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], Simon R. Harris[1,*]

[1]Pathogen Genomics, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, 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.

*Corresponding author: Simon R. Harris (sh16@sanger.ac.uk).

Abstract

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 phylogeographic analysis of 419 gonococcal isolates from across the globe. Results indicate that modern gonococci originated in Europe or Africa, possibly as late as the 16th century and subsequently disseminated globally. We provide evidence that the modern gonococcal population has been shaped by antimicrobial treatment of sexually transmitted and other infections, leading to the emergence of two major lineages with different evolutionary strategies. The well-described multi-resistant lineage is associated with high rates of homologous recombination and infection in high-risk sexual networks. A second, multi-susceptible 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 newborns, and enhanced transmission of HIV[2]. The emergence and proliferation of gonococci with resistance to front-line antimicrobials such as extended-spectrum cephalosporins (ESCs; 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 (MSM) 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 like '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 another *Neisseria* species. For example, mosaic penicillin-binding protein 2 (PBP2; encoded by the *penA* gene) alleles gained via recombination have been key in the emergence of resistance to ESCs[10,11] which led to the replacement of cefixime as the first-line treatment for gonorrhoea. The first mosaic *penA* allele causing high-level ceftriaxone resistance was seen in an isolate from a pharyngeal infection in a female sex worker in Japan in 2009[12], but similar mosaic *penA* alleles have been seen worldwide[2,11,13,14]. In fact, a number of resistances have first been 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 five continents and more than 50 years, including varying susceptibilities to

104 important antimicrobials. Our aim was to elucidate when and where modern
105 gonococcal populations emerged, evolved and dispersed, and how antimicrobial usage
106 and transmission in different sexual networks has influenced their population
107 dynamics.

108

109 **Results**

110 **Modern gonococcus not 'as old as mankind'**

111 Our collection spans a period of more than 50 years (1960 – 2013) and 58 countries
112 from five continents (Supplementary Table 1, Figure 1). A population-level analysis
113 revealed a high level of admixture among *N. gonorrhoeae* with no significant
114 differentiation between continents (Supplementary Table 2), with the exception of
115 Africa (Supplementary Figure 1, Supplementary Tables 3-5). We estimated the
116 substitution rate for the non-recombining section of the genomes in the collection
117 (Supplementary Figure 2) to be $3.74E^{-06}$ substitutions/site/year CI (confidence
118 interval) [$3.39E^{-06}$ – $4.07E^{-06}$], which is similar to previous reports[16,18] and comparable
119 to other bacteria[21]. The time of the most recent common ancestor (tMRCA) was
120 estimated to be around the 16th century (1589, CI [1544 – 1623]) (Figure 2). Although
121 high rates of recombination can lead to underestimation of tMRCAs to some extent,
122 these results are strongly at odds with the hypothesis that modern gonorrhoea has
123 existed as long as mankind and cast further doubt on the ascribing of historical
124 descriptions of gonorrhoea-like symptoms to infection with 'modern' gonococci.

125 Despite modern gonococci being globally mixed, we found strong evidence of historic
126 geographical separation, suggesting rapid mixing of populations is a relatively recent
127 phenomenon. A phylogeographic 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 Figure 3 and Supplementary Table 6). From this African root, we identified a number of change-points in the continental distribution of isolates across the tree (Supplementary Figure 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%, 30/44) from a lineage containing a high proportion of Asian isolates (92.6%, 137/148), despite the temporal sampling from the two continents being similar (Supplementary Figure 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], Figure 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 Figure 4) and the occurrence of genetic AMR determinants were significantly higher among the isolates belonging to the lineage that arose after the phylogeographic breakpoint representing the initial introduction into Asia (Wilcoxon test W=66159, p-value<0.0001) (Figures 3 and 4c). We will therefore 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 (Supplementary Table 1, Figure 3). 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 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 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) and ceftriaxone (2 mg/L) (Supplementary Figure 5), and these were found to possess *penA*42, which is a single SNP (A501P) variant from *penA*34[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} co-localized far more frequently than expected (Pearson's χ^2=97.82, df=1, p-value<0.0001), possibly reflecting the mobilization of pBlaTEM by the pConjugative plasmid[25], and were completely absent from ESC-resistant isolates (Figure 3). The Gonococcal Genomic Island (GGI) was found in 277 (67%) isolates (Supplementary Figure 6), but showed no clear association with AMR. The plasmid-encoded resistances showed no significant difference in prevalence in lineages A or B (two-sided test for equality of proportions for *tetM*: χ^2=0.01, 95% CI [-0.089, 0.110], df=1, p-value=0.92, and for bla_{TEM}: χ^2=0.88, 95% CI [-0.046, 0.147], df=1, p-value=0.35). In contrast, of the 29 chromosomally-mediated resistance substitutions examined, 18 were significantly associated with clade A (Figure 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 Figure 7).

Two strategies for gonococcal success

Overall, our data show far fewer gains of chromosomally-encoded AMR determinants in lineage B compared to A (Supplementary Figure 8). Since 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=19416, p-value<0.001; Figures 4a-b and Supplementary Figure 9). Note that the distribution of branch lengths in both clades was similar (Wilcoxon test W=14427, p-value=0.739). Similarly, the proportion of clustered SNPs, another signal of recombination, was also higher on the terminal branches in lineage A (Wilcoxon test W=16984, p-value<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-value=0.05184).

One explanation for such differences could be opportunity. For recombination to occur, donor and recipient bacteria must co-localise. Thus, recombination between gonococci would be expected to occur more frequently in high-risk host populations where coinfection 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 and other STIs[5].

