination events removed using only the PhD study sample and assessed the relationship between the SNP differences of each pairwise combination of specimens using linear regression.

7.2.2.1 European collection

The European collection contained specimens from the European Gonococcal Antimicrobial resistance Surveillance Programme (Euro-GASP).[108] Euro-GASP is a sentinel surveillance programme for monitoring antimicrobial resistant *N. gonorrhoeae* across Europe. Specimens of *N. gonorrhoeae* from women and men were collected and tested for antimicrobial susceptibility, then linked to epidemiological information. The European collection of sequences came from the study published by Harris *et al*.[142] Overall, 1,054 specimens from 2013 across 20 countries were sequenced, including 106 from England, a subset of which were duplicates of specimens in the PhD study sample (21%, 22/106). All specimens from England in the European collection were excluded leaving 948 European specimens in total. The specimens had been previously mapped to the reference genome FA1090 (the same reference used for the PhD study sample) as reported by Harris *et al*.[142] The available metadata for the European collection used in the analyses reported in this chapter were: reporting country, antimicrobial susceptibility MIC profile for ceftriaxone and cefixime (MICs were grouped to match the categories used in GRASP), and the presence of the mosaic *penAXXXIV* allele.

7.2.2.2 USA collection

The USA collection of specimens was from two previous studies investigating the association between antimicrobial resistance phenotype and genotype.[121,141] The USA collection contained 1,139 specimens collected between 2000 and 2013. For these two studies, the specimens were purposively selected so the samples included matched pairs of resistant and susceptible specimens over a 14-year period. The USA collection of specimens was mapped to the reference genome FA1090 using the method reported in Chapter 3, by Dr Leonor Sánchez-Busó at the Wellcome Sanger Institute for another study.[168] The available metadata for the USA collection used in the analyses reported in this chapter included: antimicrobial susceptibility profile for ceftriaxone, and cefixime (MICs were grouped to match the categories used in GRASP), sexual orientation and the presence of the mosaic *penAXXXIV* allele.

7.2.3 Phylogenetic analysis

The phylogenetic tree created for each combined PhD study and international dataset was visualised using Phandango[244] alongside metadata to describe and compare the distribution of specimens. SNP differences between each pair of specimens in the dataset were extracted from the phylogenetic tree using Seaview (Version 4.7).[245] Clusters that contained at least one specimen from the PhD study sample and one specimen from the international dataset were analysed together using MicrobeTrace (Version 0.1.10)[246] to identify mixed study clusters. Different SNP thresholds were used to define these clusters ranging from 5 to 100 SNPs.

7.2.3.1 Describing the PhD study specimens that clustered with specimens from other countries

The location of the PhD study specimens that had the mosaic *penAXXXIV* allele on the phylogenetic tree were compared to the location of specimens with the same allele from the international datasets to investigate whether they were part of the same lineage. Epidemiological and phenotypic characteristics of the cases in the PhD study sample that clustered with specimens from other countries were compared with those that did not. Univariate and multivariable logistic regression (using the methods reported in Chapter 5) were used to assess whether the specimens that cluster with *N. gonorrhoeae* found in other countries were more likely to have reduced susceptibility or resistance to antimicrobials, and to identify the patient subgroups more likely to be infected with *N. gonorrhoeae* that was genetically similar to *N. gonorrhoeae* found in other countries.

7.3 Results

7.3.1 Association between recent travel-associated sexual partnerships and the epidemiological and phenotypic characteristics of cases

Phenotypic and epidemiological data were available for 4,133 *N. gonorrhoeae* specimens in the GRASP archive between 2013 and 2016. The majority of specimens were from MSM (62%; 2,579/4,133). More heterosexuals than MSM reported travel-associated sexual partnerships in the three months prior to their gonorrhoea diagnosis (MSM: 10%; 264/2,579, heterosexuals: 15%; 198/1,356, P<0.001 (two-proportion Z-test).[235]

MSM who reported a recent travel-associated sexual partnership were more likely to be born outside of the UK (base: MSM born in the UK; MSM born outside the UK

adjusted odds ratio (aOR) 2.02, 95% confidence interval (CI) 1.41-2.89, P value (P) <0.001) and were less likely to have had recent sexual partnerships in the UK (base: no partners in the past three months, one partner aOR 0.02, CI 0.01-0.04, P<0.001); ≥2 partners aOR 0.007, CI 0.003-0.01, P<0.001) compared to MSM who had not travelled (Table 7.1). No association was found between travel-associated sexual partnerships in MSM and specimens with reduced susceptibility to ceftriaxone, azithromycin, cefixime or penicillin.

Heterosexual women and men who reported a recent travel-associated sexual partnership were more likely to be of Asian ethnicity (base: white ethnicity; Asian ethnicity aOR 2.30, CI 1.28-4.13, P 0.005), be born outside of the UK (base: born in the UK; born outside of the UK aOR 4.27, CI 2.91-6.26, P <0.001) or to be infected with a ciprofloxacin resistant infection (base: ciprofloxacin susceptible infection; ciprofloxacin resistant infection aOR 2.31, CI 1.61-3.30, P <0.001) (Table 7.2). They were less likely to be of black Caribbean ethnicity (aOR 0.51, CI 0.28-0.95, P 0.033). In the univariate analysis, heterosexual women and men who reported a recent travel-associated sexual partnership were more likely to be infected with a ceftriaxone reduced susceptible infection (base: ceftriaxone susceptible infection; ceftriaxone reduced susceptible infection cOR 1.65, CI 1.16-2.33, P=0.005), although this association did not remain in the multivariable analysis.

Table 7.1 Epidemiological and phenotypic characteristics of MSM infected with *N. gonorrhoeae* who had a recent travel-associated sexual partnership: all GRASP, 2013-2016
(A)Univariate analysis: cOR=crude odds ratio, L/UCI=lower/upper confidence interval

Explanatory variables		Outcome: recent travel-associated sexual partnership					
	N	n	Row %	cOR	LCI	UCI	P
Year							
2013	486	71	14.6	1.00	-	-	-
2014	887	79	8.9	**0.57**	**0.41**	**0.81**	**0.001**
2015	731	49	6.7	**0.42**	**0.29**	**0.62**	**<0.001**
2016	475	65	13.7	0.93	0.64	1.33	0.681
Age group (years)							
≤24	497	56	11.3	1.00	-	-	-
25-34	1,228	122	9.9	0.87	0.62	1.21	0.410
≥35	854	86	10.1	0.88	0.62	1.26	0.489
Ethnicity							
White	2,061	195	9.5	1.00	-	-	-
Black Caribbean	67	9	13.4	1.48	0.72	3.04	0.277
Black African	52	6	11.5	1.25	0.53	2.96	0.614
Black Other	14	2	14.3	1.59	0.35	7.18	0.540
Asian	105	6	5.7	0.58	0.25	1.34	0.197
Other	87	19	21.8	**2.67**	**1.57**	**4.55**	**<0.001**
Mixed	140	17	12.1	1.32	0.78	2.24	0.298

Table continued on the next page

Explanatory variables	Outcome: recent travel-associated sexual partnership						
	N	n	Row %	cOR	LCI	UCI	P
Country of birth							
UK	1,264	89	7.0	1.00	-	-	-
Not UK	946	135	14.3	**2.20**	**1.65**	**2.92**	**<0.001**
Symptoms							
No	899	68	7.6	1.00	-	-	-
Yes	1,358	172	12.7	**1.77**	**1.32**	**2.38**	**<0.001**
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis							
No	1,557	170	10.9	1.00	-	-	-
Yes	685	60	8.8	0.78	0.57	1.07	0.121
HIV status							
Negative/Unknown	1,972	215	10.9	1.00	-	-	-
Positive	607	49	8.1	**0.72**	**0.52**	**0.99**	**0.044**
Number of sexual partners in the UK in the three months prior to diagnosis							
0	113	96	85.0	1.00	-	-	-
1	607	73	12.0	**0.02**	**0.01**	**0.05**	**<0.001**
≥2	1,859	95	5.1	**0.01**	**0.00**	**0.02**	**<0.001**
Reduced susceptibility to ceftriaxone (MIC ≥0.015 mg/L)							
No	1,920	196	10.2	1.00	-	-	-
Yes	659	68	10.3	1.01	0.76	1.35	0.936
Reduced susceptibility to cefixime (MIC ≥0.03 mg/L)							
No	1,782	182	10.2	1.00	-	-	-
Yes	797	82	10.3	1.01	0.77	1.33	0.954
Resistance to azithromycin (MIC >0.5 mg/L)							
No	2,464	253	10.3	1.00	-	-	-
Yes	115	11	9.6	0.92	0.49	1.74	0.808
Resistance to ciprofloxacin (MIC >0.06 mg/L)							
No	1,571	159	10.1	1.00	-	-	-
Yes	1,008	105	10.4	1.03	0.80	1.34	0.809
Resistance to penicillin (MIC >1.0 mg/L)							
No	1,997	200	10.0	1.00	-	-	-
Yes	582	64	11.0	1.11	0.82	1.50	0.492

(B) Multivariable analysis: aOR = adjusted odds ratio, L/UCI = lower/upper confidence interval

Explanatory variables	Outcome: recent travel-associated sexual partnership			
	aOR	LCI	UCI	P value
Year				
2013	1.00	-	-	-
2014	**0.47**	**0.29**	**0.77**	**0.003**
2015	**0.37**	**0.22**	**0.64**	**<0.001**
2016	0.86	0.52	1.40	0.536
Country of birth				
UK	1.00	-	-	-
Not UK	**2.02**	**1.41**	**2.89**	**<0.001**
Symptoms				
No	1.00	-	-	-
Yes	**1.54**	**1.05**	**2.27**	**0.027**
Number of sexual partners in the UK in the three months prior to diagnosis				
0	1.00	-	-	-
1	**0.02**	**0.01**	**0.04**	**<0.001**
≥2	**0.007**	**0.003**	**0.01**	**<0.001**

Table 7.2 Epidemiological and phenotypic characteristics of heterosexual women and men infected with *N. gonorrhoeae* who had a recent travel-associated sexual partnership: all GRASP, 2013-2016
(A) Univariate analysis: cOR = crude odds ratio, L/UCI = lower/upper confidence interval

Explanatory variables		Outcome: recent travel-associated sexual partnership					
	N	n	Row %	cOR	LCI	UCI	P value
Year							
2013	491	51	10.4	1.00	-	-	-
2014	381	53	13.9	1.39	0.92	2.10	0.112
2015	334	32	9.6	0.91	0.57	1.46	0.706
2016	**348**	**62**	**17.8**	**1.87**	**1.25**	**2.80**	**0.002**
Age group (years)							
≤24	732	65	8.9	1.00	-	-	-
25-34	**549**	**84**	**15.3**	**1.85**	**1.31**	**2.62**	**<0.001**
≥35	**273**	**49**	**17.9**	**2.24**	**1.50**	**3.36**	**<0.001**
Ethnicity							
White	804	98	12.2	1.00	-	-	-
Black Caribbean	**233**	**14**	**6.0**	**0.46**	**0.26**	**0.82**	**0.008**
Black African	**122**	**23**	**18.9**	**1.67**	**1.01**	**2.76**	**0.042**
Black Other	52	6	11.5	0.94	0.39	2.26	0.889
Asian	**76**	**26**	**34.2**	**3.75**	**2.21**	**6.35**	**<0.001**
Other	**40**	**10**	**25.0**	**2.40**	**1.14**	**5.08**	**0.018**
Mixed	151	11	7.3	0.57	0.30	1.08	0.082
Country of birth							
UK	882	63	7.1	1.00	-	-	-
Not UK	**400**	**112**	**28.0**	**5.06**	**3.56**	**7.18**	**<0.001**
Symptoms							
No	315	28	8.9	1.00	-	-	-
Yes	**1,133**	**165**	**14.6**	**1.75**	**1.14**	**2.67**	**0.009**
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis							
No/Unknown	**1,193**	**173**	**14.5**	**1.00**	**-**	**-**	**-**
Yes	**150**	**11**	**7.3**	**0.47**	**0.25**	**0.88**	**0.016**
HIV status							
Negative/Unknown	1,532	194	12.7	1.00	-	-	-
Positive	22	4	18.2	1.53	0.51	4.58	0.441
Number of sexual partners in the UK in the three months prior to diagnosis							
0	120	104	86.7	1.00	-	-	-
1	**708**	**52**	**7.3**	**0.01**	**0.01**	**0.03**	**<0.001**
≥2	**726**	**42**	**5.8**	**0.01**	**0.00**	**0.02**	**<0.001**
Reduced susceptibility to ceftriaxone (MIC ≥0.015 mg/L)							
No	1,267	147	11.6	1.00	-	-	-
Yes	**287**	**51**	**17.8**	**1.65**	**1.16**	**2.33**	**0.005**
Reduced susceptibility to cefixime (MIC ≥0.03 mg/L)							
No	1,155	137	11.9	1.00	-	-	-
Yes	399	61	15.3	1.34	0.97	1.86	0.077
Resistance to azithromycin (MIC >0.5 mg/L)							
No	1,512	191	12.6	1.00	-	-	-
Yes	42	7	16.7	1.38	0.61	3.16	0.439
Resistance to ciprofloxacin (MIC >0.06 mg/L)							
No	1,141	98	8.6	1.00	-	-	-
Yes	**413**	**100**	**24.2**	**3.40**	**2.49**	**4.65**	**<0.001**
Resistance to penicillin (MIC >1.0 mg/L)							
No	1,269	134	10.6	1.00	-	-	-
Yes	**285**	**64**	**22.5**	**2.45**	**1.76**	**3.43**	**<0.001**

(B) Multivariable analysis: aOR = adjusted odds ratio, L/UCI = lower/upper confidence interval

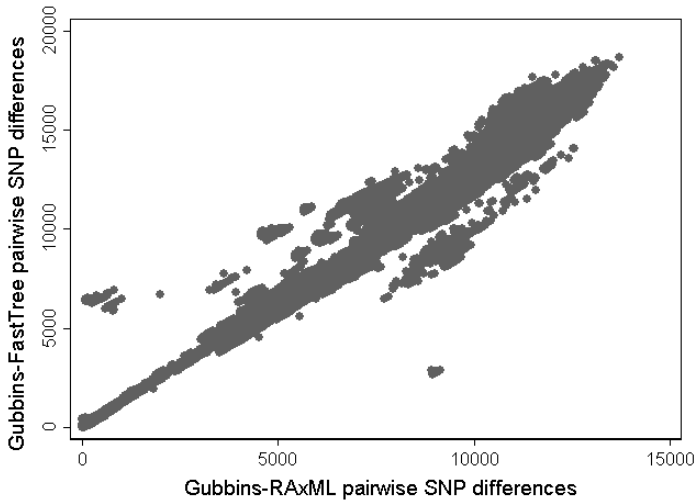
Explanatory variables	Outcome: recent travel-associated sexual partnership			
	aOR	**LCI**	**UCI**	**P value**
Ethnicity				
White	1.00	-	-	-
Black Caribbean	**0.51**	**0.28**	**0.95**	**0.033**
Black African	0.66	0.36	1.18	0.159
Black Other	0.86	0.34	2.18	0.757
Asian	**2.30**	**1.28**	**4.13**	**0.005**
Other	0.98	0.42	2.29	0.965
Mixed	0.57	0.28	1.18	0.130
Country of birth				
UK	1.00	-	-	-
Not UK	**4.27**	**2.91**	**6.26**	**<0.001**
Ciprofloxacin resistant (>0.06 mg/L)				
No	1.00	-	-	-
Yes	**2.31**	**1.61**	**3.30**	**<0.001**

7.3.2 Phylogenetic analyses: comparing FastTree and RAxML methods

To compare and validate the two phylogenetic tree building methods (FastTree and RAxML), I created phylogenetic trees with recombination events removed using only the PhD study sample and assessed the relationship between the SNP differences of each pairwise combination of specimens using linear regression.

The estimated SNP differences between specimen pairs from the PhD calculated using the FastTree and RAxML phylogenetic methods were positively correlated (R^2=0.96) (Figure 7.1). Focusing on the SNP differences that were less than 200, the results were still positively correlated (R^2=0.94). The FastTree SNP differences were estimated to be 20% higher than the RAxML SNP differences (linear regression model equation: FastTree SNP=1.19*RAxML SNP + 4.2). Consequently, a SNP difference of five between pairs using RAxML was estimated to be 10 SNPs using FastTree and 30 SNPs in RAxML was estimated to be 40 SNPs in FastTree.

(A)



(B)

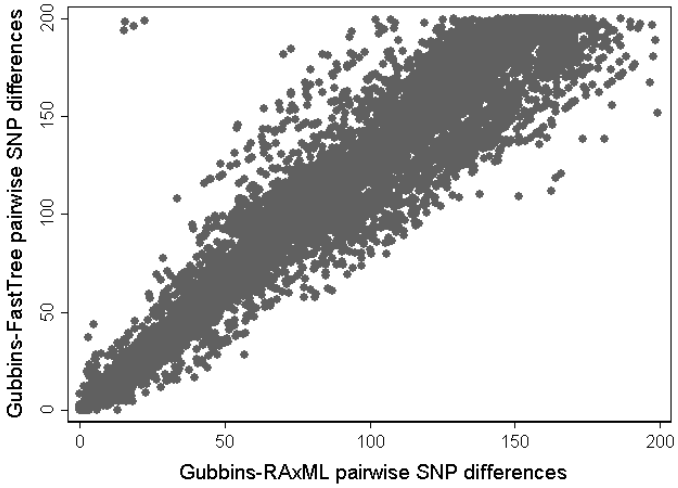


Figure 7.1 Scatterplot of pairwise SNP values between specimens in the PhD study sample using FastTree and RAxML methods
(a) all specimen pairs, (b) specimen pairs with <200 SNP differences

7.3.3 Phylogenetic analyses: comparing PhD study and European specimens

7.3.3.1 Phylogenetic tree

The PhD study and European specimens were fairly evenly dispersed throughout the tree (Figure 7.2). Of the 2,553,832 specimen pair comparisons, the largest difference between two specimens was 18,125 SNPs.

7.3.3.2 Specimens with the mosaic penAXXXIV allele

The larger group of specimens with the mosaic *penAXXXIV* allele identified in the PhD study sample that spanned 2013 to 2016 (N=57, Chapter 6) were genetically similar to the large *penAXXXIV* clade (M2) identified in the European specimens in 2013 (Figure 7.2). The other smaller *penAXXXIV* group in the PhD study (N=26) that were only found post-2013 were not genetically related to specimens in either of the *penAXXXIV* clades from the European dataset.

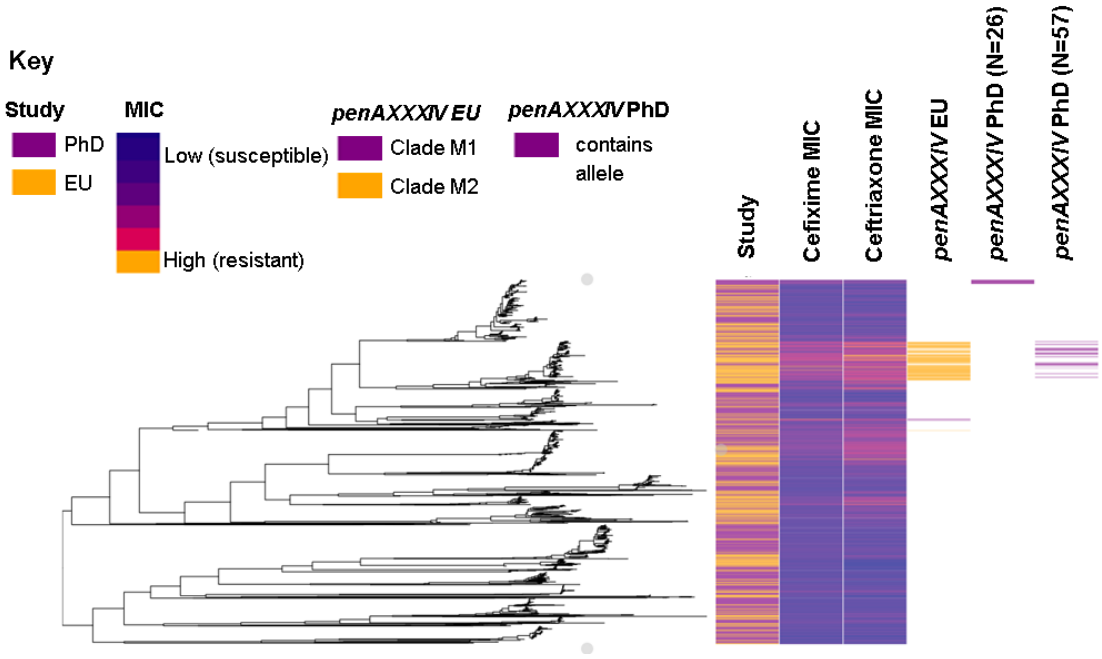


Figure 7.2 Phylogenetic tree of PhD and European specimens with metadata for study type, ceftriaxone and cefixime MIC and presence of *penAXXXIV* allele
Total number of PhD study specimens = 1,277, Total number of European study specimens = 948

7.3.3.3 Clustering by SNP differences

A larger percentage of European specimens clustered with specimens from the PhD study dataset than vice versa, indicating that some specimens from the PhD study sample clustered with more than one European specimen (Table 7.3). Using a five SNP threshold, 1% (16/1,277) of the PhD study specimens clustered with European specimens. Using SNP thresholds of 30 and 100, 17% and 36% of PhD study specimens clustered with European specimens, respectively. At the 100 SNP threshold, 65% of European specimens clustered with PhD study specimens.

