nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-019-0501-y

The impact of antimicrobials on gonococcal evolution

Leonor Sánchez-Busó[1], Daniel Golparian[2], Jukka Corander[1,3,4], Yonatan H. Grad[5,6], Makoto Ohnishi[7,8], Rebecca Flemming[9], Julian Parkhill[1], Stephen D. Bentley[1], Magnus Unemo[2] and Simon R. Harris[1*]

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[1]. 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[2]. 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[3]; however, decreased susceptibility to ceftriaxone has been reported from all continents and azithromycin resistance is on the increase globally[4], 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)[5]. However, AMR is not the only factor driving the recent success of *N. gonorrhoeae*. Dual therapy is effective against the vast majority of infections, yet since its introduction, gonorrhoea infections have continued to increase in most settings[6].

Georges Luys famously opened his medical textbook on gonorrhoea with the statement that "Gonorrhoea is as old as mankind"[7]. However, despite *N. gonorrhoeae* often being described as an ancient pathogen, there are no clear descriptions of a disease like modern gonorrhoea in the ancient sources. Some compatible symptoms do appear in the medical literature of classical Greece and Rome, but nothing decisive, and the presence or absence of modern gonorrhoea in the ancient Mediterranean has been much debated as a result[8]. Early modern terms such as 'the clap', 'the pox' or 'the venereal disease' also covered a range of conditions, and it was not until 1879 that Albert Neisser identified the bacteria that now bears his name[9]. AMR in *N. gonorrhoeae* became apparent soon after antimicrobials were first introduced for its treatment. One characteristic of *N. gonorrhoeae* that 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* (encoding penicillin-binding protein 2 (PBP2)) alleles gained via recombination have been key in the emergence of resistance to extended-spectrum cephalosporins[10,11], which led to the replacement of cefixime as the first-line treatment for gonorrhoea. The first mosaic *penA* allele 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]), but similar mosaic *penA* alleles have been seen worldwide[2,11,13,14]. 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[2].

Whole-genome sequencing has been successfully used to reveal the origins, global spread and population structure of several human pathogens[15]. However, gonococcal genome sequencing has mostly targeted specific populations and outbreaks[16–20]. Here, we report the findings of a global genomic study of 419 *N. gonorrhoeae* isolates spanning 5 continents and more than 50 years,

[1]Pathogen Genomics, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. [2]WHO Collaborating Centre for Gonorrhoea and other Sexually Transmitted Infections, National Reference Laboratory for Sexually Transmitted Infections, Department of Laboratory Medicine, Clinical Microbiology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden. [3]Helsinki Institute for Information Technology, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland. [4]Department of Biostatistics, University of Oslo, Oslo, Norway. [5]Department of Immunology and Infectious Diseases, Harvard TH Chan School of Public Health, Boston, MA, USA. [6]Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [7]Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo, Japan. [8]Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, Tokyo, Japan. [9]Faculty of Classics, University of Cambridge, Cambridge, UK. *e-mail: sh16@sanger.ac.uk

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 3.74×10^{-6} substitutions per site per year confidence interval (CI; 3.39×10^{-6} to 4.07×10^{-6}), which is similar to previous reports[16,18] and comparable to other bacteria[21]. 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 tMRCAs 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[16], there was support for an African origin (90.7% inferred ancestry) (Supplementary Fig. 3 and Supplementary Table 6). From this African root, we identified numerous change points in the continental distribution of isolates across the tree (Supplementary Fig. 3). Most of these were recent events, but the most significant change point separated a basal lineage containing a high proportion of African isolates (68.2%, 30 out of 44) from a lineage containing a high proportion of Asian isolates (92.6%, 137 out of 148), despite the temporal sampling from the two continents being similar (Supplementary Fig. 3). When combined with the dating, this can be interpreted as an early introduction of the modern gonococcus population into Asia (1617, CI: 1578–1649; Fig. 2) soon after its emergence. More recently, many re-introductions into the rest of the world have occurred from this Asian lineage, contributing to the highly mixed population observed today.

Emergence of antimicrobial-resistant gonorrhoea. Minimum inhibitory concentrations (MICs) for six antimicrobials (Supplementary Fig. 4) and the occurrence of genetic AMR determinants were significantly higher among the isolates belonging to the lineage that arose after the phylogeographical breakpoint representing the initial introduction into Asia (Wilcoxon test $W=66{,}159$, $P<0.0001$) (Figs. 3 and 4c). Thus, we refer to the 298 isolates after the breakpoint as lineage A and the 121 isolates before the breakpoint as lineage B.

Two AMR determinants, *folP* R228S, which reduces susceptibility to sulfonamides, and *rpsJ* V57M, which reduces susceptibility to tetracyclines, were carried by a large proportion of isolates, especially in lineage A (Fig. 3 and Supplementary Table 1). Fifty-one isolates contained a mosaic *penA* allele[22]. We identified three independent gains of mosaic alleles, all in lineage A. In a clade of 59 isolates with multilocus sequence typing (MLST) ST1901, a first recombination event replaced the wild-type allele with a mosaic *penA*10 allele and a subsequent event replaced *penA*10 with mosaic *penA*34. These two alleles differ by 16 single-nucleotide polymorphisms (SNPs) and a codon insertion in the last 105 bases of the nucleotide sequence. Two isolates in this clade exhibited high MICs for both cefixime (3–4 mg l^{-1}) and ceftriaxone (2 mg l^{-1}) (Supplementary Fig. 5), and these were found to possess *penA*42, which is a single SNP (A501P) variant from *penA*34 (refs. [23,24]). In another lineage, associated with MLST ST7363, most isolates possessed the *penA*10 alleles, but we again observed a case of replacement with *penA*34. Only one isolate carried the A2045G 23S rRNA mutation (A2059G in *Escherichia coli*) that confers high-level resistance to azithromycin. Six isolates carried the low-level azithromycin resistance C2597T 23S rRNA mutation (C2611T in *E. coli*). Strikingly, the plasmids carrying *tetM* and bla_{TEM} colocalized far more frequently than expected (Pearson's $\chi^2=97.82$, d.f. = 1, $P<0.0001$), possibly reflecting the mobilization of pBlaTEM by the pConjugative plasmid[25], and were completely absent from isolates resistant to extended-spectrum cephalosporins (ESCs) (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 bla_{TEM}: $\chi^2=0.88$, 95% CI: −0.046 to 0.147, d.f. = 1, $P=0.35$). By contrast, of 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[5]. 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 genomic 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