Unfortunately, due to limitations in availability of data on patient sexual behaviour, we could not adequately assess association of the lineages to risk factors in our dataset. However, we could analyse the distribution of the gender of the patients from which 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 (40/136, 29.4%) than A (69/503, 13.7%) (two-sided test for equality of proportions χ^2=17.54, 95% CI [0.070, 0.244], df=1, p-value<0.0001) (Figure 4d and Supplementary Figure 8), which would suggest that 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 (61/214, 28.5% of lineage B isolates were from women vs 100/821, 12% of lineage A; two-sided test for equality of proportions χ^2=33.21, 95% CI [0.096, 0.231], df=1, p-value<0.0001), and more importantly, of the patients that reported sexual orientation, 78.3% (94/120) of isolates in lineage B were from heterosexuals, in contrast to 52.6% (200/380) in A (two-sided test for equality of proportions χ^2=23.82, 95% CI [0.162, 0.352], df=1, p-value<0.0001) (Supplementary Figure 10). Particular sublineages within lineage B appeared particularly strongly associated with heterosexuals[27]. We suspect 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 and high-risk populations is common within lineage A, so we suspect opportunity is not the only explanation for the differential recombination rate in the two lineages. The observation that plasmid-born resistances do not 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 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 limited availability of isolates, so further study is required in this area.

We estimated an origin of modern gonococci in the 16th 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 our estimated confidence intervals 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 we now know as gonorrhoea is 'as old as mankind'. The 16th 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 phylogeographic 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 17th 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 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 (TMP-SMZ) for prophylaxis in HIV positive patients and to treat a variety of 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]. We therefore suspect 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 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 ESCs are another example. These elements were first described in *N. gonorrhoeae* around the turn of the century, but have already been independently acquired by a number of A sublineages, clearly showing that these mutations are transferring *en masse* via recombination rather than by repeated *de novo* mutation. Lineage B, on the other hand, has remained susceptible to most antimicrobials. More generally, 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.

301 The rise of AMR gonorrhoea is generally assumed to have been facilitated by
302 particular demographics who partake in high-risk sexual behaviours, particularly
303 unprotected sex with multiple partners. These groups are also more often treated with
304 antimicrobials than the general population due to frequent infection[32]. Concordantly,
305 we found that lineage A is associated with infection in MSM, one of the predominant
306 risk groups, while isolates from B are more rarely found in this demographic group.
307 Thus, lineage A isolates have the means (increased homologous recombination),
308 motive (higher antimicrobial exposure) and opportunity (higher rates of coinfection
309 with commensal *Neisseria* and other STIs) for recombination-driven gain of AMR.

310 Most recent media attention and gonococcal genomics research has focused on the
311 increasing levels of AMR in gonorrhoea. However, we have shown that a mostly
312 susceptible lineage is successfully persisting in lower-risk groups where it is probably
313 less likely to be exposed to antimicrobials. Notably, this lineage was associated with
314 heterosexual groups, and with infections in women, where rates of asymptomatic
315 infection are higher. Turner *et al.*[33] showed, using a modelling approach, that in a
316 situation where both resistant and susceptible strains are present in a population, high
317 rates of asymptomatic infection, and therefore under-treatment, can allow susceptible
318 isolates to survive and thrive. In such circumstances, rates of susceptible infection can
319 be hugely underestimated, potentially meaning that our understanding of gonococcal
320 prevalence, and rates of AMR may be biased. Interestingly, the majority of our
321 African samples were from lineage B, consistent with epidemiological studies that
322 describe a hidden epidemic of gonococcus in rural South African women, in which
323 48% of cases were asymptomatic and another 50% were symptomatic but not seeking
324 care[34]. Similarly in Namibia, prevalence of asymptomatic gonococcal infections in
325 both men and women in rural villages are 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 phylogeographic 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 16th 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 A (after the phylogenetic breakpoint) and B (before the phylogenetic breakpoint). Lineage A has gained and proliferated AMR determinants, aided by an increased rate of recombination. We hypothesise that these isolates are often transmitted in higher-risk networks, e.g. MSM, where pharyngeal infections are more common and individuals are more frequently exposed to treatment for gonorrhoea and other STIs. Lineage B, however, 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 suffering gonorrhoea in 58 countries spanning five continents. The strains were selected to represent a wide geographic, temporal, phenotypic (based on antimicrobial resistance), 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 minimum inhibitory concentrations (MIC) 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 2x100 bp platform at the Wellcome Sanger Institute.

Mapping and variant calling

Fastq files from the 413 new gonococcal strains and the *N. meningitidis* 10356_1#65 outgroup (ENA accession 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[39] after indel realignment with GATK v1.5.9[40] and further filtered as described previously[41].

Six public reference genomes were obtained from the NCBI and aligned using progressiveMAUVE v2.3.1[42] (*N. gonorrhoeae* FA1090 (NCBI NC_002946.2), FA19 (NCBI CP012026.1), FA6140 (NCBI CP012027.1), MS11 (NCBI NC_022240.1), 35/02 (NCBI CP012028.1) and NCCP11945 (NCBI NC_011035.1)). The XMFA output alignment was converted into a plain fasta format using *N. gonorrhoeae* FA1090 as 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 one resulting from mapping the 413 isolates, producing a 419-strains 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[44], which was used to remove segments that can have undergone recombination. This is done by detecting regions of the alignment

398 in which single nucleotide polymorphisms (SNPs) are densely clustered and occur on
399 the same branches of the tree. The *N. meningitidis* 10356_1#65 strain was used as
400 outgroup so that events affecting all *N. gonorrhoeae* strains were not excluded from
401 subsequent calculations.

