The impact of antimicrobials on gonococcal evolution

Leonor Sánchez-Busó, Daniel Golparian, Jukka Corander, Yonatan H. Grad, Makoto Ohnishi, Rebecca Flemming, Julian Parkhill, Stephen D. Bentley, Magnus Unemo and Simon R. Harris

The sexually transmitted pathogen Neisseria gonorrhoeae is regarded as being on the way to becoming an untreatable superbug. Despite its clinical importance, little is known about its emergence and evolution, and how this corresponds with the introduction of antimicrobials. We present a genome-based phylogeographical analysis of 419 gonococcal isolates from across the globe. Results indicate that modern gonococci originated in Europe or Africa, possibly as late as the sixteenth century and subsequently disseminated globally. We provide evidence that the modern gonococcal population has been shaped by antimicrobial treatment of sexually transmitted infections as well as other infections, leading to the emergence of two major lineages with different evolutionary strategies. The well-described multidrug-resistant lineage is associated with high rates of homologous recombination and infection in high-risk sexual networks. A second, multisusceptible lineage is more associated with heterosexual networks, with potential implications for infection control.

Almost 360 million curable sexually transmitted infections (STIs) are estimated to occur globally each year, with Neisseria gonorrhoeae, the causative agent of gonorrhoea, infecting approximately 78 million. The highest gonorrhoea burden is reported among men, although problematic infections are more common in women for whom urogenital infections are often asymptomatic. Unresolved urogenital infections can lead to severe complications and sequelae, such as reproductive problems including infertility, serious eye infections in newborn babies and enhanced transmission of HIV. The emergence and proliferation of gonococci with resistance to front-line antimicrobials, such as extended-spectrum cephalosporins (cefixime and ceftriaxone) and azithromycin, have contributed to, although do not explain, the increase in incidence of gonorrhoea. Resistance to dual therapy (injectable ceftriaxone plus oral azithromycin), the current recommended treatment in many countries, is fortunately rare; however, decreased susceptibility to ceftriaxone has been reported from all continents and azithromycin resistance is on the increase globally, raising fears that the effectiveness of this regimen will be short-lived. Much of the focus of gonococcal control is on particular high-risk sexual networks that often partake in unprotected sex with multiple partners, particularly sex workers and men who have sex with men but also young heterosexuals. These groups are more frequently exposed to both infection and antimicrobial treatment, which has led to these networks being the suspected drivers of antimicrobial resistance (AMR). However, AMR is not the only factor driving the emergence of resistance to extended-spectrum cephalosporins, which led to the replacement of cefixime as the first-line treatment for gonorrhoea. The first mosaic penA (encoding penicillin-binding protein 2 (PBP2)) alleles gained via recombination have been key in the emergence of resistance to extended-spectrum cephalosporins, which, in turn, has played an important role in its rapid gain and spread of AMR is its ability to exchange DNA via homologous recombination both within its own species and with other Neisseria species. For example, mosaic penA alleles that caused high-level ceftriaxone resistance was seen in an isolate from a pharyngeal infection in a female sex worker in Japan in 2009 (ref. 12). Other mosaic penA alleles have been seen worldwide, but similar mosaic penA alleles have been seen worldwide. In fact, a number of resistances were first identified in Japan, leading to the hypothesis that most AMR gonorrhoea originates there, or elsewhere in the WHO Western Pacific Region.

Whole-genome sequencing has been successfully used to reveal the origins, global spread and population structure of several human pathogens. However, gonococcal genome sequencing has mostly targeted specific populations and outbreaks. Here, we report the findings of a global genomic study of 419 N. gonorrhoeae isolates spanning 5 continents and more than 50 years,
including varying susceptibilities to important antimicrobials. Our aim was to elucidate when and where modern gonococcal populations emerged, evolved and dispersed, and how antimicrobial usage and transmission in different sexual networks has influenced their population dynamics.

Results

Modern gonococcus is not ‘as old as mankind’. Our collection spans a period of more than 50 years (1960–2013) and 58 countries from 5 continents (Fig. 1 and Supplementary Table 1). A population-level analysis revealed a high level of admixture among N. gonorrhoeae with no significant differentiation between continents (Supplementary Table 2), with the exception of Africa (Supplementary Fig. 1 and Supplementary Tables 3–5). We estimated the substitution rate for the non-recombining section of the genomes in the collection (Supplementary Fig. 2) to be $3.74 \times 10^{-8}$ substitutions per site per year confidence interval (CI; 3.39 $\times 10^{-8}$ to 4.07 $\times 10^{-8}$), which is similar to previous reports[19,18] and comparable to other bacteria[1]. The time of the most recent common ancestor (tMRCA) was estimated to be around the sixteenth century (1589, CI: 1544–1623) (Fig. 2). Although high rates of recombination can lead to underestimation of tMRCA to some extent, these results are strongly at odds with the hypothesis that modern gonorrhoea has existed as long as mankind and cast further doubt on the ascribing of historical descriptions of gonorrhoea-like symptoms to infection with ‘modern’ gonococci.