Table 7.3 Mixed PhD and European specimen clusters defined by SNP thresholds

SNP Threshold	Number of clusters	PhD study specimens that cluster		European study specimens that cluster	
		n	% (N=1,277)	n	% (N=948)
5	12	16	1.3	13	1.4
30	59	219	17.1	134	14.1
50	75	247	19.3	377	39.8
100	71	466	36.5	616	65.0

7.3.3.4 Epidemiological and phenotypic characteristics of specimens clustered with European specimens

There were no measured epidemiological characteristics of specimens from MSM in the PhD study sample associated with clustering with European specimens (using a 30 SNP threshold) (Table 7.4). However, specimens from MSM in the PhD study sample that clustered with specimens from the European dataset were less likely to have reduced susceptibility to azithromycin (base: susceptible to azithromycin; reduced susceptibility to azithromycin: aOR 0.62, CI 0.42-0.93, P = 0.019) or resistance to penicillin (base: susceptible to penicillin; resistance to penicillin: aOR 0.47, CI 0.25-0.88, P = 0.018).

Similarly, I did not find any epidemiological characteristics associated with clustering with European specimens for women and heterosexual men (using a 30 SNP threshold) (Table 7.5). However, specimens from women and heterosexual men in the PhD study sample that clustered were more likely to have reduced susceptibility to ceftriaxone (base: susceptible to ceftriaxone; reduced susceptibility to ceftriaxone cOR 2.65, 1.46-4.82, P<0.01).

Table 7.4 Epidemiological characteristics and antimicrobial susceptibility profile of PhD study specimens from MSM that clustered with the European study specimens compared to those that did not cluster at the 30 SNP threshold

(A) Univariate analysis: cOR = crude odds ratio, L/UCI = lower/upper confidence interval

	Outcome: Clustered with European specimens					
Explanatory variables	**n**	**Row %**	**cOR**	**LCI**	**UCI**	**P value**
Year						
2013	86	41.5	1.00	-	-	-
2014	**35**	**17.2**	**0.29**	**0.18**	**0.47**	**<0.001**
2015	**25**	**10.8**	**0.17**	**0.10**	**0.29**	**<0.001**
2016	**10**	**8.1**	**0.12**	**0.06**	**0.26**	**<0.001**
Clinic						
Bristol	17	19.3	1.00	-	-	-
Liverpool	12	24.5	1.35	0.58	3.15	0.479
West London	27	19.6	1.02	0.52	2.00	0.964
Birmingham	36	26.1	1.47	0.77	2.84	0.243
MMC	64	18.1	0.92	0.51	1.68	0.797
Age (years)						
≤24	29	21.5	1.00	-	-	-
25-34	83	24.9	1.21	0.75	1.95	0.439
≥35	44	14.8	0.64	0.38	1.07	0.087
Ethnicity						
White	124	21.3	1.00	-	-	-
Black Caribbean	8	26.7	1.35	0.58	3.10	0.483
Black African	1	6.3	0.25	0.03	1.90	0.145
Black Other	1	14.3	0.62	0.07	5.18	0.653
Asian	9	23.1	1.11	0.51	2.40	0.790
Other	5	20.8	0.97	0.36	2.66	0.959
Mixed	6	11.3	0.47	0.20	1.13	0.086
Country of birth						
UK	93	21.9	1.00	-	-	-
Not UK	60	20.9	0.94	0.65	1.36	0.744
Symptoms						
No	42	18.9	1.00	-	-	-
Yes	95	21.4	1.17	0.78	1.76	0.448
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis						
No/Unknown	108	20.8	1.00	-	-	-
Yes	48	19.6	0.93	0.64	1.36	0.706
HIV status						
Negative/unknown	115	21.1	1.00	-	-	-
Positive	41	18.6	0.86	0.58	1.28	0.451
Number of sexual partners in the UK in the three months prior to diagnosis						
0	3	27.3	1.00	-	-	-
1	23	19.3	0.64	0.16	2.62	0.530
≥2	62	23.0	0.79	0.20	3.10	0.740
Reduced susceptibility to ceftriaxone (MIC≥0.015 mg/L)						
No	124	21.7	1.00	-	-	-
Yes	32	16.7	0.72	0.47	1.11	0.11
Reduced susceptibility to cefixime (MIC≥0.03 mg/L)						
No	119	22.1	1.00	-	-	-
Yes	37	16.5	0.70	0.46	1.05	0.083

Table continued on the next page

	Outcome: Clustered with European specimens					
Explanatory variables	**n**	**Row %**	**cOR**	**LCI**	**UCI**	**P value**
Reduced susceptibility to azithromycin (MIC ≥0.125 mg/L)						
No	92	27.3	1.00	-	-	-
Yes	**64**	**15.0**	**0.47**	**0.33**	**0.68**	**<0.001**
Resistance to ciprofloxacin (MIC >0.06mg/L)						
No	97	22.0	1.00	-	-	-
Yes	59	18.3	0.80	0.55	1.14	0.214
Resistance to penicillin (MIC >1.0 mg/L)						
No	143	23.0	1.00	-	-	-
Yes	**13**	**9.2**	**0.34**	**0.19**	**0.62**	**<0.001**

(B) Multivariable analysis: aOR = adjusted odds ratio, L/UCI = lower/upper confidence interval

Explanatory variables	**Outcome: Clustered with European specimens**			
	aOR	**LCI**	**UCI**	**P value**
Year				
2013	1	-	-	-
2014	**0.33**	**0.20**	**0.52**	**<0.001**
2015	**0.23**	**0.14**	**0.39**	**<0.001**
2016	**0.14**	**0.07**	**0.28**	**<0.001**
Reduced susceptibility to azithromycin (MIC ≥0.125 mg/L)				
No	1.00	-	-	-
Yes	**0.62**	**0.42**	**0.93**	**0.019**
Resistance to penicillin (MIC >1.0 mg/L)				
No	1.00	-	-	-
Yes	**0.47**	**0.25**	**0.88**	**0.018**

Table 7.5 Epidemiological characteristics and antimicrobial susceptibility profile of PhD study specimens from women and heterosexual men that clustered with the European study specimens compared to those that did not cluster at the 30 SNP threshold

cOR = crude odds ratio, L/UCI = lower/upper confidence interval

Explanatory variables	**Outcome: clustered with European specimens**					
	n	**Row %**	**cOR**	**LCI**	**UCI**	**P value**
Year						
2013	15	12.6	1.00	-	-	-
2014	19	14.7	1.20	0.58	2.48	0.628
2015	12	8.8	0.67	0.30	1.50	0.328
2016	17	13.4	1.07	0.51	2.26	0.856
Clinic						
Bristol	10	13.3	1.00	-	-	-
Liverpool	20	22.5	1.88	0.81	4.36	0.133
West London	7	14.6	1.11	0.39	3.16	0.845
Birmingham	18	6.8	0.47	0.21	1.08	0.068
MMC	8	24.2	2.08	0.73	5.96	0.163
Age (years)						
≤24	23	9.2	1.00	-	-	-
25-34	25	14.8	1.71	0.93	3.13	0.081
≥35	15	16.1	1.89	0.93	3.82	0.072

Table continued on the next page

Explanatory variables	Outcome: clustered with European specimens					
	n	Row %	cOR	LCI	UCI	P value
Ethnicity						
White	33	13.7	1.00	-	-	-
Black Caribbean	8	7.8	0.54	0.24	1.21	0.127
Black African	4	12.9	0.93	0.31	2.85	0.904
Black Other	1	33.3	3.15	0.28	36.09	0.330
Asian	6	17.1	1.30	0.50	3.39	0.585
Other	3	37.5	3.78	0.85	16.80	0.060
Mixed	5	9.6	0.67	0.25	1.81	0.428
Country of birth						
UK	45	12.6	1.00	-	-	-
Not UK	17	14.2	1.15	0.653	2.09	0.652
Symptoms						
No	11	9.5	1.00	-	-	-
Yes	48	13.3	1.47	0.73	2.94	0.274
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis						
No/Unknown	58	12.7	1.00	-	-	-
Yes	5	9.4	0.72	0.27	1.88	0.499
HIV status						
Negative/Unknown	61	12.1	1.00	-	-	-
Positive	2	33.3	3.64	0.65	20.4	0.116
Number of sexual partners in the UK in the three months prior to diagnosis						
0	1	2.8	1.00	-	-	-
1	22	10.1	3.93	0.51	30.5	0.157
≥2	28	13.9	5.66	0.73	43.9	0.061
Reduced susceptibility to ceftriaxone (MIC≥0.015 mg/L)						
No	43	10.3	1.00	-	-	-
Yes	**20**	**23.3**	**2.65**	**1.46**	**4.82**	**<0.001**
Reduced susceptibility to cefixime (MIC≥0.03 mg/L)						
No	43	11.5	1.00	-	-	-
Yes	20	15.4	1.40	0.79	2.49	0.244
Reduced susceptibility to azithromycin (MIC ≥0.125 mg/L)						
No	29	8.1	1.00	-	-	-
Yes	**34**	**23.1**	**3.41**	**1.97***	**5.93**	**<0.001***
Resistance to ciprofloxacin (MIC >0.06 mg/L)						
No	41	11.0	1.00	-	-	-
Yes	22	16.5	1.60	0.91	2.81	0.099
Resistance to penicillin (MIC >1.0 mg/L)						
No	58	13.5	1.00	-	-	-
Yes	5	6.7	0.46	0.18	1.19	0.099

*the association between reduced susceptibility to azithromycin and clustering with EU specimens was no longer significant when reduced susceptibility to ceftriaxone was included in the model

7.3.4 Phylogenetic analyses: comparing PhD and USA specimens

7.3.4.1 Phylogenetic tree

The PhD study specimens were interspersed with the USA specimens throughout the phylogenetic tree. However, there were some large clades of solely PhD study or USA specimens (Figure 7.3). Of the 2,859,636 specimen pairs, the largest difference between two specimens was 17,142 SNPs.

7.3.4.2 Specimens with the mosaic penAXXXIV allele

As found when comparing the PhD study specimens with the European specimens, the large group of PhD study specimens with the mosaic *penAXXXIV* allele, but not the smaller group with this allele, were found to be genetically similar to a group of specimens from USA with the mosaic *penAXXXIV* allele (Figure 7.3).

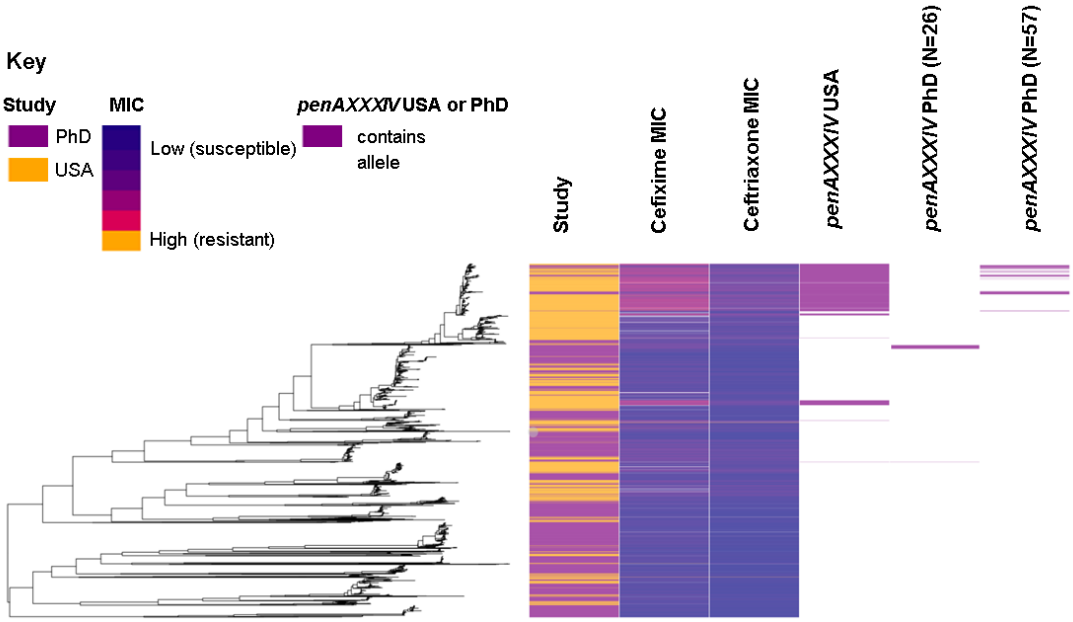


Figure 7.3 Phylogenetic tree of PhD and USA specimens with metadata for study type, ceftriaxone and cefixime MIC and presence of *penAXXXIV* allele
Total number of PhD study specimens = 1,277, Total number of USA study specimens = 1,139

7.3.4.3 Clustering by SNP differences

Fewer PhD study specimens clustered with the USA specimens than to the European specimens. Using a five SNP threshold, only one PhD study specimen clustered with a USA specimen. Using the 30 and 50 SNP thresholds, 1% and 15% of PhD specimens clustered with USA specimens, respectively. At the 100 SNP threshold, 22% of USA specimens clustered with PhD study specimens (Table 7.6).

Clusters defined by the more relaxed thresholds (30, 50 and 100 SNP) contained specimens collected over a longer period (Figure 7.4). One cluster with specimens from 2013 comprised two PhD study specimens from MSM and one USA specimen from a heterosexual man. These three specimens had low to middle range MICs to ceftriaxone (0.004 mg/L), cefixime (0.008 mg/L to 0.03 mg/L) and azithromycin (0.25 mg/L). The two PhD study specimens that clustered with the USA specimen also

clustered with three European specimens (Denmark, France and Hungary) from 2013.

Table 7.6 Clusters defined by SNP thresholds with specimens from the PhD study sample and USA collection

SNP threshold	Number of clusters	Total number of specimens that cluster	PhD study specimens that cluster		USA study specimens that cluster	
			n	% (N=1,277)	n	% (N=1,139)
5	1	2	1	0.1	1	0.1
10	10	35	15	1.2	20	1.8
50	18	114	40	3.1	74	6.5
100	21	446	191	15.0	255	22.4

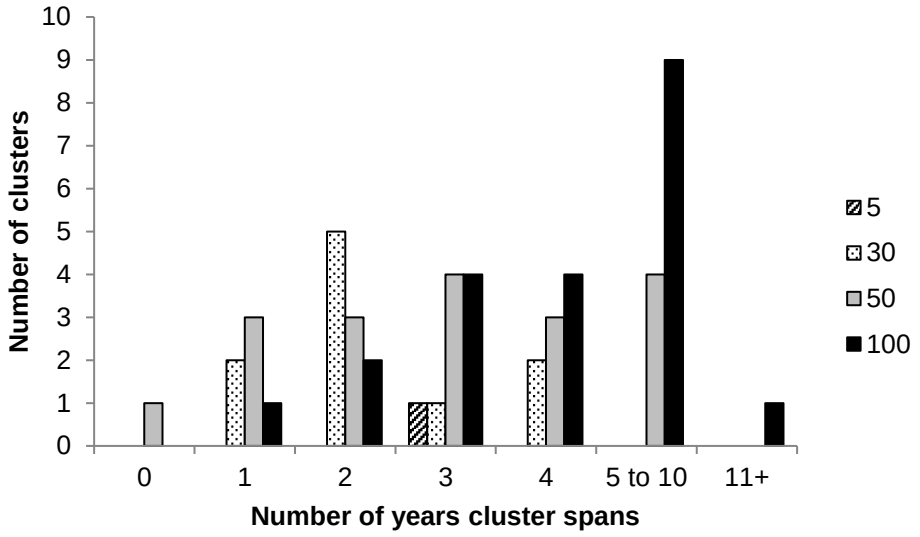


Figure 7.4 Number of years spanned by the specimens from the PhD study and the USA sample in each cluster

7.3.4.4 Epidemiological and phenotypic characteristics of specimens clustered with USA specimens

There were no measured epidemiological characteristics of specimens from MSM in the PhD study sample associated with clustering with specimens from the USA (using a 100 SNP threshold) (Table 7.7). However, specimens from MSM in the PhD study sample that clustered were more likely to have reduced susceptibility to azithromycin (base: susceptible to azithromycin; reduced susceptible to azithromycin aOR 3.95, CI 2.46-6.34, P <0.001) or resistance to ciprofloxacin (base: susceptible to ciprofloxacin; resistant to ciprofloxacin aOR 2.46, CI 1.59-3.80, P<0.001) than

those that did not (Table 7.7); they were less likely to be resistant to penicillin (base: susceptible to penicillin; resistant to penicillin aOR 0.18, CI 0.07-0.48, P = 0.001).

Similarly, I did not find any epidemiological characteristics associated with clustering with specimens from the USA for women and heterosexual men (using a 100 SNP threshold) (Table 7.8). However, specimens from women and heterosexual men in the PhD study sample that clustered were more likely to have reduced susceptibility to azithromycin (base: susceptible to azithromycin; reduced susceptibility to azithromycin aOR 12.7, CI 5.37-30.1, P <0.001) and resistance to ciprofloxacin (base: susceptible to ciprofloxacin; resistant to ciprofloxacin aOR 3.93, CI 1.77-8.73, P<0.001) than those that did not. They were also less likely to be from people diagnosed in the Birmingham clinic (base: Bristol clinic, Birmingham aOR 0.22, CI 0.08-0.65, P <0.001).

Table 7.7 Epidemiological characteristics and antimicrobial susceptibility profile of PhD study specimens from MSM that clustered with USA study specimens compared to those that did not cluster at the 100 SNP threshold (A) Univariate analysis: cOR = crude odds ratio, L/UCI = lower/upper confidence interval

Explanatory variables	Outcome: clustered with specimens from the USA					
	n	Row %	cOR	LCI	UCI	P value
Year						
2013	79	38.2	1.00	-	-	-
2014	33	16.2	**0.31**	**0.19**	**0.51**	**<0.001**
2015	10	4.3	**0.07**	**0.03**	**0.16**	**<0.001**
2016	16	12.9	**0.24**	**0.13**	**0.45**	**<0.001**
Clinic						
Bristol	21	23.9	1.00	-	-	-
Liverpool	15	30.6	1.41	0.64	3.09	0.391
West London	29	21.0	0.85	0.45	1.61	0.626
Birmingham	**16**	**11.6**	**0.42**	**0.20**	**0.87**	**0.015**
MMC	57	16.1	0.61	0.35	1.09	0.090
Age (years)						
≤24	29	21.5	1.00	-	-	-
25-34	66	19.8	0.90	0.55	1.47	0.675
≥35	43	14.5	0.62	0.37	1.05	0.071
Ethnicity						
White	107	18.4	1.00	-	-	-
Black Caribbean	6	20.0	1.11	0.44	2.79	0.821
Black African	2	12.5	0.64	0.14	2.84	0.550
Black Other	2	28.6	1.78	0.34	9.31	0.489
Asian	9	23.1	1.33	0.61	2.90	0.464
Other	3	12.5	0.64	0.19	2.17	0.466
Mixed	8	15.1	0.79	0.36	1.73	0.555
Country of birth						
UK	78	18.4	1.00	-	-	-
Not UK	55	19.2	1.05	0.72	1.54	0.717
Symptoms						
No	39	17.6	1.00	-	-	-
Yes	87	19.6	1.15	0.75	1.74	0.521

Table continued on the next page

Explanatory variables	Outcome: clustered with specimens from the USA					
	n	Row %	cOR	LCI	UCI	P value
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis						
No/Unknown	97	18.7	1.00	-	-	-
Yes	41	16.7	0.88	0.59	1.31	0.520
HIV status						
Negative/Unknown	100	18.3	1.00	-	-	-
Positive	38	17.3	0.93	0.62	1.41	0.734
Number of sexual partners in the UK in the three months prior to diagnosis						
0	1	9.1	1.00	-	-	-
1	24	20.2	2.53	0.30	21.0	0.374
≥2	52	19.3	2.39	0.30	19.1	0.399
Reduced susceptibility to ceftriaxone (MIC≥0.015 mg/L)						
No	81	14.2	1.00	-	-	-
Yes	57	29.7	**2.55**	**1.72**	**3.79**	**<0.001**
Reduced susceptibility to cefixime (MIC≥0.03 mg/L)						
No	79	14.7	1.00	-	-	-
Yes	59	26.3	**2.08**	**1.42**	**3.06**	**<0.001**
Reduced susceptibility to azithromycin (MIC ≥0.125 mg/L)						
No	41	12.2	1.00	-	-	-
Yes	97	22.8	**2.13**	**1.42**	**3.18**	**<0.001**
Resistance to ciprofloxacin (MIC >0.06 mg/L)						
No	53	12.1	1.00	-	-	-
Yes	85	26.2	**2.57**	**1.75**	**3.78**	**<0.001**
Resistance to penicillin (MIC >1.0 mg/L)						
No	133	21.4	1.00	-	-	-
Yes	5	3.5	**0.14**	**0.05**	**0.34**	**<0.001**

(B) Multivariable analysis: aOR = adjusted odds ratio, L/UCI = lower/upper confidence interval

Explanatory variables	Outcome: clustered with specimens from the USA			
	aOR	LCI	UCI	P value
Year				
2013	1.00	-	-	-
2014	**0.31**	**0.19**	**0.53**	**<0.001**
2015	**0.05**	**0.02**	**0.10**	**<0.001**
2016	**0.19**	**0.10**	**0.36**	**<0.001**
Reduced susceptibility to azithromycin (MIC ≥0.125 mg/L)				
No	1.00	-	-	-
Yes	**3.95**	**2.46**	**6.34**	**<0.001**
Resistance to ciprofloxacin (MIC >0.06 mg/L)				
No	1.00	-	-	-
Yes	**2.46**	**1.59**	**3.80**	**<0.001**
Resistance to penicillin (MIC >1.0 mg/L)				
No	1.00	-	-	-
Yes	**0.18**	**0.07**	**0.48**	**0.001**

Table 7.8 Epidemiological characteristics and antimicrobial susceptibility profile of PhD study specimens from women and heterosexual men that clustered with USA study specimens compared to those that did not cluster at the 100 SNP threshold

(A) Univariate analysis: cOR = crude odds ratio, L/UCI = lower/upper confidence interval