402 The detected recombination events and repeat regions inferred by repeat-match
403 (MUMMER v3.23)[45] using default options on the *N. gonorrhoeae* FA1090 strain
404 genome were masked in order to minimize the occurrence of false-positive SNPs.
405 Gblocks v0.91b[46] was run on the resulting alignment to further clean poorly aligned
406 regions that may introduce noise to phylogenetic analyses. Gblocks was run by
407 allowing gap positions in up to 50% of the sequences, with a minimum block length
408 of 10 and 8 as maximum number of contiguous non-conserved positions. The
409 resulting 1,211,180 bp clean alignment included 15,562 variable sites, identified by
410 snp-sites[47], and was used for population structure analysis, phylogenetic inference and
411 divergence estimation. Genetic clusters were obtained from the non-recombining
412 alignment using hierBAPS v7.3[48].

413 The final SNP alignment was used for Maximum Likelihood (ML) phylogenetic tree
414 reconstruction using RAxML v7.8.6[49] under the GTRGAMMA model of nucleotide
415 substitution and 100 bootstrap replicates. An algorithm called BOOSTER v0.1.2[50]
416 was also used to obtain an enhanced estimate of node support values (Supplementary
417 Note). Ancestral states of all SNPs before recombination removal were reconstructed
418 onto the resulting phylogenetic tree using ACCTRAN transformation in python
419 (http://scikit-learn.org). Homoplasic sites in the terminal branches of the tree were
420 detected for the whole tree and the two main lineages. It is important to note that
421 Gubbins removed 97% (33,026/34,034) of those homoplasic sites, minimizing their
422 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]. Multi-locus sequence typing (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[54], respectively. The presence of the β-lactamase (bla_{TEM}) and tetracycline (*tetM*) genes on plasmids and the Gonococcal Genomic Island (GGI) were detected using BLAST v2.3.0+[55] and ARIBA v2.4[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*, 1999[57]. To type the bla_{TEM} plasmids, primers described in Dillon *et al*, 1999[58] were used and the resulting amplicon sizes evaluated to differentiate among the Asia, Africa and Toronto/Rio types (Supplementary Table 1).

Analysis of population structure

In order to study population structure from the resulting alignment, the *poppr* R package v2.5.0[59] was used to perform an 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. In order to test if the observed differentiation between continents was significant, a randomization test (N = 1000 permutations) was performed using the *randtest* function from the *ade4* R package v1.7-11[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 over-fitting and keep enough discrimination power, the optimal number of principal components (PC) 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 PCs. This methodology resulted in 83 principal components as optimal to keep a balance between discrimination power and over-fitting. Prior assignment to continents was randomized and the DAPC analysis repeated to confirm that the observed separation among clusters does not occur by chance. Four discriminant functions were kept for the analysis, considering the number of variables was 5 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 False Discovery Rate (FDR)[65].

DAPC derives group membership probabilities from the retained discriminant functions. These results were used to evaluate the level of admixture in the dataset 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

Year of isolation for all the strains was used to calculate a root-to-tip distance regression versus 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 1000 permutations were performed with this method by randomizing the isolation dates in order to get an estimate of the statistical significance of the results. This procedure was applied to the whole dataset and to the different BAPS clusters.

In order to get an estimate of the substitution rate and the Most Recent Common Ancestor (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]. In order to compare the performance between LSD and BEAST, individual BAPS clusters were used. Specifically, Bayesian approximation using BEAST v1.8.2[69] was run to estimate the 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 1000) 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 1000. 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 dataset. LSD was also run for the BAPS clusters that reached convergence in BEAST and the results 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[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[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 US[16] was added to the global collection and the phylogeographic analyses repeated to confirm our results. To avoid biases due to different number of strains from different continents, the combined datasets were down sampled 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 is 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 1000 evaluated maps.

Evolution of antimicrobial resistance determinants

Mutations conferring antimicrobial resistance 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[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[75]: the occurrence of different antimicrobial resistance 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 lineages A and 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 from two North American studies with this information available[17,26]. ARIBA v2.4 was run for the extra isolates and the obtained results joined with the ones from the global dataset. Metadata on gender and 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 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].

In order 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 lineages A and 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[76] and running a pseudo-maximum likelihood tree with the resulting SNPs[47] with FastTree v2.1.9[77].

Visualization

Visualization of metadata in phylogenetic trees was performed using iTOL v4[78]. Mapping of the and presence/absence of antimicrobial resistance determinants detected with ARIBA were visualized using Phandango v1.1.0[79].

Code availability

The custom Perl script to convert xmfa to fasta files (xmfa2fas.pl) is available from https://gist.github.com/leosanbu/.

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.

Correspondence and requests for materials should be addressed to S.R.H (sh16@sanger.ac.uk).

Acknowledgements

We thank Hien To and Olivier Gascuel for their help with the LSD software, and the Pathogen Informatics group at the Wellcome Sanger Institute for informatics support. We also thank Prof. Simon Szreter, Prof. Tim Bayliss-Smith and Dr. Piers Mitchell from the University of Cambridge for interesting discussions on the historical evidence of gonorrhoea infection. Japanese isolates were kindly provided by Yuko Watanabe and Toshio Kuroki, Department of Microbiology, Kanagawa Prefectural Institute of Public Health, Kanagawa, Japan. This work was funded by Wellcome grant number 098051 and the Foundation for Medical Research at Örebro University Hospital, Örebro, Sweden. JC was funded by the ERC grant no. 745258. YHG is supported by The Smith Family Foundation and NIH/NIAID grant 1R01AI132606-01.

Author contributions

SRH, MU, SDB and JP conceived and managed the study. LSB and SRH analysed the data and drafted the manuscript. DG, MU and MO cultured isolates and extracted DNA. LSB, SRH, MU and YG interpreted the data. JC provided statistical analysis. RF advised on historical interpretation. All authors contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Author information

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

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

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

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, doi:10.1371/journal.pone.0143304 (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**, doi:10.1128/microbiolspec.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, doi:10.1056/NEJMc1512757 (2016).

