Whole-genome sequencing to determine *Neisseria gonorrhoeae* transmission: an observational study

*Dilrini De Silva PhD[1,2,3], *Joanna Peters MBBS [4], Kevin Cole BSc[5,7], Michelle J Cole DBMS[6], Fiona Cresswell MBChB[5], Gillian Dean MBChB[5], Jayshree Dave MD[7], Daniel Rh Thomas PhD[8], Kirsty Foster MSc[9], Alison Waldram PhD[10], Daniel J Wilson DPhil[1,2], Xavier Didelot DPhil[11], Yonatan H Grad MD PhD[12], Derrick W Crook FRCPath[1,2,3,7], Tim EA Peto FRCP[1,2,3], A Sarah Walker PhD[1,2,3], *John Paul MD[2,5,7], *David W Eyre DPhil[1,2,3]

[1]Nuffield Department of Medicine, University of Oxford, Oxford, UK
[2]National Institute for Health Research Biomedical Research Centre, Oxford, UK
[3]Oxford National Institute for Health Research Health Protection Research Unit
[4]St George's University Hospitals NHS Foundation Trust, London, UK
[5]Brighton and Sussex University Hospitals NHS Trust, Brighton, UK
[6]Public Health England Sexually Transmitted Bacterial Reference Unit, Colindale, London, UK
[7]National Infection Service, Public Health England, UK
[8]Communicable Disease Surveillance Centre, Public Health Wales, Cardiff, UK
[9]Health Protection Team, Public Health England North East, UK
[10]Field Epidemiology Service Newcastle, Public Health England, Newcastle upon Tyne, UK
[11]Department of Infectious Disease Epidemiology, Imperial College, London, UK
[12]Department of Immunology and Infectious Diseases, Harvard TH Chan School of Public Health, and Division of Infectious Diseases, Brigham and Women's Hospital, Boston, MA, USA

* Drs De Silva and Peters, and Drs Paul and Eyre contributed equally.

Corresponding author:
Dr David Eyre, Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, OX3 9DU, UK, david.eyre@ndm.ox.ac.uk, 01865 220855.

Keywords:
Neisseria gonorrhoeae, gonorrhoea, whole genome sequencing, transmission, contract tracing, epidemiology

Abstract

Background

New approaches are urgently required to address increasing rates of gonorrhoea and the emergence and global spread of antibiotic-resistant *Neisseria gonorrhoeae*. Whole genome sequencing (WGS) can be applied to study transmission and track resistance.

Methods

We performed WGS on 1659 isolates from Brighton, UK, and 217 additional isolates from other UK locations. We included WGS data (n=196) from the USA. Estimated mutation rates, plus diversity observed within patients across anatomical sites and probable transmission pairs, were used to fit a coalescent model to determine the number of single nucleotide polymorphisms (SNPs) expected between sequences related by direct/indirect transmission, depending on the time between samples.

Findings

We detected extensive local transmission. 281/1061(26%) Brighton cases were indistinguishable (0 SNPs) to ≥1 previous case(s), and 786(74%) had evidence of a sampled direct or indirect Brighton source. There was evidence of sustained transmission of some lineages. We observed multiple related samples across geographic locations. Of 1273 infections in Brighton, 225(18%) were linked to another case from elsewhere in the UK, and 115(9%) to a case from the USA. Four lineages initially identified in Brighton could be linked to 70 USA sequences, including 61 from a lineage carrying the mosaic *penA* XXXIV associated with reduced cefixime susceptibility.

Interpretation

We present a WGS-based tool for genomic contact tracing of *N. gonorrhoeae* and demonstrate local, national and international transmission. WGS can be applied across geographical boundaries to investigate gonorrhoea transmission and to track antimicrobial resistance.

Funding

Oxford NIHR Health Protection Research Unit and Biomedical Research Centre.

Introduction

Seventy-eight million cases of gonorrhoea occur annually worldwide.[1] Increasing antimicrobial resistance threatens effective treatment and control.[2] In England, 34,958 cases occurred in 2014, a 19% increase from 2013.[3] National United Kingdom (UK) guidelines recommend combined single dose ceftriaxone and azithromycin as first-line treatment.[4] Without available alternatives for empirical treatment, strategies are urgently required to address the spread of drug-resistant strains.