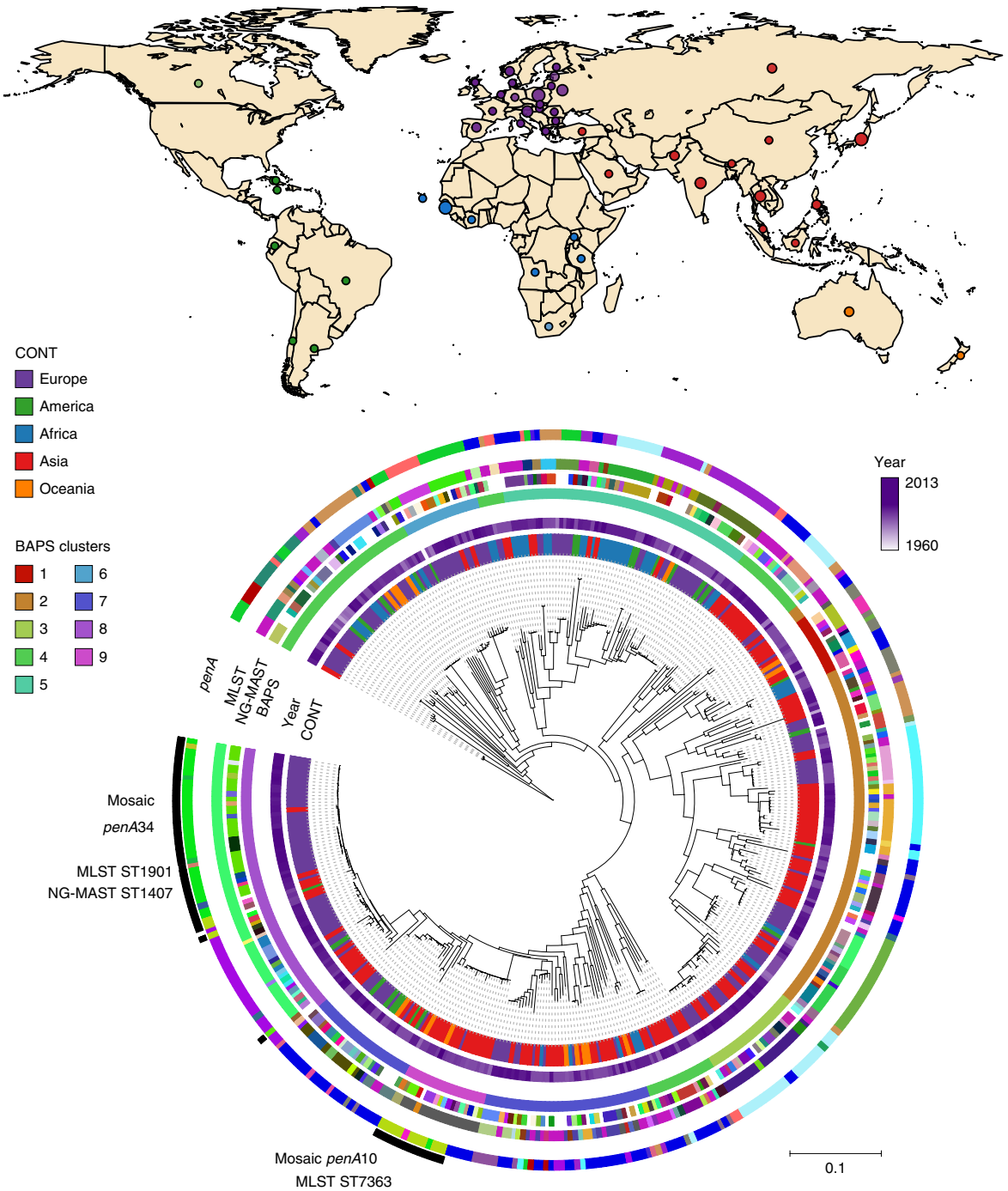

Fig. 1 | Geographical and phylogenetic distribution of *N. gonorrhoeae* isolates. The map (top) shows the countries of isolation of the strains in the collection coloured by continent. The phylogeny (bottom) shows the relationship among the strains ($n = 419$). The coloured strips show (from inside out) 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.

suggest that lineage B is more closely associated with heterosexuals. Corroborating this, data from a 2013 European-wide structured survey[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[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

