

# Genomic Analysis and Comparison of Two Gonorrhea Outbreaks

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

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

**ABSTRACT** Gonorrhea 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 gonorrhea 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 data sets were selected to have different epidemiological properties: the Sheffield data were collected over 6 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 a criterion to identify likely transmission links in both data sets. 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 behaviors. Comparison of the two data sets suggests that the London epidemic involved about ten times more cases than the Sheffield outbreak.

**IMPORTANCE** The recent increases in gonorrhea incidence and antibiotic resistance are 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 gonorrhea 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.

Received 22 March 2016 Accepted 17 May 2016 Published 28 June 2016

**Citation** Didelot X, Dordel J, Whittles LK, Collins C, Bilek N, Bishop CJ, White PJ, Aanensen DM, Parkhill J, Bentley SD, Spratt BG, Harris SR. 2016. Genomic analysis and comparison of two gonorrhea outbreaks. *mBio* 7(3):e00525-16. doi:10.1128/mBio.00525-16.

**Editor** Paul Stephen Keim, Northern Arizona University

**Copyright** © 2016 Didelot et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

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

Gonorrhea is a sexually transmitted disease (STD) caused by the bacterium *Neisseria gonorrhoeae*. In the United Kingdom, gonorrhea 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 frontline antibiotics has recently been observed (2, 3). Implementation of effective control measures to mitigate the spread of gonorrhea 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 gonorrhea can be carried asymptotically for months in about 10% of men and 50% of women (5) and in the United Kingdom infects disproportionately men who have sex with men (MSM)

and young heterosexuals of black ethnicity and features significant geographical and temporal variations (6).

Molecular epidemiology approaches—for example, *opa* typing (7), *N. gonorrhoeae* multiantigen sequence typing (NG-MAST) (8), or multilocus 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 se-

quencing, 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 United States, showing important structuring of the pathogen population with both geography and sexual orientation of the hosts (19).

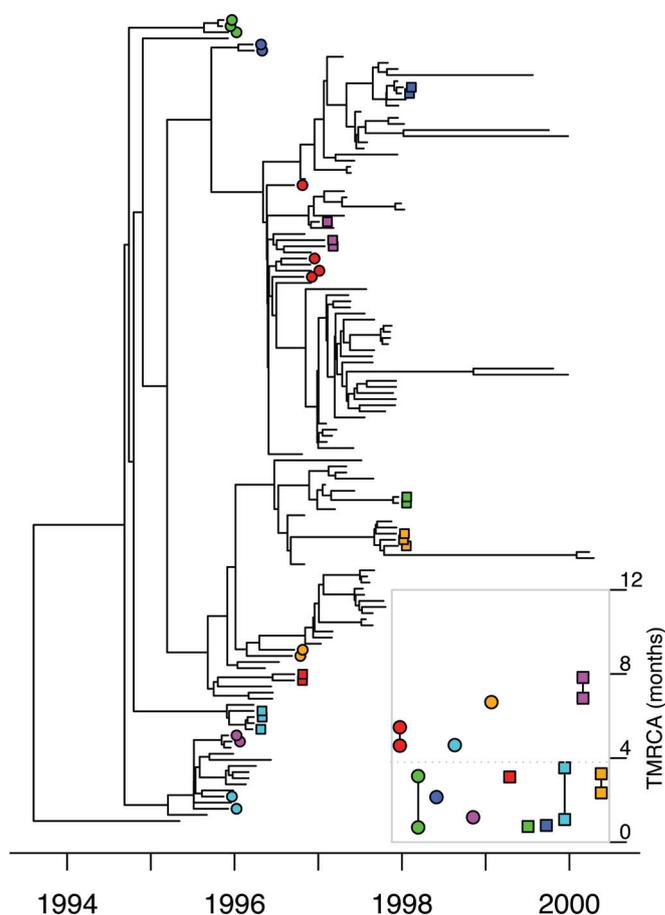
To test the usefulness of gonococcal genome sequencing to track transmission at the finer scale of outbreaks occurring within a city, we selected two local isolate collections with different epidemiological properties, both of which have been previously described without genomic data. The Sheffield isolate 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 2,045 isolates sampled between June and November 2004 from 13 major sexual health clinics throughout London, representing 54% of the 3,754 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 coexistence of a large number of strains, some of which were mostly MSM associated and others mostly heterosexual associated (23). Forty-five percent of infected individuals had one of 21 major strains, among 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 backgrounds of the two studies are very different, with the Sheffield data set coming from a single STD clinic in a city more than 10 times smaller than the London data set, 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 data sets reflect their underlying context.