In men, incubation periods until symptomatic urethritis are typically 2-5 days, and usually <2 weeks.[5,6] Prompt treatment usually limits symptomatic infection to <2 weeks.[7] However, infections in women,[8] and rectal, pharyngeal,[9] and some urethral[10] infections in men may be asymptomatic, impairing control efforts. In settings where most infections are symptomatic and rapidly treated, on-going transmission requires high rates of partner change in a sub-population, known as "core transmitters".[11] However, transmission from chronically infectious asymptomatic or untreated cases[12] is also important,[10] including in men who have sex with men (MSM), where rectal and pharyngeal carriage predominates:[13] urethral screening alone may miss up to 95% of infections.[14]

Whole genome sequencing (WGS) allows high precision investigation of pathogen transmission epidemiology. Its application to *Neisseria gonorrhoeae* is complicated by high recombination rates, which must be accounted for. WGS has been used to investigate azithromycin-resistant gonorrhoea outbreaks,[15] and the spread of strains with reduced susceptibility to cefixime and azithromycin across the United States (USA)[16], and Canada[17,18]. However, these studies selected nationwide samples based on antimicrobial susceptibility, and therefore could not quantify the extent of local transmission or what proportion of cases originated from other regions or countries.

We sequenced all available *N. gonorrhoeae* isolates from Brighton, UK, over 4-years, plus isolates from other UK locations, combining results with previous USA WGS. We aimed to define the expected genetic diversity between samples related by transmission, and to apply this to detect local, regional and international transmission.

Methods

Setting, diagnostic testing and samples

Clinical samples were collected from patients attending sexual health services (~25000 attendances/year, 25% MSM) and primary care in Brighton and Hove, UK (population 273,400). Asymptomatic sexual health screens included genital and extra-genital sites according to sexual history, using nucleic acid amplification testing (NAAT). *N. gonorrhoeae* NAAT-positive individuals were recalled for microscopy, culture and susceptibility testing (MC&S) before treatment. Symptomatic individuals were sampled and treated the same day if microscopy suggested *N. gonorrhoeae*. NAAT (BD ProbeTec, BD, Franklin Lakes, USA) and culture (VCAT selective-agar, Oxoid, Basingstoke, UK) were undertaken at the Royal Sussex County Hospital. Cefixime susceptibility testing was undertaken in selected isolates by agar incorporation.[19] We stored a sweep of colonies from culture-positive selective-agar plates between 01 January 2011 and 09 March 2015 inclusive. DNA extracted using a commercial kit (QuickGene, Fujifilm, Tokyo, Japan) was sequenced using the Illumina HiSeq platform. Sequence data were mapped to a reference genome and variants identified[20] (see Supplementary Material) and compared using single nucleotide polymorphisms (SNPs) obtained from maximum likelihood phylogenetic trees, adjusted for the impact of recombination using ClonalFrameML.[21] *N. gonorrhoeae* multi-antigen sequence typing (NG-MAST, http://www.ng-mast.net) sequence types (STs) and *penA* genotypes were determined *in silico*.

Calibration and comparison collections

Calibration samples were used to determine how much variation between sequences was compatible with transmission. Sequencing pipeline reproducibility and laboratory culture stability were assessed using repeat subculture and sequencing of 115 isolates, demonstrating an error rate of 1 false SNP per 58 genomes sequenced (Supplementary Materials). The diversity present within a single clinical sample was investigated by independent subculture and sequencing of 12-14 randomly-selected bacterial colonies from six randomly-selected patient samples (total 76 colony picks). We sequenced all isolates from patients infected at multiple anatomical sites to determine within-host variation between sites. Samples from 15 contact pairs from a low incidence setting were sequenced to assess the distribution of SNPs across highly probable transmission events.[22]

Additional sequences (Table 1) were obtained from: 94 consecutive samples from London Public Health Laboratory, UK (May–August 2013); 222 archived samples from Brighton (July 2004–September 2010); 15 samples from Wales, in addition to 30 from 15 contact pairs, 45 total (June 2005–August 2006[22]); 78 samples from a ST25 outbreak in north-east England (July 2010-May 2013[23]); 196 previously published USA sequences (January 2009–December 2010[16]).