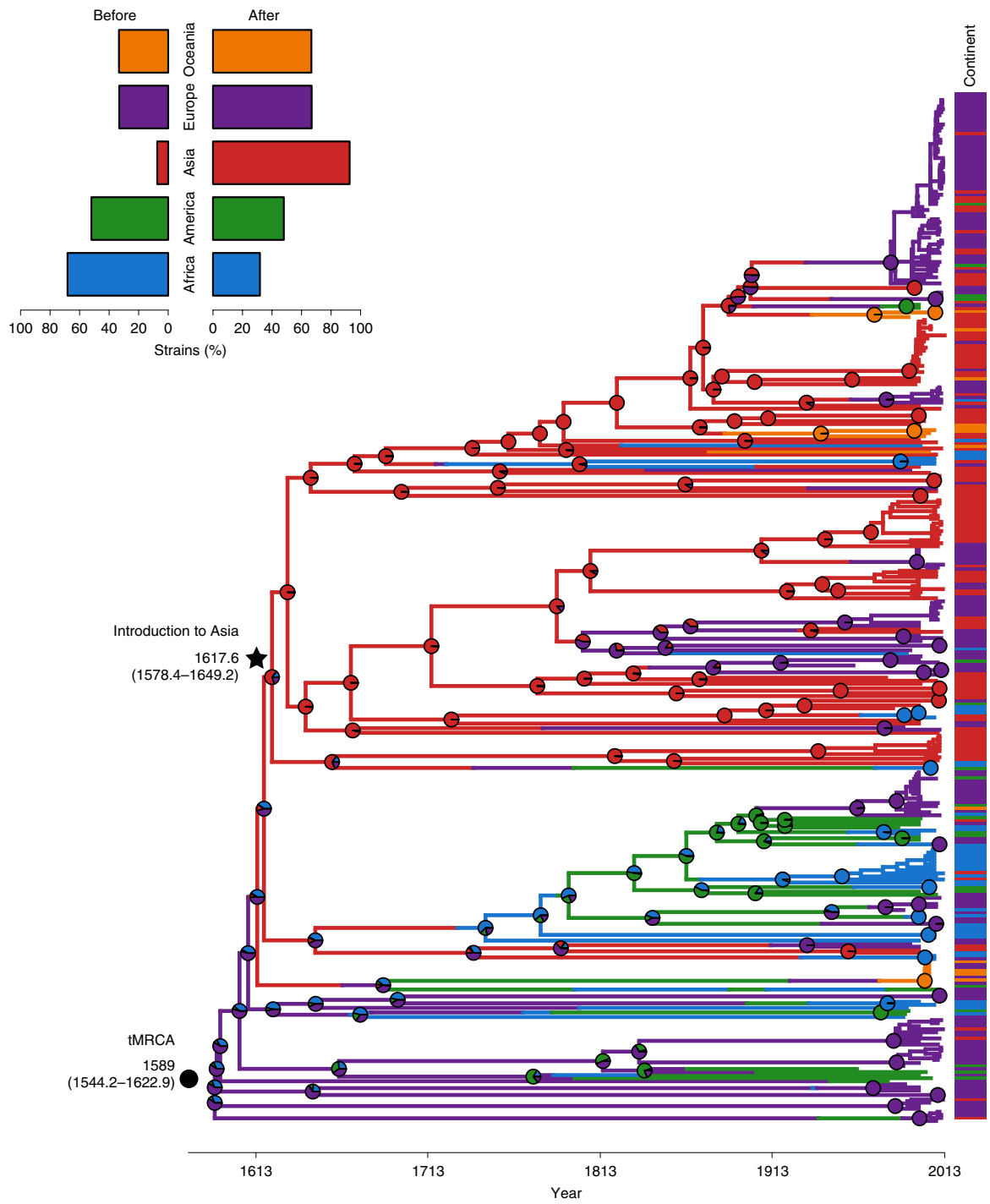

Fig. 2 | Global phylogeographical analysis. The dated maximum likelihood phylogenetic tree shows the posterior probabilities for each continent at every node (pie charts). Continents of isolation (prior) are shown as metadata next to the tips ($n = 419$). The top left graph contains information on the proportion of strains from different continents before ($n = 121$) and after ($n = 298$) the introduction to Asia.

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[2]. 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.