## RESULTS AND DISCUSSION

**Genomic analysis of the Sheffield data set.** 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 (see Table S1 in the supplemental material). 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 those of previous reports based on species-wide diversity (19, 26). This may reflect a difference in the



**FIG 1** Timed phylogeny for the Sheffield data set reconstructed using BEAST. Known sexual contacts are indicated by uniquely colored circles and squares. (Inset) Intervals for the time to the last common ancestor (TMRCAs) of each pair within a group of known sexual contacts.

effect of recombination when measured at different scales or diversity or between different lineages, as previously reported in other bacterial pathogens, such as *Staphylococcus aureus* (27), *Clostridium difficile* (15), or *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, *opa* and *pil* genes (29). After both repetitive and recombinant regions were removed, 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 ( $R^2 = 0.49$ ) (see Fig. S1 in the supplemental material).

We therefore applied the Bayesian evolutionary analysis software BEAST (30) to these data in order to reconstruct a timed phylogeny (Fig. 1). The molecular clock rate was estimated to be  $1.41 \times 10^{-6}$  single nucleotide polymorphisms (SNPs) per site per year with a 95% confidence interval (CI) of  $1.15 \times 10^{-6}$  to  $1.70 \times 10^{-6}$ . This rate was equivalent to 3.05 (95% CI, 2.47 to 3.67) mutations per year across the genome, is in good agreement with previous estimates in *N. gonorrhoeae* (19, 26) and the related species *Neisseria 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 (Fig. 1 [permutation test]; see Fig. S2 in the supplemental material [ $P < 10^{-4}$ ]). For the 25 pairs of known sexual contacts (see Table S1 in the supplemental material), 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 (Fig. 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 modeling studies (37–39). This duration is significantly longer 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 fulfill this criterion, and yet not all of them are direct transmission pairs since the data include four triplets and one quadruplet of sexual contacts (Fig. 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% (398/1,632) had a TMRCA of less than 8 months, including the known sexual contacts (see Fig. S3 in the supplemental material). Since all of these pairs were sampled within a short period of time in the same city and had the same NG-MAST type, a nongenomic 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 (Fig. 2). When looking at all pairs of individuals in the Sheffield data set 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 (see Fig. S4 in the supplemental material) (Kruskal-Wallis test [KWT],  $P < 10^{-15}$ ; mean TMRCA of 0.28 and 1.54 years for pairs linked and unlinked

