Genomic analysis and comparison of two gonorrhoea outbreaks

Xavier Didelot\textsuperscript{a,\#}, Janina Dordel\textsuperscript{b,c}, Lilith K Whittles\textsuperscript{a}, Caitlin Collins\textsuperscript{a}, Nicole Bilek\textsuperscript{a,d}, Cynthia J Bishop\textsuperscript{a,e}, Peter J White\textsuperscript{a,f,g,h}, David M Aanensen\textsuperscript{a,i}, Julian Parkhill\textsuperscript{b}, Stephen D Bentley\textsuperscript{b}, Brian G Spratt\textsuperscript{a}, Simon R Harris\textsuperscript{b,\#}

\textsuperscript{a} Department of Infectious Disease Epidemiology, Imperial College, London, UK
\textsuperscript{b} Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK
\textsuperscript{c} Department of Biology, Drexel University, Philadelphia, USA
\textsuperscript{d} South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, South Africa
\textsuperscript{e} International AIDS Vaccine Initiative, Human Immunology Laboratory, Imperial College, London UK
\textsuperscript{f} Modelling and Economics Unit, Centre for Infectious Disease Surveillance and Control, Public Health England, London, UK
\textsuperscript{g} MRC Centre for Outbreak Analysis and Modelling, School of Public Health, Imperial College London, UK
\textsuperscript{h} NIHR Health Protection Research Unit in Modelling Methodology, School of Public Health, Imperial College London, UK
\textsuperscript{i} Centre for Genomic Pathogen Surveillance, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK

Running head: Genomic analysis of gonorrhoea outbreaks

\# Address correspondence to Xavier Didelot, x.didelot@imperial.ac.uk or Simon R Harris, sh16@sanger.ac.uk
Abstract

Gonorrhoea is a sexually transmitted disease causing growing concern, with a substantial increase in reported incidence over the past few years in the United Kingdom and rising levels of resistance to a wide range of antibiotics. Understanding its epidemiology is therefore of major biomedical importance, not only on a population scale but also at the level of direct transmission. However, the molecular typing techniques traditionally used for gonorrhoea infections do not provide sufficient resolution to investigate such fine scale patterns. Here we sequenced the genomes of 237 isolates from two local collections of isolates from Sheffield and London, each of which was resolved into a single type using traditional methods. The two datasets were selected to have different epidemiological properties: the Sheffield data were collected over six years from a predominantly heterosexual population, whereas the London data were gathered within half a year and strongly associated with men who have sex with men. Based on contact tracing information between individuals in Sheffield, we found that transmission is associated with a median time to most recent common ancestor of 3.4 months with an upper bound of 8 months which we used as criterion to identify likely transmission links in both datasets. In London, we found that transmission happened predominantly between individuals of similar age, sexual orientation and location, and also with the same HIV serostatus, which may reflect serosorting and associated risk behaviours. Comparison of the two datasets suggests that the London epidemic involved about ten times more cases than the Sheffield outbreak.

Importance

The recent increases in gonorrhoea incidence and antibiotic resistance is cause for public health concern. Successful intervention requires a better understanding of transmission patterns, which is not uncovered by traditional molecular epidemiology techniques. Here we studied two outbreaks that took place in Sheffield and London, United Kingdom. We show that whole genome sequencing provides the resolution to investigate direct gonorrhoea transmission between infected individuals. Combining genome sequencing with rich epidemiological information about infected individuals reveals the importance of several transmission routes and risk factors, which can be used to design better control measures.
Introduction

Gonorrhoea is a sexually transmitted disease (STD) caused by the bacterium *Neisseria gonorrhoeae*. In the UK, gonorrhoea is one of the most common bacterial STDs, and its reported incidence has markedly increased since 2008 in both men and women, reaching a total of 35,000 diagnosed cases in 2014 (1). Treatment with antimicrobials is usually successful, but increased resistance to many front-line antibiotics has recently been observed (2, 3). Implementing effective control measures to mitigate the spread of gonorrhoea is difficult due to a lack of understanding of the importance of the complex transmission routes and reservoirs (4). Traditional epidemiological studies are complicated by the facts that gonorrhoea can be carried asymptomatically for months in about 10% of men and 50% of women (5), and in the UK infects disproportionately MSM and young heterosexuals of black ethnicity, and features significant geographical and temporal variations (6).

Molecular epidemiology approaches, for example *opa* typing (7), multiantigen sequence typing (NG-MAST) (8) or Multi-Locus Sequence Typing (MLST) (9), have proved helpful to demonstrate that multiple strains often circulate simultaneously within a host population, to detect emerging, often resistant, clones and to identify clusters of individuals infected with the same strain. Primarily, these typing schemes are useful to rule out transmission links between individuals who carry different types and do not have sufficient resolution to shed light on fine patterns of transmission between individuals carrying the same type.