Despite modern gonococci being globally mixed, we found strong evidence of historic geographical separation, suggesting that rapid mixing of populations is a relatively recent phenomenon. A phylogeographical analysis ascribed the origin of our collection to Europe (60.9% inferred ancestry). However, when corrected for biases in the number of samples from each continent, complementing with isolates from a US study[19], there was support for an African origin (90.7% inferred ancestry) (Supplementary Fig. 3 and Supplementary Table 6). From this African root, we identified a significant African origin (90.7% inferred ancestry) (Supplementary Fig. 3). The gonococcal genomic island was found in 277 (67%) isolates (Supplementary Fig. 6), but showed no clear association with AMR. The plasmid-encoded resistances showed no significant difference in prevalence in lineage A or lineage B (two-sided test for equality of proportions for tetM: $\chi^2 = 0.01$, 95% CI: $-0.089$ to 0.110, d.f. = 1, $P = 0.92$, and for blaTEM: $\chi^2 = 0.88$, 95% CI: $-0.046$ to 0.147, d.f. = 1, $P = 0.35$). By contrast, the 29 chromosomally mediated resistance substitutions examined, 18 were significantly associated with clade A (Fig. 3). Importantly, based on our phylogenetic dating, the majority of occurrences of these 29 determinants were estimated to have been acquired after the introduction of the antimicrobial against which they act (Supplementary Fig. 7).

Two strategies for gonococcal success. Overall, our data show far fewer gains of chromosomally encoded AMR determinants in lineage B than in lineage A (Supplementary Fig. 8). As these determinants primarily spread through the population via homologous recombination, such differences could be explained by differences in recombination frequency. To assess this, we compared the proportion of homoplasic sites, an indicator of recombination, in the terminal branches of the phylogenetic tree in the two lineages. This confirmed a significantly higher proportion in clade A, particularly for short branches, which represent very recent evolution (Wilcoxon test $W = 19,416$, $P < 0.001$; Fig. 4a,b and Supplementary Fig. 9). Note that the distribution of branch lengths in both clades was similar (Wilcoxon test $W = 14,427$, $P = 0.739$). Similarly, the proportion of clustered SNPs, another sign of recombination, was also higher on the terminal branches in lineage A (Wilcoxon test $W = 16,984$, $P < 0.05$). The proportion of recombination-deficient strains (those with no recombination events detected, $r = 0$) in lineage B was higher than expected, bordering on statistical significance (one-tailed test of proportions, $P = 0.05184$).

One explanation for such differences could be opportunity. For recombination to occur, donor and recipient bacteria must colocalize. Thus, recombination between gonococci would be expected to occur more frequently in high-risk host populations where co-infection with other STIs and pharyngeal infections, which allow access to commensal Neisseria species, are more common. These risk groups are also more likely to be exposed to repeated antimicrobial therapy for gonorrhoea infection and other STIs[4]. Unfortunately, due to limitations in the availability of data on patient sexual behaviour, we could not adequately assess the association of the lineages to risk factors in our data set. However, we could analyse the distribution of the gender of the patients from whom the isolates were taken. To increase the power of the analysis, we included 376 isolates from two North American genomics studies[17,26], to give a set of 639 isolates with complete gender information. Strikingly, lineage B included a significantly higher proportion of women (40 out of 136, 29.4%) than lineage A (69 out of 503, 13.7%) (two-sided test for equality of proportions $\chi^2 = 17.54$, 95% CI: 0.070–0.244, d.f. = 1, $P < 0.0001$) (Fig. 4d and Supplementary Fig. 8), which would
suggest that lineage B is more closely associated with heterosexuals. Corroborating this, data from a 2013 European-wide structured survey\textsuperscript{27} showed a similar pattern. Lineage B isolates were strongly associated with reduced MICs and female patients (61 out of 214, 28.5\% of lineage B isolates were from women versus 100 out of 821, 12\% of lineage A; two-sided test for equality of proportions $\chi^2=33.21$, 95\% CI: 0.096–0.231, d.f. = 1, $P<0.0001$), and more importantly, of the patients that reported sexual orientation, 78.3\% (94 out of 120) of isolates in lineage B were from heterosexuals, in contrast to 52.6\% (200 out of 380) in lineage A (two-sided test for equality of proportions $\chi^2=23.82$, 95\% CI: 0.162–0.352, d.f. = 1, $P<0.0001$) (Supplementary Fig. 10). Particular sublineages within lineage B appeared to be particularly strongly associated with heterosexuals\textsuperscript{27}. We suspect that lineage A, being associated with higher-risk populations, does have greater opportunity for recombination, which may explain the observed higher recombination rate. However, transmission between low-risk and high-risk populations is common within lineage A, so we suspect that opportunity is not the only explanation for the differential recombination rate in the two lineages. The observation that plasmid-born resistances do not