4 Wi, T. *et al.* Antimicrobial resistance in Neisseria gonorrhoeae: Global surveillance and a call for international collaborative action. *PLoS Med* **14**, e1002344, doi:10.1371/journal.pmed.1002344 (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, doi:10.1371/journal.ppat.1005611 (2016).

6 ECDC, European Centre for Disease Prevention and C. Annual Epidemiological Report., (Stockholm: ECDC. Available from http://ecdc.europa.eu/en/healthtopics/gonorrhoea/Pages/Annual-Epidemiological-Report-2016.aspx, 2016).

7 Luys, G. *A textbook on gonorrhoea and its complications, translated by Arthur Foerster (London: Baillière, Tindall and Cox, 1913).* (1912).

8 Grmek, M. *Diseases in the Ancient Greek World (Trans. M. and L. Mueller, Baltimore: John Hopkins University Press).* 142-9 (1989).

9 Oriel, J. D. *The Scars of Venus: A History of Venereology (London: Springer-Verlag).* (1994).

10 Spratt, B. G. Hybrid penicillin-binding proteins in penicillin-resistant strains of Neisseria gonorrhoeae. *Nature* **332**, 173-176, doi:10.1038/332173a0 (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, doi:10.1128/AAC.00325-11 (2011).

12 Ohnishi, M. Ceftriaxone-resistant Neisseria gonorrhoeae, Japan. *Emerg Infect Dis* **17**, 148-149, doi:10.3201/eid1701.100397 (2011).

13 Lefebvre, B. *et al.* Ceftriaxone-Resistant Neisseria gonorrhoeae, Canada, 2017. *Emerg Infect Dis* **24**, doi:10.3201/eid2402.171756 (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**, doi:10.2807/1560-7917.ES.2018.23.47.1800617 (2018).

15 Popovich, K. J. & Snitkin, E. S. Whole Genome Sequencing-Implications for Infection Prevention and Outbreak Investigations. *Curr Infect Dis Rep* **19**, 15, doi:10.1007/s11908-017-0570-0 (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, doi:10.1016/s1473-3099(13)70693-5 (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, doi:10.1128/JCM.02589-14 (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, doi:10.1016/S1473-3099(16)30157-8 (2016).

19 Didelot, X. *et al.* Genomic Analysis and Comparison of Two Gonorrhea Outbreaks. *MBio* **7**, doi:10.1128/mBio.00525-16 (2016).

20 Chisholm, S. A. *et al.* An outbreak of high-level azithromycin resistant Neisseria gonorrhoeae in England. *Sex Transm Infect*, doi:10.1136/sextrans-2015-052312 (2015).

21 Duchene, S. *et al.* Genome-scale rates of evolutionary change in bacteria. *Microb Genom* **2**, e000094, doi:10.1099/mgen.0.000094 (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, doi:10.1128/JCM.00100-17 (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, doi:10.1093/jac/dks162 (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, doi:10.1128/AAC.05760-11 (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, doi:10.1099/00221287-125-1-123 (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, doi:10.1128/JCM.03195-15 (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, doi:10.1016/S1473-3099(18)30225-1 (2018).

28 Hopkins, A. G. *Globalization in World History (London: Pimlico)*. (2002).

29 McNeill, W. H. *Plagues and Peoples (New York: Doubelday).* 207-241 (1977).

30 Huovinen, P. Resistance to Trimethoprim-Sulfamethoxazole. *Clin Infect Dis* **32**, 1608-1614, doi:1058-4838/2001/3211-0014$03.00 (2001).

31 Workowski, K. A. & Bolan, G. A. Sexually Transmitted Diseases Treatment Guidelines, 2015. *MMWR Recommendations and Reports* **64** (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, doi:10.3201/eid2407.172104 (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 Suppl 1**, i20-30 (2002).

34 Wilkinson, D. *et al.* Unrecognized sexually transmitted infections in rural South African women: a hidden epidemic. *Bulletin of the World Health Organization* **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, doi:10.1017/S0950268813003488 (2014).

711 36 Grad, Y. H. *et al.* Genomic epidemiology of gonococcal resistance to extended
712 spectrum cephalosporins, macrolides, and fluoroquinolones in the US, 2000-
713 2013. *J Infect Dis* **214**, 1579-1587, doi:10.1093/infdis/jiw420 (2016).
714 37 Unemo, M., Olcen, P., Berglund, T., Albert, J. & Fredlund, H. Molecular
715 epidemiology of Neisseria gonorrhoeae: sequence analysis of the porB gene
716 confirms presence of two circulating strains. *J Clin Microbiol* **40**, 3741-3749
717 (2002).
718 38 Unemo, M., Fasth, O., Fredlund, H., Limnios, A. & Tapsall, J. Phenotypic and
719 genetic characterization of the 2008 WHO Neisseria gonorrhoeae reference
720 strain panel intended for global quality assurance and quality control of
721 gonococcal antimicrobial resistance surveillance for public health purposes. *J*
722 *Antimicrob Chemother* **63**, 1142-1151, doi:10.1093/jac/dkp098 (2009).
723 39 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools.
724 *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
725 40 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework
726 for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-
727 1303, doi:10.1101/gr.107524.110 (2010).
728 41 Harris, S. R. *et al.* Evolution of MRSA during hospital transmission and
729 intercontinental spread. *Science* **327**, 469-474, doi:10.1126/science.1182395
730 (2010).
731 42 Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome
732 alignment with gene gain, loss and rearrangement. *PLoS One* **5**, e11147,
733 doi:10.1371/journal.pone.0011147 (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, doi:10.1186/1471-2180-7-66 (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, doi:10.1093/nar/gku1196 (2015).