In recent years, bacterial epidemiology has started to be transformed by the availability of fast, affordable whole genome sequencing, which can help identify transmission links sometimes even at the level of direct transmission between individuals (10–12). The value of genomic data to investigate local outbreaks has been demonstrated for several bacterial pathogens, including *Staphylococcus aureus* (13, 14), *Clostridium difficile* (15, 16) and *Mycobacterium tuberculosis* (17, 18). In a study of gonococcal epidemiology, whole genome sequencing has recently been applied to a collection of isolates from across the USA, showing important structuring of the pathogen population with both geography and sexual orientation of the hosts (19).
To test the usefulness of gonococcal genome sequencing to track transmission at the finer scale of outbreaks occurring within a city, we selected two local isolate collections with different epidemiological properties, both of which have been previously described without genomic data. The Sheffield isolate collection was assembled between 1995 and 2000, at the single genitourinary clinic in the city, from a mostly heterosexual population in which contact tracing was performed. Multiple strains of the pathogen have been found to circulate using *opa* typing (20, 21) and NG-MAST typing (22), and in both cases molecular typing was correlated with known sexual contact links. We applied whole-genome sequencing to 132 isolates with the most prevalent NG-MAST type ST12.

The London collection consists of 2045 isolates sampled between June and November 2004 from 13 major sexual health clinics throughout London, representing 54% of the 3754 cases reported in London at that time (23, 24). Contact tracing information was not available, but rich metadata were recorded for each infected individual, including ethnic background, HIV status where known, postcode and reporting clinic. All isolates were previously typed using NG-MAST, revealing the co-existence of a large number of strains, some of which were mostly MSM-associated and others mostly heterosexual-associated (23). 45% of infected individuals had one of 21 major strains, amongst which NG-MAST type ST225 was the most strongly MSM-associated (with 92% of cases being MSM) and not geographically clustered within London (24). We applied whole-genome sequencing to 105 isolates of ST225.

Our aim was to assess the potential benefits that whole-genome sequencing data could provide for epidemiology within single NG-MAST types. A relatively recent history of transmission would be expected due to the shared type and the localized sampling frames. The epidemiological background of the two studies is very different, with the Sheffield dataset coming from a single STD clinic in a city more than ten times smaller than the London dataset which was collected at 13 clinics. Furthermore, the Sheffield isolates are heterosexual-associated whereas the London isolates are MSM-associated. Consequently, this will also allow us to investigate how differences in variation within the two datasets reflect their underlying context.
Results and Discussion

Genomic analysis of the Sheffield dataset

We sequenced the genomes of 132 isolates from Sheffield dated from 1995 to 2000, all of which had the same genotype, ST12, as defined by NG-MAST (Table S1). Few recombination events were detected using Gubbins (25), with a relative effect of recombination compared to mutation of $r/m=0.04$ which was much lower than previous reports based on species-wide diversity (19, 26). This may reflect a difference in the effect of recombination when measured at different scales or diversity, or between different lineages, as previously reported in other bacterial pathogens such as \textit{Staphylococcus aureus} (27), \textit{Clostridium difficile} (15) or \textit{Streptococcus pneumoniae} (28). The few detected recombination events occurred mainly in genes coding for outer membrane proteins that undergo antigenic variation and have previously been described as highly recombinant, for example \textit{opa} and \textit{pil} genes (29). After removing repetitive regions and recombinant regions, a total of 156 variable sites were found to distinguish the genomes from each other. A maximum likelihood phylogeny was constructed using these data and showed a strong temporal signal based on the correlation between root-to-tip distances and isolation dates for all leaves in the phylogeny (Figure S1, $R^2=0.49$).

We therefore applied the Bayesian evolutionary analysis software BEAST (30) to this data in order to reconstruct a timed phylogeny (Figure 1). The molecular clock rate was estimated to be $1.41\times10^{-6}$ SNPs per site per year with 95% credibility interval $[1.15\times10^{-6};1.70\times10^{-6}]$. This rate was equivalent to $3.05\ [2.47;3.67]$ mutations per year across the genome and is in good agreement with previous estimates in \textit{N. gonorrhoeae} (19, 26) and the related species \textit{N. meningitidis} (31), and was around the middle of the range of reported values for other bacterial pathogens (32, 33).

Known sexual contacts were significantly clustered on the timed phylogeny (Figure 1; Permutation test; Figure S2; $p$-value<$10^{-4}$). For the 25 pairs of known sexual contacts (Table S1), the median time to the most recent common ancestor (TMRCA) of the isolated bacteria was 3.4 months, with an interquartile range from 2.3 to 5.1 months (Figure 1 inset). When considering direct transmission between individuals sampled at roughly the same time, with no diversity being transmitted (due to a strong transmission bottleneck), the most recent common ancestor of the two
sampled genomes would have existed within the pathogen population of the infector (34–36). The TMRCA is therefore a lower bound for the time from infection to sampling of the donor, and an upper bound for the time from infection to sampling of the recipient, so that the mean TMRCA for pairs of contacts is expected to be approximately equal to the average duration of infection. Our result is in good agreement with previous estimates ranging from 2 to 6 months for the average duration of gonorrhoea infection in several modelling studies (37–39). This duration is significantly more than the few days reported for the incubation period in male experimental challenges (40), even accounting for the few additional days taken from symptom onset to care-seeking (41). However, since the Sheffield population is predominantly heterosexual, the mean duration of infection is an average between men and women, and takes into account the fact that a large fraction of women as well as a smaller proportion of men can remain asymptomatic for extended periods of time (42).