**Fig. 1 | Geographical and phylogenetic distribution of *N. gonorrhoeae* isolates.** The map (top) shows the countries of isolation of the strains in the collection coloured by continent. The phylogeny (bottom) shows the relationship among the strains ($n=419$). The coloured strips show (from inside out) the continent of isolation (CONT), year and further typing information (BAPS clusters, NG-MAST, MLST and penA types; the colours represent different types or alleles). Mosaic penA types are marked in the outermost black strip.
show the same difference in frequency between the two lineages also supports this view.

**Discussion**
Gonorrhoea is one of the most clinically important STIs worldwide. Its rapid mode of transmission, especially among high-risk groups, and the emergence of resistance to many antimicrobials, has made the control of *N. gonorrhoeae* of primary importance for public health. In recent years, there has been an understandable focus on AMR gonorrhoea, with resistance to all classes of antimicrobials used to treat the infection having been reported\(^3\). However, the increase in prevalence of gonorrhoea has continued in many settings\(^6\) despite resistance to dual therapy being extremely rare.

Our genomic analysis revealed a contemporary global population with little geographical structure, suggesting that rapid recent intercontinental transmission is occurring. In particular, introductions from Asia into the rest of the world appear common, consistent with the observation that a number of recent resistant gonococcal clones have emerged from this region\(^2\). The one exception was Africa, where the sampled gonococcus was less diverse.
However, our African sample size was small due to the limited availability of isolates, so further study is required in this area.

We estimated an origin of modern gonococci in the sixteenth century (1544–1623), which contrasts with historical interpretations of modern gonorrhoea as an ancient disease. Although we are keen to stress that high rates of recombination make accurate estimates difficult and that our estimated CIs are probably too narrow, this dating suggests that ancient accounts of gonorrhoea are likely to have been affected by posthumous contamination or are evidence of an ancient population distinct from that observed today. It certainly disputes the view that the disease that we now know as gonorrhoea is ‘as old as mankind’. The sixteenth century was, nonetheless, an opportune time for the global dissemination of pathogens. It was a period of early modern globalization marked by the initiation and intensification of many intercontinental trade links, particularly by sea. This period was of utmost importance for globalization due to an increase in the import of crops from the Americas to Europe. Increased movement of people around the world also spawned local epidemics and pandemics, and may well have played an important role in the evolution of modern gonorrhoea. A phylogeographical analysis using several subsampled sets of strains from different continents to avoid bias placed the origin of the current global gonococcal population in Europe or Africa. We identified a subsequent introduction into Asia in the early seventeenth century (1578–1649), which expanded rapidly throughout the continent. Much more recently, this lineage has been repeatedly transmitted back to the rest of the world.

A major finding is a strong association between isolates from the lineage that evolved from this early introduction to Asia and the development of AMR. Nearly all isolates in this lineage A, but only 50% of those in lineage B, harboured resistance to sulfonamides (folP R228S mutation) and tetracyclines (rpsJ V57M mutation). Sulfonamides were the first antimicrobials introduced to treat gonorrhoea in 1935, with initial efficacies of around 90%. By the mid-to-late 1940s sulfonamide resistance was common and it was discarded as a treatment for gonorrhoea. However, sulfonamides are still widely used in combination with trimethoprim for prophylaxis in patients who are HIV positive and to treat various bacterial infections. Doxycycline (a tetracycline) is still used to treat gonococcal or presumptively non-gonococcal urethritis/cervicitis and is recommended treatment for lymphogranuloma venereum. Thus, we suspect that the high incidence of sulfonamide and tetracycline resistance in modern gonorrhoea is due to historic treatment of the disease itself followed by continued use of these drug classes for other purposes. The high proportion of diverse circulating strains carrying the folP and rpsJ mutations could be used as evidence that they were in the gonococcal population long before the introduction of antimicrobials. However, this seems unlikely. More plausibly, the use of sulfonamides and tetracyclines has produced a strong selective pressure over an extended period of time, which has led to many independent acquisitions of resistance mutations and convergent gains of resistance via homologous recombination. In the more recombinogenic lineage A, this has resulted in these mutations sweeping through the entire clade. Furthermore, other AMR determinants that have entered the gonococcal population more recently appear to be undergoing the same process, particularly in
we found that lineage A is associated with infection in men who have sex with men, one of the predominant risk groups, whereas isolates from lineage B are more rarely found in this demographic group. Thus, lineage A isolates have the means (increased homologous recombination), motive (higher antimicrobial exposure) and opportunity (higher rates of co-infection with commensal Neisseria and other STIs) for recombination-driven gain of AMR.