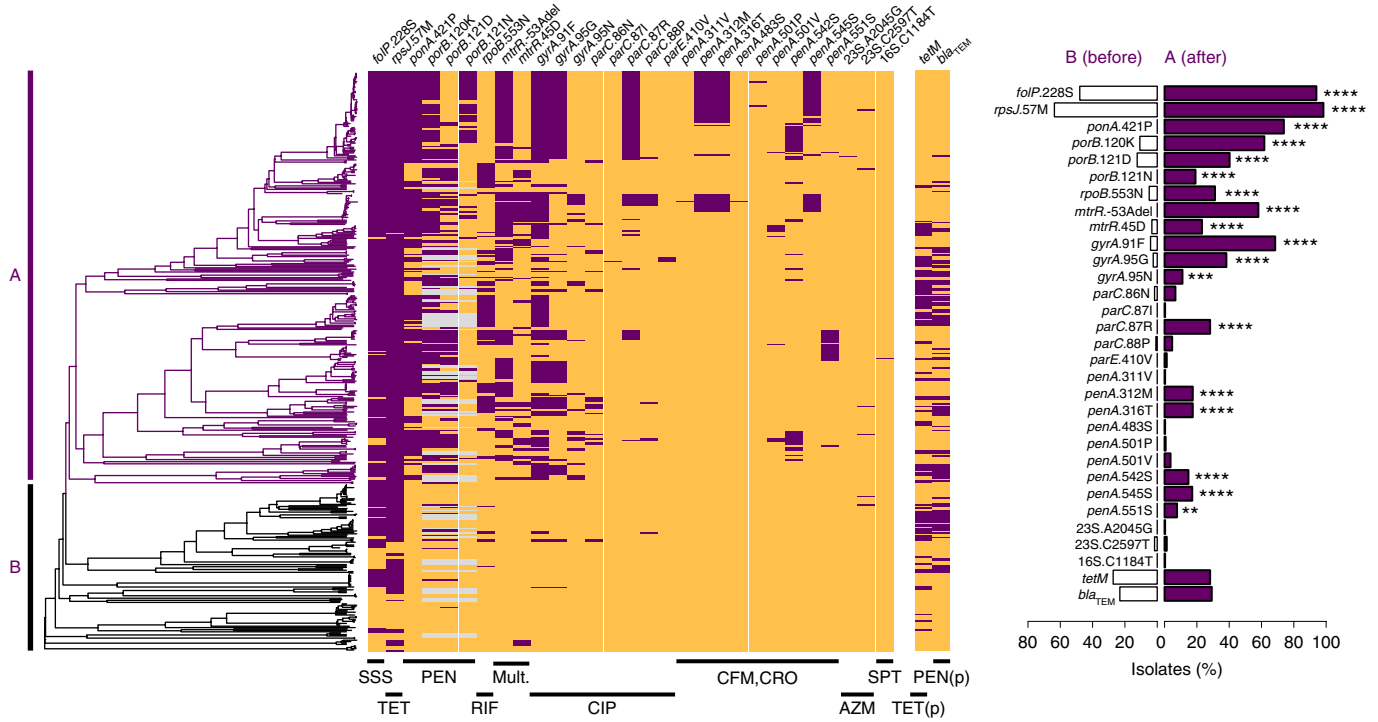

Fig. 3 | Evolution of AMR genetic determinants in *N. gonorrhoeae*. AMR determinants (chromosomal mutations and the presence or absence of the *tetM* and bla_{TEM} genes on plasmids (p)) detected in the new 413 strains included in this study using ARIBA[56] and mapped on the maximum-likelihood-dated tree. Purple represents the presence of the determinant and orange represents the absence of the determinant. Grey indicates the isolates possessing *porB1a* rather than *porB1b*. The two main lineages are marked as A ($n=294$) and B ($n=119$). The left graph shows the proportion of strains with each resistance determinant for both lineages. Black bars below the left graph indicate the antimicrobials associated with each resistance determinant. Statistical significance from a two-sided test for equality of proportions is also shown in the right graph: $^{**}P<0.01$, $^{***}P<0.001$ and $^{****}P<0.0001$. SSS, sulfonomides; TET, tetracyclines; PEN, penicillin; RIF, rifampicin; CIP, ciprofloxacin; CFM, cefixime; CRO, cephtriaxone; AZM, azithromycin; SPT, spectinomycin; (p), plasmid-encoded resistance determinants, Mult., multiple.

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 gonorrhoeal-like symptoms may have been caused by other pathogens or are evidence of an ancient *N. gonorrhoeae* 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[28]. This period was of utmost importance for globalization due to an expeditious increase in exchange of goods, including the import of crops from the Americas to Europe. Increased movement of people around the world also spawned local epidemics and pandemics[29], 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[2]. However, sulfonamides are still widely used in combination with trimethoprim for prophylaxis in patients who are HIV positive and to treat various bacterial infections[30]. Doxycycline (a tetracycline) is still used to treat gonococcal or presumptively non-gonococcal urethritis/cervicitis and is the recommended treatment for lymphogranuloma venereum[31]. 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 *rspJ* 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