Explanatory variables	Outcome: clustered with specimens from the USA					
	n	Row %	cOR	LCI	UCI	P value
Year						
2013	26	21.8	1.00	-	-	-
2014	15	11.6	**0.47**	**0.23**	**0.95**	**0.031**
2015	6	4.4	**0.17**	**0.06**	**0.43**	**<0.001**
2016	6	4.7	**0.18**	**0.07**	**0.46**	**<0.001**
Clinic						
Bristol	16	21.3	1.00	-	-	-
Liverpool	14	15.7	0.69	0.31	1.53	0.357
West London	9	18.8	0.85	0.34	2.13	0.729
Birmingham	11	4.1	**0.16**	**0.07**	**0.37**	**<0.001**
MMC	3	9.1	0.37	0.10	1.39	0.126
Age (years)						
≤24	23	9.2	1.00	-	-	-
25-34	18	10.7	1.17	0.61	2.25	0.634
≥35	12	12.9	1.46	0.69	3.07	0.320
Ethnicity						
White	34	14.1	1.00	-	-	-
Black Caribbean	6	20.0	0.38			
Black African	0	0.0	-	-	-	-
Black Other	0	0.0	-	-	-	-
Asian	4	7.5	0.78	0.26	2.37	0.668
Other	2	3.8	2.03	0.39	10.5	0.390
Mixed	5	9.4	0.89	0.33	2.45	0.830
Country of birth						
UK	34	9.5	1.00	-	-	-
Not UK	18	15.0	1.68	0.91	3.11	0.094
Symptoms						
No	10	8.6	1.00	-	-	-
Yes	40	11.1	1.33	0.64	2.74	0.447
Diagnosed with an STI (excluding HIV) in the year prior to gonorrhoea diagnosis						
No/Unknown	49	10.7	1.00	-	-	-
Yes	4	7.5	0.68	0.24	1.97	0.477
HIV status						
Negative/Unknown	53	10.5	-	-	-	-
Positive	0	0.0	-	-	-	-
Number of sexual partners in the UK in the three months prior to diagnosis						
0	7	19.4	1.00	-	-	-
1	18	8.3	0.37	0.14	0.98	0.037
≥2	23	11.4	0.54	0.21	1.37	0.184
Reduced susceptibility to ceftriaxone (MIC≥0.015 mg/L)						
No	24	5.7	1.00	-	-	-
Yes	29	33.7	**8.37**	**4.38**	**16.0**	**<0.001**
Reduced susceptibility to cefixime (MIC≥0.03 mg/L)						
No	24	6.4	1.00	-	-	-
Yes	29	22.3	**4.20**	**2.30**	**7.66**	**<0.001**

Table continued on the next page

Explanatory variables	Outcome: clustered with specimens from the USA					
	n	Row %	cOR	LCI	UCI	P value
Reduced susceptibility to azithromycin (MIC ≥0.125 mg/L)						
No	12	3.4	1.00	-	-	-
Yes	41	27.9	**11.1**	**5.38**	**23.1**	**<0.001**
Resistance to ciprofloxacin (MIC >0.06 mg/L)						
No	18	4.9	1.00	-	-	-
Yes	35	25.9	**6.84**	**3.60**	**13.0**	**<0.001**
Resistance to penicillin (MIC >1.0 mg/L)						
No	50	11.6	1.00	-	-	-
Yes	3	4.0	**0.32**	**0.10**	**1.05**	**0.047**

(B) Multivariable analysis: aOR = adjusted odds ratio, L/UCI = lower/upper confidence interval

Explanatory variables	Outcome: Clustered with specimens from the USA			
	aOR	LCI	UCI	P value
Year				
2013	1.00	-	-	-
2014	**0.27**	**0.10**	**0.70**	**0.007**
2015	**0.04**	**0.01**	**0.12**	**<0.001**
2016	**0.09**	**0.03**	**0.30**	**<0.001**
Clinic				
Bristol	1.00	-	-	-
Liverpool	0.92	0.32	2.62	0.871
West London	1.22	0.36	4.13	0.746
Birmingham	**0.22**	**0.08**	**0.65**	**0.006**
MMC	0.38	0.08	1.85	0.230
Reduced susceptibility to azithromycin (MIC ≥0.125 mg/L)				
No	1.00	-	-	-
Yes	**12.7**	**5.37**	**30.1**	**<0.001**
Resistance to ciprofloxacin (MIC >0.06 mg/L)				
No	1.00	-	-	-
Yes	**3.93**	**1.77**	**8.73**	**0.001**

7.4 Discussion and conclusions

In this chapter, I explored the phylogenetic relationships between *N. gonorrhoeae* diagnosed in England and internationally. Some lineages of *N. gonorrhoeae* in England were genetically similar to those found in Europe and the USA and specimens that clustered with international specimens were more likely to have reduced susceptibility to antimicrobials. Frequent transmission between countries is likely to undermine efforts in England to prevent and control antimicrobial resistance in *N. gonorrhoeae*. The success of antimicrobial stewardship policies in England, such as good compliance with treatment guidelines, that aim to slow the development of resistance in the endemic gonococcal population, could be diminished by the importation and spread of resistant strains. These data support the need to promote STI prevention messages and testing to travellers, particularly

those visiting an area where the prevalence of *N. gonorrhoeae* antimicrobial resistance is high, such as South East Asia.

7.4.1 The genetic similarity of *N. gonorrhoeae* specimens in England to those circulating globally

I identified only a few clusters that likely represented recent transmission of *N. gonorrhoeae* between people from England and Europe or the USA. This number may have been higher had there been more overlap between the studies (period of studies: PhD study: 2013-2016, European study: 2013, USA study: 2000-2013). One cluster was found in people from all three locations with diagnoses in the same year and in different sexual risk groups (MSM and heterosexual men). Additionally, the large group of specimens in England with the mosaic *penAXXXIV* allele were genetically similar to specimens from Europe and the USA with the same allele. The majority of the mosaic *penAXXXIV* allele specimens belonged to the NG-MAST 1407 lineage, which was a widely-disseminated clone associated with elevated ceftriaxone and cefixime MICs, and the catalyst for changing national treatment guidelines from cefixime as first-line therapy to ceftriaxone.[99]

The genetic similarity of specimens across all three datasets demonstrates how strains of *N. gonorrhoeae* with concerning genotypic and phenotypic features are present across multiple countries and probably represents frequent transmission across borders and continents driven by travel-associated sexual partnerships. Furthermore, the clustering analysis found that specimens with reduced susceptibility to antimicrobials in England were more likely to be genetically similar to specimens from Europe or the USA than susceptible specimens.

By contrast, my analysis using epidemiological data alone found only one example of an association between reporting recent travel-associated sex and infection with a reduced susceptible or resistant strain of *N. gonorrhoeae* (ciprofloxacin resistant *N. gonorrhoeae* in heterosexuals), providing a concrete example where the WGS data provide additional insights.

7.4.2 Epidemiological characteristics of people reporting travel-associated sexual partnerships or clustering with international specimens

A similar percentage of MSM and heterosexuals reported a recent travel-associated sexual partnership. People who were not born in the UK were more likely to report a sexual partner whilst travelling abroad, which may reflect the greater likelihood of

travel and connection to people from other countries, such as their country of birth, and therefore strains of *N. gonorrhoeae* from outside of the UK. There was no difference in the epidemiological characteristics of cases that clustered with specimens from Europe and the USA compared to those that did not, suggesting that either the sexual networks that span between countries are epidemiologically diverse or further investigation using more specific sexual behaviour or larger datasets is required to identify risk groups or factors associated with international dissemination of *N. gonorrhoeae*.

7.4.3 Methodological and data source limitations

The international datasets used in these analyses were the largest datasets of sequenced *N. gonorrhoeae* outside of the UK available to date (mid-2018). However, these datasets were neither comprehensive nor representative of the geographical locations they covered. Each dataset was collected for a specific purpose, and the sampling strategies varied. The direction of spread i.e. importation or exportation to or from England was not easily identified due to differing study periods. No dataset collected consecutive specimens across the whole study period or geographical area and these gaps in coverage could lead to misinterpretation. The sampling strategy of the USA study included matched resistant and sensitive pairs.[141] Therefore, the conclusion that significantly more resistant than susceptible specimens in England clustered with USA specimens could be because of a higher proportion of the resistant gonococcal population in the USA was available for comparison to the PhD study sample compared to the susceptible gonococcal population.

In order to compare the international datasets with the PhD study sample I used a different method (FastTree instead of RAxML) for creating the phylogenetic trees that was computationally achievable with the large number of sequences being processed. However, FastTree is considered to be a less accurate method for phylogeny reconstruction than RAxML.[114] Using FastTree resulted in larger SNP differences between specimen pairs, which could potentially have underestimated the number of specimens that clustered. However, I analysed the data using a range of SNP thresholds that should incorporate the increase in SNPs as a result of using FastTree instead of RAxML. Although some of these SNPs thresholds were more relaxed than other analyses in previous chapters, the purpose was to identify specimens from England that were more genetically similar to international specimens relative to other specimens in England. Genetic similarity can be defined by any SNP threshold; the only difference would be the magnitude of similarity

concluded from the analysis. The smaller the SNP threshold the more closely related the specimens are interpreted to be, therefore, the use of wider SNP thresholds in this chapter means there may be more indirect links between people grouped in a sexual network.

There is considerable concern about antimicrobial resistance in *N. gonorrhoeae* believed to have emerged in East and South East Asia, not least because the most recently identified treatment failures have derived from these areas.[139,164] While it would have been interesting to include specimens from these areas, there was a lack of publicly available sequences datasets. More data from East and South East Asia may become available as sequencing becomes cheaper and there are already programmes in place supported by CDC and the WHO to improve the surveillance of *N. gonorrhoeae* in East and South East Asia.[286]

7.4.4 Public health application

7.4.4.1 Research

Quantifying the relationship between *N. gonorrhoeae* circulating in England and internationally may help parameterise mathematical models that aim to predict the prevalence and distribution of antimicrobial resistant infections and the degree to which they are due to importation as compared with *de novo* development. Previous studies concluded that importation events probably drive the spread of antimicrobial resistance in a population with low population prevalence such as the UK[287-290] but other recent evidence from WGS data has identified *de novo* development of resistance to azithromycin in England.[140]

7.4.4.2 Surveillance

These analyses emphasise the degree of international dissemination of *N. gonorrhoeae* and the data support the need for collaboration across countries in efforts to prevent and control the spread of infection and AMR. The value of WGS data to understand how sexual networks that cross international borders support and facilitate the transmission of *N. gonorrhoeae* would be maximised by globally standardised data collection and analysis methodologies, and open access to international datasets. Although WGS data are usually freely available in online archives when studies are published, these data are not easily analysable by bioinformatics novices and often lack epidemiological and phenotypic data, thereby limiting the opportunity for epidemiological analyses. A useful recent development

has been the production of an online tool that provides a quick and easy way to analyse WGS data and associated metadata from a range of countries.[270] One of the most valuable features of the tool is that no specialist bioinformatics training is required to complete analyses and extract information, enabling use in the wider public health field. Therefore, this new online tool provides a much-needed facility to help improve the accessibility of WGS data and translate findings from molecular epidemiological studies into public health action.

7.4.5 Chapter summary

To answer the research questions of this chapter, I have; (i) described the relationship between people reporting recent travel-associated sexual partnerships and infection with a reduced susceptible strain of *N. gonorrhoeae*; (ii) described the genetic similarity of *N. gonorrhoeae* in England to *N. gonorrhoeae* circulating in Europe and the USA; and (iii) characterised the patient groups more likely to be infected with an international strain of *N. gonorrhoeae*. In summary:

- WGS data identified cases of *N. gonorrhoeae* in England that were genetically similar to cases found in Europe and the USA regardless of whether the patient reported a recent travel-associated sexual partnership, showing international dissemination, particularly of *N. gonorrhoeae* with reduced susceptibility to antimicrobials, is common

- Computational limitations restricted the analyses of datasets containing thousands of samples but progress in other areas of information technology have improved the accessibility of bioinformatics analyses for non-bioinformaticians working in public health

8 Discussion

8.1 Chapter summary

In this chapter, I revisit the background and rationale for the PhD study and summarise the findings of each chapter. I present a summary of how WGS data could be used to support existing and develop new methods of *N. gonorrhoeae* prevention, surveillance and control, and reflect on how WGS might be implemented more routinely in public health practice. I also outline the overarching limitations of the PhD study, which impact the interpretation and utility of the results. Finally, I propose future areas of research that could address these limitations and extend the value of using WGS to inform the control of *N. gonorrhoeae*.

8.2 Background and rationale of the thesis

The purpose of this PhD was to explore how WGS data could be used to inform the control of *N. gonorrhoeae*, which is a STI associated with poor sexual health outcomes. *N. gonorrhoeae* is of growing concern globally because of the development of antimicrobial resistance and the impending threat of untreatable infections. Interventions to prevent and control *N. gonorrhoeae* rely on understanding transmission patterns at local, national and international levels. Epidemiological investigations of sexual mixing patterns have relied on reported data on time, person and place to infer links between cases and better understand the drivers of gonorrhoea transmission. However, epidemiological data are limited by reporting bias associated with sensitive sexual behaviour information and the challenge of knowing whether infected individuals with similar characteristics or from the same geographic area truly share a sexual network. More recently, molecular technologies for comparing specimens at a genome level have been used to explore the epidemiology of *N. gonorrhoeae*. In theory, WGS should provide an additional and potentially invaluable layer of evidence about transmission and the spread of infection between different population groups. However, there are numerous limitations and analytical challenges, and most molecular epidemiology studies to date have used only a few genes to distinguish clusters of infection. Although improvements in technology and reductions in cost mean that more recent studies use the whole genome, it remains unclear whether or exactly how WGS should be incorporated into routine public health practice.

In isolation, molecular typing data can determine the genetic distance between specimens and group cases into probable sexual transmission networks. However, this is of limited value unless the networks can be characterised by adding time and demographic, clinical and behavioural information to the analyses. Few *N. gonorrhoeae* WGS studies have described sexual networks using patient data, and none have investigated infection transmission across multiple locations in England during the same period. Many WGS studies have focused on studying the genetic determinants of AMR by comparing genomic and phenotypic data. These studies contribute evidence towards the development of rapid diagnostic tests for antimicrobial susceptibility, which may inform treatment decisions. Detecting AMR infection quickly and preferably before any treatment is given is of increasing importance given the ongoing dissemination of AMR strains. However, there are problems with this approach; including that *de novo* mutation may render the tests less sensitive over time.

My thesis aimed to extend the knowledge of how WGS can be used to understand sexual mixing patterns and thereby inform the development and delivery of interventions to control gonorrhoea and antimicrobial resistant infections. I created and analysed a unique dataset combining WGS, epidemiological and phenotypic data and used it to describe the genetic variation of *N. gonorrhoeae* and antimicrobial resistance in England, characterise sexual networks and identify opportunities for infection control and public health intervention.

8.3 Summary of thesis chapters and findings

I described the biology and epidemiology of *N. gonorrhoeae* and outlined the PhD topic and structure of the thesis in Chapter 1. A systematic review (Chapter 2) of the published literature was used to consolidate and critique existing *N. gonorrhoeae* studies that combined molecular and epidemiological data.[1] The review identified few WGS studies of gonorrhoea. Most studies used a convenience sampling strategy or restricted the study to specimens with a specific phenotype or NG-MAST type, which limited the interpretation and extrapolation of results. In contrast, the dataset created in this PhD (Chapter 3 and Chapter 4) consisted of consecutive specimens from five clinics in England during a three-month period each year for four years, and was broadly representative of *N. gonorrhoeae* circulating in communities served by these clinics during that timeframe.

Clustering of genetically similar specimens was used to infer sexual transmission networks and the clusters were characterised using the epidemiological data available (Chapter 5). A considerable number of sexual networks contained specimens from heterosexual men and MSM only, suggesting there may be groups of heterosexual-identifying men who could benefit from enhanced sexual health testing and prevention messages that would normally be targeted to MSM. Within the MSM population there was evidence of extensive sexual mixing between men living with HIV and those who were HIV-negative/unknown within networks where gonorrhoea was being transmitted. These men may be at considerable and unrecognised risk of HIV infection. These findings highlight opportunities for the delivery of STI and HIV prevention and control measures, as outlined in the sections below.

I used the molecular data combined with AMR phenotype to investigate to what extent the *N. gonorrhoeae* AMR in England might be due to the emergence of different strains or dissemination of one or a small number of clonal strains (Chapter 6). I observed genetically distinct clusters with similar phenotypes and genotypic markers of resistance, indicating that antimicrobial resistance had developed and spread simultaneously in multiple discrete sexual networks. I was also able to compare the PhD study sample with specimens from the rest of Europe and the USA to assess international transmission of *N. gonorrhoeae* (Chapter 7). I observed many specimens in the PhD dataset that were genetically similar to specimens isolated in the USA and Europe, including specimens with reduced susceptibility to antimicrobials, highlighting the importance of international travel in connecting sexual networks and facilitating dissemination of *N. gonorrhoeae* AMR. The frequent transmission between countries may undermine efforts in England to prevent and control antimicrobial resistance in *N. gonorrhoeae* because new introductions from areas with less advanced AMR reduction and prevention initiatives are likely to continue.

8.4 Improving the control of *N. gonorrhoeae* using WGS data

8.4.1 Clinical care: sexual health risk assessment and service provision

8.4.1.1 Limitations of current practice

National guidelines for STI and HIV testing and prevention services vary for different populations and their associated sexual behaviours. For example, MSM are routinely offered Hepatitis A (HAV) and Hepatitis B (HBV) testing and vaccination,[262] while this

is not the case for most heterosexual men. To determine the appropriate guidelines to follow and services to offer, clinicians use self-reported behavioural data from the patient and any available information on the characteristics of the sexual network the patient is participating in. However, patients may be reluctant to report their true sexual behaviour due to concerns about confidentiality and stigma[291-293] and clinicians may not have specific information about the characteristics of a patient's sexual network because it requires intensive contact tracing, which is time consuming and often unsuccessful.[39,45]

8.4.1.2 Improvements from using WGS data

WGS data could, when combined with epidemiological data, be used to gain better understanding about a person's sexual network and risk of STIs including HIV. Evidence from the thesis demonstrated two examples of where WGS data could be used to tailor the sexual health services provided, if the information could be shared sensitively and without deductive disclosure:[248]

> (i) Sexual networks consisting only of MSM and men reporting heterosexual behaviour were identified; these heterosexual men may benefit from being offered more tailored sexual health services similar to the services offered to MSM – e.g. HIV pre-exposure prophylaxis (PrEP) and HAV and HBV testing and vaccination.
>
> (ii) Sexual networks consisting of both HIV-positive and HIV-negative/unknown MSM, implying potential risk of HIV infection in HIV-negative MSM belonging to these sexual networks who might be offered HIV PrEP and HIV prevention advice. Indeed, heterosexual women and men, who are in mixed HIV status sexual networks might also be offered PrEP and HIV prevention advice.

8.4.2 Clinical care: treatment

8.4.2.1 Limitations of current practice

In England, many people diagnosed with gonorrhoea in sexual health clinics are treated empirically with the first-line recommended therapy in the national guideline and often without information about the antimicrobial susceptibility profile of the infection.[11] The guideline states that if the infection is known to be susceptible to antimicrobials other than the first-line therapy, such as ciprofloxacin, these can be used instead. However, phenotypic testing for antimicrobial susceptibility by primary

diagnostic laboratories serving sexual health clinics might only be performed for the first-line therapies because it is expensive and time consuming.[294] Therefore, at the time of treatment, clinicians often do not know whether the infection could be treated with another antimicrobial.

8.4.2.2 Improvements from using WGS data

If the relationship between the genotypic markers of resistance and phenotypic antimicrobial susceptibility were refined, and data were available in a clinically relevant timeframe (discussed further in section 8.45), WGS could provide susceptibility data for multiple antimicrobials and be used to inform decisions about the drug most likely to be active against the infecting strain, which could assist in the effort to preserve ceftriaxone for instances where no alternative is available. Reducing the use of first line drugs should reduce selection pressures on the gonococcal population and the development of resistance, thereby helping to preserve antimicrobial treatment options for the future.[99,295] Although rapid molecular tests for genetic markers of resistance are at different stages of development, these tests mostly only provide data on one antimicrobial at a time. A benefit of using WGS data is that information about all the genetic determinants of resistance are available simultaneously.