Most recent media attention and gonococcal genomics research has focused on the increasing levels of AMR in gonorrhoea. However, we have shown that a mostly susceptible lineage is successfully persisting in lower-risk groups where it is probably less likely to be exposed to antimicrobials. Notably, this lineage was associated with heterosexual groups and with infections in women, where rates of asymptomatic infection are higher. Turner et al.\(^\text{13}\) showed, using a modelling approach, that in a situation where both resistant and susceptible strains are present in a population, high rates of asymptomatic infection, and therefore undertreatment, can allow susceptible isolates to survive and thrive. In such circumstances, rates of susceptible infection can be hugely underestimated, potentially meaning that our understanding of gonococcal prevalence and rates of AMR may be biased. Interestingly, the majority of our African samples were from lineage B, consistent with
epidemiological studies that describe a hidden epidemic of gonococcal disease in rural South African women, in which 48% of cases were asymptomatic and another 50% were symptomatic but not seeking care. Similarly, in Namibia, the prevalence of asymptomatic gonococcal infections in both men and women in rural villages is high. This may suggest that lineage B is associated with asymptomatic infection more fundamentally than simply being more often found in women. In such a situation, if compensatory mutations are not developed, gain of AMR determinants may be detrimental as these elements may come with an associated general cost to fitness. Grad et al. reported, for example, that 235 rRNA mutations resulting in resistance to azithromycin resistance were associated with reduced ESC MICs in isolates with mosaic penA alleles. Similarly, we have observed that the tetM and blaTEM-containing plasmids are negatively associated with isolates with mosaic penA alleles.

In conclusion, in the first phylogeographical analysis of a global collection of gonococci, we have shown that, although the modern gonococcal population is highly mixed, this mixing is relatively recent. This gonococcal population originated as late as the sixteenth century, most likely in Europe or Africa, and an early single introduction into Asia led to a rapid spread throughout the continent and the rest of the world. Despite most recent focus being on gonococcal AMR, we have demonstrated that N. gonorrhoeae has adapted to sexual networks with different risk profiles and exposures to antimicrobial treatment. Modern global gonococcal data can be divided into two lineages, which we term lineage A (after the phylogenetic breakpoint) and lineage B (before the phylogenetic breakpoint). Lineage A has gained and proliferated AMR determinants, aided by an increased rate of recombination. We hypothesize that these isolates are often transmitted in higher-risk networks, for example, men who have sex with men, where pharyngeal infections are more common and individuals are more frequently exposed to treatment for gonorrhoea and other STIs. However, lineage B has not gained AMR so rapidly, with 26% of isolates containing no known AMR determinants, and is potentially being silently transmitted in undertreated groups where levels of asymptomatic infection are higher. Thus, our results have shown that the effect of antimicrobial treatment on the gonococcal population has been more complex than simply initiating an inexorable progression towards AMR.

Methods

Global N. gonorrhoeae strains and antimicrobial susceptibility testing. A total of 413 N. gonorrhoeae strains without known epidemiological relatedness were collected from patients with gonorrhoea in 58 countries spanning 5 continents. The strains were selected to represent a wide geographical, temporal, phenotypic (based on AMR) and genetic diversity, that is, to represent as much as feasible of the N. gonorrhoeae species phylogeny (Supplementary Table 1). Six genome references were also included in the study, spanning a range of isolation dates between 1960 and 2013 in total. Bacterial isolation from the corresponding samples, preservation and transportation was performed following standard microbiological procedures. β-Lactamase production and MICs were tested for a range of antimicrobials as described previously: spectinomycin, tetracycline, penicillin G, ciprofloxacin, azithromycin, ceftriaxone and cefixime.