Analysis

Rates of *N. gonorrhoeae* mutation were estimated with BEAST[24] from time-scaled phylogenies. Mutation rates and the diversity observed across anatomical sites and probable transmission pairs, were used together to fit a coalescent theory-based model of the number of SNPs expected between sequences related by either direct (sampled case to sampled case) or indirect (via ≥1 intermediate [unsampled] hosts) transmission (see Supplementary Materials). We determined the plausibility of direct/indirect transmission

between any pair of samples, based on the time between samples, and the 99% prediction interval for the expected number of SNPs.

Ethics

Individual patient consent for use of anonymised bacterial isolates was not required. Research Ethics Committee (14/LO/0435) approval was obtained to collect anonymised data from patients in Brighton.

Role of the funding source

Funders had no role in study design, data collection, analysis, or writing of the report. The corresponding author had full access to the study data and final responsibility for the decision to submit for publication.

Results

Samples

Between 01 January 2011 and 09 March 2015, 3512/248627 samples were NAAT-positive for *N. gonorrhoeae*. 1267 ~~(1033 from sexual health clinics)~~/ ~~of~~ 21785 cultures were positive. Including multiple colony picks and quality control replicates, 1407/1437(98%) isolates were successfully sequenced (Table 1). Sequenced isolates were obtained from the urethra 578(41%), rectum 518(37%), pharynx 239(17%), cervix 68(5%), eye 1(0.1%), not recorded 3(0.2%).

Considering sequences >60 SNPs different from any other as distinct infections (see below), 1061 infections were identified from 907 patients (839[93%] men, 66[7%] women, 2 no gender recorded). Over 4 years, 791(87%) patients had a single infection, 91(10%) had 2 different infections, either over time or at different body sites, 17(2%) 3 infections, 5(0.6%) 4 infections, 2(0.2%) 5 infections and 1 patient 7 infections. All multiple infections were in men, apart from 1 woman with 2 infections. The median (inter-quartile range) [range] age of patients infected was 31 (24-40) [15-76] years, and 1026/1061(97%) of infections were identified by hospital or community-based sexual health clinics (Table 2). NG-MAST STs were determined *in silico* for 978/1061 (92%) infections, the most common STs were 2992, 1407, 26, 292 and 2400. The mean SNPs between isolates within these STs ranged from 29-496 (Table S1).

Transmission calibration samples

Independent subculture and sequencing of multiple colony picks from single clinical samples showed minimal diversity present within patients at the same anatomical site (Figure 1A, Supplementary Materials). Variation across anatomical sites in the same patient was assessed using 206 pairs of samples obtained within 30 days (203 pairs obtained on the same day). 171/206(83%) and 175(85%) pairs were within ≤3 and ≤6 SNPs respectively, consistent with within-host variation arising from one infection, 26(13%) were ≥1938 SNPs different, had different STs, and varying antimicrobial susceptibilities, consistent with multiple infections at different anatomical sites in a significant minority (Figure 1B, Table S2, Supplementary Materials).

Samples from 15 patient pairs (11 heterosexual, four MSM) identified through contact tracing in a low-incidence setting,[22] a median(IQR)[range] 5(1-15)[0-38] days apart, were sequenced to assess SNPs across probable transmission events with no alternative likely source of infection. 10(67%) pairs were indistinguishable and all were within ≤6 SNPs (Figure 1C).

113 Brighton patients were sampled at >1 time point, median(IQR)[range] 423(254-829)[44-2353] days apart. Only 6(5%) patients were convincingly infected with one strain over time, e.g. resulting from re-infection from an untreated partner or delay in re-attending for treatment (Figure 1D). As few patients had evidence of chronic infection, rates of *N. gonorrhoeae* mutation were estimated from time-scaled phylogenies as 3.55 (95% credibility interval 3.27-3.83) SNPs/genome/year.

To estimate the expected SNPs between direct or indirect transmission pairs, based on the time between them, this mutation rate was combined with the estimated within-host

diversity (determined from diversity across anatomical sites in the same host and the highly probable transmission pairs). The resulting Transmission Nomogram (Figure 2), shows the SNP range, for any given time interval, expected to contain 99% of all direct or indirect transmission pairs, e.g. 0-9 SNPs for samples obtained on the same day, 0-11 SNPs for samples 6 months apart, and 0-14 SNPs for samples a year apart.