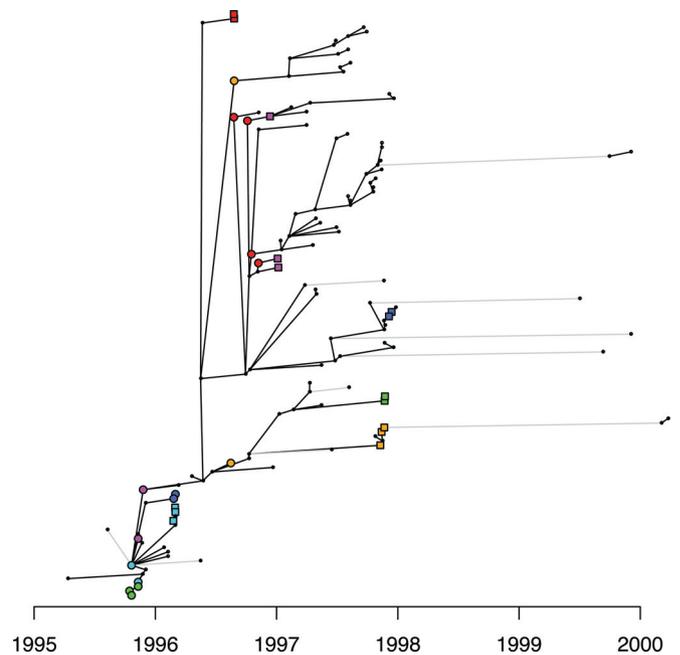


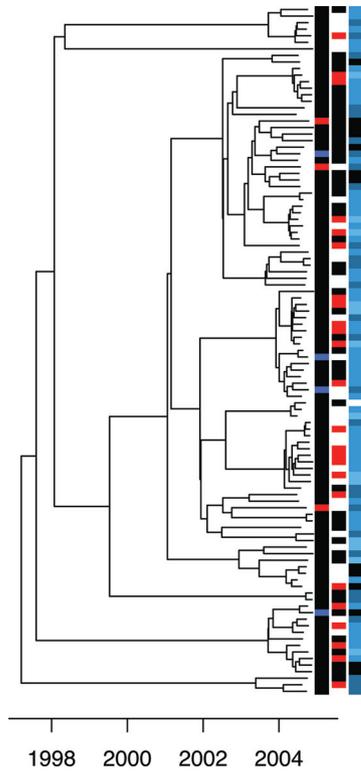
FIG 2 Transmission tree for the Sheffield data set reconstructed using Outbreaker. Cases are indicated by black dots, except for known sexual contacts, who are indicated using the same markers as in Fig. 1. Each case is aligned on the x axis with its reporting date, and the y axis is arbitrary. Black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

by Outbreaker, respectively), indicating a good agreement between the two approaches.

**Genomic analysis of the London data set.** 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 (see Table S2 in the supplemental material). 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 repetitive and recombinant regions were removed.

The temporal signal was weak in this data set 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 data set. The distribution of root-to-tip distance versus sampling dates was, however, compatible with that expected under the rate estimated for the Sheffield data set (see Fig. S5 in the supplemental material). 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 (Fig. 3). We also tried to apply Outbreaker to this data set as we did for the Sheffield data set, 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 come from



**FIG 3** Timed phylogeny for the London data set 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 United Kingdom partners in the last 3 months, with black for zero, gray for one, dark blue for two to five, and light blue for six or more.

a predominantly MSM population, but the same threshold should be applicable since the asymptomatic frequency of anorectal cases among MSM is  $\sim 80\%$  (44), similar to the  $\sim 75\%$  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% (165/4,251) of such pairs (see Fig. S6 in the supplemental material). Likely transmission links were significantly associated with shorter geographical distances between postcodes of residence in London (KWT,  $P = 6 \times 10^{-3}$ ; mean, 10.3 versus 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,  $P = 1.3 \times 10^{-4}$ ; mean, 7.5 versus 9.8 years), likely reflecting a tendency for sexual partners to be about the same age, as found by behavioral surveys such as Natsal-3 (the third of the National Surveys of Sexual Attitudes and Lifestyles) (45). Linked individuals reported the same sexual orientation more often than expected by chance for this sample of individuals (Fisher's exact test [FET],  $P = 6 \times 10^{-3}$ ). 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 the correct transmission links. Among 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 individual and an HIV-negative indi-