DNA preparation and whole-genome sequencing. All isolates were confirmed to be N. gonorrhoeae and genomic DNA was extracted from the isolates using the Promega Wizard DNA purification kit, following the instructions from the manufacturer. Purified DNAs were multiplexed and sequenced using two lanes of the HiSeq 2500 2x 100 bp platform at the Wellcome Sanger Institute (Hinxton, UK).

Mapping and variant calling. Fastq files from the 413 new gonococcal strains and the N. meningitidis 10356_1_465 outgroup (ENA accession number: ER5248641) were mapped to a common reference, N. gonorrhoeae FA1090 (NCBI accession NC_002496.1, 2,153,922 bp) using SMALT v0.7.4 (http://www.sanger.ac.uk/science/tools/svmap). Variants were called using SAMtools and BCFTools v1.2 (ref. 38) after indel (insertion or deletion) realignment with GATK v5.5.9 (ref. 39) and further filtered as described previously.

Six public reference genomes were obtained from the NCBI and aligned using progressiveMAUVE v2.3.1 (ref. 40). (N. gonorrhoeae FA1090 (NCBI accession NC_002496.3), FA19 (NCBI accession CP012026.1), FA6140 (NCBI accession CP012027.1), MS11 (NCBI accession NC_022440.1), 35/02 (NCBI accession CP012028.1) and NCCP11945 (NCBI accession NC_011035.1)). The XMAF output alignment was converted into a plain fasta format using N. gonorrhoeae FA1090 as a reordering reference through a custom Perl script (see ‘Code availability’). Positions with gaps in this reference were removed, so that the resulting alignment had homologous positions to the 2,153,922 bp in the FA1090 genome. This alignment was added into the alignment resulting from mapping the 413 isolates, producing a 419-strain alignment containing the core genome and accessory sites from FA1090 that are shared by any other strain in the collection.

Recombination removal and phylogenetic reconstruction. Prophages described in the N. gonorrhoeae FA1090 strain were masked in the alignment before running Gubbins v1.4.10 (ref. 41), which was used to remove segments that can have undergone recombination. We detected 72,114,190 bp alignment contained 15,562 variable sites, identified by snp-sites4 and was used for population structure analysis, phylogenetic inference and divergence estimation. Genetic clusters were obtained from the non-recombining alignment using hierBAPS v7.3 (ref. 42).

The final SNP alignment was used for maximum likelihood phylogenetic tree reconstruction using RAxML v8.2.6 (ref. 43) under the GTR+GAMMA model of nucleotide substitution and 100 bootstrap replicates. An algorithm called BOOSTER v0.1.2 (ref. 44) was also used to obtain an enhanced estimate of nodal support values (Supplementary Note). Ancestral states of all SNPs before recombination removal were reconstructed onto the resulting phylogenetic tree using ACCTRAN transformation in python (https://scikit-learn.org). Homologous sites in the terminal branches of the tree were detected for the two main lineages. It is important to note that Gubbins removed 97% (33,026 out of 34,034) of those homoplastic sites, minimizing their effect on subsequent analyses.

 Genome de novo assembly and in silico typing. In parallel to the mapping process, reads were assembled using the assembly and improvement iterative pipeline developed at the Wellcome Sanger Institute37. MLST42 and N. gonorrhoeae multi-antigen sequence typing (NG-MAST)45 typing schemes were retrieved directly from the sequences using the get_sequence_type script (https://github.com/sanger-pathogens/mlst_check/blob/master/blast/mlst_sequence_type) and NG-MAST v2.4.4 (ref. 46), respectively. The presence of β-lactamase (tetM) genes on plasmids and the gonococcal genomic island were detected using BLAST v2.3.0+ and ARIBA v2.4.2 (ref. 47). Typing was performed for the conjugal plasmid and the blaTEM plasmids using an in silico PCR (https://github.com/simonnharris/in_silico_pcr). Primers to differentiate between the Dutch and the American tetM-containing plasmids were obtained from Turner et al.48 and type the blaTEM plasmids using the primers described in Dinh et al.49 were used and the resulting amplicon sizes were evaluated to differentiate among the Asia, Africa and Toronto/Rio types (Supplementary Table 1).