Diversity in wider population

SNP differences between all pairs of first isolates from Brighton patients between 2011-2015 are shown in Figure 1E. Assessing the specificity of our Transmission Nomogram, the probability of two randomly chosen isolates being compatible with direct/indirect transmission was 0.95%(5336/562330), and restricting to isolate pairs obtained within 1 year: 1.6%(3846/246463), 90 days: 2.6%(1739/67072), and 28 days: 3.6%(856/23848). Hence, even with a conservative 99% prediction interval and samples obtained close in time, high discriminatory power was achieved. In contrast, using NG-MAST 5.2%(24669/ 477753) of all pairs of isolates shared the same ST; 8.8%(1675/19071) restricting to isolates obtained within 28 days. Where the first isolate of a pair was one of the five most common STs (42%(410/978) of all samples), the chance of a second isolate within 28 days sharing the same ST was 16.2%(1330/8204).

Genetic links between cases in Brighton

We detected extensive local transmission between Brighton cases. Comparing 1061 infections (2011-2015) to all previous sampled Brighton cases (2004 onwards), 281(26%) were indistinguishable (0 SNPs) to a previous case, and 786(74%) had evidence of a sampled direct/indirect Brighton source using our Transmission Nomogram. Most linked cases occurred close in time, suggesting possible direct transmission: of 786 linked cases in Brighton, 414(53%) were sampled within 30 days of each other, and 565(72%) within 90 days (Figure 3). However, 96/786(12%) were genetically related but sampled >1 year apart, suggesting indirect transmission or long-term asymptomatic (i.e. untreated) carriage in the source or recipient. Despite sampling all culture-positive cases in Brighton over 4 years, 275/1061(26%) infections lacked a genetically plausible Brighton source. This is not explained simply by unsampled sources for earlier cases: restricting to cases from January 2012 onwards, 205/867(24%), and January 2013 onwards, 142/628(23%), lacked a genetically plausible Brighton source.

Brighton cases related by SNP distances and time consistent with transmission were grouped into 305 clusters. Inclusion in a cluster required a case to be related to ≥1 other case in the cluster, but not necessarily to all cases in the cluster. There was evidence of sustained transmission of some lineages. 520/1061(49%) cases belonged to clusters containing ≥10 patients, the largest clusters including 110, 58, 52, 38, and 32 patients, with ST2992, ST292, ST26, ST2400, and ST2992 the dominant genotypes in each cluster respectively. Similar numbers of patients, 433(41%), belonged to smaller clusters containing ≤5 cases (Figure 4A). Sexual orientation data were not available; however, 14/21(67%) of clusters with ≥10 patients were exclusively male, including 3 of the largest clusters with 110, 52, and 32 patients.

For clusters with ≥2 patients, the first and last sampled case were median(IQR)[range] 156(31-486)[1-1425] days apart, and individual cases were 34(9-73)[0-415] days apart. In

some clusters there was evidence for multiple short-term transmissions; restricting clustering to cases diagnosed within 30 days of at least one other case, 122/1061(11%) cases were part of clusters with ≥10 cases (Figure 4B). After an initial period of sampling, the number of actively circulating lineages (defined as having ≥1 isolate in the last six months) was relatively constant (50-70)(Figure S4).

Comparison of samples across geographic locations

Unaccounted sources for Brighton cases probably include asymptomatic and unsampled NAAT/microscopy-positive-culture-negative cases. As acquisition outside Brighton is also likely, we compared sequences from Brighton at any time with sequences from other UK locations and the USA (Figures 3 and 5) We observed multiple links across geographical boundaries. Of 1273 Brighton infections (2004-2015), 225(18%) were linked using the Transmission Nomogram to another non-Brighton UK case, and 115(9%) to a USA case. Combining Brighton and comparison samples, we identified 494 clusters of genetically linked cases. Of 60 clusters including ≥1 of 94 cases sampled cross-sectionally in London (May-August 2013), 22(37%) included Brighton cases, consistent with extensive exchange of infections between these cities 50 miles apart. Sixteen clusters were isolated first in Brighton, and six first in London. Of 76 samples from an ST25 outbreak in north-east England, the majority, 52(68%), were plausibly part of a single transmission cluster with other samples from north-east England; one subsequent case of the same genetic cluster was isolated from Brighton.