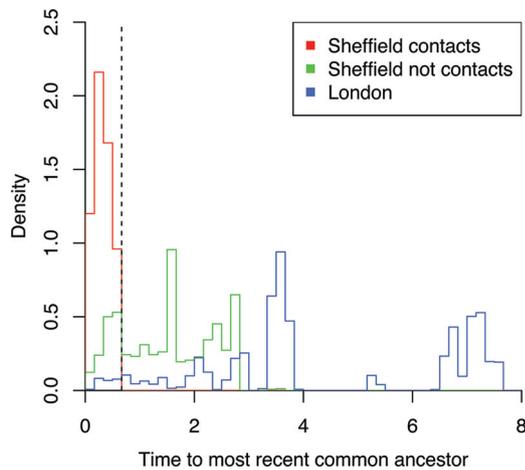
vidual, there was an excess of transmission links found between pairs of individuals who were both HIV positive (FET,  $P = 2.6 \times 10^{-3}$ ), and simultaneously there was a dearth of transmission links when both individuals were HIV negative (FET,  $P = 0.03$ ).

Among 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,  $P = 0.13$ ; mean, 3.19 versus 1.28 links), reported sex abroad (KWT,  $P = 0.41$ ; mean, 2.37 versus 3.41 links), or reported previous gonorrhoea (KWT,  $P = 0.33$ ; mean, 3.52 versus 2.89 links) but was significantly increased for individuals who were HIV positive (KWT,  $P = 1.6 \times 10^{-4}$ ; mean, 5.25 versus 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 in susceptibility and transmissibility of gonorrhoea (46, 47), may be caused by both HIV and gonorrhoea infection being linked with the same high-risk behaviors, or both. It is also probable that HIV-positive MSM have on average more partners and engage in serosorting and riskier behaviors (48, 49). A significant correlation was found between the number of transmission links and the reported number of sexual partners in the United Kingdom in the last 3 months (Spearman's rank correlation test,  $\rho = 0.4$ ;  $P = 2 \times 10^{-5}$ ). This correlation was partly explained by a higher number of reported partners for HIV-positive individuals (KWT,  $P = 2.2 \times 10^{-3}$ ; mean, 8.03 versus 2.61), but a correlation was also suggested between number of links and number of partners when analyzing 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 data sets.

Having analyzed the Sheffield and London data sets separately, we now turn to comparative analysis between them. The temporal sampling frames are different in the two data sets 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 frequencies with which a putative transmission donor is found for cases in both data sets, 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 data sets (Fig. 1 and 3), we can compare the distributions of TMRCA for all pairs of isolates sampled within 3 months of each other (Fig. 4). As previously noted, a larger proportion of pairs have a TMRCA under 8 months in Sheffield relative to London (24% versus 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,  $P < 1 \times 10^{-15}$ ). 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 (Fig. 4).

This difference between the Sheffield and London data sets 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,



**FIG 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.

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 6 years among 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 3 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 were found to be likely using genomic comparisons were 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 type, ST12. Table S1 in the supplemental material contains the list of these isolates and associated metadata. From the previously described London collection (23, 24), which consists of 2,045 isolates sampled between June and November 2004, we included 105 out of 124 isolates from the most MSM-associated NG-MAST type, 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 data sets, 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 1/4 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 the 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 *inverted* 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 SMALTv0.7.5 (<http://www.sanger.ac.uk/science/tools/smalt-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 *phym1* (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. Comparison of root-to-tip distances with isolation dates for each genome revealed a strong temporal signal in the Sheffield data (see Fig. S1 in the supplemental material) but not in London (see Fig. S5 in the supplemental material) due to a short sampling interval. The Bayesian evolutionary analysis software BEAST (30) was applied to both data sets, with the molecular clock of the London data set being forced equal to the rate estimated for the Sheffield data set. The resulting timed phylogenies are shown in Fig. 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 data sets, 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 data set, 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 Fig. S3 and S6 in the supplemental material for the Sheffield and London data sets, 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 Fig. 2 and compared with the BEAST-based approach in Fig. S4 in the supplemental material. 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.

**Nucleotide sequence accession numbers.** Sequence data have been deposited in the European Nucleotide Archive. Accession numbers for the Illumina data from each isolate can be found in Tables S1 and S2 in the supplemental material.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00525-16/-/DCSupplemental>.