Analysis of population structure. To study population structure from the resulting alignment, the poppr R package v2.5.0 (ref. 50) was used to perform an analysis of molecular variance (AMOVA) test on the non-recombining section of the genome51 on three geographical hierarchies—continent, subcontinent and country—to calculate the percentage of observed variance within and between groups. To test whether the observed differentiation between continents was significant, a randomization test (n = 1,000 permutations) was performed using the randtest function from the ade4 R package v1.7-11 (ref. 61), which randomly tries to maximize the discrimination between the predefined groups. To avoid overfitting and to keep enough discrimination power, the optimal number of principal components to retain was determined using the a-score optimization test, which uses randomized groups to calculate the proportion of successful reassignments corrected by the number of retained principal components. This methodology resulted in 83 principal components as optimal to keep a balance between discrimination power and overfitting. Prior assignment to continents was randomized and the DAPC analysis was repeated to confirm that the observed separation among clusters does not occur by chance. Four discriminant functions
were kept for the analysis, considering that the number of variables was five continents. A multivariate analysis of variance (MANOVA) test was applied to test whether there were differences between the means of the different clusters (continents) on the discriminant functions. Wilks’ lambda was used to test the significance of this MANOVA test. Resulting $P$ values were adjusted for multiple tests using the false discovery rate.

DAPC derives group membership probabilities from the retained discriminant functions. These results were used to evaluate the level of admixture in the data set under study. Isolates assigned with 80% posterior probability to a continent different from the prior assignment were interpreted as intercontinental transmission cases. Isolates with continent different from the prior assignment were interpreted as admixed.

To get an estimate of the substitution rate and tMRCA for the whole N. gonorrhoeae global collection, the least-square dating (LSD) v0.3 software was used. This approach has been shown to be robust to uncorrelated changes of the molecular clock and to give similar results to BEAST.

To compare the performance between LSD and BEAST, individual BAPS clusters were used. Specifically, Bayesian approximation using BEAST v1.8.2 (ref. 68) was run to estimate tMRCA and the substitution rate of the genetic clusters determined by hierBAPS.

Three chains were run per cluster up to 100 million generations by using a GTRGAAM model of nucleotide substitution with 4 categories, strict molecular clock with a gamma distribution (shape 0.001 and 0.000) and a constant population size as priors. Default priors were used. For models using relaxed clocks, the ucl mean prior was set to a gamma distribution with shape 0.001 and scale 1.000. The same configuration was used to run two different chains with the whole collection, which did not reach proper convergence because of the complexity of the data set. LSD was also run for the BAPS clusters that reached convergence in BEAST and the results were compared (Supplementary Note). The obtained tMRCA was further confirmed using the Wald statistic (Supplementary Note).

Phylogeography with stochastic character mapping. The continent of isolation was used as a discrete trait to study changes in the distribution over the phylogenetic tree using treeBreaker v1.1. (ref. 68) (https://github.com/ansariazim/treeBreaker). This program calculates the per-branch posterior probability of having a change in the distribution of a discrete character.

Stochastic character mapping with a symmetric transition model (SYM) was applied to the phylogenetic tree to get posterior probabilities for each continent at every node using the make.simmmap function implemented in the phytools R package v0.6-44 (ref. 69). Given a phylogeny and a set of tip states (‘continent’ in this study), this method uses an MCMC approach to sample character histories from their posterior probability distribution consistent with those states given a model of evolution for the mapped character.

This procedure was applied to the prior and posterior continent assignments excluding the admixed individuals to reduce noise from the prior metadata.

An extra set of 236 isolates from the United States was added to the global collection and the phylogenotypical analyses were repeated to confirm our results. To avoid biases due to a different number of strains from different continents, the combined data sets were downsampled 100 times to n = 41 (the maximum number of strains with a posterior assignment to the continent with the least number of strains, Africa), except for Oceania, from which there are not more data in the public databases to include, generating 100 subtrees. Ten stochastic maps were inferred for each of those subtrees and posteriorly combined using phytools, resulting in a total of 1,000 evaluated maps.