Since all pairs of known sexual contacts have a TMRCA of less than 8 months, we decided to use this value as the maximum threshold for the TMRCA between two individuals who have directly infected each other. This threshold should conservatively rule out transmission for pairs of genomes with a higher TMRCA, because all pairs of known sexual contacts fulfil this criterion and yet not all of them are direct transmission pairs since the data include four triplets and one quadruplet of sexual contacts (Figure 1 inset). Pairs of genomes sampled more than 8 months apart are unlikely to be transmission links. We compared all pairs of genomes sampled within 3 months of each other, which applies to all 25 pairs of known sexual contacts, and found that 24% (398/1632) had a TMRCA less than 8 months, including the known sexual contacts (Figure S3). Since all these pairs were sampled within a short period of time in the same city, and had the same NG-MAST type, a non-genomic analysis could not rule out transmission for any of them, but with the help of genomic data we can confidently rule out transmission for the majority.

To complement the phylogenetic analysis of the Sheffield data, we applied Outbreaker, which allows the direct reconstruction of a transmission tree representing transmission pathways within a sample, including the possibility of unsampled missing links in the transmission chains (43). Known sexual contacts
clustered together on the reconstructed transmission tree, with at most three intermediates in the transmission chain between contacts (Figure 2). When looking at all pairs of individuals in the Sheffield dataset sampled within 3 months of each other, a strong correlation was found between their TMRCA in the BEAST tree and whether or not Outbreaker inferred them to be linked (Figure S4; Kruskal-Wallis Test KWT; p-value<10^{-15}; means TMRCA of 0.28 and 1.54 years for pairs linked and unlinked by Outbreaker, respectively), indicating a good agreement between the two approaches.

Genomic analysis of the London dataset
We sequenced the genomes of 105 isolates sampled in London between June and November 2004, all of which were representatives of NG-MAST type ST225 (Table S2). Significant recombination was detected using Gubbins (25) with r/m = 1.17, which was higher and in better agreement with previous reports (19, 26) than the value for the Sheffield data. Most of the recombination events occurred on deep branches and mainly affected outer membrane and pilus genes, repeat regions and a prophage. A total of 167 variable sites were found after removing repetitive and recombinant regions.

The temporal signal was weak in this dataset due to the short sampling span of only 6 months, so that the molecular clock rate could not be directly estimated with confidence from this dataset. The distribution of root-to-tip distance versus sampling dates was, however, compatible with that expected under the rate estimated for the Sheffield dataset (Figure S5). A timed tree was therefore produced using BEAST (30) but forcing the molecular clock rate to be equal to that estimated for the Sheffield data (Figure 3). We also tried to apply Outbreaker to this dataset as we did for the Sheffield dataset, but the results were not biologically meaningful due to the shorter sampling interval for the London data which is incompatible with Outbreaker’s assumption of a constant sampling density from beginning to end of an outbreak (43).

The analysis of the Sheffield data suggested a threshold of at most 8 months for the TMRCA compatible with direct transmission from one individual to another. The London data comes from a predominantly MSM population, but the same threshold
should be applicable since the asymptomatic frequency of anorectal cases among
MSM is ~80% (44) similar to the ~75% in heterosexual females (42). We therefore
applied the same criterion to all pairs of individuals in the London data who were
sampled within 3 months of each other. We found that transmission was a possibility
for 4% (165/4251) of such pairs (Figure S6). Likely transmission links were
significantly associated with shorter geographical distances between postcodes of
residence in London (KWT; p=6x10^{-3}; means 10.3 vs 19.1 km) as would be expected
if sexual partnerships tend to be geographically clustered. The age difference was
significantly lower for linked individuals than for unlinked individuals (KWT; p=1.3x10^{-4};
means 7.5 vs 9.8 years), likely reflecting a tendency for sexual partners to be
about the same age, as found by behavioural surveys such as Natsal-3 (45). Linked
individuals reported the same sexual orientation more often than expected by
chance for this sample of individuals (Fisher Exact Test FET; p=6x10^{-3}). This
significant similarity in location, age and sexuality of linked individuals suggests that
the transmission analysis which was based only on pathogen genome similarity
successfully captured correct transmission links. Amongst the 105 individuals, 24
were HIV positive, 47 were HIV negative and 33 had unknown HIV status. Relative
to the number of transmission links involving an HIV positive and an HIV negative
individual, there was an excess of transmission links found between pairs of
individuals who were both HIV positive (FET; p=2.6x10^{-3}) and simultaneously a
dearth of transmission links when both individuals were HIV negative (FET; p=0.03).