8.4.3 Surveillance and outbreak detection

8.4.3.1 Limitations of current practice

In England, outbreaks of gonorrhoea and antimicrobial resistant infection are usually detected by formal and informal exceedance reports that detect and flag observed increases in cases beyond the expected number, based on past trends.[44,45,140,296] Once an outbreak is suspected, further investigations usually seek to determine the connections between cases and to describe the sexual network involved in the spread of infection. However, no routine surveillance systems and few local information sources have the ability to recreate the sexual network. Instead, efforts to identify connections between cases occur through contact tracing, which can be time consuming, resource intensive and have a low success rate.[40,43-45] In a recent outbreak of *N. gonorrhoeae* with high-level azithromycin resistance, contact tracing was only successful in 13% of cases.[45]

8.4.3.2 Improvements from using WGS data

My data suggest that WGS data might be used to support outbreak investigations by providing additional data that may be incomplete or unavailable using more traditional epidemiological methods. The benefits would depend on the timing of WGS analysis during the investigation:

- *pre-outbreak investigation:*
    - If WGS becomes part of routine *N. gonorrhoeae* surveillance, the background pattern of transmission in sexual networks would be known, therefore novel and/or expanding clusters could be detected as possible outbreaks that require further investigation

- *during outbreak investigation:*
    - Once an outbreak is declared, WGS data could be used in the outbreak case definition and where available would confirm whether suspected cases are part of the same cluster

    - WGS analysis could support case finding by identifying cases that may not have been identified through contact tracing or other means

    - WGS data could be used to determine the sexual networks in which the outbreak is being transmitted and help monitor the spread of infection between different population groups, geographic areas and over time

    - By comparing the WGS data of outbreak cases to international datasets, the question of whether the outbreak was linked to *N. gonorrhoeae* strains in other countries could be investigated

- *during/end of outbreak investigation:*
    - WGS data could be used to evaluate whether an outbreak has been successfully controlled by tracking whether the outbreak strain persists in the population over time

8.4.4 Research

8.4.4.1 Limitations of current practice

Research studies investigating the spread of *N. gonorrhoeae* and antimicrobial resistance typically use self-reported data from people about their sexual behaviours

to understand sexual mixing patterns associated with STI transmission.[80,253,297,298] However, self-reported data on sexual behaviour can be subject to social desirability and recall bias.[299,300] Despite these limitations, data from epidemiological studies are frequently used as estimates for sexual mixing parameters within mathematical modelling studies, which are used to estimate the impact of gonorrhoea prevention and control interventions.[56,58,273,301,302]

8.4.4.2 Improvements from using WGS data

As demonstrated in this PhD, WGS data can be used to construct sexual networks and describe sexual mixing patterns. The complexity of the networks identified through WGS analysis highlights aspects of sexual mixing in the population that are important to include in mathematical models, such as the mixing of some heterosexual men and MSM, and the importation of *N. gonorrhoeae* from other countries. Estimates for the likelihood of contact between different groups in the population could be extracted from the WGS cluster analysis and used within mathematical models. If WGS data were routinely collected and analysed as part of gonorrhoea surveillance, then sexual mixing patterns of the population could be updated frequently.

8.5 Implementing *N. gonorrhoeae* WGS in public health and clinical care settings

There are several practical and cost considerations to take into account if WGS data are to be used routinely to inform the control of *N. gonorrhoeae*.[2,303] There would need to be an expansion in facilities and expertise to conduct, analyse and share WGS data at the local and national level. At present, only a few local laboratories, research institutes and the national reference laboratory at PHE have the equipment and personnel in place to perform WGS. Standard protocols for data collection, analysis and reports would need to be developed specifically for *N. gonorrhoeae* and tailored for clinical care and outbreak control. Results would need to be available in enough time to be relevant for clinical and/or public health decision making: existing methods for detection and antimicrobial susceptibility testing require about two weeks. WGS would ideally be available within this timeframe at a minimum, but would have greater value as fast as a near patient test that detects *N. gonorrhoeae* within hours.[304] Conducting WGS in real-time would require direct sequencing from urine or swabs, rather than from culture. This would have an additional benefit of enabling more specimens to be analysed, as culture is only completed on around

50% of cases.[219] Phenotypic testing for antimicrobial susceptibility using culture specimens would still be required to understand the relationship between the genotype and phenotype of new strains.

Implementation might be guided by systems developed for other pathogens, such as *Mycobacterium tuberculosis* (TB) and bacterial gastro-intestinal infections including *Shigella sonnei* and *Salmonella enterica*.[305-308] For TB, WGS improved the turn-around times for reporting antimicrobial resistance phenotypes from one month to one week. For the gastro-pathogens, WGS has largely replaced existing typing methods used to identify transmission networks and emerging outbreaks. Other countries are also using WGS data routinely in their public health services: in British Columbia, Canada, all new HIV diagnoses are genotyped and automatically analysed within a HIV phylogenetic monitoring system. This system identifies and describes phylogenetic clusters of infection using a range of variables including drug resistance profile, viral load, geographical location, and injecting drug use and sexual behaviour.[309] A monthly report is provided to public health practitioners who use the information to help prioritise public health resources, usually by focusing on following up cases in clusters that continue to increase in size.

8.6 Limitations of the PhD study

8.6.1 Sampling strategy

The sampling strategy of the PhD study aimed to capture both the diversity of the *N. gonorrhoeae* population and sufficient representation of sexual networks by selecting all available specimens from each clinic over a three-month period for four consecutive years. Research datasets, such as the dataset I created in this PhD, will always be incomplete in representing sexual networks because (i) many infections are never diagnosed (particularly where asymptomatic),[30] (ii) only around half of all diagnosed cases are cultured,[219] and (iii) not all diagnosed cases cultured were included in the PhD study sample. Cases within a sexual network that are not included may have different characteristics to the sampled population, which might alter the results and interpretation. However, when I compared the PhD study sample to gonorrhoea diagnoses in England (Chapter 4), the specimens included in the PhD study sample were broadly representative of the wider pathogen population, indicating that the findings from the PhD are probably reliable in providing estimates of *N. gonorrhoeae* transmission and clustering in England.

Prior knowledge about the clustering of *N. gonorrhoeae* from different locations and population groups in England using WGS data are necessary to complete formal sample size and power calculations. However, data on WGS *N. gonorrhoeae* in England were limited and it was not known in advance how many clusters of *N. gonorrhoeae* there would be or the characteristics of the cases in these clusters. Therefore, sample size and power calculations couldn't be used to determine how many specimens to include in the study to detect differences in clustering between population groups.

8.6.2 Extracting WGS data

In the PhD study, sexual networks were constructed using one genome sequence to represent each infection. This sequence was extracted from an archive specimen, which may differ from the actual infection in the patient because of the series of subcultures taken to create the archive and re-extract for sequencing. Genotypic changes may occur with every subculture, meaning the genome of the infection becomes different to the genome of infection that was transmitted in the network.[118] Depending on the extent of these genetic changes, specimens in the same sexual network may no longer cluster. To address this for future WGS studies, sequencing should be completed on clinical specimens taken directly from the patient.

Infections with mixed strains of *N. gonorrhoeae* may also cause people in a sexual network to be inaccurately separated in the clustering analysis. Only one colony of *N. gonorrhoeae* is selected from culture to represent each person's infection. If mixed strains are transmitted between two people, the same strain may not be selected for each person and their infections may appear genetically different. More studies are required to assess the variability of *N. gonorrhoeae* in the host. Some studies have found that there is minimal *N. gonorrhoeae* genomic diversity within a patient.[118,310]

8.6.3 Defining sexual networks

The estimated SNP differences between specimens were taken from phylogenetic trees and used to group specimens together into clusters that represented sexual networks with gonorrhoea transmission. SNP differences were measured on a continuous scale and SNP difference thresholds were used to categorise specimens into clusters. This allows comparisons to be made but also brings analytical challenges because specimens with a SNP difference slightly higher than the threshold would not be considered in the same network, even though they may be

genetically very similar. To mitigate this problem, sensitivity analyses using a range of SNP thresholds could examine how clusters change and the effect this has on the conclusions made and public health action taken. Further discussion of this topic along with the effect of recombination on analyses of the gonococcal genome when using SNPs to define clusters is presented in Chapter 5.

8.6.4 Epidemiological data

The epidemiological data available on the cases in the PhD study sample were more detailed than those available in most *N. gonorrhoeae* WGS studies to date[118,121,142] but were nonetheless limited compared to other sexual behaviour studies on *N. gonorrhoeae*, such as Natsal.[298] More specific data on sexual partnerships, such as where and how partnerships were formed, may help further characterise the sexual network and explain sexual mixing patterns. For example, in the PhD study (Chapter 5) there was no measurable difference in the characteristics of HIV-negative/unknown MSM that were in sexual networks with HIV-positive MSM, compared to those that were not.

The variables most notably missing data were travel-associated sexual partnerships and number of sexual partners in the UK (both 35%; 422/1,277). These variables were predominantly missing for all the specimens from the MMC clinic (N=386). To improve data completeness, I requested the data from the clinic but unfortunately it was not possible to access these data. Consequently, the analyses using these variables only included specimens from the other four clinics. As the population attending MMC clinic is predominantly MSM, analyses using data on travel-associated sexual partnerships and number of sexual partnerships may be less representative of the MSM population.

Clinical errors in the recording of the epidemiological data items may have occurred, which would affect the accuracy of the data analysed in the PhD. However, this was unlikely to be a significant problem because each year GRASP data are routinely checked for consistency at PHE and reported back to each clinic. As some of the epidemiological data items are self-reported by the patient, they are subject to recall bias. However, this is also not expected to be a significant problem because patients were asked to recall information from the recent past (three months).

8.7 Future research

The PhD study has provided novel insights into how WGS data can be used to inform the control of *N. gonorrhoeae* but additional research in the following areas is likely to enhance our understanding.

i. Implement and evaluate interventions using WGS

WGS data could be used in clinical care and public health surveillance as described earlier (see sections 8.4.2.2 and 8.4.3.2) and these could be prospectively evaluated to determine whether WGS data has a measurable benefit in controlling *N. gonorrhoeae* cases and/or other infections, such as HIV. The evaluation could be informed by evaluations of molecular typing services for other pathogens, such as TB.[311-313]

ii. Compare data from England to other international datasets

Analyses to explore the transmission patterns between England and other countries known to have a higher prevalence of AMR, such as those in South East Asia, could be repeated with *N. gonorrhoeae* WGS datasets as these become available.

iii. Refine the definition of sexual networks from WGS data

To further understand the expected number of SNP differences between specimens from cases in the same transmission network, WGS studies of known partners could be completed. Additionally, phylogenetic analyses that incorporate time to create what is known as a timed phylogeny could also be performed to determine the temporal relationship between cases and predict the transmission tree, which estimates the direction of transmission between cases in the sample.[314] Timed phylogenies determine with more precision than other phylogenetic methods the temporal relationship between cases and are used to estimate the mutation rate of the pathogen and the number of unsampled cases. However, current methods require a large percentage of the infections in a transmission network to be sampled.[267,268]

iv. Genome wide association studies to identify markers of antimicrobial resistance

Further exploration to identify the regions of the genome associated with antimicrobial resistance is required if genotypic data are to be used to inform treatment. These studies would need to be repeated frequently with recent specimens to account for the rapid change of the gonococcal genome.

9 References

1. Town K, Bolt H, Croxford S, et al. *Neisseria gonorrhoeae* molecular typing for understanding sexual networks and antimicrobial resistance transmission: A systematic review. *J Infect* 2018; 76(6): 507-14.

2. Kirkcaldy RD, Town K, Gernert KM, et al. Considering the potential application of whole genome sequencing to gonorrhea prevention and control. *Sex Transm Dis* 2018; 45(6): e29-e32.

3. Town K. Harnessing molecular technologies to inform our understanding of HIV and STI epidemics. Fourth Joint Conference of BHIVA with BASHH. Edinburgh, United Kingdom; 2018.

4. Town K, Field N, Furegato M, Cole M, Harris S, Hughes G. The emergence and spread of antimicrobial resistance *Neisseria gonorrhoeae* in HIV positive men who have sex with men. International Society for Sexually Transmitted Disease Research STI & HIV World Congress. Rio de Janeiro, Brazil: *Sex Transm Infect*; 2017. p. A27-A8.

5. Town K, Field N, Cole M, Harris S, Hughes G. What is the role of travel-associated sexual partnerships on the importation and spread of AMR *Neisseria gonorrhoeae* in the UK? International Society for Sexually Transmitted Disease Research STI & HIV World Congress. Rio de Janeiro, Brazil: *Sex Transm Infect*; 2017. p. A11-A2.

6. Town K, Field N, Harris S, Hughes G. A transdisciplinary approach to identify and delay the spread of antimicrobial resistance *Neisseria gonorrhoeae* using epidemiology, microbiology and genomics. UCL Populations and Lifelong Health Domain Symposium - The future health of the public: Towards transdisciplinary research. London, United Kingdom; 2017.

7. Centers for Disease Control and Prevention. Combatting the Threat of Antibiotic-Resistant Gonorrhea. 2018. https://www.cdc.gov/std/gonorrhea/arg/carb.htm (accessed 20/11/2018).

8. Barlow D, Phillips I. Gonrrhoea in women. *Lancet* 1978; 1(8067): 761-4.

9. Sherrard J, Barlow D. Gonorrhoea in men: clinical and diagnostic. *Genitourin Med* 1996; 72: 422-6.

10. Bignell C, Fitzgerald M. UK national guideline for the management of gonorrhoea in adults, 2011. *Int J STD AIDS* 2011; 22(10): 541-7.

11. Fifer H, Saunders J, Sunneta S, Sadiq ST, Fitzgerald M. 2018 UK national guidelines for the management of infection with *Neisseria gonorrhoeae* (draft). https://www.bashh.org/guidelines (accessed 20/11/2018).

12. Tsevat DG, Wiesenfeld HC, Parks C, Peipert JF. Sexually transmitted diseases and infertility. *Am J Obstet Gynecol* 2017; 216(1): 1-9.

13. Marri PR, Paniscus M, Weyand NJ, et al. Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. *PLoS One* 2010; 5(7): e11835.

14. Unemo M, Golparian D, Sanchez-Buso L, et al. The novel 2016 WHO *Neisseria gonorrhoeae* reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization. *J Antimicrob Chemother* 2016; 71(11): 3096-108.

15. Cehovin A, Lewis SB. Mobile genetic elements in *Neisseria gonorrhoeae*: movement for change. *Pathog Dis* 2017; 75(6).

16. Pachulec E, van der Does C. Conjugative plasmids of *Neisseria gonorrhoeae*. *PLoS One* 2010; 5(4): e9962.

17. Davidsen T, Rodland EA, Lagesen K, Seeberg E, Rognes T, Tonjum T. Biased distribution of DNA uptake sequences towards genome maintenance genes. *Nucleic Acids Res* 2004; 32(3): 1050-8.

18. Dillard JP. Genetic Manipulation of *Neisseria gonorrhoeae*. *Curr Protoc Microbiol* 2011; Chapter 4: Unit4A 2.

19. Goodman SD, Scocca JJ. Indentification and arrangement of the DNA sequence recognised in specific transformation of *Neisseria gonorrhoeae*. *Proc Natl Acad Sci USA* 1988; 85: 6982-6.

20. Graves JF, Biswas GD, Sparling PF. Sequence-specific DNA uptake in transformation of *Neisseria gonorrhoeae*. *J Bacteriol* 1982; 152(3): 1071-7.

21. Seth-Smith HM, Thomson NR. Whole-genome sequencing of bacterial sexually transmitted infections: implications for clinicians. *Curr Opin Infect Dis* 2013; 26(1): 90-8.

22. Walcher M, Skvoretz R, Montgomery-Fullerton M, Jonas V, Brentano S. Description of an unusual *Neisseria meningitidis* isolate containing and expressing *Neisseria gonorrhoeae*-Specific 16S rRNA gene sequences. *J Clin Microbiol* 2013; 51(10): 3199-206.

23. Schenck LP, Surette MG, Bowdish DM. Composition and immunological significance of the upper respiratory tract microbiota. *FEBS Lett* 2016; 590(21): 3705-20.

24. Weyand NJ. *Neisseria* models of infection and persistence in the upper respiratory tract. *Pathog Dis* 2017; 75(3).

25. Hughes G, Field N. The epidemiology of sexually transmitted infections in the UK: impact of behavior, services and interventions. *Future Microbiol* 2015; 10(1): 35-51.

26. Mohammed H, Blomquist P, Ogaz D, et al. 100 years of STIs in the UK: a review of national surveillance data. *Sex Transm Infect* 2018; 94(8): 553-8.

27. Nicoll A, Hughes G, Donnelly M, et al. Assessing the impact of national anti-HIV sexual health campaigns: trends in the transmission of HIV and other sexually transmitted infections in England. *Sex Transm Infect* 2001; 77: 242-7.

28. Public Health England. Sexually transmitted infections and screening for chlamydia in England, 2017. *Health Protection Report* 2018; 12(20).

29. Ogaz D, Furegato M, Nash S, et al. Exploring the drivers of the recent decline in gonorrhoea diagnoses in MSM attending GUM clinics in England. Fourth Joint Conference of BHIVA with BASHH. Edinburgh, United Kingdom: *HIV Med*; 2018. p. s5-s20.

30. Sonnenberg P, Clifton S, Beddows S, et al. Prevalence, risk factors, and uptake of interventions for sexually transmitted infections in Britain: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). *Lancet* 2013; 382(9907): 1795-806.

31. Town K, Furegato M, Field N, Hughes G. Estimating gonorrhoea prevalence in young heterosexual men and women attending community-based sexual health services to inform decisions on gonorrhoea testing. *Epidemiol Infect* 2017; 145(8): 1682-7.

32. Adimora AA, Schoenbach VJ. Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. *J Infect Dis* 2005; 191 Suppl 1: S115-22.

33. Doherty IA, Padian NS, Marlow C, Aral SO. Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *J Infect Dis* 2005; 191(Suppl 1): S42-54.

34. Gomez GB, Ward H, Garnett GP. Risk pathways for gonorrhea acquisition in sex workers: can we distinguish confounding from an exposure effect using a priori hypotheses? *J Infect Dis* 2014; 210 Suppl 2: S579-85.

35. Ward H. Prevention strategies for sexually transmitted infections: importance of sexual network structure and epidemic phase. *Sex Transm Infect* 2007; 83 Suppl 1: i43-9.

36. Wasserheit JN, Aral SO. The dynamic topology of sexually transmitted disease epidemics: implications for prevention strategies. *J Infect Dis* 1996; 174(Suppl 2): S201-13.

37. De P, Singh AE, Wong T, Yacoub W, Jolly AM. Sexual network analysis of a gonorrhoea outbreak. *Sex Transm Infect* 2004; 80(4): 280-5.

38. Fichtenberg CM, Muth SQ, Brown B, Padian NS, Glass TA, Ellen JM. Sexual network position and risk of sexually transmitted infections. *Sex Transm Infect* 2009; 85(7): 493-8.

39. Althaus CL, Turner KM, Mercer CH, et al. Effectiveness and cost-effectiveness of traditional and new partner notification technologies for curable sexually transmitted infections: observational study, systematic reviews and mathematical modelling. *Health Technol Assess* 2014; 18(2): 1-100, vii-viii.

40. Ferreira A, Young T, Mathews C, Zunza M, Low N. Strategies for partner notification for sexually transmitted infections, including HIV. *Cochrane Database Syst Rev* 2013; (10): CD002843.

41. Low N, Broutet N, Adu-Sarkodie Y, Barton P, Hossain M, Hawkes S. Global control of sexually transmitted infections. *Lancet* 2006; 368(9551): 2001-16.

42. Ward H, Bell G. Partner notification. *Medicine (Abingdon)* 2014; 42(6): 314-7.

43. Bell G, Ward H, Day S, et al. Partner notification for gonorrhoea: a comparative study with a provincial and a metropolitan UK clinic. *Sex Transm Infect* 1998; 74: 409-14.

44. Foster K, Cole M, Hotonu O, et al. How to do it: lessons identified from investigating and trying to control an outbreak of gonorrhoea in young heterosexual adults. *Sex Transm Infect* 2016; 92(5): 396-401.

45. Smolarchuck C, Wensley A, Padfield S, Fifer H, Lee A, Hughes G. Persistence of an outbreak of gonorrhoea with high-level resistance to azithromycin in England, November 2014-May 2018. *Euro Surveill* 2018; 23(23): 1800287.

46. Anderson RM, Garnett GP. Mathematical models of the transmission and control of sexually transmitted diseases. *Sex Transm Dis* 2000; 27(10): 636-43.

47. Eames KT, Keeling MJ. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc Natl Acad Sci U S A* 2002; 99(20): 13330-5.

48. Garnett GP. An introduction to mathematical models in sexually transmitted disease epidemiology. *Sex Transm Infect* 2002; 78(1): 7-12.

49. Ghani AC, Swinton J, Garnett GP. The role of sexual partnership networks in the epidemiology of gonorrhea. *Sex Transm Dis* 1997; 24(1): 45-56.

50. Heesterbeek H, Anderson RM, Andreasen V, et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science* 2015; 347(6227): aaa4339.

51. May RM, Anderson RM. Transmission dynamics of HIV infection. *Nature* 1987; 326: 137-42.

52. Garnett GP, Anderson RM. Contact tracing and the estimation of sexual mixing patterns: the epidemiology of gonococcal infections. *Sex Transm Dis* 1993; 20(4): 181-91.

53. Yorke JA, Hethcote HW, Nold A. Dynamics and control of the transmission of gonorrhea. *Sex Transm Dis* 1978; 5(2): 51-6.

54. Chow EPF, Camilleri S, Ward C, et al. Duration of gonorrhoea and chlamydia infection at the pharynx and rectum among men who have sex with men: a systematic review. *Sex Health* 2016; 13(3): 199-204.

55. Farley TA, Cohen DA, Elkins W. Asymptomatic sexually transmitted diseases: the case for screening. *Prev Med* 2003; 36(4): 502-9.

56. Turner K. Mathematical models of gonorrhoea and chlamydia: biology, behaviour and interventions: Imperial College London; 2004.

57. Turner KME, Garnett GP. The impact of the phase of an epidemic of sexually transmitted infection on the evolution of the organism. *Sex Transm Infect* 2002; 78(Suppl 1): i20-i30.

58. Hui B, Fairley CK, Chen M, et al. Oral and anal sex are key to sustaining gonorrhoea at endemic levels in MSM populations: a mathematical model. *Sex Transm Infect* 2015; 91(5): 365-9.

59. Zhang L, Regan DG, Chow EPF, et al. *Neisseria gonorrhoeae* transmission among men who have sex with men: an anatomical site-specific mathematical model evaluating the potential preventive impact of mouthwash. *Sex Transm Dis* 2017; 44(10): 586-92.

60. O'Brien A, Sherrard-Smith E, Sile B, Watts C, Simms I. Spatial clusters of gonorrhoea in England with particular reference to the outcome of partner notification: 2012 and 2013. *PLoS One* 2018; 13(4): e0195178.

61. Savage EJ, Mohammed H, Leong G, Duffell S, Hughes G. Improving surveillance of sexually transmitted infections using mandatory electronic clinical reporting: the genitourinary medicine clinic activity dataset, England, 2009-2013. *Euro Surveill* 2014; 19(48): 20981.