Evolution of AMR determinants. Mutations conferring AMR in known genetic determinants (165 rRNA, 235 rRNA, rpsL, rpsJ, mtrC, folP, gyrA, parC, parE, penA, penA, penA, parC, parE) as well as the presence of the β-lactamase (Blαmase) and tetM genes were obtained for the 413 strains sequenced in this study using ARIMA v2.4 (ref. 70) (Supplementary Table 1) with a custom database created for N. gonorrhoeae (precomputed version available in https://github.com/martinghun/ariba-publication/tree/master/N_gonorrhoeae/Ref). ARIBA searches for the presence of bacterial AMR genes and their associated known mutations using reference sequences as a subject database and the fastq files of the strains in the collection as queries. Subsequent analyses were performed using R v3.1.2 (ref. 71): the occurrence of different AMR determinants before and after the change point detected by treeBreaker on the distribution of continents and the distribution of MIC values for penicillin G, tetracycline, ciprofloxacin, ceftriaxone, ceftizoxime and azithromycin against N. gonorrhoeae in the phylogenetic tree using treeBreaker on the distribution of continents and the distribution of MIC values for penicillin G, tetracycline, ciprofloxacin, ceftriaxone, ceftizoxime and azithromycin against the different continents of the genetic determinants. The average number of changes from a susceptible to a resistant state was inferred for each of the resistant determinants under study in both lineage A and lineage B independently using stochastic mapping (100 simulations) with the make.simmmap function implemented in the phytools R package.

The inferred number was corrected by the number of edges in each lineage: n = 586 in lineage A and n = 236 in lineage B. To confirm our hypothesis on the two lineages being associated to different risk groups and antimicrobial susceptibilities, we downloaded the phylogenetic tree of 1,054 European isolates from a 2013 Euro-GASP survey from the Pathogenwatch N. gonorrhoeae scheme (https://pathogenwatch.collection/eurogasp2013). The breakpoint between lineage A and lineage B was detected by obtaining a combined core genome alignment of this and our global set (1,473 strains in total) using Roary v3.11.3 (ref. 72) and running a pseudo-maximum likelihood tree with the resulting SNPs with FastTree v2.1.9 (ref. 73).

Visualization. Visualization of metadata in phylogenetic trees was performed using iTOL v4 (ref. 74). Mapping and the presence or absence of AMR determinants detected with ARIBA were visualized using Phandango v1.1.0 (ref. 74).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All genomic data have been deposited in the European Nucleotide Archive (ENA) under project number PRJEB6203. Accession numbers for the particular strains are indicated in Supplementary Table 1. All other data supporting the findings of this study are available within the paper and its Supplementary Information files.

Code availability

The custom Perl script to convert xma to fasta files (xma2fas.pl) is available from https://gist.github.com/leosanbu/.

Received: 4 June 2018; Accepted: 30 May 2019; Published online: 29 July 2019

References

1. Newman, L. et al. Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. PLoS ONE 10, e0143304 (2015).
2. Unemo, M., Del Rio, C. & Shafer, W. M. Antimicrobial resistance expressed by Neisseria gonorrhoeae: a major global public health problem in the 21st century. Microb. Spectr. 4, E1109-0095 (2016).
3. Fifer, H. et al. Failure of dual antimicrobial therapy in treatment of gonorrhea. N. Engl. J. Med. 374, 2504–2506 (2016).
4. Wi, T. et al. Antimicrobial resistance in Neisseria gonorrhoeae: global surveillance and a call for international collaborative action. PLoS Med. 14, e1002344 (2017).
5. Fingerhuth, S. M., Bonhoeffer, S., Low, N. & Althaus, C. L. Antibiotic-resistant Neisseria gonorrhoeae spread faster with more treatment, not more sexual partners. PLoS Pathog. 12, e1005611 (2016).
6. Annual Epidemiological Report for 2016 (ECDC, 2018); https://ecdc.europa.eu/en/publications-data/gonorrhoea-annual-epidemiological-report-2016
7. Layt, G. A Textbook on Gonorrhoea and its Complications (Ballière, Tindall and Cox, 1913).
8. Grmek, M. Diseases of the Ancient Greek World (John Hopkins Univ. Press, 1989).
9. Oriel, J. D. The Scars of Venus: A History of Venereology (Springer, 1994).
10. Spratt, B. G. Hybrid penicillin-binding proteins in penicillin-resistant strains of Neisseria gonorrhoeae. Nature 332, 173–176 (1988).
11. Oknishi, M. et al. Is Neisseria gonorrhoeae initiating a future era of untreatable gonorrhea? Detailed characterization of the first strain with high-level resistance to ceftriaxone. Antimicrob. Agents Chemother. 55, 3538–3545 (2011).
12. Oknishi, M. Ceftriaxone-resistant Neisseria gonorrhoeae, Japan. Emerg. Infect. Dis. 17, 148–149 (2011).
13. Lefebvre, B. et al. Ceftriaxone-resistant Neisseria gonorrhoeae, Canada, 2017. Emerg. Infect. Dis. 24, 381–383 (2018).
14. Golparian, D. et al. Multidrug-resistant Neisseria gonorrhoeae isolates, belonging to the internationally spreading Japanese FC428 clone, with ceftriaxone resistance and intermediate resistance to azithromycin, Ireland, August 2018. Euro Surveill. 23, 1800617 (2018).
Acknowledgements
We thank H. To and O. Gascuel for their help with the LSD software, and the Pathogen Informatics group at the Wellcome Sanger Institute for informatics support. We also thank S. Szreter, T. Bayliss-Smith and P. Mitchell from the University of Cambridge, Cambridge, UK, for interesting discussions on the historical evidence of gonorrhoea infection. The Japanese isolates were kindly provided by Y. Watanabe and T. Kuroki, Department of Microbiology, Kanagawa Prefectural Institute of Public Health, Kanagawa, Japan. This work was funded by the Wellcome grant number 098051 and the Foundation for Medical Research at Örebro University Hospital, Örebro, Sweden. J.C. was funded by the ERC grant number 745258. Y.H.G. is supported by The Smith Family Foundation and the NIH/NIAID grant 1R01AI132606-01.