45 Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12, doi:10.1186/gb-2004-5-2-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, doi:http://dx.doi.org/10.1099/mgen.0.000056 (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, doi:10.1093/molbev/mst028 (2013).

49 Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690, doi:10.1093/bioinformatics/btl446 (2006).

50 Lemoine, F. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452-456, doi:10.1038/s41586-018-0043-0 (2018).

51 Page, A. J. *et al.* Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb Genom* **2**, e000083, doi:10.1099/mgen.0.000083 (2016).

52 Jolley, K. A. & Maiden, M. C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**, 595, doi:10.1186/1471-2105-11-595 (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, doi:10.1086/383047 (2004).

54 Kwong, J. C. *et al.* NGMASTER: in silico Multi-Antigen Sequence Typing for Neisseria gonorrhoeae. *Microb Genom* **2**, e000076, doi:10.1101/057760 (2016).

55 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).

56 Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* **3**, doi:10.1099/mgen.0.000131 (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 gonorrhoeaebeta-lactamase-producing plasmids. *Mol Cell Probes* **13**, 89-92, doi:10.1006/mcpr.1998.0216 (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, doi:10.7717/peerj.281 (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. *Journal of Statistical Software.* **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, doi:10.1186/1471-2156-11-94 (2010).

63 Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405, doi:10.1093/bioinformatics/btn129 (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. *Statistics in Medicine* **8**, doi:10.1002/sim.4780081218 (1989).

65 Benjamini, Y. Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 405-416, doi:10.1111/j.1467-9868.2010.00746.x (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, doi:10.1111/2041-210X.12466 (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, doi:10.1093/molbev/msv056 (2015).

68 To, T. H., Jung, M., Lycett, S. & Gascuel, O. Fast Dating Using Least-Squares Criteria and Algorithms. *Syst Biol* **65**, 82-97, doi:10.1093/sysbio/syv068 (2016).

69 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214, doi:10.1186/1471-2148-7-214 (2007).

70 Ansari, M. A. & Didelot, X. Bayesian Inference of the Evolution of a Phenotype Distribution on a Phylogenetic Tree. *Genetics* **204**, 89-98, doi:10.1534/genetics.116.190496 (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. & Graham Reynolds, R. A new Bayesian method for fitting evolutionary models to comparative data with intraspecific variation. *Evolution* **66**, 2697-2707, doi:10.1111/j.1558-5646.2012.01645.x (2012).

73 Bollback, J. P. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**, 88, doi:10.1186/1471-2105-7-88 (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, doi:10.1128/CMR.00010-14 (2014).

75 R Development Core Team, R. l. R: A Language and Environment for Statistical Computing. (2008).

76 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691-3693, doi:10.1093/bioinformatics/btv421 (2015).

77 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490, doi:10.1371/journal.pone.0009490 (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, doi:10.1093/nar/gkw290 (2016).

79 Hadfield, J. *et al.* Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics*, 1-2, doi: 10.1093/bioinformatics/btx610 (2017).

80 Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* **3**, 217-223, doi:10.1111/j.2041-210X.2011.00169.x (2012).

Figure legends

Figure 1 | Geographic and phylogenetic distribution of *Neisseria gonorrhoeae* isolates. The map shows the countries of isolation of the strains in the collection coloured by continent. The phylogeny shows the relationship among the strains (n=419). Coloured strips show (from inside out) the continent of isolation (CONT), year and further typing information (BAPS clusters, NG-MAST, MLST and *penA* types; colours represent different types or alleles). Mosaic *penA* types are marked in the outermost black strip.

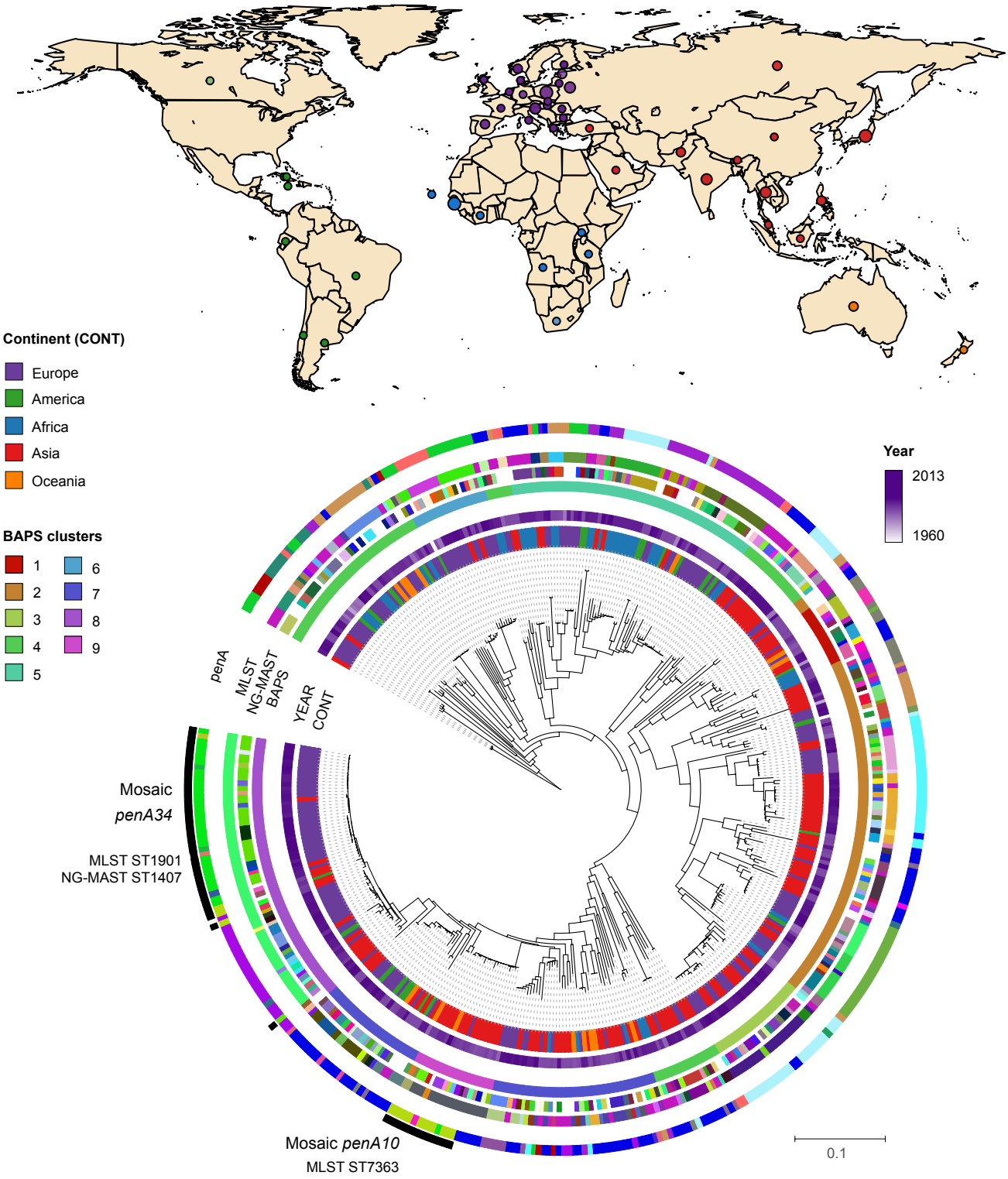
Figure 2 | Global phylogeographic analysis. The dated maximum likelihood phylogenetic tree shows the posterior probabilities for each continent in every node