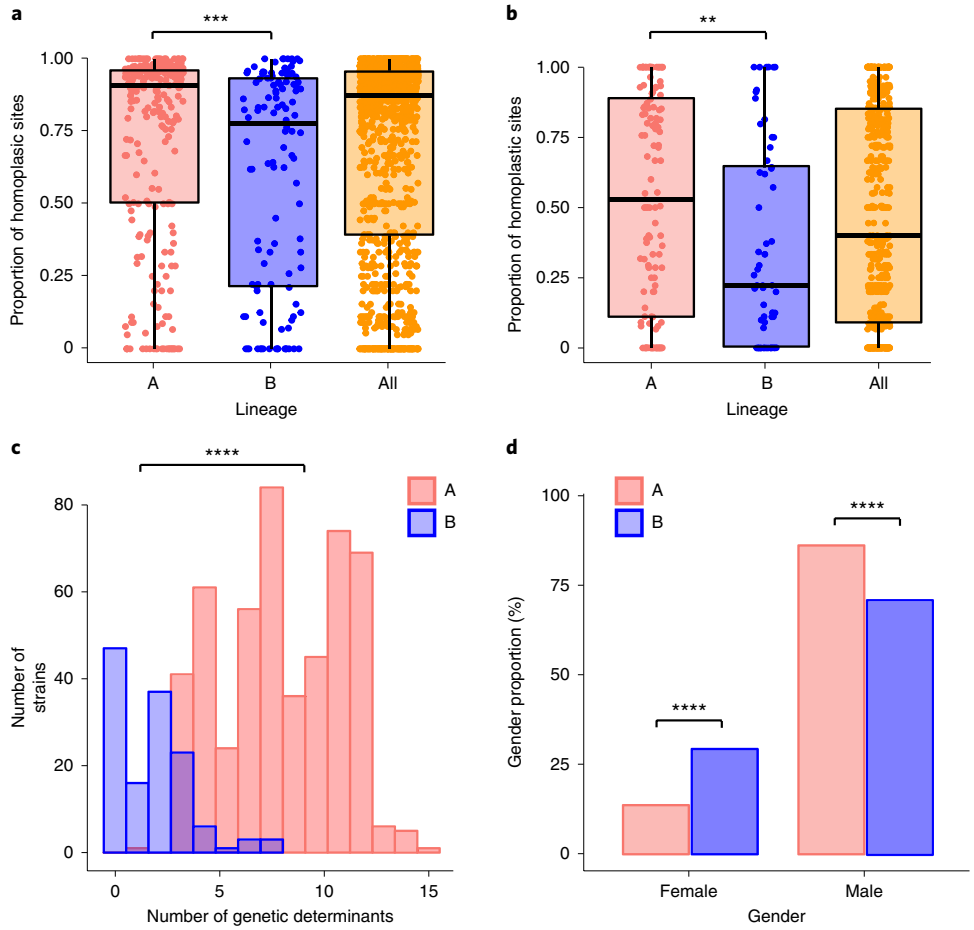

Fig. 4 | Characterization of the lineages of *N. gonorrhoeae*. **a**,**b**, Distribution of the proportions of homoplasic sites in all terminal branches (**a**) and in short terminal branches (≤100 SNPs) (**b**) in lineage A ($n=298$), lineage B ($n=121$) and all strains ($n=419$) represented as boxplots. Each point represents the proportion of homoplasies in one branch drawn from the total variation found in that branch. The box represents the first and third quartiles, the thick solid line in the box represents the median, and the whiskers extend from the first quartile −1.5 times the interquartile range to the third quartile +1.5 times the interquartile range. Statistical significance between lineage A and lineage B was assessed using a two-sided Wilcoxon test: $^{**}P<0.01$ and $^{***}P<0.001$. **c**, Distribution of the total number of AMR genetic determinants in the strains of each lineage as detected using ARIBA[56] (lineage A: $n=294$; lineage B: $n=119$). **d**, Proportion of strains isolated from female ($n=114$) and male ($n=566$) patients in each lineage obtained from combining the global data set with 376 isolates from two North American genomic studies[17,26] (total: $n=639$; lineage A: $n=503$; lineage B: $n=136$). In **c**,**d**, the asterisks show the significance level from a two-sided test for equality of proportions with continuity correction: $^{****}P<0.0001$.

lineage A. The DNA gyrase A S91F substitution, which provides resistance to ciprofloxacin, is one of many resistance mutations that show extremely high levels of homoplasy in lineage A, consistent with a combination of de novo mutation and rapid dissemination via recombination. The mosaic *penA* alleles, which reduce susceptibility to extended-spectrum cephalosporins are another example. These elements were first described in *N. gonorrhoeae* around the turn of the century, but have already been independently acquired by numerous A sublineages, clearly showing that these mutations are transferring *en masse* via recombination rather than by repeated de novo mutation. Conversely, lineage B has remained susceptible to most antimicrobials. More generally, the levels of homoplasy and SNP clustering were found to be significantly higher in clade A, supporting the hypothesis that higher rates of recombination in this lineage have played a role in its high levels of AMR.

The rise of AMR gonorrhoea is generally assumed to have been facilitated by particular demographics who partake in high-risk sexual behaviours, particularly unprotected sex with multiple partners. These groups are also more often treated with antimicrobials than the general population due to frequent infection[32]. Concordantly, 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.[33] 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 gonococcus in rural South African women, in which 48% of cases were asymptomatic and another 50% were symptomatic but not seeking care[34]. Similarly, in Namibia, the prevalence of asymptomatic gonococcal infections in both men and women in rural villages is high[35]. 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.[36] reported, for example, that 23S rRNA mutations resulting in azithromycin resistance were associated with reduced ESC MICs in isolates with mosaic *penA* alleles. Similarly, we have observed that the *tetM* and bla_{TEM}-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 gonorrhoea 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[37]. β-Lactamase production and MICs were tested for a range of antimicrobials as described previously[38]: spectinomycin, tetracycline, penicillin G, ciprofloxacin, azithromycin, cefixime and ceftriaxone.

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 2×100 bp platform at the Wellcome Sanger Institute (Hinxton, UK).

Mapping and variant calling. Fastq files from the 413 new gonococcal strains and the *Neisseria meningitidis* 10356_1#65 outgroup (ENA accession number: ERS248641) were mapped to a common reference, *N. gonorrhoeae* FA1090 (NCBI accession NC_002946; 2,153,922 bp) using SMALT v0.7.4 (http://www.sanger.ac.uk/science/tools/smalt-0). Variants were called using SAMtools and BCFtools v1.2 (ref. [39]) after indel (insertion or deletion) realignment with GATK v1.5.9 (ref. [40]) and further filtered as described previously[41].

Six public reference genomes were obtained from the NCBI and aligned using progressiveMAUVE v2.3.1 (ref. [42]) (*N. gonorrhoeae* FA1090 (NCBI accession NC_002946.2), FA19 (NCBI accession CP012026.1), FA6140 (NCBI accession CP012027.1), MS11 (NCBI accession NC_022240.1), 35/02 (NCBI accession CP012028.1) and NCCP11945 (NCBI accession NC_011035.1)). The XMFA 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[43] were masked in the alignment before running Gubbins v1.4.10 (ref. [44]), which was used to remove segments that can have undergone recombination. This is done by detecting regions of the alignment in which SNPs are densely clustered and occur on the same branches of the tree. The *N. meningitidis* 10356_1#65 strain was used as the outgroup so that events affecting all *N. gonorrhoeae* strains were not excluded from subsequent calculations.