Author contributions
S.R.H., M.U., S.D.B. and J.P. conceived and managed the study. L.S.B. and S.R.H. analysed the data and drafted the manuscript. D.G., M.U. and M.O. cultured isolates and extracted DNA. L.S.B., S.R.H., M.U. and Y.H.G. interpreted the data. J.C. provided statistical analysis. R.E. advised on historical interpretation. All authors contributed to the writing of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0501-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.R.H.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019
**Reporting Summary**

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

**Statistics**

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  
  *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  
  *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated
  
  *Our web collection on statistics for biologists contains articles on many of the points above.*

**Software and code**

Policy information about availability of computer code

Data collection

No software was used.

Data analysis

SMALT v0.7.4, SAMtools v1.2, BCFtools v1.2, GATK v1.5.9, progressiveMAUVE v2.3.1, Gubbins v1.4.10, MUMMER v3.23, Gblocks v0.91b, hierBAPS v7.3, RAxML v7.8.6, get_sequence_type https://github.com/sanger-pathogens/mlst_check/blob/master/bin/get_sequence_type, ngmaster v0.4, BLAST v2.3.0+, ARIBA v2.4, in_silico_pcr https://github.com/simonrharris/in_silico_pcr, BOOSTER v0.1.2, Python scikit-learn module (ACCTRAN) http://scikit-learn.org, Pathogenwatch https://pathogen.watch/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

**Data**

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All genomic data has been deposited in the European Nucleotide Archive (ENA) under project number PRJEB4024. Accession numbers for the particular strains are indicated in Supplementary Table 1. All other data supporting the findings of this study are available within the paper and its supplementary information files.
Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Study description | Evolution of Neisseria gonorrhoeae |
|-------------------|------------------------------------|
| Research sample   | Genome sequencing of 419 Neisseria gonorrhoeae isolates from 58 countries around the world isolated between 1960 and 2013. The collection was meant to represent most of the modern circulating Neisseria gonorrhoeae lineages worldwide, including strains with resistant and susceptible phenotypes to several antimicrobials. |
| Sampling strategy | Samples were collected where available to maximise geographic and temporal distribution. The number of collected samples was sufficient to get an overview of the circulating strains worldwide. |
| Data collection   | The isolate collection was put together by Magnus Unemo at Örebro University. Strains were isolated in each country, sent to Örebro University, and then a batch containing all strains was shipped to the Wellcome Sanger Institute for sequencing. |
| Timing and spatial scale | Samples were collected between 1960 and 2013 and from 58 different countries to maximise geographic and temporal distribution. |
| Data exclusions   | No data were excluded. |
| Reproducibility  | Due to the nature of the study reproducibility is not assessable because the sequenced DNAs are not available due to the destructive nature of sequencing. Strains are available from the authors on request. |
| Randomization     | Randomization was performed on the continents of isolation to test if the observed differentiation between continents was significant. It was also applied on the isolation dates to get an estimate of the statistical significance of the coefficient of determination ($r^2$) between the sample dates and the root-to-tip distances. |
| Blinding          | Blinding was not relevant as all available samples were included. |

Did the study involve field work?  Yes  No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

**Materials & experimental systems**

- n/a Involved in the study
  - ☒ Antibodies
  - ☒ Eukaryotic cell lines
  - ☒ Palaeontology
  - ☒ Animals and other organisms
  - ☒ Human research participants
  - ☒ Clinical data

**Methods**

- n/a Involved in the study
  - ☒ ChIP-seq
  - ☒ Flow cytometry
  - ☒ MRI-based neuroimaging