62. Thomas JC, Tucker MJ. The development and use of the concept of a sexually transmitted disease core. *J Infect Dis* 1996; 174(Suppl 2): S134-S42.

63. Aral SO, Hughes JP, Stoner B, et al. Sexual mixing patterns in the spread of gonococcal and chlamydial infections. *Am J Public Health* 1999; 89(6): 825-33.

64. Giguere K, Alary M. Targeting core groups for gonorrhoea control: feasibility and impact. *Sex Transm Infect* 2015; 91(4): 241-4.

65. Prah P, Hickson F, Bonell C, et al. Men who have sex with men in Great Britain: comparing methods and estimates from probability and convenience sample surveys. *Sex Transm Infect* 2016; 92(6): 455-63.

66. Mercer CH, Prah P, Field N, et al. The health and well-being of men who have sex with men (MSM) in Britain: Evidence from the third National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *BMC Public Health* 2016; 16: 525.

67. Malek R, Mitchell H, Furegato M, et al. Contribution of transmission in HIV-positive men who have sex with men to evolving epidemics of sexually transmitted infections in England: an analysis using multiple data sources, 2009-2013. *Euro Surveill* 2015; 20(15): 21093.

68. Aghaizu A, Wayal S, Nardone A, et al. Sexual behaviours, HIV testing, and the proportion of men at risk of transmitting and acquiring HIV in London, UK, 2000–13: a serial cross-sectional study. *Lancet HIV* 2016; 3(9): e431-e40.

69. Daskalopoulou M, Rodger AJ, Phillips AN, et al. Condomless sex in HIV-diagnosed men who have sex with men in the UK: prevalence, correlates, and implications for HIV transmission. *Sex Transm Infect* 2017; 0: 1-9.

70. Grewal R, Allen VG, Gardner S, et al. Serosorting and recreational drug use are risk factors for diagnosis of genital infection with chlamydia and gonorrhoea among HIV-positive men who have sex with men: results from a clinical cohort in Ontario, Canada. *Sex Transm Infect* 2017; 93(1): 71-5.

71. Hotton AL, Gratzer B, Mehta SD. Association between serosorting and bacterial sexually transmitted infection among HIV-negative men who have sex with men at an urban lesbian, gay, bisexual, and transgender health center. *Sex Transm Dis* 2012; 39(12): 959-64.

72. Kennedy CE, Bernard LJ, Muessig KE, et al. Serosorting and HIV/STI infection among HIV-negative MSM and transgender people: a systematic review and meta-analysis to inform WHO guidelines. *J Sex Transm Dis* 2013; 2013: 583627.

73. Lattimore S, Thornton A, Delpech V, Elford J. Changing patterns of sexual risk behavior among London gay men: 1998-2008. *Sex Transm Dis* 2011; 38(3): 221-9.

74. Paz-Bailey G, Mendoza MC, Finlayson T, et al. Trends in condom use among MSM in the United States: the role of antiretroviral therapy and seroadaptive strategies. *AIDS* 2016; 30(12): 1985-90.

75. Ronn M, White PJ, Hughes G, Ward H. Developing a conceptual framework of seroadaptive behaviors in HIV-diagnosed men who have sex with men. *J Infect Dis* 2014; 210 Suppl 2: S586-93.

76. Vallabhaneni S, Li X, Vittinghoff E, Donnell D, Pilcher CD, Buchbinder SP. Seroadaptive practices: association with HIV acquisition among HIV-negative men who have sex with men. *PLoS One* 2012; 7(10): e45718.

77. Ward H, Ronn M. Contribution of sexually transmitted infections to the sexual transmission of HIV. *Curr Opin HIV AIDS* 2010; 5(4): 305-10.

78. Furegato M, Chen Y, Mohammed H, Mercer C, Savage EJ, Hughes G. Examining the role of socioeconomic deprivation in ethnic differences in sexually transmitted infection diagnosis rates in England: evidence from surveillance data. *Epidemiol Infect* 2016; 144(15): 3253-62.

79. Low N, Sterne JA, Barlow D. Inequalities in rates of gonorrhoea and chlamydia betwen black ethnic groups in south east London: cross sectional study. *Sex Transm Infect* 2001; 77: 15-20.

80. Aicken C, Wayal S, Blomquist P, et al. Ethnic variations in the characteristics of sexual partnerships reported by behaviourally-heterosexual people attending sexual health clinics in England: do they explain ethnic differences in STI diagnoses? International Union against Sexually Transmitted Infections World and European Congress. Dublic, Republic of Ireland; 2018. p. 281.

81. Coyle RM, Miltz AR, Lampe FC, et al. Ethnicity and sexual risk in heterosexual people attending sexual health clinics in England: a cross-sectional, self-administered questionnaire study. *Sex Transm Infect* 2018; 0: 1-8.

82. Fenton KA, Mercer CH, McManus S, et al. Ethnic variations in sexual behaviour in Great Britain and risk of sexually transmitted infections: a probability survey. *Lancet* 2005; 365(9466): 1246-55.

83. Wayal S, Hughes G, Sonnenberg P, et al. Ethnic variations in sexual behaviours and sexual health markers: findings from the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *Lancet Public Health* 2017; 2(10): e458-e72.

84. Public Health England. Guidance for the detection of gonorrhoea in England, 2014. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/405293/170215_Gonorrhoea_testing_guidance_REVISED__2_.pdf (accessed 20/11/2018)

85. Clutterbuck D, Asboe D, Barber T, et al. 2016 United Kingdom national guideline on the sexual health care of men who have sex with men, 2018. https://www.bashhguidelines.org/media/1162/msm-2016.pdf (accessed 20/11/2018)

86. Palmer HM, Mallinson H, Wood RL, Herring AJ. Evaluation of the specificities of five DNA amplification methods for the detection of *Neisseria gonorrhoeae*. *J. Clin. Microbiol.* 2003; 41(2): 835-7.

87. Field N, Clifton S, Alexander S, et al. Confirmatory assays are essential when using molecular testing for *Neisseria gonorrhoeae* in low-prevalence settings: insights from the third National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *Sex Transm Infect* 2015; 91(5): 338-41.

88. Field N, Kennedy I, Folkard K, et al. Screening for gonorrhoea using samples collected through the English national chlamydia screening programme and risk of false positives: a national survey of local authorities. *BMJ Open* 2014; 4(10): e006067.

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

90. Van Slyke CJ, Wolcutt RR, Mahoney JF. The chemotherapy of gonococcal infections. *JAMA* 1941; 22+(5): 276-80.

91. Mahoney JF, Ferguson C, Buchholtz M, Van Slyke CJ. The use of penicillin sodium in the treatment of sulfonamide-resistant gonorrhea in men. A preliminary report. *J Am Vener Dis Assoc* 1943; 27: 525-8.

92. Fox KK, Thomas JC, Weiner DH, Davis RH, Sparling PF, Cohen MS. Longitudinal evaluation of serovar-specific immunity to *Neisseria gonorrhoeae*. *Am J Epidemiol* 1999; 149(4): 353-8.

93. Plummer FA, Chubb H, Simonsen JN, et al. Antibodies to opacity proteins (Opa) correlate with a reduced risk of gonococcal salpingitis. *J Clin Invest* 1994; 83: 1748-55.

94. Vincent LR, Jerse AE. Biological feasibility and importance of a gonorrhea vaccine for global public health. *Vaccine* 2018.

95. Wellcome Trust, Boston Consulting Group. Vaccines to tackle drug resistant infections: an evaluation of R&D opportunities, 2018. https://vaccinesforamr.org/read-the-report/ (accessed 20/11/2018)

96. Gottlieb SL, Johnston C. Future prospects for new vaccines against sexually transmitted infections. *Curr Opin Infect Dis* 2017; 30(1): 77-86.

97. Kuhdari P, Stefanati A, Lupi S, Valente N, Gabutti G. Meningococcal B vaccination: real-world experience and future perspectives. *Pathog Glob Health* 2016; 110(4-5): 148-56.

98. Petousis-Harris H, Paynter J, Morgan J, et al. Effectiveness of a group B outer membrane vesicle meningococcal vaccine against gonorrhoea in New Zealand: a retrospective case-control study. *Lancet* 2017; 390(10102): 1603-10.

99. Unemo M, Nicholas RA. Emergence of multidrug-resistant, extensively drug-resistant and untreatable gonorrhea. *Future Microbiol* 2012; 7(12): 1401-22.

100. European Society of Clinical Microbiology and Infectious Diseases. Clinical breakpoints. http://www.eucast.org/clinical_breakpoints/ (accessed 20/12/2018)

101. Tapsall JW. Antimicrobial resistance in *Neisseria gonorrhoeae*: World Health Organisation, 2001. http://apps.who.int/iris/handle/10665/66963 (accessed 20/11/2018)

102. World Health Organisation. WHO guidelines for the treatment of *Neisseria gonorrhoeae*, 2016. https://www.who.int/reproductivehealth/publications/rtis/gonorrhoea-treatment-guidelines/en/ (accessed 20/11/2018)

103. World Health Organisation. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics, 2017. https://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/ (accessed 20/11/2018)

104. Public Health England. Gonococcal resistance to antimicrobials surveillance programme (GRASP): protocol. 2014. https://www.gov.uk/government/publications/gonococcal-resistance-to-antimicrobials-surveillance-programme-grasp-protocol/gonococcal-resistance-to-antimicrobials-surveillance-programme-grasp-protocol (accessed 20/11/2018).

105. Health Protection Scotland. Surveillance report: Gonococcal antibiotic surveillance in Scotland (GASS): prevalence, patterns and trends in 2017, 2018. https://www.hps.scot.nhs.uk/resourcedocument.aspx?id=6722 (accessed 11/01/2019)

106. World Health Organisation. Baseline report on global sexually transmitted infection surveillance: monitoring gonococcal antimicrobial susceptibility, 2012. https://www.who.int/reproductivehealth/publications/rtis/9789241505895/en/ (accessed 20/11/2018)

107. Wi T, Lahra MM, Ndowa F, et al. Antimicrobial resistance in *Neisseria gonorrhoeae*: Global surveillance and a call for international collaborative action. *PLoS Med* 2017; 14(7): e1002344.

108. Spiteri G, Cole M, Unemo M, Hoffmann S, Ison C, van de Laar M. The European Gonococcal Antimicrobial Surveillance Programme (Euro-GASP)--a sentinel approach in the European Union (EU)/European Economic Area (EEA). *Sex Transm Infect* 2013; 89 Suppl 4: iv16-8.

109. Kirkcaldy RD, Zaidi A, Hook EW, et al. *Neisseria gonorrhoeae* antimicrobial resistance among men who have sex with men and men who have sex exclusively with women: the gonococcal isolate surveillance project, 2005-2010. *Ann. Intern. Med.* 2013; 158: 321-8.

110. Lahra MM, Enriquez R, George CRR. Australian gonococcal surveillance programme annual report, 2017, 2017. http://www.health.gov.au/internet/main/publishing.nsf/content/cda-pubs-annlrpt-gonoanrep.htm (accessed 20/11/2018)

111. Yin YP, Han Y, Dai XQ, et al. Susceptibility of *Neisseria gonorrhoeae* to azithromycin and ceftriaxone in China: A retrospective study of national surveillance data from 2013 to 2016. *PLoS Med* 2018; 15(2): e1002499.

112. Cole MJ, Spiteri G, Jacobsson S, et al. Overall low extended-spectrum cephalosporin resistance but high azithromycin resistance in *Neisseria gonorrhoeae* in 24 European Countries, 2015. *BMC Infect Dis* 2017; 17(1): 617.

113. Field N, Cohen T, Struelens MJ, et al. Strengthening the reporting of molecular epidemiology for infectious diseases (STROME-ID): an extension of the STROBE statement. *Lancet Infect Dis* 2014; 14(4): 341-52.

114. Lees JA, Kendall M, Parkhill J, Colijn C, Bentley SD, Harris SR. Evaluation of phylogenetic reconstruction methods using bacterial whole genomes: a simulation based study. *Wellcome Open Res* 2018; 3: 33.

115. Yang Z, Rannala B. Molecular phylogenetics: principles and practice. *Nat Rev Genet* 2012; 13(5): 303-14.

116. Rieux A, Balloux F. Inferences from tip-calibrated phylogenies: a review and a practical guide. *Mol Ecol* 2016; 25(9): 1911-24.

117. Croucher NJ, Didelot X. The application of genomics to tracing bacterial pathogen transmission. *Curr Opin Microbiol* 2015; 23: 62-7.

118. De Silva D, Peters J, Cole K, et al. Whole-genome sequencing to determine transmission of *Neisseria gonorrhoeae*: an observational study. *Lancet Infect Dis* 2016; 16(11): 1295-303.

119. Didelot X, Dordel J, Whittles LK, et al. Genomic analysis and comparison of two gonorrhea outbreaks. *mBio* 2016; 7(3).

120. Ezewudo MN, Joseph SJ, Castillo-Ramirez S, et al. Population structure of *Neisseria gonorrhoeae* based on whole genome data and its relationship with antibiotic resistance. *PeerJ* 2015; 3: e806.

121. Grad YH, Kirkcaldy RD, Trees D, et al. Genomic epidemiology of *Neisseria gonorrhoeae* with reduced susceptibility to cefixime in the USA: A retrospective observational study. *Lancet Infect Dis* 2014; 14(3): 220-6.

122. Eldholm V, Balloux F. Antimicrobial resistance in *Mycobacterium tuberculosis*: The Odd One Out. *Trends Microbiol* 2016; 24(8): 637-48.

123. Croucher NJ, Harris SR, Grad YH, Hanage WP. Bacterial genomes in epidemiology--present and future. *Philos Trans R Soc Lond B Biol Sci* 2013; 368(1614): 20120202.

124. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015; 43(3): e15.

125. Hedge J, Wilson DJ. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *MBio* 2014; 5(6): e02158.

126. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; 74(12): 5463-7.

127. Martin IMC, Ison CA, Aanensen DM, Fenton KA, Spratt BG. Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. *J Infect Dis* 2004; 189(8): 1497-505.

128. Maiden MC, Bygraves JA, Feil E, et al. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998; 95: 3140-5.

129. Unemo M, Dillon JA. Review and international recommendation of methods for typing *Neisseria gonorrhoeae* isolates and their implications for improved knowledge of gonococcal epidemiology, treatment, and biology. *Clin Microbiol Rev* 2011; 24(3): 447-58.

130. Palmer HM, Young H, Graham C, Dave J. Prediction of antibiotic resistance using *Neisseria gonorrhoeae* multi-antigen sequence typing. *Sex Transm Infect* 2008; 84(4): 280-4.

131. Illumina. An introduction to next-generation sequencing technology, 2018. https://www.illumina.com/science/technology/next-generation-sequencing.html (accessed 20/11/2018)

132. Abrams AJ, Kirkcaldy RD, Pettus K, Fox JL, Kubin G, Trees DL. 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* 2017; 44(8): 492-4.

133. Al Suwayyid BA, Coombs GW, Speers DJ, Pearson J, Wise MJ, Kahler CM. Genomic epidemiology and population structure of *Neisseria gonorrhoeae* from remote highly endemic Western Australian populations. *BMC Genomics* 2018; 19(1): 165.

134. Costa-Lourenco APR, Abrams JA, Barros dos Santos KT, et al. Reduced susceptibility to cefixime but not ceftriaxone: an uncertain perspective for the treatment of gonorrhoea in Brazil. *Int J Antimicrob Agents* 2017; 49(4): 515-6.

135. Demczuk W, Lynch T, Martin I, et al. Whole-genome phylogenomic heterogeneity of *Neisseria gonorrhoeae* isolates with decreased cephalosporin susceptibility collected in Canada between 1989 and 2013. *J Clin Microbiol* 2015; 53(1): 191-200.

136. Demczuk W, Martin I, Peterson S, et al. Genomic epidemiology and molecular resistance mechanisms of azithromycin-resistant *Neisseria gonorrhoeae* in Canada from 1997 to 2014. *J Clin Microbiol* 2016; 54(5): 1304-13.

137. Demczuk W, Sidhu S, Unemo M, et al. *Neisseria gonorrhoeae* sequence typing for antimicrobial resistance, a novel antimicrobial resistance multilocus typing scheme for tracking global dissemination of *N. gonorrhoeae* Strains. *J Clin Microbiol* 2017; 55(5): 1454-68.

138. Eyre DW, De Silva D, Cole K, et al. WGS to predict antibiotic MICs for *Neisseria gonorrhoeae*. *J Antimicrob Chemother* 2017; 72(7): 1937-47.

139. Eyre DW, Sanderson ND, Lord E, et al. Gonorrhoea treatment failure caused by a *Neisseria gonorrhoeae* strain with combined ceftriaxone and high-level azithromycin resistance, England, February 2018. *Euro Surveill* 2018; 23(27).

140. Fifer H, Cole M, Hughes G, et al. Sustained transmission of high-level azithromycin-resistant *Neisseria gonorrhoeae* in England: an observational study. *Lancet Infect Dis* 2018; 18(5): 573-81.

141. Grad YH, Harris SR, Kirkcaldy RD, et al. Genomic epidemiology of gonococcal resistance to extended-spectrum cephalosporins, macrolides, and fluoroquinolones in the United States, 2000-2013. *J Infect Dis* 2016; 214(10): 1579-87.

142. Harris SR, Cole MJ, Spiteri G, et al. Public health surveillance of multidrug-resistant clones of *Neisseria gonorrhoeae* in Europe: a genomic survey. *Lancet Infect Dis* 2018; 18(7): 758-68.

143. Harrison OB, Clemence M, Dillard JP, et al. Genomic analyses of *Neisseria gonorrhoeae* reveal an association of the gonococcal genetic island with antimicrobial resistance. *J Infect* 2016; 73(6): 578-87.

144. Jacobsson S, Golparian D, Cole M, 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* 2016; 71(11): 3109-16.

145. Johnson SR, Grad Y, Abrams AJ, Pettus K, Trees DL. Use of whole-genome sequencing data to analyze 23S rRNA-mediated azithromycin resistance. *Int J Antimicrob Agents* 2017; 49(2): 252-4.

146. Lahra MM, Martin I, Demczuk W, et al. Cooperative recognition of internationally disseminated ceftriaxone-resistant *Neisseria gonorrhoeae* strain. *Emerg Infect Dis* 2018; 24(4).

147. Mac Aogain M, Fennelly N, Walsh A, et al. Fourteen Draft Genome Sequences for the First Reported Cases of Azithromycin-Resistant Neisseria gonorrhoeae in Ireland. *Genome Announc* 2017; 5(23).

148. Papp JR, Abrams AJ, Nash E, et al. Azithromycin resistance and decreased ceftriaxone susceptibility in *Neisseria gonorrhoeae*, Hawaii, USA. *Emerg Infect Dis* 2017; 23(5): 830-2.

149. Peterson SW, Martin I, Demczuk W, et al. Molecular assay for detection of ciprofloxacin resistance in *Neisseria gonorrhoeae* isolates from cultures and clinical nucleic acid amplification test specimens. *J Clin Microbiol* 2015; 53(11): 3606-8.

150. Peterson SW, Martin I, Demczuk W, et al. Molecular assay for detection of genetic markers associated with decreased susceptibility to cephalosporins in *Neisseria gonorrhoeae*. *J Clin Microbiol* 2015; 53(7): 2042-8.

151. Ryan L, Golparian D, Fennelly N, 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* 2018: 1-12.

152. Vidyaprakash E, Abrams AJ, Shafer WM, Trees DL. 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* 2017; 61(4).

153. Whiley DM, Jennison A, Pearson J, Lahra MM. Genetic characterisation of *Neisseria gonorrhoeae* resistant to both ceftriaxone and azithromycin. *Lancet Infect Dis* 2018; 18(7): 717-8.

154. Whiley DM, Trembizki E, Buckley C, et al. Molecular antimicrobial resistance surveillance for *Neisseria gonorrhoeae*, Northern Territory, Australia. *Emerg Infect Dis* 2017; 23(9): 1478-85.

155. Yahara K, Nakayama S-I, Shimuta K, et al. Genomic surveillance of *Neisseria gonorrhoeae* to investigate the distribution and evolution of antimicrobial-resistance determinants and lineages. *Microb Genom* 2018; 4.

156. Allan-Blitz LT, Humphries RM, Hemarajata P, et al. Implementation of a rapid genotypic assay to promote targeted ciprofloxacin therapy of *Neisseria gonorrhoeae* in a large health system. *Clin Infect Dis* 2017; 64(9): 1268-70.

157. Dona V, Low N, Golparian D, Unemo M. Recent advances in the development and use of molecular tests to predict antimicrobial resistance in *Neisseria gonorrhoeae. Expert Rev Mol Diagn* 2017; 17(9): 845-59.

158. Dona V, Smid JH, Kasraian S, et al. Mismatch Amplification Mutation Assay-Based Real-Time PCR for Rapid Detection of Neisseria gonorrhoeae and Antimicrobial Resistance Determinants in Clinical Specimens. *J Clin Microbiol* 2018; 56(9).

159. Low N, Unemo M. Molecular tests for the detection of antimicrobial resistant *Neisseria gonorrhoeae*: when, where, and how to use? *Curr Opin Infect Dis* 2016; 29(1): 45-51.

160. Wong LK, Hemarajata P, Soge OO, Humphries RM, Klausner JD. Real-Time PCR targeting the *penA mosaic XXXIV* type for prediction of extended-spectrum-cephalosporin susceptibility in clinical *Neisseria gonorrhoeae* isolates. *Antimicrob Agents Chemother* 2017; 61(11).

161. Buckley C, Forde BM, Trembizki E, Lahra MM, Beatson SA, Whiley DM. Use of whole genome sequencing to investigate an increase in *Neisseria gonorrhoeae* infection among women in urban areas of Australia. *Sci Rep* 2018; 8(1): 1503.