The detected recombination events and repeat regions inferred by repeat-match (MUMMER v3.23)[45] using default options on the *N. gonorrhoeae* FA1090 strain genome were masked to minimize the occurrence of false-positive SNPs. Gblocks v0.91b[46] was run on the resulting alignment to further clean poorly aligned regions that may introduce noise to phylogenetic analyses. Gblocks was run by allowing gap positions in up to 50% of the sequences, with a minimum block length of 10 and 8 as the maximum number of contiguous non-conserved positions. The resulting 1,211,180-bp clean alignment included 15,562 variable sites, identified by snp-sites[47], 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. [48]).

The final SNP alignment was used for maximum likelihood phylogenetic tree reconstruction using RAxML v7.8.6 (ref. [49]) under the GTRGAMMA model of nucleotide substitution and 100 bootstrap replicates. An algorithm called BOOSTER v0.1.2 (ref. [50]) was also used to obtain an enhanced estimate of node support values (Supplementary Note). Ancestral states of all SNPs before recombination removal were reconstructed onto the resulting phylogenetic tree using ACCTRAN transformation in python (http://scikit-learn.org). Homoplasic sites in the terminal branches of the tree were detected for the whole tree and the two main lineages. It is important to note that Gubbins removed 97% (33,026 out of 34,034) of those homoplasic 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 Institute[51]. MLST[52] and *N. gonorrhoeae* multi-antigen sequence typing (NG-MAST)[53] typing schemes were retrieved directly from the sequences using the get_sequence_type script (https://github.com/sanger-pathogens/mlst_check/blob/master/bin/get_sequence_type) and NGMASTER v0.4 (ref. [54]), respectively. The presence of the β-lactamase (bla_{TEM}) and tetracycline (*tetM*) genes on plasmids and the gonococcal genomic island were detected using BLAST v2.3.0+[55] and ARIBA v2.4 (ref. [56]). Typing was performed for the conjugative plasmid and the bla_{TEM} plasmids using an in silico PCR (https://github.com/simonrharris/in_silico_pcr). Primers to differentiate between the Dutch and the American *tetM*-containing plasmids were obtained from Turner et al.[57]. To type the bla_{TEM} plasmids, the primers described in Dillon et al.[58] 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. [59]) was used to perform an analysis of molecular variance (AMOVA) test on the non-recombining section of the genome[60] 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 permutes the population structure to assess the observed signal of differentiation.

To further study population structure, a discriminant analysis of principal components (DAPC; adegenet R package v2.1.1)[62,63] analysis was applied to the non-recombining 15,562 SNPs alignment using continent of isolation as population. The procedure followed by this multivariate discriminant analysis 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[64] was applied to test whether there were differences between the means of the different clusters (continents) on the discriminant clustering. 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[65].

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 <80% of posterior assignment to any of the continents were considered as admixed.

Divergence estimation with LSD and BEAST. The year of isolation for all of the strains was used to calculate a root-to-tip distance regression versus the time to make an estimate of the temporal signal in the data. To do this, a 'clustered permutation' approach was used as described[66,67], which considers potential confounding temporal and genetic structure in the data. A total of 1,000 permutations were performed with this method by randomizing the isolation dates to get an estimate of the statistical significance of the results. This procedure was applied to the whole data set and to the different BAPS clusters.

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[68] was used. This approach has been shown to be robust to uncorrelated changes of the molecular clock and to give similar results to BEAST[68]. To compare the performance between LSD and BEAST, individual BAPS clusters were used. Specifically, Bayesian approximation using BEAST v1.8.2 (ref. [69]) was run to estimate tMRCA and the substitution rate of the genetic clusters determined by hierBAPS[48]. Three chains were run per cluster up to 100 million generations by using a GTRGAMMA model of nucleotide substitution with 4 categories, strict molecular clock with a diffuse gamma distribution (shape 0.001 and scale 1,000) and a constant population size as priors. Default priors were used. For models using relaxed clocks, the ucld 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. [70]) (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[71] 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.simmap function implemented in the phytools R package v0.6-44 (ref. [72]). 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[73]. 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[16] was added to the global collection and the phylogeographical 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[72], resulting in a total of 1,000 evaluated maps.

Evolution of AMR determinants. Mutations conferring AMR in known genetic determinants (16S rRNA, 23S rRNA, *rpoB*, *rpsJ*, *mtrR*, *folP*, *gyrA*, *parC*, *parE*, *penA*, *ponA* and *porB*) as well as the presence of the β-lactamase (bla_{TEM}) and *tetM* genes[74] were obtained for the 413 strains sequenced in this study using ARIBA v2.4 (ref. [56]) (Supplementary Table 1) with a custom database created for *N. gonorrhoeae* (precomputed version available in https://github.com/martinghunt/ariba-publication/tree/master/N_gonorrhoeae/Ref). ARIBA searches for the presence of particular AMR genes and 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. [75]): 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, cefixime and azithromycin against different combinations 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.simmap function implemented in the phytools R package[72]. The inferred number was corrected by the number of edges in each lineage: $n=586$ in lineage A and $n=236$ in lineage B.