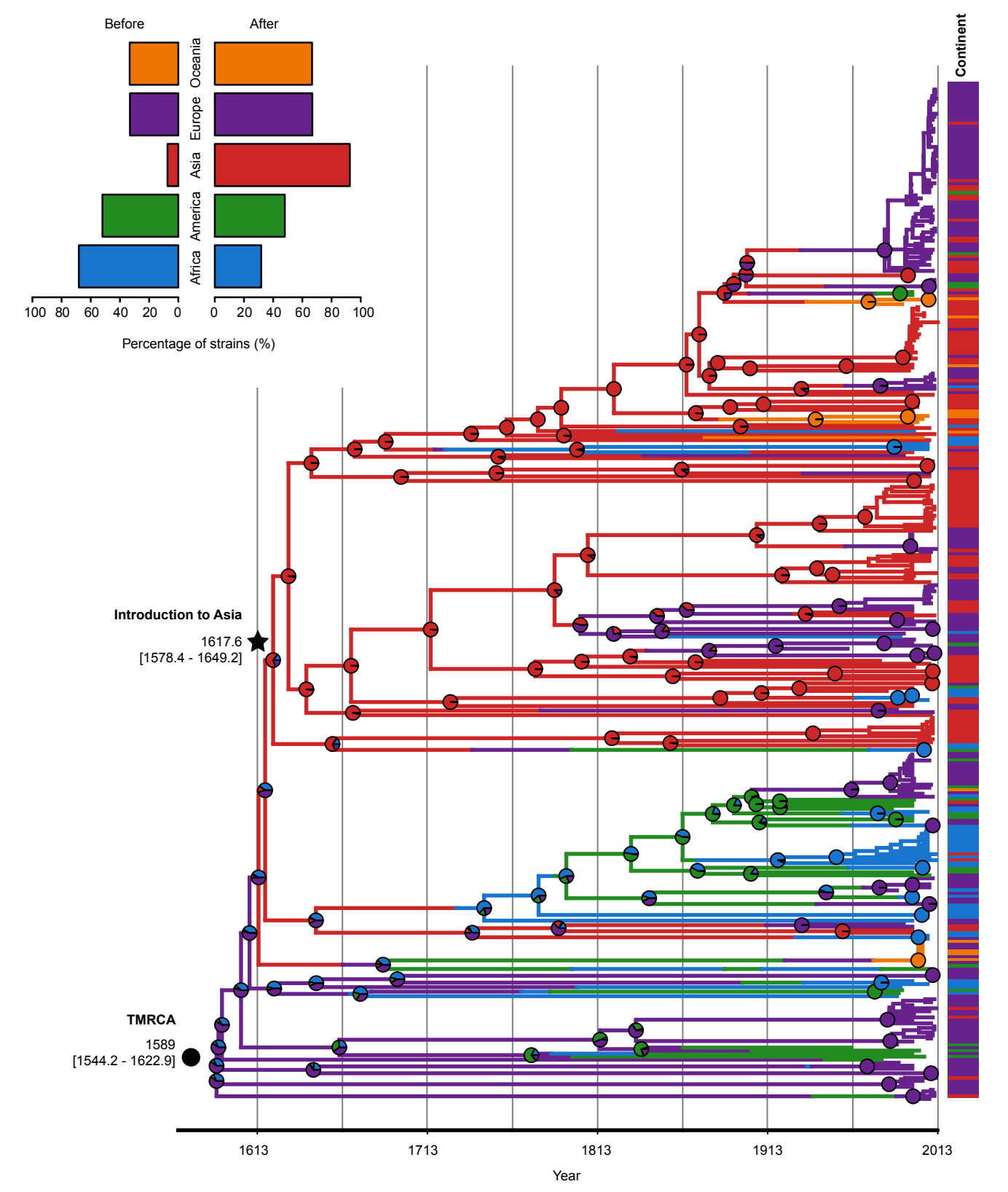
(pie charts). Continents of isolation (prior) are shown as metadata next to the tips (n=419). The top left legend contains information on the proportion of strains from different continents before (n=121) and after (n=298) the introduction to Asia.