Amongst the 105 individuals in the London sample, the genomic analysis above
suggested that 29 did not have any likely transmission link, 19 had one, 8 had two,
14 had three and the remaining 35 had between four and eleven links. This number
of likely transmission links for a given individual was not significantly associated with
sexual orientation (KWT; p=0.13; means 3.19 vs 1.28 links), reported sex abroad
(KWT; p=0.41; means 2.37 vs 3.41 links) or reported previous gonorrhoea (KWT;
p=0.33; means 3.52 vs 2.89 links) but was significantly increased for individuals who
were HIV positive (KWT; p=1.6x10^{-4}; means 5.25 vs 2.08 links), in accordance with
the observation above of a higher number of links between HIV positive individuals.
This observation may indicate a role of HIV infection on susceptibility and
transmissibility of gonorrhoea (46, 47), or may be caused by both HIV and
gonorrhoea infection being linked with the same high-risk behaviours, or both. It is
also probable that HIV positive MSM have on average more partners and engage in serosorting and riskier behaviours (48, 49). A significant correlation was found between the number of transmission links and the reported number of sexual partners in the UK in the last three months (Spearman’s rank correlation test; rho=0.4; p=2x10^{-5}). This correlation was partly explained by a higher number of reported partners for HIV positive individuals (KWT; p=2.2x10^{-3}; means 8.03 vs 2.61), but a correlation was also suggested between number of links and number of partners when analysing HIV positive and HIV negative individuals separately (rho=0.5; p=0.01 and rho=0.3; p=0.04, respectively).

**Comparison between the Sheffield and London datasets**

Having analysed the Sheffield and London datasets separately, we now turn to comparative analysis between them. The temporal sampling frames are different in the two datasets since sampling in Sheffield happened over 6 years whereas the London data cover only 6 months. It is therefore not possible for example to compare the frequency with which a putative transmission donor is found for cases in both datasets, because this frequency would be expected to be smaller in London since donors would more often have been reported before sampling started. However, direct comparisons become possible by focusing on pairs of isolates sampled at approximately the same time, since the relationships between such pairs should not be affected by differences in sampling frame duration. Having reconstructed dated phylogenies for both datasets (Figures 1 and 3), we can compare the distributions of TMRCA for all pairs of isolates sampled within 3 months of each other (Figure 4). As previously noted, a larger proportion of pairs have a TMRCA under 8 months in Sheffield relative to London (24% vs 4%), and overall the London pairs have an average TMRCA of 4.4 years compared to 1.5 years for the Sheffield pairs, with the difference being highly significant (KWT; p<1x10^{-15}). More pairs of individuals in Sheffield are therefore likely to have infected each other compared to in London, and this result is robust to the choice of other TMRCA thresholds than 8 months for when to rule out transmission (Figure 4).

This difference between the Sheffield and London datasets can be explained from a population genetics viewpoint by a pathogen effective population size ($N_e$) roughly three times higher in London relative to Sheffield (50). Using a metapopulation
analogy, the effective population size of a pathogen can be shown to be proportional to the number of infected individuals and inversely proportional to the transmission rate (51–53). A recent study estimated that gonorrhoea spreads two to three times faster in MSM than heterosexual networks (54), which is consistent with a lower proportion of asymptomatic cases in men compared to women, and symptoms that are more likely to result in rapid care seeking (5, 42). This difference should contribute to a lower $N_e$, suggesting that the number of infected individuals in London at a given time may have been six to nine times higher than in Sheffield, for the two specific lineages of gonorrhoea studied here, namely NG-MAST ST12 in Sheffield and ST225 in London. This estimation is in broad agreement with the similar number of reported cases in Sheffield and London, despite the sampling interval being about ten times longer in the former.

**Conclusions**

We have presented the first genomic epidemiology investigation of two large, localized outbreaks of gonorrhoea – one including 132 isolates from Sheffield collected over six years amongst a mostly heterosexual population and the other including 105 isolates from London gathered over half a year from a mainly MSM population. We showed that whole genome sequencing can be used to predict person-to-person transmission events, which we combined with epidemiological information about the infected individuals to reveal patterns of transmission and infection risk factors. Importantly, both isolate collections had previously been assigned to single NG-MAST types, highlighting again the superior resolution provided by genomic data for molecular epidemiology. Pairs of cases detected in the same town, within three months of each other and carrying the same molecular type would be considered to be 'linked' using traditional molecular epidemiology, and yet we showed that genome sequencing ruled out transmission for the majority of such pairs in both outbreaks. Furthermore, the proportions of pairs of cases for which transmission was found to be likely using genomic comparisons was very different in Sheffield (24%) and in London (4%), and we estimated that the number of cases was six to nine times higher in London compared to Sheffield. Molecular typing using NG-MAST remains useful to identify cases that are part of the same local outbreak, but our results clearly show that a type does not represent a uniform epidemiological entity. As cost effectiveness and turnaround time of whole genome sequencing
continue to improve, it will become an increasingly important tool in the investigation and control of gonococcal outbreaks, but unlocking its full potential requires the simultaneous application of traditional epidemiological techniques such as the use of questionnaires or contact tracing.