162. Chisholm SA, Wilson J, Alexander S, et al. An outbreak of high-level azithromycin resistant *Neisseria gonorrhoeae* in England. *Sex Trans Infect* 2016; 92(5): 365-7.

163. de Curraize C, Kumanski S, Micaelo M, et al. Ceftriaxone-resistant *Neisseria gonorrhoeae* isolates (2010 to 2014) in France characterized by using whole-genome sequencing. *Antimicrob Agents Chemother* 2016; 60(11): 6962-4.

164. Fifer H, Natarajan U, Jones L, et al. Failure of dual antimicrobial therapy in treatment of gonorrhea. *N Engl J Med* 2016; 374(25): 2504-6.

165. Kwong JC, Chow EPF, Stevens K, et al. Whole-genome sequencing reveals transmission of gonococcal antibiotic resistance among men who have sex with men: an observational study. *Sex Transm Infect* 2018; 94(2): 151-7.

166. Lee RS, Seemann T, Heffernan H, et al. Genomic epidemiology and antimicrobial resistance of *Neisseria gonorrhoeae* in New Zealand. *J Antimicrob Chemother* 2018; 73(2): 353-64.

167. Peters J, Cresswell F, Amor L, et al. Whole genome sequencing of *Neisseria gonorrhoeae* reveals transmission clusters involving patients of mixed HIV serostatus. *Sex Transm Infect* 2018; 94(2): 138-43.

168. Sánchez-Busó L, Golparian D, Corander J, et al. Antimicrobial exposure in sexual networks drives divergent evolution in modern gonococci. *bioRxiv* 2018.

169. European Centre for Disease Prevention and Control. Molecular typing of *Neisseria gonorrhoeae* - a study of 2013 isolates: technical report, 2017. https://ecdc.europa.eu/en/publications-data/molecular-typing-neisseria-gonorrhoeae-study-2013-isolates (accessed 20/11/2018)

170. Town K, Bolt H, Harris S, Field N, Hughes G. A systematic review of *Neisseria gonorrhoeae* (NG) DNA typing techniques to inform public health understanding of NG transmission in sexual networks and the risk factors for NG acquisition. PROSPERO; 2016.

171. Web of Science. http://wok.mimas.ac.uk/ (accessed 20/11/2018)

172. Scopus. https://www.scopus.com/ (accessed 20/11/2018)

173. MedLine. https://www.nlm.nih.gov/bsd/pmresources.html (accessed 20/11/2018)

174. EMBASE. https://www.elsevier.com/solutions/embase-biomedical-research (accessed 20/11/2018)

175. Cochrane Library. https://www.cochranelibrary.com/ (accessed 20/11/2018)

176. EndNote. X8 ed. https://endnote.com/ (accessed 20/11/2018)

177. Abu-Rajab K, Palmer HM, Scoular A, et al. To what extent does *Neisseria gonorrhoeae* multiantigen sequence typing of gonococcal isolates support information derived from patient interviews? *Int J STD AIDS* 2009; 20(6): 414-7.

178. Bernstein KT, Marcus JL, Barry PM, et al. Characteristics of males infected with common neisseria gonorrhoeae sequence types in the gonococcal isolate surveillance project, san francisco, California, 2009. *Am J Epidemiol* 2013; 178(8): 1289-95.

179. Bilek N, Martin IM, Bell G, Kinghorn GR, Ison CA, Spratt BG. Concordance between *Neisseria gonorrhoeae* genotypes recovered from known sexual contacts. *J Clin Microbiol* 2007; 45(11): 3564-7.

180. Carannante A, Latini A, Cusini M, et al. Update on antimicrobial susceptibility and genotype of *Neisseria gonorrhoeae* isolated in Italy. *Diagn Microbiol Infect Dis* 2012; 72(3): 288-90.

181. Carannante A, Renna G, Conte ID, et al. Changing antimicrobial resistance profiles among *Neisseria gonorrhoeae* isolates in Italy, 2003 to 2012. *Antimicrob. Agents Chemother* 2014; 58(10): 5871-6.

182. Chen CC, Yen MY, Wong WW, et al. Tracing subsequent dissemination of a cluster of gonococcal infections caused by an ST1407-related clone harbouring mosaic penA alleles in Taiwan. *J Antimicrob Chemother* 2013; 68(7): 1567-71.

183. Chen SC, Yin YP, Dai XQ, Unemo M, Chen XS. First nationwide study regarding ceftriaxone resistance and molecular epidemiology of *Neisseria gonorrhoeae* in China. *J Antimicrob Chemother* 2016; 71(1): 92-9.

184. Cheng CW, Li LH, Su CY, Li SY, Yen MY. Changes in the six most common sequence types of *Neisseria gonorrhoeae*, including ST4378, identified by surveillance of antimicrobial resistance in northern Taiwan from 2006 to 2013. *J Microbiol Immunol Infect* 2016; 49(5): 708-716.

185. Chisholm SA, Alexander S, Desouza-Thomas L, et al. Emergence of a *Neisseria gonorrhoeae* clone showing decreased susceptibility to cefixime in England and Wales. *J Antimicrob Chemother* 2011; 66(11): 2509-12.

186. Chisholm SA, Neal TJ, Alawattegama AB, Birley HD, Howe RA, Ison CA. Emergence of high-level azithromycin resistance in *Neisseria gonorrhoeae* in England and Wales. *J Antimicrob Chemother* 2009; 64(2): 353-8.

187. Chisholm SA, Unemo M, Quaye N, et al. Molecular epidemiological typing within the European Gonococcal Antimicrobial Resistance Surveillance Programme reveals predominance of a multidrugresistant clone. *Euro Surveill* 2013; 18(3).

188. Choudhury B, Risley CL, Ghani AC, et al. Identification of individuals with gonorrhoea within sexual networks: a population-based study. *Lancet* 2006; 368(9530): 139-46.

189. Cole MJ, Thomas DR, Chisholm SA, et al. Molecular epidemiology of gonorrhoea in Wales (UK). *Sex Trans Infect* 2013; 89(3): 267-72.

190. Fernando I, Palmer HM, Young H. Characteristics of patients infected with common *Neisseria gonorrhoeae* NG-MAST sequence type strains presenting at the Edinburgh genitourinary medicine clinic. *Sex Trans Infect* 2009; 85(6): 443-6.

191. Florindo C, Pereira R, Boura M, et al. Genotypes and antimicrobial-resistant phenotypes of *Neisseria gonorrhoeae* in Portugal (2004-2009). *Sex Trans Infect* 2010; 86(6): 449-53.

192. Hjelmevoll SO, Golparian D, Dedi L, et al. Phenotypic and genotypic properties of *Neisseria gonorrhoeae* isolates in Norway in 2009: Antimicrobial resistance warrants an immediate change in national management guidelines. *Eur J Clin Microbiol Infect Dis* 2012; 31(6): 1181-6.

193. Horn NN, Kresken M, Korber-Irrgang B, et al. Antimicrobial susceptibility and molecular epidemiology of *Neisseria gonorrhoeae* in Germany. *Int J Med Microbiol* 2014; 304(5-6): 586-91.

194. Ison CA, Town K, Obi C, et al. Decreased susceptibility to cephalosporins among gonococci: Data from the Gonococcal Resistance to Antimicrobials Surveillance Programme (GRASP) in England and Wales, 2007-2011. *Lancet Infect Dis* 2013; 13(9): 762-8.

195. Jeverica S, Golparian D, Maticic M, Potocnik M, Mlakar B, Unemo M. Phenotypic and molecular characterization of *Neisseria gonorrhoeae* isolates from Slovenia, 2006-12: Rise and fall of the multidrug-resistant NG-MAST genogroup 1407 clone? *J Antimicrob Chemother* 2014; 69(6): 1517-25.

196. Lahra MM, Ward A, Trembizki E, et al. Treatment guidelines after an outbreak of azithromycin-resistant *Neisseria gonorrhoeae* in South Australia. *Lancet Infect Dis* 2017; 17(2): 133-4.

197. Lundback D, Fredlund H, Berglund T, Wretlind B, Unemo M. Molecular epidemiology of *Neisseria gonorrhoeae*- identification of the first presumed Swedish transmission chain of an azithromycin-resistant strain. *Apmis* 2006; 114(1): 67-71.

198. Martin IMC, Ison CA, Aanensen DM, Fenton KA, Spratt BG. Changing epidemiologic profile of quinolone-resistant *Neisseria gonorrhoeae* in London. *J Infect Dis* 2005; 192(7): 1191-5.

199. Monfort L, Caro V, Devaux Z, Delannoy AS, Brisse S, Sednaoui P. First *Neisseria gonorrhoeae* genotyping analysis in France: Identification of a strain cluster with reduced susceptibility to ceftriaxone. *J Clin Microbiol* 2009; 47(11): 3540-5.

200. Ni C, Xue J, Zhang C, Zhou H, van der Veen S. High prevalence of *Neisseria gonorrhoeae* with high-level resistance to azithromycin in Hangzhou, China. *J Antimicrob Chemother* 2016; 71(8): 2355-7.

201. Ota KV, Ng LK, Melano RG, et al. Identification of sexual networks through molecular typing of quinolone-resistant *Neisseria gonorrhoeae* in Ontario, Canada. *Sex Trans Dis* 2011; 38(9): 811-4.

202. Palmer HM, Young H. Dramatic increase in a single genotype of TRNG ciprofloxacin-resistant *Neisseria gonorrhoeae* isolates in men who have sex with men. *Int J STD AIDS* 2006; 17(4): 254-6.

203. Serra-Pladevall J, Barberá MJ, Callarisa AE, Bartolomé-Comas R, Andreu A. Differences in *Neisseria gonorrhoeae* population structure and antimicrobial resistance pattern between men who have sex with men and heterosexuals. *Epidemiol Infect* 2017; 145(2): 379-85.

204. Singh AE, Gratrix J, Read R, et al. *Neisseria gonorrhoeae* multiantigen sequence typing is beneficial in further characterizing gonococcal populations in Alberta, Canada. *Sex Trans Dis* 2013; 40(9): 744-50.

205. Starnino S, Dal Conte I, Matteelli A, et al. Trend of ciprofloxacin resistance in *Neisseria gonorrhoeae* strains isolated in Italy and analysis of the molecular determinants. *Diagn Microbiol Infect Dis* 2010; 67(4): 350-4.

206. Starnino S, Stefanelli P. Azithromycin-resistant *Neisseria gonorrhoeae* strains recently isolated in Italy. *J Antimicrob Chemother* 2009; 63(6): 1200-4.

207. Starnino S, Suligoi B, Regine V, et al. Phenotypic and genotypic characterization of *Neisseria gonorrhoeae* in parts of Italy: detection of a multiresistant cluster circulating in a heterosexual network *Clin Microbiol Infect* 2008; 14(10): 949-54.

208. Stevens K, Zaia A, Tawil S, et al. *Neisseria gonorrhoeae* isolates with high-level resistance to azithromycin in Australia. *J Antimicrob Chemother* 2015; 70(4): 1267-8.

209. Unemo M, Sjostrand A, Akhras M, et al. Molecular characterization of *Neisseria gonorrhoeae* identifies transmission and resistance of one ciprofloxacin-resistant strain. *Apmis* 2007; 115(3): 231-40.

210. Wong WW, Huang CT, Li LH, Chiang CC, Chen BD, Li SY. Molecular epidemiological identification of *Neisseria gonorrhoeae* clonal clusters with distinct susceptibility profiles associated with specific groups at high risk of contracting human immunodeficiency virus and syphilis. *J Clin Microbiol* 2008; 46(12): 3931-4.

211. Yuan LF, Yin YP, Dai XQ, et al. Resistance to azithromycin of *Neisseria gonorrhoeae* isolates from 2 cities in China. *Sex Trans Dis* 2011; 38(8): 764-8.

212. Palmer HM, Young H, Martin IMC, Ison CA, Spratt BG. The epidemiology of ciprofloxacin resistant isolates of *Neisseria gonorrhoeae* in Scotland 2002: A comparison of phenolypic and genotypic analysis. *Sex Trans Infect* 2005; 81(5): 403-7.

213. Perez-Losada M, Crandall KA, Bash MC, Dan M, Zenilman J, Viscidi RP. Distinguishing importation from diversification of quinolone-resistant *Neisseria gonorrhoeae* by molecular evolutionary analysis. *BMC Evol Biol* 2007; 7 (84).

214. Trembizki E, Wand H, Donovan B, et al. The Molecular Epidemiology and Antimicrobial Resistance of *Neisseria gonorrhoeae* in Australia: A Nationwide Cross-Sectional Study, 2012. *Clin Infect Dis* 2016; 63(12): 1591-8.

215. Risley CL, Ward H, Choudhury B, et al. Geographical and demographic clustering of gonorrhoea in London. *Sex Trans Infect* 2007; 83(6): 481-7.

216. Paine TC, Fenton KA, Herring A, et al. GRASP: a new national sentinel surveillance initiative for monitoring gonococcal antimicrobial resistance in England and Wales. *Sex Transm Infect* 2001; 77: 398-400.

217. Hughes G, Nichols T, Ison CA. Estimating the prevalence of gonococcal resistance to antimicrobials in England and Wales. *Sex Transm Infect* 2011; 87(6): 526-31.

218. Public Health England. Surveillance of antimicrobial resistance in *Neisseria gonorrhoeae* in England and Wales: Key findings from the Gonococcal Resistance to Antimicrobials Surveillance Programme (GRASP), 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746261/GRASP_2017_report.pdf (accessed 11/01/2019)

219. Mohammed H, Ison CA, Obi C, et al. Frequency and correlates of culture-positive infection with *Neisseria gonorrhoeae* in England: a review of sentinel surveillance data. *Sex Transm Infect* 2015; 91(4): 287-93.

220. Wellcome Trust Sanger Institute. Illumina High Throughput Scientific Operations. https://www.sanger.ac.uk/science/groups/illumina-high-throughput (accessed 06/01/2019)

221. Ison CA. Laboratory methods in genitourinary medicine: Methods of diagnosing gonorrhoea. *Genito Med* 1990; 66: 453-9.

222. Sanders ER. Aseptic laboratory techniques: plating methods. *J Vis Exp* 2012; (63): e3064.

223. Ng LK, Martin IE. The laboratory diagnosis of *Neisseria gonorrhoeae*. *Can J Infect Dis Med Microbiol* 2005; 16(1): 15-25.

224. Cowan and Steel. Manual for the Identification of Medical Bacteria. 3rd ed: Cambridge University Press; 1993.

225. Luys G. A Text-Book on Gonorrhea and it's Complications. New York: William Wood & Company; 1913.

226. Qiagen. QIAsymphony DNA Handbook; 2010. https://www.qiagen.com/us/resources/download.aspx?id=e5247873-ee4c-4cf4-98ab-0239a3a5dc8e&lang=en (accessed 20/11/2018)

227. Illumina. Specification Sheet: Sequencing - HiSeq X Series of Sequencing Systems. 2018. https://www.illumina.com/documents/products/datasheets/datasheet-hiseq-x-ten.pdf (accessed 20/11/2018)

228. NCBI. Genome Assembly and Annotation report: *Neisseria gonorrhoeae* FA1090. https://www.ncbi.nlm.nih.gov/genome/864?genome_assembly_id=300414 (accessed 06/01/2019).

229. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013; 1303(3997).

230. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25(16): 2078-9.

231. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20(9): 1297-303.

232. Government of the United Kingdom. National Health Service, England and Wales: The Health Service (Control of Patient Information) Regulations. In: Government of the United Kingdom, editor. 1438. http://www.legislation.gov.uk/uksi/2002/1438/contents/made; 2002.

233. European Genome-Phenome Archive. https://ega-archive.org/ (accessed 06/01/2019)

234. UNAIDS. UNAIDS Terminology Guidelines. 2015. www.unaids.org (accessed 06/01/2019)

235. Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd ed. Malden, Mass.: Wiley-Blackwell; 2003.

236. Page AJ, De Silva N, Hunt M, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb Genom* 2016; 2(8): e000083.

237. Illumina. Technical Note: Sequencing - Quality Scores for Next-Generation Sequencing. 2011. https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf (accessed 20/11/2018)

238. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 2014; 15(3): R46.

239. NCBI. RefSeq. https://www.ncbi.nlm.nih.gov/refseq/ (accessed 20/11/2018)

240. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; 30(9): 1312-3.

241. Piekarowicz A, Klyz A, Majchrzak M, Adamczyk-Poplawska M, Maugel TK, Stein DC. Characterization of the dsDNA prophage sequences in the genome of *Neisseria gonorrhoeae* and visualization of productive bacteriophage. *BMC Microbiol* 2007; 7: 66.

242. Rambaut. FigTree. 1.4.3 ed; 2012. http://tree.bio.ed.ac.uk/software/figtree/ (accessed 20/11/2018)

243. Lemoine F, Domelevo Entfellner JB, Wilkinson E, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 2018; 556(7702): 452-6.

244. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* 2017.

245. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010; 27(2): 221-4.

246. Centers for Disease Control and Prevention. MicrobeTrace. GitHub. https://github.com/CDCgov/MicrobeTRACE/wiki (accessed 20/11/2018)

247. Goldbaum G, Perdue T, Wolitski R, et al. Differences in risk behavior and sources of AIDS information among gay, bisexual, and straight-identified men who have sex with men. *AIDS Behav* 1998; 2(1): 13-21.

248. Newman CE, Persson A, Manolas P, et al. "So Much Is at Stake": Professional views on engaging heterosexually identified men who have sex with men with sexual health care in Australia. *Sex Res Soc Policy* 2018; 15: 302-11.

249. Persson A, Newman CE, Manolas P, et al. Challenging perceptions of “straight”: heterosexual men who have sex with men and the cultural politics of sexual identity categories. *Men Masculinities* 2017: 1-22.

250. Reback CJ, Larkins S. HIV risk behaviors among a sample of heterosexually identified men who occasionally have sex with another male and/or a transwoman. *J Sex Res* 2013; 50(2): 151-63.

251. Carrillo H, Hoffman A. From MSM to heteroflexibilities: Non-exclusive straight male identities and their implications for HIV prevention and health promotion. *Glob Public Health* 2016; 11: 7-8.

252. Pathela P, Hajat A, Schillinger J, Blank S, Randall S, Mostashari F. Discordance between sexual behavior and self-reported sexual identity: a population-based survey of New York City men. *Ann Intern Med* 2006; 145: 416-25.

253. Mercer CH, Copas AJ, Sonnenberg P, et al. Who has sex with whom? Characteristics of heterosexual partnerships reported in a national probability survey and implications for STI risk. *Int J Epidemiol* 2009; 38(1): 206-14.

254. Prah P, Copas AJ, Mercer CH, Nardone A, Johnson AM. Patterns of sexual mixing with respect to social, health and sexual characteristics among heterosexual couples in England: analyses of probability sample survey data. *Epidemiol Infect* 2015; 143(7): 1500-10.

255. Smid JH, Garcia V, Low N, Mercer CH, Althaus CL. Age difference between heterosexual partners in Britain: Implications for the spread of *Chlamydia trachomatis*. *Epidemics* 2018; 24: 60-6.

256. Ragonnet-Cronin M, Hué S, Hodcroft EB, et al. Non-disclosed men who have sex with men in UK HIV transmission networks: phylogenetic analysis of surveillance data. *Lancet HIV* 2018; 5(6): e309-e16.

257. Brook G, Bacon L, Evans C, et al. 2013 UK national guideline for consultations requiring sexual history taking. Clinical Effectiveness Group British Association for Sexual Health and HIV. *Int J STD AIDS* 2014; 25(6): 391-404.

258. Shankar AG, Mandal S, Ijaz S. An outbreak of hepatitis B in men who have sex with men but identify as heterosexual. *Sex Transm Infect* 2016; 92(3): 227.

259. Hue S, Brown AE, Ragonnet-Cronin M, et al. Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. *AIDS* 2014; 28(13): 1967-75.

260. Geary RS, Tanton C, Erens B, et al. Sexual identity, attraction and behaviour in Britain: The implications of using different dimensions of sexual orientation to estimate the size of sexual minority populations and inform public health interventions. *PLoS One* 2018; 13(1): e0189607.

261. British Association for Sexual Health and HIV. 2017 interim update of the 2015 BASHH national guidelines for the management of the viral hepatitides. 2017. https://www.bashhguidelines.org/media/1161/viral-hepatitides-2017-update-18-12-17.pdf (accessed 08/01/2019)

262. British Association for Sexual Health and HIV. 2015 BASHH CEG guidance on tests for sexually transmitted infections. 2015. https://www.bashhguidelines.org/media/1084/sti-testing-tables-2015-dec-update-4.pdf (accessed 08/01/2019)

263. Public Health England. Towards elimination of HIV transmission, AIDS and HIV-related deaths in the UK. 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/675809/Towards_elimination_of_HIV_transmission_AIDS_and_HIV_related_deaths_in_the_UK.pdf (08/01/2019)

264. Traeger MW, Schroeder SE, Wright EJ, et al. Effects of ore-exposure prophylaxis for the prevention of Human Immunodeficiency Virus infection on sexual risk behavior in men who have sex with men: a systematic review and meta-analysis. *Clin Infect Dis* 2018; 67(5): 676-86.

265. Gafos M, Horne R, Nutland W, et al. The context of sexual risk behaviour among men who have sex with men seeking PrEP, and the impact of PrEP on sexual behaviour. *AIDS Behav* 2018.

266. Bavinton BR, Holt M, Grulich AE, Brown G, Zablotska IB, Prestage GP. Willingness to act upon beliefs about 'Treatment as Prevention' among Australian gay and bisexual Men. *PLoS One* 2016; 11(1): e0145847.

267. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007; 7: 214.

268. To TH, Jung M, Lycett S, Gascuel O. Fast dating using least-squares criteria and algorithms. *Syst Biol* 2016; 65(1): 82-97.