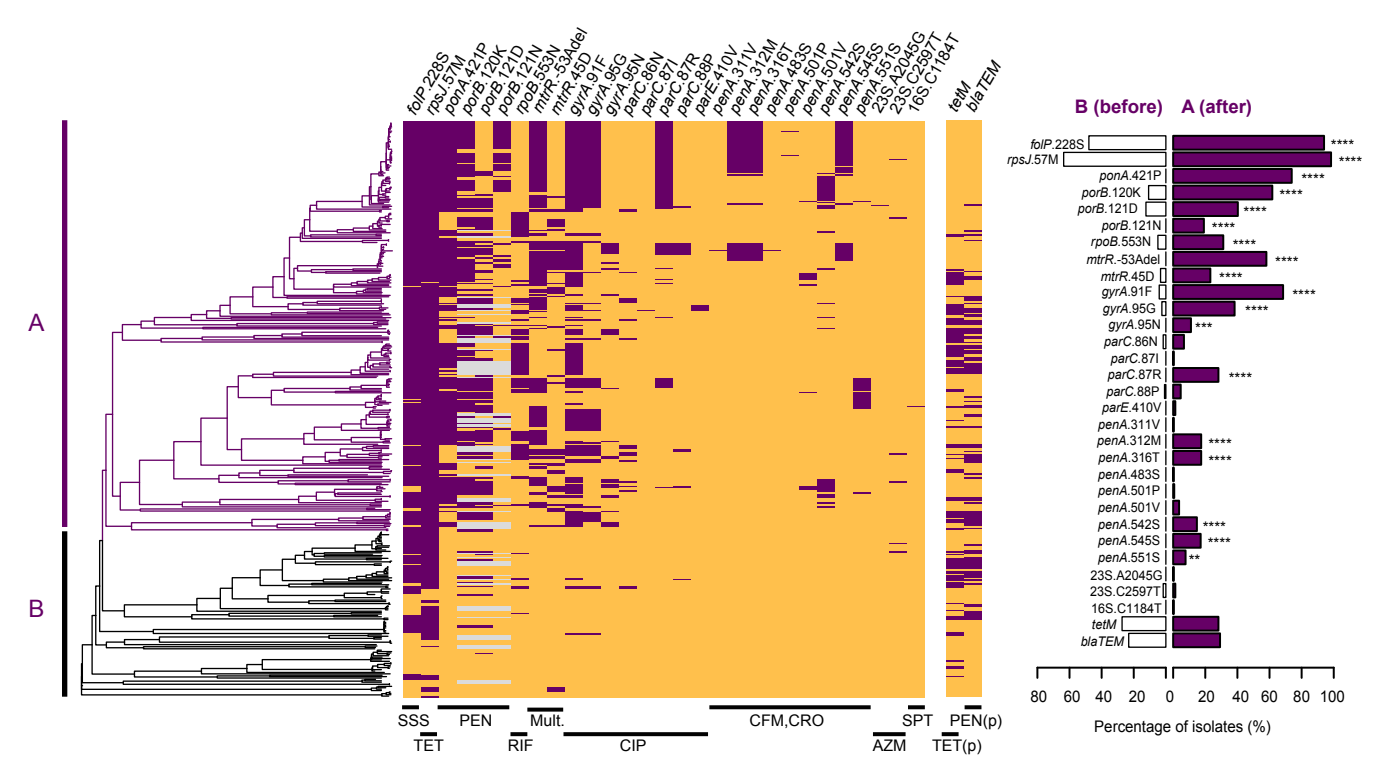
Figure 3 | Evolution of antimicrobial resistance genetic determinants in *Neisseria gonorrhoeae*. Antimicrobial resistance determinants (chromosomal mutations and presence/absence of the *tetM* and bla_{TEM} genes on plasmids (p)) detected in the new 413 strains included in this study using ARIBA[56] and mapped on the maximum likelihood dated tree. Purple represents presence of the determinant and orange its absence. Grey indicates 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. Statistical significance from a two-sided test for equality of proportions is also shown in the graph with asterisks. ****p-value<0.0001, ***p-value<0.001, **p-value<0.01, *p-value<0.05.