**Materials and Methods**

*Bacterial isolates*

From the previously described Sheffield collection (20–22) which was sampled between 1995 and 2000, we included 132 out of 140 isolates of the most prevalent NG-MAST ST12. Table S1 contains the list of these isolates and associated metadata. From the previously described London collection (23, 24) which consists of 2045 isolates sampled between June and November 2004, we included 105 out of 124 isolates from the most MSM-associated NG-MAST ST225. Table S2 contains the list of these isolates and associated metadata. Exclusion of 8 and 19 isolates from the Sheffield and London collections respectively was due to loss of samples, loss of associated metadata or failure to grow or sequence.

*Whole genome sequencing*

For each of the two datasets the oldest available isolate was used for the production of a reference genome using 454. A 3 kb library was prepared and sequenced on a ¼ plate run. For the London reference isolate this produced 139.2 Mbp. *De novo* assembly was carried out using Newbler resulting in 12 scaffolds with a total length of 2.14 Mbp. The sequencing of the Sheffield isolate produced 91.1 Mbp and 13 scaffolds with a total length of 2.14 Mbp. The length of these two reference genomes is in good agreement with previous reports of lengths of *N. gonorrhoeae* genomes (19, 26, 55). Both assemblies were annotated using the RAST annotation pipeline (56) and manually curated. Prophages were detected using the online tool PHAST (57). Further mobile genetic elements were detected through genome comparisons with already published *N. gonorrhoeae* genomes and BLAST analysis. Repeats were detected using the EMBOSS tools einverted and equicktandem (58).

The genomes of all Sheffield isolates and all London isolates were sequenced on an Illumina HiSeq2000 with 100 bp paired-end reads. Paired-end reads were mapped against the corresponding reference genome using SM AlTv0.7.5
(http://www.sanger.ac.uk/science/tools/smal-0) with subsequent realignment around indels using GATKv1.5.9 (59). Single nucleotide polymorphisms (SNPs) were called as previously described (60). Recombination was detected using Gubbins (25) and recombinant SNPs were excluded. Furthermore, SNPs within mobile genetic elements as well as repetitive regions were also excluded. This resulted in alignments containing 156 and 167 SNPs for Sheffield and London, respectively.

**Phylogenetic analysis**

Maximum-likelihood phylogenies were constructed using phyml (61). In the Sheffield tree we included a single genome from London, and in the London tree we included a single genome from Sheffield, to be used as outgroups in order to root the trees. Comparing root-to-tip distances with isolation dates for each genome revealed a strong temporal signal in the Sheffield data (Figure S1) but not in London (Figure S5) due to a short sampling interval. The Bayesian evolutionary analysis software BEAST (30) was applied to both datasets, with the molecular clock of the London dataset being forced equal to the rate estimated for the Sheffield dataset. The resulting timed phylogenies are shown in Figures 1 and 3. We used BEAST version 1.8.2 with the default HKY substitution model, the default coalescent model with constant population size and a strict clock model with rate prior distribution Exponential(1) for the Sheffield data and fixed rate for the London data. For each of the two datasets we performed four runs of $10^7$ iterations which were compared in Tracer (http://beast.bio.ed.ac.uk/tracer) to confirm convergence.

**Transmission analysis**

We sought a value for the threshold on the time to the most recent common ancestry of two genomes beyond which direct transmission between the two corresponding individuals can be discounted (15). Based on known sexual contacts in the Sheffield dataset, we estimated that 8 months was an appropriate value for this threshold. We applied this criterion to all pairs of genomes sampled within 3 months of each other in both Sheffield and London to determine for each pair whether transmission was likely or not. Pairs of genomes for which transmission was likely are shown in Figures S3 and S6 for the Sheffield and London datasets respectively. We also used Outbreaker, a Bayesian approach for reconstructing a transmission tree from dated genetic data (43). For the generation time we used a discretized Gamma distribution
with a mean of 90 days and a standard deviation of 40 days, which has high variance reflecting our lack of knowledge of exact values (62). The Outbreaker output for the Sheffield data is shown in Figure 2 and compared with the BEAST-based approach in Figure S4. Outbreaker did not produce meaningful results for the London data due to the short sampling frame which implies that the donor of many cases would have occurred and been reported before the sampling frame.
Data availability

Sequence data were deposited to the European Nucleotide Archive. Accession numbers for the Illumina data of each isolate can be found in Tables S1 and S2.