269. Hunt M, Mather AE, Sanchez-Buso L, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017; 3(10): e000131.

270. Centre for Genomic Pathogen Surveillance. Pathogenwatch: a global platform for genomic surveillance. https://pathogen.watch/ (accessed 20/11/2019)

271. Ohnishi M, Watanabe Y, Ono E, et al. Spread of a chromosomal cefixime-resistant *penA* gene among different *Neisseria gonorrhoeae* lineages. *Antimicrob Agents Chemother* 2010; 54(3): 1060-7.

272. Kenyon CR. Association between intensity of STI screening and development of antimicrobial resistance in *N. gonorrhoeae* in 12 cities in the USA: An ecological study. *F1000Research* 2018; 7: 1237.

273. Fingerhuth SM, Bonhoeffer S, Low N, Althaus CL. Antibiotic-resistant *Neisseria gonorrhoeae* spread faster with more treatment, not more sexual partners. *PLoS Pathog* 2016; 12(5): e1005611.

274. Xiridou M, Soetens LC, Koedijk FDH, Van Der Sande MAB, Wallinga J. Public health measures to control the spread of antimicrobial resistance in *Neisseria gonorrhoeae* in men who have sex with men. *Epidemiol Infect* 2015; 143: 1575-84.

275. Vincent LR, Kerr SR, Tan Y, et al. In Vivo-Selected Compensatory mutations restore the fitness cost of mosaic *penA* alleles that confer ceftriaxone resistance in *Neisseria gonorrhoeae*. *MBio* 2018; 9(2).

276. Tomberg J, Unemo M, Davies C, Nicholas RA. Molecular and structural analysis of mosaic variants of Penicillin-Binding Protein 2 conferring decreased susceptibility to expanded-spectrum cephalosporins in *Neisseria gonorrhoeae*: role of epistatic mutations. *Biochemistry* 2010; 49(37): 8062-70.

277. Unemo M, Golparian D, Nicholas R, Ohnishi M, Gallay A, Sednaoui P. 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* 2012; 56(3): 1273-80.

278. Whiley DM, Goire N, Lambert SB, et al. Reduced susceptibility to ceftriaxone in *Neisseria gonorrhoeae* is associated with mutations G542S, P551S and P551L in the g

onococcal Penicillin-Binding Protein 2. *J Antimicrob Chemother* 2010; 65(8): 1615-8.

279. Zhao S, Duncan M, Tomberg J, Davies C, Unemo M, Nicholas RA. Genetics of chromosomally mediated intermediate resistance to ceftriaxone and cefixime in *Neisseria gonorrhoeae*. *Antimicrob Agents Chemother* 2009; 53(9): 3744-51.

280. Town K, Obi C, Quaye N, Chisholm S, Hughes G, Group GC. Drifting towards ceftriaxone treatment failure in gonorrhoea: risk factor analysis of data from the Gonococcal Resistance to Antimicrobials Surveillance Programme in England and Wales. *Sex Transm Infect* 2017; 93(1): 39-45.

281. Holmes AH, Moore LSP, Sundsfjord A, et al. Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* 2016; 387(10014): 176-87.

282. Zhao L, Liu A, Li R, Zhao S. Multiplex TaqMan real-time PCR platform for detection of *Neisseria gonorrhoeae* with decreased susceptibility to ceftriaxone. *Diagn Microbiol Infect Dis* 2018.

283. Mortimer TD, Grad YH. Applications of genomics to slow the spread of multidrug-resistant *Neisseria gonorrhoeae*. *Ann N Y Acad Sci* 2018.

284. Tanaka M, Furuya R, Kobayashi I, Kanesaka I, Ohno A, Katsuse AK. Antimicrobial resistance and molecular characterization of *Neisseria gonorrhoeae* isolates in Fukuoka, Japan, from 1996 to 2016. *J Glob Antimicrob Resist* 2018.

285. Liu K, Linder CR, Warnow T. RAxML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation. *PLoS One* 2011; 6(11): e27731.

286. Sirivongrangson P, Girdthep N, Sukwicha W, et al. The first year of the global Enhanced Gonococcal Antimicrobial Surveillance Programme (EGASP) in Bangkok, Thailand, 2015-2016. *PLoS One* 2018; 13(11): e0206419.

287. Goldstein E, Kirkcaldy RD, Reshef D, et al. Factors related to increasing prevalence of resistance to ciprofloxacin and other antimicrobial drugs in *Neisseria gonorrhoeae*, United States. *Emerg Infect Dis* 2012; 18(8): 1290-7.

288. Hui BB, Whiley DM, Donovan B, Law MG, Regan DG, Investigators GS. Identifying factors that lead to the persistence of imported gonorrhoeae strains: a modelling study. *Sex Transm Infect* 2017; 93(3): 221-5.

289. Iverson CJ, Wang SA, Lee MV, et al. Fluoroquinolone Resistance Among Neisseria gonorrhoeae Isolates in Hawaii, 1990-2000. *Sex Trans Dis* 2004; 31(12): 702-8.

290. Tapsall JW, Limnios EA, Murphy D, Australian Gonococcal Surveillance P. Analysis of trends in antimicrobial resistance in *Neisseria gonorrhoeae* isolated in Australia, 1997-2006. *J Antimicrob Chemother* 2008; 61(1): 150-5.

291. Fortenberry JD, McFarlane M, Bleakley A, et al. Relationships of stigma and shame to gonorrhea and HIV screening. *Research and Practice* 2002; 92(3): 378-81.

292. Pachankis JE, Hatzenbuehler ML, Hickson F, et al. Hidden from health: structural stigma, sexual orientation concealment, and HIV across 38 countries in the European MSM Internet Survey. *AIDS* 2015; 29(10): 1239-46.

293. World Health Organisation. Prevention and treatment of HIV and other sexually transmitted infections among men who have sex with men and transgender people, 2011. https://www.who.int/hiv/pub/guidelines/msm_guidelines2011/en/ (accessed 20/11/2018)

294. Public Health England. Surveillance of antimicrobial resistance in *Neisseria gonorrhoeae* in England and Wales: Key findings from the Gonococcal Resistance to Antimicrobials Surveillance Programme (GRASP), 2015. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567602/GRASP_Report_2016.pdf (accessed 20/11/2018)

295. Buono SA, Watson TD, Borenstein LA, Klausner JD, Pandori MW, Godwin HA. Stemming the tide of drug-resistant *Neisseria gonorrhoeae*: the need for an individualized approach to treatment. *J Antimicrob Chemother* 2015; 70(2): 374-81.

296. Public Health England. Managing outbreaks of sexually transmitted infections, 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/584723/PHE_STI_outbreak_guidelinesNB180117.pdf (accessed 20/11/2018)

297. Klovdahl AS, Potterat JJ, Woodhouse DE, Muth JB, Muth SQ, Darrow WW. Social networks and infectious disease: the Colorado Springs Study. *Soc Sci Med* 1994; 38(1): 79-88.

298. Mercer CH, Tanton C, Prah P, et al. Changes in sexual attitudes and lifestyles in Britain through the life course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). *Lancet* 2013; 382(9907): 1781-94.

299. DiClemente RJ, Swartzendruber AL, Brown JL. Improving the validity of self-reported sexual behavior: no easy answers. *Sex Transm Dis* 2013; 40(2): 111-2.

300. Fenton KA, Johnson AM, McManus S, Erens B. Measuring sexual behaviour: methodological challenges in survey research. *Sex Transm Infect* 2001; 77(2): 84-92.

301. Buyze J, Vanden Berghe W, Hens N, Kenyon C. Current levels of gonorrhoea screening in MSM in Belgium may have little effect on prevalence: a modelling study. *Epidemiol Infect* 2018; 146(3): 333-8.

302. Fingerhuth SM, Low N, Bonhoeffer S, Althaus CL. Detection of antibiotic resistance is essential for gonorrhoea point-of-care testing: a mathematical modelling study. *BMC Med* 2017; 15(1): 142.

303. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. *Pathology* 2015; 47(3): 199-210.

304. Whitlock GG, Gibbons DC, Longford N, Harvey MJ, McOwan A, Adams EJ. Rapid testing and treatment for sexually transmitted infections improve patient care and yield public health benefits. *Int J STD AIDS* 2018; 29(5): 474-82.

305. Ashton PM, Nair S, Peters TM, et al. Identification of *Salmonella* for public health surveillance using whole genome sequencing. *PeerJ* 2016; 4: e1752.

306. Dallman TJ, Chattaway MA, Mook P, Godbole G, Crook PD, Jenkins C. Use of whole-genome sequencing for the public health surveillance of *Shigella sonnei* in England and Wales, 2015. *J Med Microbiol* 2016; 65(8): 882-4.

307. Public Health England. Tuberculosis in England: 2018, 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/742782/TB_Annual_Report_2018.pdf (accessed 08/01/2019)

308. Satta G, Lipman M, Smith GP, Arnold C, Kon OM, McHugh TD. *Mycobacterium tuberculosis* and whole-genome sequencing: how close are we to unleashing its full potential? *Clin Microbiol Infect* 2018; 24(6): 604-9.

309. Poon AFY, Gustafson R, Daly P, et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *The Lancet HIV* 2016; 3(5): e231-e8.

310. Goire N, Kundu R, Trembizki E, et al. Mixed gonococcal infections in a high-risk population, Sydney, Australia 2015: implications for antimicrobial resistance surveillance? *J Antimicrob Chemother* 2017; 72(2): 407-9.

311. Mears J. Evaluation of the Tuberculosis Strain Typing Service in England: University College London.

312. Mears J, Abubakar I, Crisp D, et al. Prospective evaluation of a complex public health intervention: lessons from an initial and follow-up cross-sectional survey of the tuberculosis strain typing service in England. *BMC Public Health* 2014; 14: 1023.

313. Mears J, Vynnycky E, Lord J, et al. The prospective evaluation of the TB strain typing service in England: a mixed methods study. *Thorax* 2016; 71(8): 734-41.

314. Didelot X, Gardy J, Colijn C. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol Biol Evol* 2014; 31(7): 1869-79.

10 Appendix

10.1 Systematic review extended tables

Table 10.1 Summary of studies included in the systematic review: study aims and conclusions

Reference	Aim	Study conclusion
Studies using NG-MAST		
Martin (2004)[127]	Demonstrate the method and application of NG-MAST to identify individuals infected by the same gonococcal strain	New approach to typing can identify linked cases of gonorrhoea; different sized sequence type clusters indicate local transmission or repeated importation; large clusters in cities are indicative of an outbreak and can identify people for interventions
Martin (2005)[198]	Analyse the transmission of ciprofloxacin resistant NG by combining genotypic information with demographic and behavioural characteristics of the people infected	In earlier years ciprofloxacin resistance predominated in heterosexuals who had sex abroad but in later years was more likely to be found in MSM. Clusters increased in size as there was less infection associated with sex abroad and more ciprofloxacin resistance identified in MSM
Palmer (2005)[212]	Compare the discrimination of auxotype/serovar typing with NG-MAST using NG isolates exhibiting reduced susceptibility or resistance to ciprofloxacin, and describe the epidemiology of these isolates	NG-MAST provided more detail to the epidemiological data regarding clusters of infection; sequence types showed multiple transmission networks exist; sequence type clusters were transmitted exclusively among separate sexual networks but sequence type also found within the same networks; unique sequence type associated with travel suggesting importation
Palmer (2006)[202]	Investigate ciprofloxacin-resistant isolates with serogroup WI to determine whether one or more strain types accounted for the increase in these isolates	Most cases were from MSM who acquired their infection in the UK. Onward transmission of this sequence type facilitated by dense sexual networks as people reported multiple sexual partners. Bridging to heterosexual populations may have occurred as some people reported as bisexual, but partner tracing did not substantiate this
Lundback (2006)[197]	Characterise Swedish azithromycin-resistant NG to examine the genetic homogeneity/heterogeneity of azithromycin resistant NG	All high-level azithromycin resistant isolates were the same sequence type; this sequence type is domestically transmitted; possibly also linked to importation

Reference	Aim	Study conclusion
Choudhury (2006)[188]	Identify clusters of linked people using typing data in the absence of contact tracing data	Clusters of people infected with the same strain showed similarities in behavioural and demographic features. Heterosexuals with unique strains had a higher proportion of individuals that had sex outside of the UK, suggesting importation of novel strains to London. Bisexual men usually have strains similar to MSM so suggest gonococcal infection is spreading between men more than between men and women
Bilek (2007)[179]	Examine the concordance between the genotypes of isolates of NG from recent sexual contacts	Concordance between sequence types of known sexual contacts was high, authors indicate that this means NG-MAST is a suitable tool as an adjunct to contact tracing for reconstructing sexual networks
Unemo (2007)[209]	Investigate whether one strain of NG is responsible for the widespread transmission of ciprofloxacin-resistant infection, particularly among young heterosexuals	Confirmation of domestic spread of one ciprofloxacin-resistant NG strain in Sweden in young heterosexuals
Risley (2007)[215]	Explore the geographical distribution of NG across London, sexual orientation networks, ethnicity and social deprivation, and assess if there is local transmission according to molecular typing and geographical clustering of cases	Sequence types clustered more in heterosexual men and women than in MSM, the authors suggest that MSM may be less likely to acquire infections locally
Palmer (2008)[130]	Investigate the characteristics of NG isolates with decreased susceptibility to azithromycin and associated patient demographics	Most sequence types contained a mix of azithromycin resistant and susceptible isolates and there was no demographic separation between these patients. However, the distribution of sequence types over time suggests that the development of resistance to azithromycin may have occurred several times within these different sequence types within discrete sexual networks
Wong (2008)[210]	Assess the genotype distribution of NG in Taiwan to identify core groups at high risk of infection and investigate the domestic and international transmission of antibiotic resistant infections in specific sexual networks	Large sequence type clusters suggested that multiple clonal transmissions existed in Taiwan. Some sequence types only identified within specific sub-groups of the populations such as HIV-positive MSM and these sequence types are where most of the AMR isolates are. For example, ST457 is only found in HIV-positive MSM and this ST457 makes up 65% of ciprofloxacin resistant isolates.

Reference	Aim	Study conclusion
Starnino (2008)[207]	Assess: the epidemiological characteristics of people diagnosed with gonorrhoea and the rate of HIV co-infection; the serovar of NG isolates and the antimicrobial susceptibility to five antimicrobial drugs; the correlation between molecular types and specific antimicrobial profile	HIV status not associated with AMR; large sequence type cluster associated with heterosexuals; clonal dissemination of AMR through one sequence type likely; high number of different sequence type found in only one isolate
Abu-Rajab (2009)[177]	Assess the strain type similarities between isolates from sexual contacts and the association between strain types and patient demographic and behavioural data	Sequence type associated with sexual orientation but no other epidemiological variables that might be indicative of more specific sexual networks such as geographical clustering
Starnino (2009)[206]	Describe the prevalence of azithromycin resistant NG and characteristics of infected people	Some clustering of specific sequence types by residence, sexual orientation or AMR
Fernando (2009)[190]	Describe the characteristics of people infected with common NG strain types	People repeatedly infected had different NG sequence types at repeat attendances. The authors suggest that this is evidence that these individuals were part of more than sexual network. Identical sequence types found in 94% of known sexual partners. People with unique sequence types significantly more likely to have had sex outside of the study area. Possible evidence of sexual networks that extend beyond the study area as the common sequence types were also reported in a London study previously. Distinct transmission networks were present in Lothian as unique sequence types found here. Some clustering of sequence types amongst different patient characteristics including gender, sexual orientation and HIV status
Chisholm (2009)[186]	Investigate the origin of high-level azithromycin resistance in NG in England and Wales	High-level azithromycin resistance was identified in six closely related sequence types
Monfort (2009)[199]	Analyse circulating NG to identify clusters of isolates from high-risk groups and clusters with particular antibiotic resistance phenotypes	No geographic clustering of sequence types was observed and no predominant sequence types in MSM
Starnino (2010)[205]	Identify the sequence types most represented among the ciprofloxacin-resistant NG isolates	Exclusive transmission of sequence types within MSM or heterosexual networks

Reference	Aim	Study conclusion
Florindo (2010)[191]	Determine the NG antibiotic phenotype, genotype distribution and association with sexual orientation and age	Substantial differences were identified in sequence type distribution by age and sexual orientation. However, the authors state that as there was such a large diversity of sequence types, they could not identify specific sexual networks that spread NG. 8/9 sexual partners included in this study were infected by isolates with the same sequence type
Chisholm (2011)[185]	Explore the characteristics and molecular epidemiology of NG isolates with decreased susceptibility to cefixime	Cefixime decreased susceptibility is clonally spread and circulating among MSM and heterosexual people, some of whom reported sex abroad suggesting possible importation
Yuan (2011)[211]	Describe baseline data from nationwide surveillance of NG susceptibility to azithromycin and genetically characterise azithromycin resistant strains	NG-MAST provided additional information linking cases; different sequence type identified for azithromycin resistance than previous studies
Ota (2011)[201]	Describe demographic characteristics and genotypic clustering of quinolone resistant NG	This study identified multiple QRNG clusters that were associated with specific sexual networks. There was also evidence of QRNG spread between MSM and non-MSM networks, as the same sequence types were identified in both populations
Carannante (2012)[180]	Present the circulating strain types and antimicrobial resistant profile of NG	Common sequence type identified in MSM but also high variability in sequence type among MSM suggest lack of conditions favouring clonal spread of AMR or high recombination rate
Hjelmevoll (2012)[192]	Investigate the genotypic and phenotypic properties including antimicrobial resistance determination of NG	Sequence type circulating are heterogeneous; sequence type variety related to widespread importation of NG identified from epidemiological data
Cole (2013)[189]	Understand the factors affecting the transmission of NG in South East Wales, UK	Common sequence types widely disseminated in different sexual networks (MSM, heterosexuals); most localised clustering of sequence type in young heterosexuals; sequence type persistence over time within these networks demonstrates a lack of bridging between networks
Bernstein (2013)[178]	Use molecular epidemiology to monitor the local epidemiology of NG	Specific sociodemographic, behavioural, and phenotypic markers of antibiotic resistance clustered by sequence type. Among MSM, sequence type was significantly associated with ethnicity, number of sex partners, and reporting oral sex as the only recent urogenital exposure
Chen (2013)[182]	Describe a cluster of 47 NG cases belonging to NG-MAST ST4378 and MLST ST1901	Sequence type along with AMR and geographical information confirmed isolates were in the same distinct high-risk MSM sexual network; this is different to existing data that suggested this clone was circulating in heterosexual networks; identified possible bridging between networks

Reference	Aim	Study conclusion
Ison (2013)[194]	Describe changes in AMR and treatment options for NG	Cefixime decreased susceptible isolates were associated with genotype 1407 clones, which were found predominantly in MSM
Chisholm (2013)[187]	Assess the public health benefit of NG-MAST molecular epidemiological typing of NG within Euro-GASP	Considerable diversity of gonococcal sequence types exist both within and between countries, some sequence types predominate. Some sequence types are associated with specific sexual networks, such as MSM, but these strains are also identified in heterosexual networks as well
Singh (2013)[204]	Define populations of circulating NG populations, identify clusters of infection and investigate the role of particular strains in gonococcal transmission	Significant diversity of sequence type; multiple importations of cefixime decreased susceptible NG; separate sequence type clusters identified among MSM and heterosexuals
Horn (2014)[193]	Investigate the epidemiology of AMR NG in Germany	A variety of different sequence types were identified. G1407 was significantly more common among males than females. Genogroups G25 and G387 less common in male groups
Carannante (2014)[181]	Determine the sequence type of multi-drug resistant NG	Genogroup G1407 is the main genogroup among the MDR gonococci in Italy and the majority of these were from MSM
Jeverica (2014)[195]	Phenotypically and genetically characterize NG isolates to elucidate the molecular epidemiology of the emergence and spread of NG with decreased susceptibility and resistance to extended spectrum cephalosporins	AMR trends due to spread of several discrete sequence types; specific associations between sequence type and sexual orientation identified indicate spread of sequence type within sexual networks
Stevens (2015)[208]	Describe high-level azithromycin resistant NG in Australia using national surveillance data	Five different sequence types with azithromycin resistant suggests independent generation of AMR; sequence type likely imported from China but also domestic transmission
Cheng (2016)[184]	Describe longitudinal trends in NG antimicrobial resistance	Significant difference between sequence types circulating between heterosexual men and MSM. Change in the sequence types in MSM indicate different sources of sequence types
Chen (2016)[183]	Describe antimicrobial resistance, molecular epidemiology and genetic determinants of ceftriaxone resistance	Sequence types from China different from other countries indicating that domestic sexual networks are not strongly linked to sexual networks abroad. Sequence types varied by geographic area within China
Foster (2016)[44]	Describe the approaches for investigating and implementing control measures during an NG outbreak, including using molecular typing	Different distribution of NG-MAST sequence types identified between cases in outbreak compared to cases not in the outbreak leading to the conclusion that there was an association in outbreak cases between genogroup G25 and young heterosexuals