As an approximation of studying the risk groups characterizing the defined A and B lineages, 263 isolates from the global collection with information on the gender of the patient were combined with 376 isolates from two North American studies with this information available[17,26]. ARIBA v2.4 was run for the extra isolates and the obtained results were joined with the results from the global data set. Metadata on gender and the number of total resistance determinants detected per strain was plotted on a recombination-free phylogenetic tree obtained as described above and differences between the two lineages were evaluated using two-sided tests for equality of proportions with continuity correction (prop.test) and two-sided Wilcoxon rank sum tests (wilcox.test) with R[75].

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[27] (https://pathogen.watch/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. [76]) and running a pseudo-maximum likelihood tree with the resulting SNPs[47] with FastTree v2.1.9 (ref. [77]).

Visualization. Visualization of metadata in phylogenetic trees was performed using iTOL v4 (ref. [78]). Mapping and the presence or absence of AMR determinants detected with ARIBA were visualized using Phandango v1.1.0 (ref. [79]).

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 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.

Code availability

The custom Perl script to convert xmfa to fasta files (xmfa2fas.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. *Microbiol. Spectr.* **4**, EI10-0009-2015 (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. Luys, G. *A Textbook on Gonorrhoea and its Complications* (Baillière, Tindall and Cox, 1913).
8. Grmek, M. *Diseases in 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. Ohnishi, 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. Ohnishi, 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* isolate, belonging to the internationally spreading Japanese FC428 clone, with ceftriaxone resistance and intermediate resistance to azithromycin, Ireland, August 2018. *Euro Surveill.* **23**, 1800617 (2018).