Funding information

This work was supported by the UK National Institute for Health Research Health Protection Research Unit in Modelling Methodology at Imperial College London in partnership with Public Health England (grant HPRU-2012-10080 to DMA, XD and PJW), the UK Medical Research Council (grant MR/N010760/1 to XD and grant MR/K010174/1 to PJW) and the Wellcome Trust (grant 089472 to BGS). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.
1. Public Health England. 2015. Sexually transmitted infections and chlamydia screening in England: 2014. Public Heal Engl Infect Rep 9:1–20.

2. Goire N, Lahra MM, Chen M, Donovan B, Fairley CK, Guy R, Kaldor J, Regan D, Ward J, Nissen MD, Sloots TP, Whiley DM. 2014. Molecular approaches to enhance surveillance of gonococcal antimicrobial resistance. Nat Rev Microbiol 12:223–9.

3. Chisholm SA, Wilson J, Alexander S, Tripodo F, Al-Shahib A, Schaefer U, Lythgow K, Fifer H. 2015. An outbreak of high-level azithromycin resistant Neisseria gonorrhoeae in England. Sex Transm Infect sextrans–2015–052312.

4. Grad YH, Goldstein E, Lipsitch M, White PJ. 2016. Improving control of antibiotic resistant gonorrhea by integrating research agendas across disciplines: key questions arising from mathematical modeling. J Infect Dis 213:883–890.

5. Bignell C, FitzGerald M. 2011. UK national guideline for the management of gonorrhoea in adults, 2011. Int J STD AIDS 22:541–547.

6. Le Polain De Waroux O, Harris RJ, Hughes G, Crook PD. 2014. The epidemiology of gonorrhoea in London: a Bayesian spatial modelling approach. Epidemiol Infect 142:211–20.

7. O’Rourke M, Ison CA, Renton AM, Spratt BG. 1995. Opa-typing: a high resolution tool for studying the epidemiology of gonorrhoea. Mol Microbiol 17:865–875.

8. Martin IMC, Ison CA, Aanensen DM, Fenton KA, Spratt BG. 2004. Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. J Infect Dis 189:1497–1505.

9. Maiden MC, Bygraves JA, Feil EJ, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 95:3140–5.

10. Didelot X, Gardy J, Colijn C. 2014. Bayesian inference of infectious disease transmission from whole genome sequence data. Mol Biol Evol 31:1869–1879.

11. Croucher NJ, Didelot X. 2015. The application of genomics to tracing bacterial pathogen transmission. Curr Opin Microbiol 23:62–67.

12. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. 2016. Within-host evolution of bacterial pathogens. Nat Rev Microbiol 14:150–162.

13. Harris SR, Cartwright EJP, Torok ME, Holden MTG, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J, Peacock SJ, Török ME, Holden MTG, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J, Peacock SJ. 2013. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis 13:130–136.

14. Uhlemann A-C, Dordel J, Knox JR, Raven KE, Parkhill J, Holden MTG,
Peacock SJ, Lowy FD. 2014. Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. Proc Natl Acad Sci U S A 111:6738–43.

15. Didelot X, Eyre DW, Cule M, Ip CLCC, Ansari MA, Griffiths D, Vaughan A, O’Connor L, Golubchik T, Batty EM, Piazza P, Wilson DJ, Bowden R, Donnelly PJ, Dingle KE, Wilcox M, Walker AS, Crook DW, Peto TE, Harding RM. 2012. Microevolutionary analysis of Clostridium difficile genomes to investigate transmission. Genome Biol 13:R118.

16. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O’Connor L, Ip CLCC, Golubchik T, Batty EM, Finney JM, Wyllie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TEA, Walker AS. 2013. Diverse Sources of C. difficile Infection Identified on Whole-Genome Sequencing. N Engl J Med 369:1195–1205.

17. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TEA. 2013. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 13:137–146.

18. Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, Parker L, Churchill S, Bennett K, Golubchik T, Giess AP, Del Ojo Elias C, Jeffery KJ, Bowler ICJW, Laurensen IF, Barrett A, Drobniewski F, McCarthy ND, Anderson LF, Abubakar I, Thomas HL, Monk P, Smith EG, Walker AS, Crook DW, Peto TEA, Conlon CP. 2014. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. Lancet Respir Med 2:285–292.

19. Grad YH, Kirkcaldy RD, Trees D, Dordel J, Harris SR, Goldstein E, Weinstock H, Parkhill J, Hanage WP, Bentley S, Lipsitch M. 2014. Genomic epidemiology of Neisseria gonorrhoeae with reduced susceptibility to cefixime in the USA: a retrospective observational study. Lancet Infect Dis 14:220–26.