Reference	Aim	Study conclusion
Ni (2016)[200]	Identify the prevalence of azithromycin resistance in NG	Azithromycin resistance prevalence was 21%; high-level azithromycin resistance prevalence was 18%. Majority of high-level azithromycin resistant cases were the same NG-MAST sequence type and were from men and women
Lahra (2017)[196]	Describe an outbreak of azithromycin resistant NG in South Australia	Majority of isolates had the same NG-MAST sequence type and all isolates from heterosexual people. One person was a sex worker and two had sex worker contacts but unclear if these people had different NG-MAST sequence types to majority of sample
Serra-Pladevall (2017)[203]	Describe NG antimicrobial resistance and strain type differences between MSM and heterosexual populations	Significant differences between the NG strains circulating amongst MSM and heterosexuals in terms of antimicrobial resistance and NG-MAST sequence type
Studies using MLST		
Perez-Losada (2007)[213]	Infer the evolutionary history of isolates to determine if genetically variant strains of NG diverged before or after quinolone resistant NG first appeared in Israel	Phylogenetic analyses suggested that quinolone resistant were imported into Israel on at least three separate occasions.
Trembezki (2016)[214]	Describe the genetics and distribution of gonococcal AMR in Australia	The ratio of males to females differed noticeably between certain sequence types. Although many different gonococcal strains may be circulating within a population at any given time, NG infections on a population level are dominated by a relatively small number of strains
Studies using WGS		
Grad (2014)[121]	Describe the relatedness between NG isolates to reconstruct the likely spread of lineages through sexual networks according to geography and sexual orientation, and to identify likely emergence events of resistant phenotypes	Cefixime resistance spreads eastward from California, predominantly within MSM networks, with a small number of introductions into heterosexual networks
Demczuk (2015)[135]	Describe the dissemination, relatedness, and emergence of NG isolates with elevated extended spectrum cephalosporin MICs	NG is genetically heterogeneous and broadly distributed; association identified between different lineages and patient characteristics suggesting link of specific lineages to heterosexual populations
Chisholm (2016)[162]	Describe an outbreak of high-level azithromycin resistant NG	WGS and NG-MAST showed genomes of outbreak cases to be very similar to each other (≤1 nucleotide difference) and different to the control isolates. Outbreak cases one nucleotide different to sequence type associated with previous high level azithromycin resistant cases

Reference	Aim	Study conclusion
De Silva (2016)[118]	Define the genetic diversity between samples related by transmission, and apply this information to detect local, regional, and international transmission of NG	Identified a method of determining the plausibility of direct or indirect transmission between two cases using WGS data. Within Brighton found sustained local transmission between cases often within 1-3 months apart, possible reflection of frequent partner changes involving many infected individuals. Comparison to other datasets found evidence of possible link to cases elsewhere in the UK in 18% of samples and 9% to cases in the US
Didelot (2016)[119]	Investigate how WGS data can be used to track transmission of NG within a city	Used WGS data to estimate time between transmission events to support prediction of linkage between two cases; cases sampled more than 8 months apart are unlikely to be in the same transmission chain. Likely transmission links were significantly associated with shorter geographical distances in London and there were more links between HIV-positive people than HIV-negative people
Grad (2016)[141]	Define the distribution of resistance markers to the three most clinically relevant classes of antimicrobials for treatment of NG	Cefixime and ceftriaxone reduced susceptibility has spread predominantly through clonal expansions and is highly although not exclusively associated with the mosaic *penA*XXXIV allele and its derivatives. Azithromycin reduced susceptibility arises through multiple mechanisms that are less clonal. Quinolone resistance is mainly clonal and has emerged several times.
Jaccobson (2016)[144]	Describe the genomics of azithromycin resistance	Identified clonal spread of *NG* strains accounted for the majority of the azithromycin resistance in Europe. Future molecular prediction of clinical resistant to azithromycin should focus on detection of A2059G and C2611T mutations in the 23S rRNA gene

Table 10.2 Summary of risk of bias assessment for included studies

Reference	Is the study generalizable to the study population?	How many eligible isolates were included? How much missing patient epidemiological data were there?
Studies using NG-MAST		
Martin (2004)[127]	Subset of national sentinel surveillance that collects consecutive isolates during 3-month period from 13 London clinics (N=195). Other sample contained all isolates resistant to antibiotics specified and that had previously been used to evaluate typing methods (N=268).	All isolates eligible for typing typed. No information on completeness of epidemiological data presented.
Martin (2005)[198]	All quinolone resistant isolates over four-year period selected from national sentinel surveillance programme. Represent 3.6% (192/5260) of NG collected in programme. Unclear how much of all NG this represents overall.	All isolates eligible for typing included. Sexual orientation available for all isolates. Ethnicity missing for 21.3% (41/192). No information on missing concurrent STI.

Palmer (2005)[212]	Selected isolates that were reduced susceptible or resistant to ciprofloxacin representing 13% (106/818) of all isolates submitted. This is all of Scotland as all isolates are submitted to reference laboratory.	All isolates eligible were typed (n=106). Sexual orientation reported for 90.6% (96/106) of cases. Area of acquisition known for all isolates.
Palmer (2006)[202]	Selected subset of all isolates in Scotland in one year that were of particular serotype and ciprofloxacin resistant. This represented 44% (56/126) of all ciprofloxacin resistant isolates and 7% (56/824) of all NG in Scotland in this year.	All eligible isolates sequenced. Epidemiological data reported for 86% (42/56) of isolates
Lundback (2006)[197]	Study aim is to characterise all azithromycin resistant NG in Sweden in 2004, all of these isolates selected for inclusion.	All isolates eligible sequenced. All epidemiological data included reported.
Choudhury (2006)[188]	Subset of isolates from national sentinel surveillance that have come from London clinics covering 77% of all reported gonorrhoea in London during study period.	81.1% eligible isolates successfully typed (2345/2891). Epidemiological data available for 87% of isolates typed (2045/2345). 11.6% (238/2045) missing ethnicity. 11.2% (230/2045) missing concurrent STI status. 7.5% (154/2045) missing previous gonorrhoea information. 7.3% (150/20458) missing symptoms information. 42.3% (866/2045) missing HIV status information. Further breakdown by sexual orientation available in Table 1 of paper.
Bilek (2007)[179]	Isolates from sexual contact pairs in a selected clinic during one year (Sheffield's main GUM clinic) included in the sample for sequencing (167). Unclear how many samples this represents for the whole year. Study is about concurrency of sequence type between sexual contacts so generalisability to location or patient group less relevant. Study conducted exhaustive contact tracing of all NG diagnoses during the study period.	All sexual contact pairs identified sampled. Gender data only provided for subset of people in the sample that represented non-concordant genotypes.
Unemo (2007)[209]	Selected isolates of particular serotype and ciprofloxacin resistant thought to be part of a core group (N=20) and other selected ciprofloxacin resistant isolated (N=26) from 10 cities. Two ciprofloxacin susceptible isolates with same serovar type also used and was found circulating at the same time. Unclear representativeness of all NG.	One isolate unavailable for typing from core group so 95% (19/20) successfully typed. Epidemiological data available for all.
Risley (2007)[215]	All NG cases diagnosed in 13 London GUM clinics during five-month period. These clinics identify 80% of NG in London.	Of all eligible isolates, 70% successfully retrieved and typed (2045/2891). Postcode data reported for 65% (1871/2891). Other epidemiological data complete.
Palmer (2008)[130]	All isolates reported to reference laboratory during period, which would be all NG in Scotland (N=3,326).	All eligible isolates sequenced. Epidemiological data only reported for common sequence types so unclear how much missing epidemiological data.

Wong (2008)[210]	Subset of male diagnoses from all diagnoses made in one clinic during period (93%: 139/149). Unclear how many isolates in city the isolates represent.	All eligible samples sequenced. Sexual orientation reported for 95% (132/139) of isolates.
Starnino (2008)[207]	Isolates from 7/12 STI clinics in large cities participating in sentinel surveillance of STIs in Italy. No information about how many clinics there are in Italy. All isolates diagnosed during study period eligible (N=514) but 50% of isolates with a culture and patient sexual orientation data were eligible typed NGMAST (164/514). Unclear what proportion of NG in Italy this represents.	All eligible isolates typed. All epidemiological data complete.
Abu-Rajab (2009)[177]	All isolates from selected clinic (Glasgow's main GUM clinic) included in sample to be sequenced (178). 4.5% (8/178) unable to be sequenced because of inviable sample or not matched to clinical data. 59% (105/178) not included in sexual network analysis as questionnaire not completed. Information on the representativeness of the included 41% (65/178) not provided.	59% of eligible sample missing from sexual network analysis because no questionnaire data, sample not viable or sample not matched to clinical records. No information provided on these missing people to assess if different from the people included in the sexual network analysis. Unable to assess if this is missing systematically or at random.
Starnino (2009)[206]	Subset of azithromycin resistant isolates (n=22) selected from four large STI clinics that represent ~70% of all NG cases (n=219, estimated denominator =313). Therefore, in total sample represent ~7% (22/313) of isolates in Italy.	Sequenced all eligible isolates. No missing epidemiological data.
Fernando (2009)[190]	Over the study period, all 370 culture-positive NG diagnoses at the Edinburgh GUM clinic were eligible Estimated representativeness of all NG in Scotland: 20.5% ((370/2)/900). Estimated representativeness of all NG in Edinburgh: 90.2% ((370/2)/205). No information on number of isolates during study period with no culture	100% eligible isolates typed. Most epidemiological variables complete except location of recent sexual contacts missing for 12.2% (45/370)
Chisholm (2009)[186]	Subset of isolates from the national sentinel surveillance programme selected for typing from one year (N=75). Isolates selected as either resistant to azithromycin (n=34) (no information about how many this represents from one year) or from two clinics were high level azithromycin resistance has been identified (n=41) (unclear how many diagnoses form all cases in these clinics these represent).	All eligible isolates successfully typed. 98.7% (74/75) reported sexual orientation. Age only reported for subset of high-level azithromycin resistant isolates.

Monfort (2009)[199]	Aim of study is to describe NG circulating in France. Selected a subset of isolates from voluntary laboratory reporting surveillance reported during 6-month period (31.5%; 93/295). Within this 72% (67/93) selected based on AMR profile (mix of resistant and susceptible to variable antimicrobials), geographic organs, gender and site of infection. Remaining selected based on site of infection (rectal) and from MSM 23% (26/93).	All selected isolates successfully typed. Sexual orientation and site of infection known for all selected isolates. Age not reported for 6 isolates.
Starnino (2010)[205]	Subset of larger sample, selecting isolates that were resistant to ciprofloxacin and patient sexual orientation data is available. Represented 56% (137/244) of ciprofloxacin resistant isolates and 25% (137/599) of all isolates in sample. Unsure if total samples (n=599) is a consecutive from laboratories/clinics.	All isolates eligible for sequencing sequenced. No further missing data than the missing data for sexual orientation of ciprofloxacin resistant isolates 44% (107/244).
Florindo (2010)[191]	25/100 laboratories sent isolates over four years. Unclear representativeness compared to all NG circulating in Portugal. Most labs sending isolates from one city (Lisbon). Some years sampled more than others (only 17 in 2004).	86% eligible samples successfully typed (236/274). Age reported for all. Gender reported for all. Sexual orientation missing for 39% (92/236).
Chisholm (2011)[185]	Subset of isolates from sentinel surveillance or isolates referred to national reference laboratory were typed if exhibiting decreased susceptibility to cefixime (N=97). Representativeness of sentinel surveillance: 1.2% (55/4649). Unknown denominator of isolates selected from the reference laboratory sample.	99% of isolates eligible for typing successfully tested (96/97). Gender reported for 98% (94/96). Age reported for 100%. Sexual orientation only known for subset of sentinel surveillance sample (81%: 44/54). Ethnicity or Information about travel-associate sexual partnerships known for subset of sentinel surveillance sample (79.6%: 43/54).
Yuan (2011)[211]	Selected azithromycin resistant isolates representing 5% (17/318) of NG in two cities during period. Unclear how many this would be for whole country.	All eligible isolates sequenced. All epidemiological data complete.
Ota (2011)[201]	Selected a subset of consecutive quinolone resistant isolates during two-month period in 2006, representing 15%: (104/695) of all quinolone resistant isolates. This represents 4.2% (104/2482) of all NG in province studied in 2006	All eligible isolates were typed. Sexual orientation and age data available for all isolates. 3% of isolates missing residence information.

Carannante (2012)[180]	Isolates from 6/12 STI clinics participating in sentinel surveillance of STIs in Italy. No information about how many clinics there are in Italy. Six clinics selected as they are in large Italian cities. NG-MAST performed on 19.3% (120/620) of all NG diagnoses during study period (5 years) based on isolates resistant to antimicrobials (unspecified which antimicrobial) and from people with known sexual orientation. Unclear what proportion of NG in Italy this represents.	No information about epidemiological profile of all isolates collected (N=620).
Hjelmevoll (2012)[192]	All viable isolates diagnosed in one year for six hospitals in Norway. Study covers 42% (114/269) NG in Norway. Representative of all NG in Norway with regard to gender, sexual orientation and place of contracting infection.	126 eligible cases, 114 successfully typed (90.5%). Majority of people reported sexual orientation 98% (112/114).
Cole (2013)[189]	All isolates collected at three main laboratories in region included. Unclear what proportion of all diagnoses in this region this covers.	93% (475/511) of eligible isolates typed. Epidemiological data available for 99.2% (507/511) of isolates. Gender reported for all isolates. Additional epidemiological data available for GUM patients only (69.9%: 357/511).
Bernstein (2013)[178]	Sample includes first 25 male urethral specimens from symptomatic people attending one clinic each month for one year. Unclear on the representativeness of these samples for this clinic. Only analysed sexual network data from MSM as low number of specimens from heterosexual men (20%; 53/265).	All isolates eligible for typing were included. Missing epidemiological data is reported in paper table 1. Only variable with missing data is HIV status (4.3%).
Chen (2013)[182]	Aim of study was to describe subset of NG diagnoses with specific NGMAST (N=47). Subset from surveillance programme that covers 40 clinics in Taiwan (N=2,357). Unclear how many clinics in Taiwan.	All eligible isolates included. 70% of isolates included in study linked to medical records (33/47) for additional epidemiological information. For these isolates, no further missing data.
Ison (2013)[194]	Subset of isolates from sentinel surveillance programme that tests ~7-10% of all circulating NG. Isolates exhibiting resistance to cefixime selected for typing, 7.4% (547/7378) of isolates collected in the programme during study period eligible for typing.	97.8% (535/547) eligible isolates successfully typed. Sexual orientation reported for 92% (505/547) of eligible isolates. Ethnicity reported for 89% (490/547) of eligible isolates. Number of sexual partners in the UK reported for 89% (486/547) of eligible isolates. Travel-associated sexual partnerships reported for 73% (399/547) of eligible isolates. Symptomatic information reported for 65% (357/547) of eligible isolates. Chlamydia status reported for 92% (501/547) of eligible isolates. HIV status reported for 85% (463/547) of eligible isolates.

Chisholm (2013)[187]	Subset of NG across Europe from sentinel surveillance programme that collects 55 isolates from consecutive patients twice annually to be representative of national distribution of cases and have AMR data (overall representation of all NG in Europe: 3.3%; 1,066/32,028).	All eligible isolates typed. 96.6% (1,030/1,066) reported age. 98.5% (1,050/1,066) reported gender. 55% (586/1,066) reported sexual orientation
Singh (2013)[204]	All culture cases in area sent to regional lab for testing. This is 26.4% (2,250/8,535) of all NG in area during 2011-2007. Of these, isolates with resistance to penicillin tetracycline, ciprofloxacin, cefixime decreased susceptibility and/or azithromycin were typed (n=238: 2.8% of all NG (238/8,535) and 10.6% (235/2,250 of all culture diagnoses). Majority of isolates in study from 2009-2011 (95.8%; 228/238).	Unclear for earlier years how many isolates eligible for NGMAST but not typed. Missing data not reported for samples typed.
Horn (2014)[193]	23 laboratories reporting consecutive NG isolates. Unclear how representative of Germany the NG identified in these laboratories.	Unclear if more isolates eligible than successfully typed. Age and gender reported for all isolates. Site of infection not reported for 18.8% of isolates (40/213).
Carannante (2014)[181]	Typed isolates that were multi-drug resistant (MDR) from a subset of NG diagnoses (66.5%; 1,777/2,671) available in Italy over 14-year period (3.0%; 81/2,671). Study aim was to investigate strain types of MDR so sample generalizable to MDR NG cases.	All eligible isolates were typed. No missing epidemiological data for typed isolates.
Jeverica (2014)[195]	74% of all NG cases in Slovenia selected during period.	Assume all NG cases eligible so only typed 74% (N=194). Age and site of infection known for all isolates. Sexual orientation reported for 89% (179/194)
Stevens (2015)[208]	Subset of isolates from national surveillance over three-year period that exhibited high-level azithromycin resistance. Not representative of all NG cases.	All isolates eligible for sequencing sequenced. Complete information on gender, age, site for all. No data on completeness of country of acquisition but missing data for one acknowledged.
Cheng (2016)[184]	Per year, 1,500 NG cases in Taiwan; 500 in Taipei City. Study conducted isolates from one hospital over 7 years in Taipei city collected 1,111 isolates, ~10% of all NG in Taiwan.	Nearly all isolates eligible typed (98.1%; 1,090/1,111). 100% gender reported. 98.2% reported sexual orientation (1,070/1,090). 94.5% reported HIV status (1,030/1,090). 89% reported syphilis status (971/1,090).
Chen (2016)[183]	One isolate per patient from 11 sentinel sites across China selected. Sites selected as detect the majority of STIs in each respective geographical location. In China in 2013, 103,085 NG cases, so report covers 0.9% of circulating NG.	73.2% of eligible isolates typed (920/1,257). Epidemiological data available for eligible isolates (N=1,257), unclear which isolates typed had missing epidemiological data. 99.7% (1254/1257) reported gender. 92.3% (1,168/1,257) reported sexual orientation. 94.3% (1,160/1,257) reported location of infection or whether they had sex abroad. 28.4% (357/1,257) reported previous STI. 30.9% (388/1,257) reported previous drug use.

Foster (2016)[44]	Isolates from an outbreak and controls from the same area and time were period were eligible for typing (N=623). Additional historical sentinel surveillance samples also selected (N=123).	Of the eligible samples, 27% (169/623) of the outbreak/non-outbreak cases were successfully typed and matched to epidemiological data for analysis. 93% (115/123) of the historical sample sequenced.
Ni (2016)[200]	All samples from two years from one hospital in China (n=118).	All eligible samples typed. Gender and age only reported for high-level azithromycin resistant cases. Unclear if epidemiological data missing for others.
Lahra (2017)[196]	Selected all azithromycin resistant isolates collected as part of sentinel surveillance. Unclear denominator for all NG cases in this period.	All isolates eligible sequenced. Epidemiological data missing for 3.5% (1/28).
Serra-Pladevall (2017)[203]	All samples from heterosexuals and random selection of MSM attending one clinic (only STI clinic in city) in Spain in one year (N=107). Representative of 7.4% (53/715) MSM and 8.9% (54/606) of heterosexual men and women.	All eligible samples sequenced. Gender, age and sexual orientation complete for all. Ethnicity missing for 9.4% (5/53) MSM, 27.8% (15/54) heterosexuals. Education status missing for 32.1% (17/53) MSM, 4.3% (25/54) heterosexuals. HIV status missing for 15.1% (8/53) MSM, 35.2% (19/54) heterosexuals. Sexual behaviour missing for 3.2% (6/53) MSM, 38.9% (21/54) heterosexuals.
Studies using MLST		
Perez-Losada (2007)[213]	Selected subset of samples (N=48) half resistant to fluoroquinolones and half sensitive to fluoroquinolones from previous genotype study (60%; 48/80). This previous study was a random sample of male urethral isolates. This represents 9% (48/525) of all NG in Tel Aviv: the area of analysis.	All isolates selected were successfully typed. All epidemiological data included complete.
Trembezki (2016)[214]	Isolates from surveillance programme representing 34.4% of all isolates (2,452/7,128) and 98% of all isolates cultured. Surveillance programme covers 30% of all gonorrhoea isolates in Australia.	90.4% of eligible isolates sequenced (2,218/2,452). Missing data for gender low: only 15 unknown.

Studies using WGS		
Grad (2014)[121]	Purposive sampling of cefixime resistant and susceptible isolates matched on location, date of isolation and sexual orientation. One year of study from sentinel surveillance programme, which selects first 25 male urethral samples from participating clinics each month. Cefixime resistant isolates represented 97% of all cefixime resistant isolates identified by surveillance programme in this period. Susceptible isolates likely under represent all circulating NG in USA.	Eligible isolates successfully sequenced (N=236). All epidemiological data complete.
Demczuk (2015)[135]	Isolates selected based on decreased susceptibility to ceftriaxone in archives spanning ~15 years (N=65). Isolates chosen are specific subset from different years to provide broad range of geographical distribution and antimicrobial susceptibilities (N=105). Unclear how representative of all NG during period.	All selected isolates for sequencing successfully sequenced. Gender reported for 99% of samples (168/169). Age reported for 81% (137/169) of isolates.
Chisholm (2016)[162]	Selected all isolates considered part of outbreak (N=7)) plus additional cases (N=6) with range of antimicrobial susceptibility data from same period as controls but unclear how representative these controls are.	All eligible isolates sequenced. Epidemiological data provided for the outbreak cases only.
De Silva (2016)[118]	Main sample included all consecutive culture NG samples from STI clinic in one city over four-year period.	98% eligible isolates successfully sequenced (1407/1437). Gender missing for 2 people. Sexual orientation only available for 15 pairs of cases.
Didelot (2016)[119]	Two datasets (London and Sheffield), both restricted to isolates with most common ST therefore not representative of all NG circulating. Represents 3.4% of NG in London (127/3,754) but in Sheffield unclear how much sequenced isolates represent of all gonorrhoea circulating (n=140).	Missing eligible sample in London 15% (19/127), in Sheffield (8/140). Linked partner data available for Sheffield dataset but limited epidemiological data. London dataset reported if HIV status missing (31%; 33/105).
Grad (2016)[141]	Sample collected based on antimicrobial susceptibility profile. Isolates from 14 years (N=1,102) of surveillance programme, which collects first 25 male urethral isolates each month. Unclear representativeness of all NG circulating.	All eligible isolates sequenced. All have sexual orientation data.
Jaccobson (2016)[144]	Subset of international surveillance: collects 55 consecutive isolates during two periods annually from 17 countries. Azithromycin resistant isolates from four-year period selected (N=66) and nine additional isolates between 2013 and 2014	All eligible isolates successfully typed. Gender missing for 4% (3/75). Sexual orientation missing for 47% (30/64) of male cases. Site of infection missing for 4% (3/75)