Of 78 clusters including USA samples, 9(12%) also included Brighton cases. Five clusters were identified first in the USA, with 157 subsequent cases in Brighton. Four clusters found first in Brighton could be linked to 70 USA isolates, including 61 USA isolates from a cluster (cluster 65 in Figure 5, predominantly ST1407, also including 82 Brighton and 4 London isolates) carrying the mosaic *penA* XXXIV associated with reduced susceptibility to cefixime (minimum inhibitory concentration, MIC, ≥0.25mg/L[16]). Overall 121 Brighton isolates in 34 transmission clusters contained this particular mosaic *penA* allele. The earliest sample from Brighton with the mosaic *penA* XXXIV allele (in the USA linked cluster) dated from August 2007, i.e. 1.3 years before the first sequenced USA sample. We estimated the most recent common ancestor of this Brighton/USA cluster with *penA* XXXIV to be earlier: 1997 (95% credibility interval 1994-1999) and restricting to the lineage that subsequently dispersed throughout the UK and USA, 2001 (95% credibility interval 1999 – 2003; see Supplementary Material for details). Cefixime MICs determined as part of local/national surveillance for a subset of Brighton *penA* XXXIV carrying strains were 3/38(8%) ≤0.06mg/L, 29(76%) 0.125mg/L, 6(16%) 0.25mg/L (Table S3).

Discussion

Here we apply WGS to investigate gonorrhoea transmission across multiple geographic scales. We present a genomic contact tracing tool, a Transmission Nomogram, for determining plausibility of direct or indirect transmission between any two cases. It accounts for the genetic differences between cases, but also how this varies with the time between cases, providing greater precision than fixed SNP thresholds to determine transmission.

By sequencing consecutive cases in a single city, over 4 years we demonstrate significant local transmission; 74% of 1061 infections could be linked by direct/indirect transmission to an earlier Brighton case. Most transmission links related to cases sampled in the prior 90 days (72%), many (53%) within 30 days. WGS had excellent discriminatory power, even over short time periods: only 2.6% of randomly chosen pairs of cases occurring within 90 days were related using our Transmission Nomogram. We show that WGS offers increased resolution to determine transmission over NG-MAST.

Similar numbers of cases belonged to large (≥10 patients) genetic clusters (49%) and small (≤5 patients) clusters (41%), the largest cluster containing 110 patients. Many large clusters represent on-going transmission of the same lineage over long periods (Figure 5). Sustained local transmission may relate to limited numbers of "core transmitters", but might also reflect frequent partner changes involving numerous infected individuals. The most common NG-MAST types in Brighton, ST2992 and ST1407, matched those in Europe[25]; ST1407 is associated with reduced susceptibility to cefixime[26] and other antimicrobials.[25]

26% of cases were not linked to any previous case (including the initial case in each of the smaller clusters), indicating the existence of unsampled sources of infection. Several possible explanations exist. 13% of cases had mixed infections across different body sites, i.e. patients could be part of two different transmission chains simultaneously. It is possible that not all infected sites were sampled in some patients, missing transmissions where the source had a mixed infection. Other explanations include transmission from NAAT/microscopy-positive-culture-negative cases, patients not presenting despite symptoms, and asymptomatic patients.

Transmission from patients outside the immediate geographic area is another important source of infections. Although relatively few samples sequenced were from outside Brighton, 18% of Brighton infections were linked to another case elsewhere in the UK and 9% to a USA case. Previous WGS studies[16] explored the dispersion of the mosaic *penA* XXXIV, a particular mosaic allele first described in California in 2008.[27] Intriguingly we find evidence of this allele in Brighton in August 2007, as part of a large cluster of USA and UK isolates, with evidence for an earlier common ancestor, suggesting a possible origin elsewhere before it spread to the USA. Prior studies[16,28] associated this allele with reduced susceptibility to cefixime (MIC, ≥0.25mg/L), as originally described in other mosaic *penA* alleles.[29] In our dataset presence of the allele was most commonly associated with an MIC of 0.125mg/L.

Several potential applications arise from this study. This study clearly shows that efforts to control gonorrhoea should be coordinated across regional and national boundaries. WGS provides a discriminatory typing scheme, producing exchangeable data, making real-time global transmission network tracing potentially tractable. WGS can track the spread of specific resistant lineages, rather than the spread of drug-resistant phenotypes as a whole.

WGS detects links between cases not detected by traditional partner notification, e.g. between patients with multiple anonymous sexual partners. Genomic links between cases may highlight particular risk factors, enabling targeted population-based and individual interventions, including notification of contacts, e.g. via mobile phone apps used to facilitate encounters.