20. Ward H, Ison CA, Day SE, Martin I, Ghani AC, Garnett GP, Bell G, Kinghorn G, Weber JN. 2000. A prospective social and molecular investigation of gonococcal transmission. Lancet 356:1812–1817.

21. Martin IMC, Ghani A, Bell G, Kinghorn G, Ison CA. 2003. Persistence of Two Genotypes of Neisseria gonorrhoeae during Transmission. J Clin Microbiol 41:5609–5614.

22. Bilek N, Martin IM, Bell G, Kinghorn GR, Ison CA, Spratt BG. 2007. Concordance between Neisseria gonorrhoeae genotypes recovered from known sexual contacts. J Clin Microbiol 45:3564–3567.

23. Choudhury B, Risley CL, Ghani AC, Bishop CJ, Ward H, Fenton KA, Ison CA, Spratt BG. 2006. Identification of individuals with gonorrhoea within sexual networks: a population-based study. Lancet 368:139–146.

24. Risley CL, Ward H, Choudhury B, Bishop CJ, Fenton KA, Spratt BG, Ison CA, Ghani AC. 2007. Geographical and demographic clustering of gonorrhoea in London. Sex Transm Infect 83:481–487.
25. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res 43:e15.

26. Ezewudo MM, Joseph S, Castillo-Ramirez S, Dean D, Rio C Del, Didelot X, Selden RF, Shafer WM, Turingan RS, Unemo M, Read TD. 2015. Population structure of Neisseria gonorrhoeae based on whole genome data and its relationship with antibiotic resistance. PeerJ 3:e806.

27. Everitt RG, Didelot X, Batty EM, Miller RR, Knox K, Young BC, Bowden R, Auton A, Votintseva A, Larner-Svensson H, Charlesworth J, Golubchik T, Ip CLC, Godwin H, Fung R, Peto TEA, Walker AS, Crook DW, Wilson DJ. 2014. Mobile elements drive recombination hotspots in the core genome of Staphylococcus aureus. Nat Commun 5:3956.

28. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, Pessia A, Aanensen DM, Mather AE, Page AJ, Salter SJ, Harris D, Nosten F, Goldblatt D, Corander J, Parkhill J, Turner P, Bentley SD. 2014. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet 46:305–309.

29. Bilek N, Ison CA, Spratt BG. 2009. Relative contributions of recombination and mutation to the diversification of the opa gene repertoire of Neisseria gonorrhoeae. J Bacteriol 191:1878–1890.

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

31. Lamelas A, Harris SR, Röltgen K, Dangy J-P, Hauser J, Kingsley RA, Connor TR, Sie A, Hodgson A, Dougan G, Parkhill J, Bentley SD, Pluschke G. 2014. Emergence of a new epidemic Neisseria meningitidis serogroup A Clone in the African meningitis belt: high-resolution picture of genomic changes that mediate immune evasion. MBio 5:e01974–14.

32. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13:601–612.

33. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. 2015. Measurably evolving pathogens in the genomic era. Trends Ecol Evol 30:306–313.

34. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303:327–32.

35. Pybus OG, Rambaut A. 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet 10:540–50.

36. Didelot X, Nell S, Yang I, Woltemate S, van der Merwe S, Suerbaum S. 2013. Genomic evolution and transmission of Helicobacter pylori in two South African families. Proc Natl Acad Sci USA 110:13880–13885.

37. Garnett GP, Mertz KJ, Finelli L, Levine WC, St Louis ME. 1999. The transmission dynamics of gonorrhoea: modelling the reported behaviour of infected patients from Newark, New Jersey. Philos Trans R Soc Lond B Biol
38. Garnett GP, Bowden FJ. 2000. Epidemiology and control of curable sexually transmitted diseases: opportunities and problems. Sex Transm Dis.

39. White PJ, Ward H, Cassell JA, Mercer CH, Garnett GP. 2005. Vicious and virtuous circles in the dynamics of infectious disease and the provision of health care: gonorrhea in Britain as an example. J Infect Dis 192:824–836.

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

41. Mercer CH, Sutcliffe L, Johnson AM, White PJ, Brook G, Ross JDC, Dhar J, Horner P, Keane F, Jungmann E, Sweeney J, Kinghorn G, Garnett GG, Stephenson JM, Cassell JA. 2007. How much do delayed healthcare seeking, delayed care provision, and diversion from primary care contribute to the transmission of STIs? Sex Transm Infect 83:400–5.

42. Satterwhite CL, Torrone E, Meites E, Dunne EF, Mahajan R, Ocfemia MCB, Su J, Xu F, Weinstock H. 2013. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. Sex Transm Dis 40:187–93.

43. Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. 2014. Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data. PLoS Comput Biol 10:e1003457.

44. Peters RPH, Verweij SP, Nijsten N, Ouburg S, Mutsaers J, van Leeuwen a P, Morré S a. 2011. Evaluation of sexual history-based screening of anatomic sites for Chlamydia trachomatis and Neisseria gonorrhoeae infection in men having sex with men in routine practice. BMC Infect Dis 11:203.