- Figure S1, PDF file, 0.01 MB.
- Figure S2, PDF file, 0 MB.
- Figure S3, PDF file, 0.1 MB.
- Figure S4, PDF file, 0 MB.
- Figure S5, PDF file, 0.01 MB.
- Figure S6, PDF file, 0.01 MB.
- Table S1, PDF file, 0.03 MB.
- Table S2, PDF file, 0.04 MB.

## 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 D.M.A., X.D., and P.J.W.), the UK Medical Research Council (grant MR/N010760/1 to X.D. and grant MR/K010174/1 to P.J.W.), and the Wellcome Trust (grant 089472 to B.G.S.).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

1. Public Health England. 2015. Sexually transmitted infections and chlamydia screening in England: 2014. *Public Health 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, Whitley DM. 2014. Molecular approaches to enhance surveillance of gonococcal antimicrobial resistance. *Nat Rev Microbiol* 12:223–229. <http://dx.doi.org/10.1038/nrmicro3217>.
3. Chisholm SA, Wilson J, Alexander S, Tripodo F, Al-Shahib A, Schaefer U, Lythgow K, Fifer H. 24 November 2015. An outbreak of high-level azithromycin resistant *Neisseria gonorrhoeae* in England. *Sex Transm Infect* <http://dx.doi.org/10.1136/sextrans-2015-052312>.
4. Grad YH, Goldstein E, Lipsitch M, White PJ. 2016. Improving control of antibiotic-resistant gonorrhoea by integrating research agendas across disciplines: key questions arising from mathematical modeling. *J Infect Dis* 213:883–890. <http://dx.doi.org/10.1093/infdis/jiv517>.
5. Bignell C, FitzGerald M, Guideline Development Group, British Association for Sexual Health and HIV UK. 2011. UK national guideline for the management of gonorrhoea in adults, 2011. *Int J STD AIDS* 22: 541–547. <http://dx.doi.org/10.1258/ijsa.2011.011267>.
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* 142:211–220. <http://dx.doi.org/10.1017/S0950268813000745>.
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. [http://dx.doi.org/10.1111/j.1365-2958.1995.mmi\\_17050865.x](http://dx.doi.org/10.1111/j.1365-2958.1995.mmi_17050865.x).
8. Martin IM, 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. <http://dx.doi.org/10.1086/383047>.
9. Maiden MC, Bygraves JA, Feil E, 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 U S A* 95:3140–3145. <http://dx.doi.org/10.1073/pnas.95.6.3140>.
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. <http://dx.doi.org/10.1093/molbev/msu121>.
11. Croucher NJ, Didelot X. 2015. The application of genomics to tracing bacterial pathogen transmission. *Curr Opin Microbiol* 23:62–67. <http://dx.doi.org/10.1016/j.mib.2014.11.004>.
12. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. 2016. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol* 14:150–162. <http://dx.doi.org/10.1038/nrmicro.2015.13>.
13. Harris SR, Cartwright EJ, Török ME, Holden MT, 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 methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis* 13:130–136. [http://dx.doi.org/10.1016/S1473-3099\(12\)70268-2](http://dx.doi.org/10.1016/S1473-3099(12)70268-2).
14. Uhlemann A-C, Dordel J, Knox JR, Raven KE, Parkhill J, Holden MT, 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* 111:6738–6743. <http://dx.doi.org/10.1073/pnas.1401006111>.
15. Didelot X, Eyre DW, Cule M, Ip CL, 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. <http://dx.doi.org/10.1186/gb-2012-13-12-r118>.
16. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CL, Golubchik T, Batty EM, Finney JM, Wyllie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TE, Walker AS. 2013. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 369:1195–1205. <http://dx.doi.org/10.1056/NEJMoa1216064>.
17. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE. 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 13:137–146. [http://dx.doi.org/10.1016/S1473-3099\(12\)70277-3](http://dx.doi.org/10.1016/S1473-3099(12)70277-3).
18. Walker TM, Lalor MK, Broda A, Saldana Ortega L, Morgan M, Parker L, Churchill S, Bennett K, Golubchik T, Giess AP, Del Ojo Elias C, Jeffery KJ, Bowler IC, Laurenson IF, Barrett A, Drobniowski F, McCarthy ND, Anderson LF, Abubakar I, Thomas HL, Monk P, Smith EG, Walker AS, Crook DW, Peto TE, 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. [http://dx.doi.org/10.1016/S2213-2600\(14\)70027-X](http://dx.doi.org/10.1016/S2213-2600(14)70027-X).
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: 220–226. [http://dx.doi.org/10.1016/S1473-3099\(13\)70693-5](http://dx.doi.org/10.1016/S1473-3099(13)70693-5).
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 inves-