The short interval between cases in some clusters suggests a need for more frequent screening in high risk populations. Current UK guidelines[30] recommend 3-monthly testing for MSM at high risk, but we observed many transmission links occurred in <30 days. The number of cases without an identified source also raises questions about the proportion of high risk patients participating in screening for asymptomatic carriage.

Given its retrospective and laboratory based nature, this study has limitations. Our Transmission Nomogram cannot distinguish between direct case-to-case and indirect transmission, even where two cases have zero SNPs between them transmission via ≥1 intermediate host is still possible. Therefore, the prevalence of each genetic subtype, the time between cases, and available contact data need to be used with the Nomogram to determine the likelihood of direct, as opposed to indirect, transmission. We lack data on patient symptomatology and lack complete epidemiological contact data. Sample comparisons from outside of Brighton opportunistically use sequences obtained for other reasons, and do not systematically assess regional or international transmission. However, national antimicrobial susceptibility surveillance samples could be used for this purpose. Routine use of pathogen WGS in sexual health raises potential ethical issues. For example, WGS may allow linkage of cases without explicit consent for contact tracing. However, in existing contact tracing those notified have not explicitly consented to be approached either; and WGS is essentially an additional tool in the armamentarium of techniques available to those conducting contact tracing. The handling of WGS datasets, in particular those with patient identifiable information, must be robust to maintain patient confidentiality.

We have shown genomic contact tracing for gonorrhoea is possible. We provide a Transmission Nomogram to enable other investigators and health professionals to apply its use. WGS provides a powerful tool to guide interventions to stop the spread of drug-resistant *N. gonorrhoeae*.

Contributions

DDS, JPe, DWC, TEAP, ASW, JPa and DWE designed the study. JPe, KC, MC, FC, GD, JD, DRT, KF, AW, YD collected the isolates and provided sample metadata. JPe and KC did laboratory work and DNA extraction. DDS and DWE did the bioinformatic analysis. DDS, JPe, DJW, XD, TEAP, ASW, JPa and DWE analysed the data. DDS and DWE prepared the figures. DDS and DWE prepared the first draft of the manuscript which was revised by all authors.

Funding

The research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with Public Health England (PHE) [HPRU-2012-10041], and the NIHR Oxford Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or PHE. DJW is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (grant 101237/Z/13/Z). DWC and TEAP are NIHR senior investigators. DWE is a NIHR clinical lecturer.

Declaration of interests

The authors have no conflict of interest.

Acknowledgements

Angela Dunne and Andrew Bexley for assisting with case recruitment in Brighton; Judith Stonebridge, Oluseyi Hotonu, Jayanta Sarma, Samuel Moses for assistance with study samples from north-east England; Gwenda Hughes and Cathy Ison for assistance from Public Health England; David Trees and Gail Bolan for assistance from the Centers for Disease Control and Prevention.

Data deposition

Sequences generated during this study can be found on the NCBI short read archive under BioProject PRJNA315363, http://www.ncbi.nlm.nih.gov/bioproject/315363.

Figure Legends

Figure 1. Transmission calibration sampling frames. Panel A shows the genetic variation within six randomly chosen clinical samples, 12-14 colonies were sequenced independently. Within each clinical sample sequences from the first colony chosen were compared to all other colonies sequenced. On the right-hand side, each colour represents a different clinical sample. The area of the circles is proportional to the number of colonies with identical genome sequences. Lines between circles represent the numbers of SNPs between colonies. In 5 samples all sequences were identical, shown as a single circle. Panel B shows the diversity present across different anatomical sites in the same patient. Panel C shows the diversity present between highly probable transmission pairs. Panel D shows the variation in the same patient over time. Panel E shows the diversity between different patients in Brighton. All first samples from each infection in each patient were compared pairwise.

Figure 2. Transmission Nomogram. SNPs expected between direct or indirect transmission pairs for varying time between samples are shaded (99% prediction interval). The dotted line shows the mean number of SNPs. The upper panel shows expected numbers of SNPs over the longest interval possible between samples in the study. Of 1061 distinct infections, only 2 (0.2%) had a potential source with lower than the expected number of SNPs, 0 SNPs after 466 days, and 1 SNP after 686 days. The lower panel shows the expected number of SNPs over a time between samples of up to 1 year.