45. Mercer CH, Tanton C, Prah P, Erens B, Sonnenberg P, Clifton S, Macdowall W, Lewis R, Field N, Datta J, Copas AJ, Phelps A, Wellings K, Johnson AM. 2013. 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 382:1781–1794.

46. Mayer KH, Krakower DS. 2015. Uncoupling Epidemiological Synergy: New Opportunities for HIV Prevention for Men Who Have Sex With Men. Clin Infect Dis 61:288–290.

47. Wasserheit J. 1992. Epidemiological synergy: interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. Sex Transm Dis 19:61–77.

48. Golden MR, Stekler J, Hughes JP, Wood RW. 2008. HIV serosorting in men who have sex with men: is it safe? J Acquir Immune Defic Syndr 49:212–218.

49. Malek R, Mitchell H, Furegato M, Simms I, Mohammed H, Nardone A, Hughes G. 2015. 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. Eurosurveillance 20:7–15.
50. Donnelly P, Tavare S. 1995. Coalescents and genealogical structure under neutrality. Annu Rev Genet 29:401–21.

51. Wilson DJ, Falush D, McVean G. 2005. Germs, genomes and genealogies. Trends Ecol Evol 20:39–45.

52. Koelle K, Rasmussen DA. 2012. Rates of coalescence for common epidemiological models at equilibrium. J R Soc Interface 9:997–1007.

53. Dearlove B, Wilson D. 2013. Coalescent inference for infectious disease: meta-analysis of hepatitis C. Philos Trans R Soc B Biol Sci 368:20120314.

54. Fingerhuth SM, Bonhoeffer S, Low N, Althaus CL. 2015. Antibiotic-resistant Neisseria gonorrhoeae spread faster with more treatment, not more sexual partners. Peer Journal Prepr 3:e1805.

55. Snyder LAS, Cole JA, Pallen MJ. 2009. Comparative analysis of two Neisseria gonorrhoeae genome sequences reveals evidence of mobilization of Correia Repeat Enclosed Elements and their role in regulation. BMC Genomics 10:70.

56. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek R a, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.

57. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Res 39:347–352.

58. Rice P. 2000. The European Molecular Biology Open Software Suite EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16:276–277.

59. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, others. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303.

60. Harris SRR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD. 2010. Evolution of MRSA During Hospital Transmission and Intercontinental Spread. Science 327:469–474.

61. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–21.

62. Cori A, Ferguson NM, Fraser C, Cauchemez S. 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 178:1505–12.
Figure Legends

Figure 1. Timed phylogeny for the Sheffield dataset reconstructed using BEAST. Known sexual contacts are indicated by uniquely coloured circles and squares. Inset: intervals for the time to the last common ancestor of each pair within a group of known sexual contacts.

Figure 2. Transmission tree for the Sheffield dataset reconstructed using Outbreaker. Cases are indicated by black dots, except for known sexual contacts who are indicated using the same markers as in Figure 1. Each case is aligned on the X axis with its reporting date, and the Y axis is arbitrary. Black links between cases indicate inferred direct transmission, and grey links indicate indirect transmission through at least one unsampled case.

Figure 3. Timed phylogeny for the London dataset reconstructed using BEAST. Each isolate is annotated on the right hand side as follows. First column: black for MSM, blue for heterosexual men and red for heterosexual women. Second column: black for HIV negative and red for HIV positive. Third column: number of reported UK partners in the last three months, with black for zero, grey for one, dark blue for two to five and light blue for six or more.

Figure 4. Density histograms of the time to the most recent common ancestor for all pairs of cases sampled within 3 months of each other. The Sheffield data are shown in red when a sexual contact was reported and in green otherwise, whereas the London data are shown in blue.
Supplementary Material Legends

Figure S1. (A) Maximum likelihood tree reconstructed for the Sheffield genomes. (B) Temporal signal in the tree. The signal was strong enough to estimate the evolutionary rate shown by the solid line.

Figure S2. Permutation test comparing the mean TMRCA for known sexual contacts (in red) with values that would be obtained in the sexual contact labels were permuted at random (histogram).

Figure S3. Links between the Sheffield genomes, based on the 8 months maximum TMRCA criterion.

Figure S4. Distribution of TMRCA between pairs of genomes in the Sheffield dataset that have been assessed to be directly linked by Outbreaker (red) or not (blue).

Figure S5. (A) Maximum likelihood tree reconstructed for the London genomes. (B) Temporal signal in the tree. The signal was not strong enough to estimate the evolutionary rate, and instead the solid line represents the rate estimated in Figure S1 based on the Sheffield data.

Figure S6. Links between the London genomes, based on the 8 months maximum TMRCA criterion.

Table S1. List of 132 genomes in the Sheffield dataset.

Table S2. List of 105 genomes in the London dataset.