- tigation of gonococcal transmission. *Lancet* 356:1812–1817. [http://dx.doi.org/10.1016/S0140-6736\(00\)03234-7](http://dx.doi.org/10.1016/S0140-6736(00)03234-7).
21. Martin IM, Ghani A, Bell G, Kinghorn G, Ison CA. 2003. Persistence of two genotypes of *Neisseria gonorrhoeae* during transmission. *J Clin Microbiol* 41:5609–5614. <http://dx.doi.org/10.1128/JCM.41.12.5609-5614.2003>.
  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. <http://dx.doi.org/10.1128/JCM.01453-07>.
  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. [http://dx.doi.org/10.1016/S0140-6736\(06\)69003-X](http://dx.doi.org/10.1016/S0140-6736(06)69003-X).
  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. <http://dx.doi.org/10.1136/sti.2007.026021>.
  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. <http://dx.doi.org/10.1093/nar/gku1196>.
  26. Ezewudo MN, Joseph SJ, Castillo-Ramirez S, Dean D, Del Rio C, Didelot X, Selden RF, Shafer WM, Turingan RS, Nemo M, Read TD. 2015. Population structure of *Neisseria gonorrhoeae* based on whole genome data and its relationship with antibiotic resistance. *PeerJ* 3:e806. <http://dx.doi.org/10.7717/peerj.806>.
  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. <http://dx.doi.org/10.1038/ng.2895>.
  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. <http://dx.doi.org/10.1128/JB.01518-08>.
  30. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214. <http://dx.doi.org/10.1186/1471-2148-7-214>.
  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. <http://dx.doi.org/10.1128/mBio.01974-14>.
  32. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 13:601–612. <http://dx.doi.org/10.1038/nrg3226>.
  33. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. 2015. Measurably evolving pathogens in the genomic era. *Trends Ecol Evol* 30:306–313. <http://dx.doi.org/10.1016/j.tree.2015.03.009>.
  34. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303:327–332. <http://dx.doi.org/10.1126/science.1090727>.
  35. Pybus OG, Rambaut A. 2009. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* 10:540–550. <http://dx.doi.org/10.1038/nrg2583>.
  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 U S A* 110:13880–13885. <http://dx.doi.org/10.1073/pnas.1304681110>.
  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 Sci* 354:787–797. <http://dx.doi.org/10.1098/rstb.1999.0431>.
  38. Garnett GP, Bowden FJ. 2000. Epidemiology and control of curable sexually transmitted diseases: opportunities and problems. *Sex Transm Dis* 27:588–599. <http://dx.doi.org/10.1097/00007435-200011000-00007>.
  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: gonorrhoea in Britain as an example. *J Infect Dis* 192:824–836. <http://dx.doi.org/10.1086/432004>.
  40. Cohen MS, Cannon JG. 1999. Human experimentation with *Neisseria gonorrhoeae*: progress and goals. *J Infect Dis* 179(Suppl 2):S375–S379. <http://dx.doi.org/10.1086/513847>.
  41. Mercer CH, Sutcliffe L, Johnson AM, White PJ, Brook G, Ross JD, 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* 83:400–405. <http://dx.doi.org/10.1136/sti.2006.024554>.
  42. Satterwhite CL, Torrone E, Meites E, Dunne EF, Mahajan R, Ocfemia MC, Su J, Xu F, Weinstock H. 2013. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. *Sex Transm Dis* 40:187–193. <http://dx.doi.org/10.1097/OLQ.0b013e318286bb53>.