Figure 3. Percentage of Brighton infections genetically linked to a previous sampled case by maximum time between cases. Brighton vs. Brighton compares cases in Brighton (2011-2015) to all previous Brighton cases (2004 onwards). To avoid double counting of cases, cases were only compared to previous cases, accepting sampling dates may not indicate the direction of transmission. In the Brighton vs. UK and Brighton vs. USA plots all cases from Brighton (2004-2015) were compared to all cases from the rest of the UK or USA respectively, independent of the order of sampling.

Figure 4. Brighton clusters of genetically linked cases. Cases within Brighton were clustered based on those related by SNP distances and time compatible with transmission. Panel A shows clusters for 1061 cases between January 2011 and March 2015. Panel B restricts clustering to where sampling of consecutive cases within a cluster occurred within 30 days.

Figure 5. Genetic clusters within Brighton, UK and USA. Each genetic cluster contains all cases related by a number of SNPs and time compatible with transmission. Each genetic cluster is plotted on its own horizontal line, with individual cases indicated as dots. For ease of visualisation, clusters arising from January 2011 are shown separately on the right-hand side. Samples obtained in Brighton in 2004 and 2005 were collected within a 2-month interval, but the exact collection dates were not available. These samples have been randomly distributed throughout the 2 months of sampling. Similarly, only the month and year of collection was known for the USA samples, and a random day has been assigned.

Research in Context

Evidence before this study

We searched PubMed for publications up until 15 March 2016 with the terms ((Neisseria gonorrhoeae) OR Gonorrhoea) AND (sequencing OR (molecular epidemiology)), references and subsequent citations (identified using Google Scholar) were also reviewed.

Previous studies have used whole genome sequencing (WGS) of *Neisseria gonorrhoeae* to investigate the spread of drug resistant strains at a national level in the USA and Canada, and used WGS to investigate relatively small local outbreaks.

No study to date has systematically applied WGS to quantify the extent of local transmission and what proportion of cases might have originated from other regions or countries.

Added value of this study

We present a tool for genomic contact tracing of *N. gonorrhoeae*: based on multiple sampling frames, we derive a Transmission Nomogram that can be used to determine if direct or indirect transmission between any two cases is plausible using genetic data and the time between the cases being diagnosed.

From sequencing all culture-positive *N. gonorrhoeae* infections from a single city, Brighton, UK, over a 4 year period, we demonstrate extensive local transmission, with sustained transmission of some lineages, and related cases typically occurring a few days or weeks apart. However, a quarter of cases could not be linked to a local direct/indirect source.

We observed multiple related samples across geographic locations, linking samples from Brighton to other UK locations and to cases from the USA, including to a lineage carrying the mosaic *penA* XXXIV associated with reduced cefixime susceptibility.

We show that 13% of cases have distinct strains at different anatomical sites, i.e. that these patients simultaneously belong to multiple transmission networks.

Implications of all the available evidence

Genomic contact tracing has the potential to inform control of gonorrhoea transmission at a local, national and international level.

Improved local control may depend on more regular screening and treatment of high risk individuals.

Genomic led contact tracing has significant potential amongst patients with multiple anonymous sexual partners, where traditional partner notification is very difficult.

WGS provides a discriminatory typing scheme, producing readily exchangeable data, making global contact tracing and tracking of specific resistant lineages possible.

References

1 Newman L, Rowley J, Vander Hoorn S, *et al.* Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting. *PLoS ONE* 2015; **10**: e0143304.

2 Barbee LA. Preparing for an Era of Untreatable Gonorrhea. *Curr Opin Infect Dis* 2014; **27**: 282–7.

3 Public Health England. Health Protection Report: Sexually transmitted infections and chlamydia screening in England, 2014. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/437433/hpr2215_STI_NCSP_v6.pdf (accessed March 13, 2016).

4 Bignell C, FitzGerald M, (Guideline Development Group). UK national guideline for the management of gonorrhoea in adults, 2011. *Int J STD AIDS* 2011; **22**: 541–7.

5 Harrison WO, Hooper RR, Wiesner PJ, *et al.* A trial of minocycline given after exposure to prevent gonorrhea. *N Engl J Med* 1979; **300**: 1074–8.