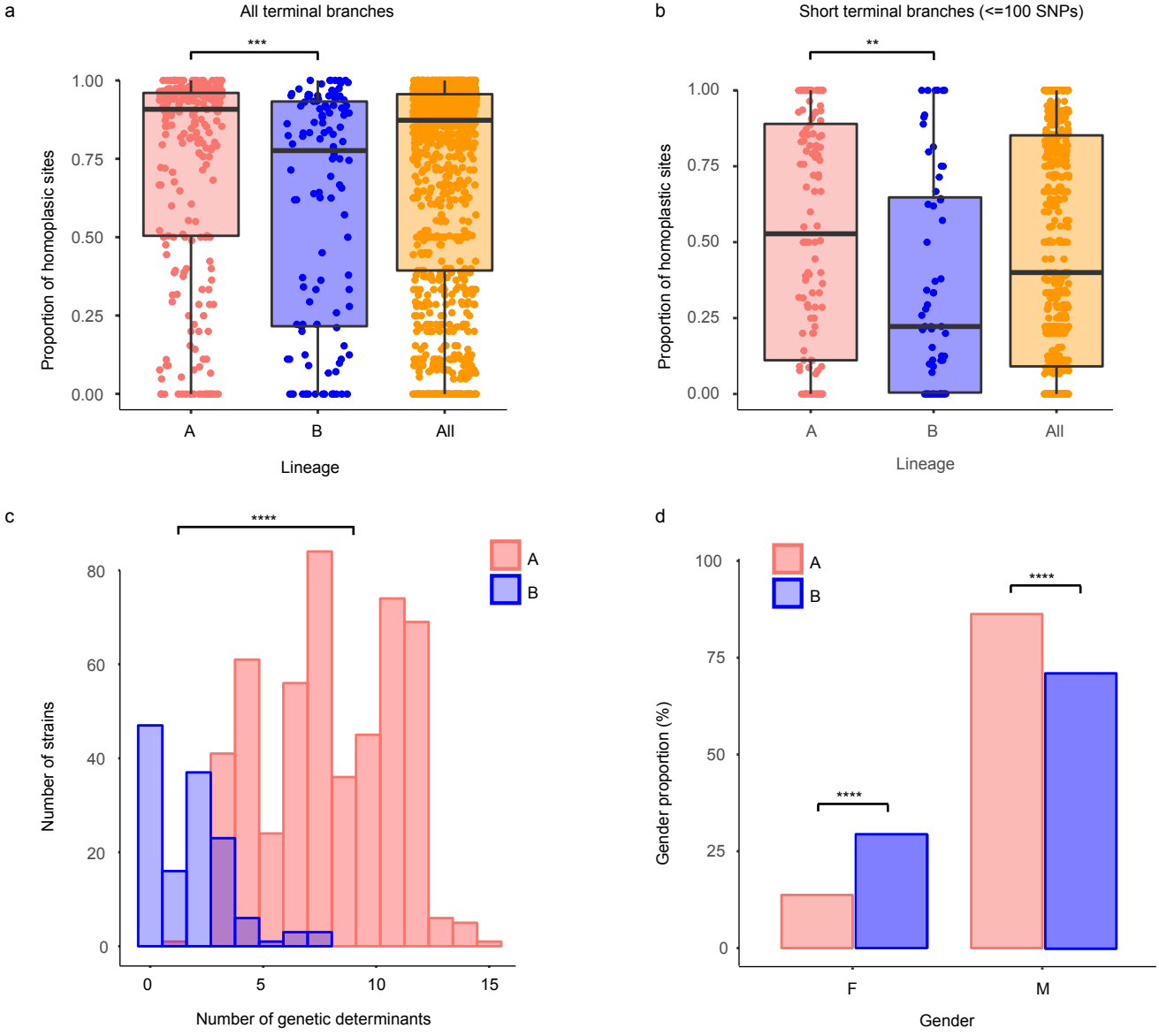
Figure 4 | Characterization of the lineages of *Neisseria gonorrhoeae*. Distribution of the proportions of homoplasic sites in all terminal (a) and short terminal branches (<=100 SNPs) (b) in lineages A (n=298), B (n=121) and all strains (n=419) represented as boxplots. Each point represents the proportion of homoplasies in one branch drawn from the total variation found in that branch. Horizontal box lines represent the first quartile, the median and the third quartile. Whiskers extend from the first quartile – 1.5x the interquartile range and the third quartile + 1.5x the interquartile range. Statistical significance between lineages A and B was assessed using a two-sided Wilcoxon test and shown as asterisks. (c) Distribution of the total number of antimicrobial resistance 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 (F, n=114) and male (M, n=566) patients in each lineage

obtained from combining the global dataset with 376 isolates from two North-American genomic studies[17,26] (total n=639; lineage A n=503; lineage B n=136). Asterisks show the significance level from a two-sided test for equality of proportions with continuity correction (c-d). ****p-value<0.0001, ***p-value<0.001, **p-value<0.01.


Continent (CONT)
Europe
America
Africa
Asia
Oceania
BAPS clusters
1
2
3
4
5
6
7
8
9
Year
2013
1960
penA
MLST
NG-MAST
BAPS
YEAR
CONT
Mosaic
penA34
MLST ST1901
NG-MAST ST1407
Mosaic penA10
MLST ST7363
0.1


Before
After
Oceania
Europe
Asia
America
Africa
100 80 60 40 20 0
0 20 40 60 80 100
Percentage of strains (%)
Continent
Introduction to Asia
1617.6
[1578.4 - 1649.2]
TMRCA
1589
[1544.2 - 1622.9]
1613
1713
1813
1913
2013
Year


folP.228S
rpsJ.57M
ponA.421P
porB.120K
porB.121D
porB.121N
rpoB.553N
mtrR.-53Adel
mtrR.45D
gyrA.91F
gyrA.95G
gyrA.95N
parC.86N
parC.87I
parC.87R
parC.88P
parE.410V
penA.311V
penA.312M
penA.316T
penA.483S
penA.501P
penA.501V
penA.542S
penA.545S
penA.551S
23S.A2045G
23S.C2597T
16S.C1184T
tetM
blaTEM
A
B
SSS
TET
PEN
RIF
Mult.
CIP
CFM,CRO
AZM
SPT
TET(p)
PEN(p)
B (before)
A (after)
80 60 40 20 0 20 40 60 80 100
Percentage of isolates (%)


a
All terminal branches
***
1.00
0.75
0.50
0.25
0.00
Proportion of homoplasic sites
A
B
All
Lineage
b
Short terminal branches (<=100 SNPs)
**
1.00
0.75
0.50
0.25
0.00
Proportion of homoplastic sites
A
B
All
Lineage
c
****
A
B
80
60
40
20
0
Number of strains
0
5
10
15
Number of genetic determinants
d
A
B
100
75
50
25
0
Gender proportion (%)
****
****
F
M
Gender