  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. <http://dx.doi.org/10.1371/journal.pcbi.1003457>.
  44. Peters RP, Nijsten N, Mutsaers J, Jansen CL, Morré SA, van Leeuwen AP. 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. <http://dx.doi.org/10.1186/1471-2334-11-203>.
  45. Mercer CH, Tanton C, Prah P, Erens B, Sonnenberg P, Clifton S, Maccowall 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. [http://dx.doi.org/10.1016/S0140-6736\(13\)62035-8](http://dx.doi.org/10.1016/S0140-6736(13)62035-8).
  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. <http://dx.doi.org/10.1093/cid/civ291>.
  47. Wasserheit JN. 1992. Epidemiological synergy: interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. *Sex Transm Dis* 19:61–77. <http://dx.doi.org/10.1097/00007435-199219020-00001>.
  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. <http://dx.doi.org/10.1097/QAI.0b013e31818455e8>.
  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. *Euro Surveill* 20:7–15.
  50. Donnelly P, Tavaré S. 1995. Coalescents and genealogical structure under neutrality. *Annu Rev Genet* 29:401–421. <http://dx.doi.org/10.1146/annurev.ge.29.120195.002153>.
  51. Wilson DJ, Falush D, McVean G. 2005. Germs, genomes and genealogies. *Trends Ecol Evol* 20:39–45. <http://dx.doi.org/10.1016/j.tree.2004.10.009>.
  52. Koelle K, Rasmussen DA. 2012. Rates of coalescence for common epidemiological models at equilibrium. *J R Soc Interface* 9:997–1007. <http://dx.doi.org/10.1098/rsif.2011.0495>.
  53. Dearlove B, Wilson DJ. 2013. Coalescent inference for infectious disease: meta-analysis of hepatitis C. *Philos Trans R Soc Lond B Biol Sci* 368:20120314. <http://dx.doi.org/10.1098/rstb.2012.0314>.
  54. Fingerhuth SM, Bonhoeffer S, Low N, Althaus CL. 2015. Antibiotic-resistant *Neisseria gonorrhoeae* spread faster with more treatment, not more sexual partners. *PeerJ PrePrints* 3:e1805. <http://dx.doi.org/10.1371/journal.ppat.1005611>.
  55. Snyder LA, Cole JA, Pallen MJ. 2009. Comparative analysis of two *Neisseria gonorrhoeae* genome sequences reveals evidence of mobilization of Correira repeat enclosed elements and their role in regulation. *BMC Genomics* 10:70. <http://dx.doi.org/10.1186/1471-2164-10-70>.
  56. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formisano K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, 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. <http://dx.doi.org/10.1186/1471-2164-9-75>.
57. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PFAST: a fast phage search tool. *Nucleic Acids Res* 39:W347–W352. <http://dx.doi.org/10.1093/nar/gkr485>.
58. Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16:276–277. [http://dx.doi.org/10.1016/S0168-9525\(00\)02024-2](http://dx.doi.org/10.1016/S0168-9525(00)02024-2).
59. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <http://dx.doi.org/10.1101/gr.107524.110>.
60. Harris SR, 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. <http://dx.doi.org/10.1126/science.1182395>.
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* 59:307–321. <http://dx.doi.org/10.1093/sysbio/syq010>.
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* 178:1505–1512. <http://dx.doi.org/10.1093/aje/kwt133>.