15. Popovich, K. J. & Snitkin, E. S. Whole genome sequencing—implications for infection prevention and outbreak investigations. *Curr. Infect. Dis. Rep.* **19**, 15 (2017).
16. Grad, Y. H. et al. Genomic epidemiology of *Neisseria gonorrhoeae* with reduced susceptibility to cefixime in the USA: a retrospective observational study. *Lancet Infect. Dis.* **14**, 220–226 (2014).
17. Demczuk, W. et al. Whole-genome phylogenomic heterogeneity of *Neisseria gonorrhoeae* isolates with decreased cephalosporin susceptibility collected in Canada between 1989 and 2013. *J. Clin. Microbiol.* **53**, 191–200 (2015).
18. De Silva, D. et al. Whole-genome sequencing to determine transmission of *Neisseria gonorrhoeae*: an observational study. *Lancet Infect. Dis.* **16**, 1295–1303 (2016).
19. Didelot, X. et al. Genomic analysis and comparison of two gonorrhea outbreaks. *mBio* **7**, e00525-16 (2016).
20. Chisholm, S. A. et al. An outbreak of high-level azithromycin resistant *Neisseria gonorrhoeae* in England. *Sex Transm. Infect.* **92**, 365–367 (2016).
21. Duchene, S. et al. Genome-scale rates of evolutionary change in bacteria. *Microb. Genom.* **2**, e000094 (2016).
22. Demczuk, W. et al. *Neisseria gonorrhoeae* sequence typing for antimicrobial resistance, a novel antimicrobial resistance multilocus typing scheme for tracking global dissemination of *N. gonorrhoeae* strains. *J. Clin. Microbiol.* **55**, 1454–1468 (2017).
23. Cámara, J. et al. Molecular characterization of two high-level ceftriaxone-resistant *Neisseria gonorrhoeae* isolates detected in Catalonia, Spain. *J. Antimicrob. Chemother.* **67**, 1858–1860 (2012).
24. Unemo, M. et al. High-level cefixime- and ceftriaxone-resistant *Neisseria gonorrhoeae* in France: novel penA mosaic allele in a successful international clone causes treatment failure. *Antimicrob. Agents Chemother.* **56**, 1273–1280 (2012).
25. Flett, F., Humphreys, G. O. & Saunders, J. R. Intraspecific and intergeneric mobilization of non-conjugative resistance plasmids by a 24.5 megadalton conjugative plasmid of *Neisseria gonorrhoeae*. *J. Gen. Microbiol.* **125**, 123–129 (1981).
26. Demczuk, W. et al. Genomic epidemiology and molecular resistance mechanisms of azithromycin-resistant *Neisseria gonorrhoeae* in Canada from 1997 to 2014. *J. Clin. Microbiol.* **54**, 1304–1313 (2016).
27. Harris, S. R. et al. Public health surveillance of multidrug-resistant clones of *Neisseria gonorrhoeae* in Europe: a genomic survey. *Lancet Infect. Dis.* **18**, 758–768 (2018).
28. Hopkins, A. G. *Globalization in World History* (Pimlico, 2002).
29. McNeill, W. H. *Plagues and Peoples* (Doubleday, 1977).
30. Huovinen, P. Resistance to trimethoprim–sulfamethoxazole. *Clin. Infect. Dis.* **32**, 1608–1614 (2001).
31. Workowski, K. A. & Bolan, G. A. Sexually transmitted diseases treatment guidelines, 2015. *MMWR Recomm. Rep.* **64**, 1–137 (2015).
32. Kenyon, C. R. & Schwartz, I. S. Effects of sexual network connectivity and antimicrobial drug use on antimicrobial resistance in *Neisseria gonorrhoeae*. *Emerg. Infect. Dis.* **24**, 1195–1203 (2018).
33. Turner, K. M. & Garnett, G. P. The impact of the phase of an epidemic of sexually transmitted infection on the evolution of the organism. *Sex. Transm. Infect.* **78**, i20–i30 (2002).
34. Wilkinson, D. et al. Unrecognized sexually transmitted infections in rural South African women: a hidden epidemic. *Bull. World Health Organ.* **77**, 22–28 (1999).
35. Hazel, A., Ponnaluri-Wears, S., Davis, G. S., Low, B. S. & Foxman, B. High prevalence of *Neisseria gonorrhoeae* in a remote, undertreated population of Namibian pastoralists. *Epidemiol. Infect.* **142**, 2422–2432 (2014).
36. Grad, Y. H. et al. Genomic epidemiology of gonococcal resistance to extended spectrum cephalosporins, macrolides, and fluoroquinolones in the US, 2000–2013. *J. Infect. Dis.* **214**, 1579–1587 (2016).
37. Unemo, M., Olcen, P., Berglund, T., Albert, J. & Fredlund, H. Molecular epidemiology of *Neisseria gonorrhoeae*: sequence analysis of the porB gene confirms presence of two circulating strains. *J. Clin. Microbiol.* **40**, 3741–3749 (2002).
38. Unemo, M., Fasth, O., Fredlund, H., Limnios, A. & Tapsall, J. Phenotypic and genetic characterization of the 2008 WHO *Neisseria gonorrhoeae* reference strain panel intended for global quality assurance and quality control of gonococcal antimicrobial resistance surveillance for public health purposes. *J. Antimicrob. Chemother.* **63**, 1142–1151 (2009).
39. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
40. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
41. Harris, S. R. et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–474 (2010).
42. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147 (2010).
43. Piekarowicz, A. et al. Characterization of the dsDNA prophage sequences in the genome of *Neisseria gonorrhoeae* and visualization of productive bacteriophage. *BMC Microbiol.* **7**, 66 (2007).
44. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
45. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
46. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
47. Page, A. J. et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* **2**, e000056 (2016).
48. Cheng, L., Connor, T. R., Siren, J., Aanensen, D. M. & Corander, J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* **30**, 1224–1228 (2013).
49. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
50. Lemoine, F. et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
51. Page, A. J. et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb. Genom.* **2**, e000083 (2016).
52. Jolley, K. A. & Maiden, M. C. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**, 595 (2010).
53. Martin, I. M., Ison, C. A., Aanensen, D. M., Fenton, K. A. & Spratt, B. G. Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. *J. Infect. Dis.* **189**, 1497–1505 (2004).
54. Kwong, J. C. et al. NGMASTER: in silico multi-antigen sequence typing for *Neisseria gonorrhoeae*. *Microb. Genom.* **2**, e000076 (2016).
55. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
56. Hunt, M. et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genom.* **3**, e000131 (2017).
57. Turner, A., Gough, K. R. & Leeming, J. P. Molecular epidemiology of tetM genes in *Neisseria gonorrhoeae*. *Sex. Transm. Infect.* **75**, 60–66 (1999).
58. Dillon, J. R., Li, H., Yeung, K. & Aman, T. A. A PCR assay for discriminating *Neisseria gonorrhoeae* β-lactamase-producing plasmids. *Mol. Cell. Probes* **13**, 89–92 (1999).
59. Kamvar, Z. N., Tabima, J. F. & Grunwald, N. J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281 (2014).
60. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
61. Dray, S. & Dufour, A. B. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).
62. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
63. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
64. Louis, T. A. Multivariate analysis of variance and repeated measures, a practical approach for behavioural scientists. D. J. Hand and C. C. Taylor, Chapman & Hall, London. *Stat. Med.* https://doi.org/10.1002/sim.4780081218 (1989).
65. Benjamini, Y. Discovering the false discovery rate. *J. R. Stat. Soc. Series B Stat. Methodol.* **72**, 405–416 (2010).
66. Murray, G. G. et al. The effect of genetic structure on molecular dating and tests for temporal signal. *Methods Ecol. Evol.* **7**, 80–89 (2016).
67. Duchene, S., Duchene, D., Holmes, E. C. & Ho, S. Y. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Mol. Biol. Evol.* **32**, 1895–1906 (2015).
68. To, T. H., Jung, M., Lycett, S. & Gascuel, O. Fast dating using least-squares criteria and algorithms. *Syst. Biol.* **65**, 82–97 (2016).
69. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
70. Ansari, M. A. & Didelot, X. Bayesian inference of the evolution of a phenotype distribution on a phylogenetic tree. *Genetics* **204**, 89–98 (2016).
71. Huelsenbeck, J. P., Nielsen, R. & Bollback, J. P. Stochastic mapping of morphological characters. *Syst. Biol.* **52**, 131–158 (2003).
72. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
73. Bollback, J. P. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**, 88 (2006).
74. Unemo, M. & Shafer, W. M. Antimicrobial resistance in *Neisseria gonorrhoeae* in the 21st century: past, evolution, and future. *Clin. Microbiol. Rev.* **27**, 587–613 (2014).
75. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2008).

76. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
77. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
78. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
79. Hadfield, J. et al. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* **34**, 292–293 (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.F. 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.

nature research

Corresponding author(s): Simon R. Harris

Last updated by author(s): May 14, 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.

n/a	Confirmed	
☐	☒	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, poppr R package v2.5.0, ade4 R package v1.7-11, adegenet R package v2.1.1, Least-Square Dating (LSD) v0.3, BEAST v 1.8.2, treeBreaker v1.1 https:// github.com/ansariazim/treeBreaker, phytools R package v0.6-44, R v3.1.2, Roary v3.11.3, FastTree v2.1.9, iTOL v4 http://itol.embl.de/, Phandango v1.1.0 www.phandango.net, custom perl script xmfa2fas.pl https://gist.github.com/leosanbu/, 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.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

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 (r2) 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