6 Cohen MS, Cannon JG. Human experimentation with Neisseria gonorrhoeae: progress and goals. *J Infect Dis* 1999; **179 Suppl 2**: S375–9.

7 Marrazzo JM, Apicella MA. 214 - Neisseria gonorrhoeae (Gonorrhea). In: Blaser JEBDJ, ed. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition). Philadelphia: Churchill Livingstone 2015: 2446–2462.e3.

8 Korenromp EL, Sudaryo MK, de Vlas SJ, *et al.* What proportion of episodes of gonorrhoea and chlamydia becomes symptomatic? *Int J STD AIDS* 2002; **13**: 91–101.

9 Kent CK, Chaw JK, Wong W, *et al.* Prevalence of rectal, urethral, and pharyngeal chlamydia and gonorrhea detected in 2 clinical settings among men who have sex with men: San Francisco, California, 2003. *Clin Infect Dis* 2005; **41**: 67–74.

10 Handsfield HH, Lipman TO, Harnisch JP, Tronca E, Holmes KK. Asymptomatic gonorrhea in men. Diagnosis, natural course, prevalence and significance. *N Engl J Med* 1974; **290**: 117–23.

11 Garnett GP, Mertz KJ, Finelli L, Levine WC, Louis MES. The transmission dynamics of gonorrhoea: modelling the reported behaviour of infected patients from Newark, New Jersey. *Philos Trans R Soc Lond B Biol Sci* 1999; **354**: 787–97.

12 Hazel A, Marino S, Simon C. An anthropologically based model of the impact of asymptomatic cases on the spread of Neisseria gonorrhoeae. *J R Soc Interface* 2015; **12**: 20150067–7.

13 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**: 365–9.

14 Marcus JL, Bernstein KT, Kohn RP, Liska S, Philip SS. Infections Missed by Urethral-Only Screening for Chlamydia or Gonorrhea Detection Among Men Who Have Sex With Men. *Sex Transm Dis* 2011; **38**: 922–4.

15 Chisholm SA, Wilson J, Alexander S, *et al.* An outbreak of high-level azithromycin resistant Neisseria gonorrhoeaein England. *Sex Transm Infect* 2015; : sextrans–2015–052312.

16 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**: 220–6.

17 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* 2014; **53**: 191–200.

18 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; : JCM.03195–15.

19 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**: 2509–12.

20 Eyre DW, Cule ML, Wilson DJ, *et al.* Diverse Sources of C. difficileInfection Identified on Whole-Genome Sequencing. *N Engl J Med* 2013; **369**: 1195–205.

21 Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 2015; **11**: e1004041.

22 Cole MJ, Thomas DR, Chisholm SA, *et al.* Molecular epidemiology of gonorrhoea in Wales (UK). *Sex Transm Infect* 2012; **89**: sextrans–2012–050663–272.

23 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; : sextrans–2015–052303.

24 Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012; **29**: 1969–73.

25 Chisholm SA, Unemo M, Quaye N, *et al.* Molecular epidemiological typing within the European Gonococcal Antimicrobial Resistance Surveillance Programme reveals predominance of a multidrug-resistant clone. *Euro Surveill* 2013; **18**.

26 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**: 762–8.

27 Pandori M, Barry PM, Wu A, *et al.* Mosaic penicillin-binding protein 2 in Neisseria

gonorrhoeae isolates collected in 2008 in San Francisco, California. *Antimicrob Agents Chemother* 2009; **53**: 4032–4.

28 Hess D, Wu A, Golparian D, *et al.* Genome sequencing of a Neisseria gonorrhoeae isolate of a successful international clone with decreased susceptibility and resistance to extended-spectrum cephalosporins. *Antimicrob Agents Chemother* 2012; **56**: 5633–41.

29 Ameyama S, Onodera S, Takahata M, *et al.* Mosaic-like structure of penicillin-binding protein 2 Gene (penA) in clinical isolates of Neisseria gonorrhoeae with reduced susceptibility to cefixime. *Antimicrob Agents Chemother* 2002; **46**: 3744–9.

30 British Association for Sexual Health and HIV. Recommendations for testing for sexually transmitted infections in men who have sex with men. http://www.bashh.org/documents/BASHH%20Recommendations%20for%20testing%20for%20STIs%20in%20MSM%20-%20FINAL.pdf (accessed Feb 18, 2016).