Use of whole genome sequencing to investigate an outbreak of gonorrhoea among females in urban New South Wales, Australia, 2012 to 2014.

Cameron Buckley[1], Brian M. Forde[2,3,4], Ella Trembizki[1], Monica M. Lahra[5,6], Scott A. Beatson[2,3,4] and David M. Whiley[1,2,7*].

[1] The University of Queensland, UQ Centre for Clinical Research, Herston, Queensland 4029, Australia.

[2] Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, Queensland 4072, Australia.

[3] Australian Centre for Ecogenomics, The University of Queensland, Brisbane, Queensland 4072, Australia.

[4] School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, 4072 Australia.

[5] WHO Collaborating Centre for STD, Microbiology Department, New South Wales Health Pathology East, Prince of Wales Hospital, Sydney, New South Wales 2031, Australia.

[6] School of Medical Sciences, The University of New South Wales, Sydney, New South Wales 2052, Australia.

[7] Pathology Queensland, Microbiology Department, Herston, Queensland 4029, Australia.

Corresponding Author

***** David. M. Whiley. The University of Queensland, UQ Centre for Clinical Research, Herston, Queensland 4029, Australia. (d.whiley@uq.edu.au)

Abstract

Increasing rates of gonorrhoea have been observed among urban heterosexuals within the Australian state of New South Wales (NSW). Here, we applied whole genome sequencing (WGS) to better understand transmission dynamics. Ninety-four isolates of a particular *N. gonorrhoeae* genotype (G122) associated with female patients (years 2012 to 2014) underwent phylogenetic analysis using core single nucleotide polymorphisms (SNPs). Context for genetic variation was provided by including an unbiased selection of 1,870 *N. gonorrhoeae* genomes from a recent United Kingdom (UK) study. NSW genomes formed a single clade, with the majority of isolates belonging to one of five clusters, and comprised patients of varying age groups. Intra-patient variability was less than 7 core SNPs. Several patients had indistinguishable core SNPs, suggesting a common infection source. These data have provided an enhanced understanding of transmission of *N. gonorrhoeae* among urban heterosexuals in NSW, Australia, and highlight the value of using WGS in *N. gonorrhoeae* outbreak investigations.

Keywords

Neisseria gonorrhoeae; genomics; epidemiology; Australia; gonorrhoea; molecular.

Background

Notification rates for *Neisseria gonorrhoeae* infection in Australia have almost doubled in the last 5 years (1). Gonorrhoea is primarily concentrated in two key populations in Australia: men who have sex with men (MSM) living in urban areas and Indigenous heterosexuals in regional and remote settings (2). Until recently, there have been relatively low rates of gonorrhoea among heterosexuals living in urban

areas of Australia (3, 4). A retrospective study involving two sexual health clinics in urban Sydney, New South Wales (NSW), showed increasing rates among heterosexuals from 2008 to 2012 (3) that were associated with unprotected oral sex and commercial sex work (CSW). Our subsequent nationwide study of *N. gonorrhoeae* genotypes conducted using isolates from throughout Australia in 2012 revealed the most common gonococcal strain was associated with heterosexuals, and was particularly prevalent among females in NSW (5). A further investigation (6) showed this strain remained prevalent within NSW in 2014 and, although not remarked upon in the article, was prevalent among females. Combined, the above studies are highly suggestive of an emerging epidemic of gonorrhoea among urban heterosexuals in Australia.

A limitation of the genotyping protocol used in our previous studies was that whilst it targets 25 informative single nucleotide polymorphisms (SNPs) across the gonococcal genome, it lacks the resolution offered by whole genome sequencing (WGS). The limitations of traditional genotyping methods have been highlighted by several *N. gonorrhoeae* sequencing studies (7, 8). WGS now has the ability to identify *N. gonorrhoeae* transmission links in the absence of traditional contact-tracing data (8), as well as to reveal geographical and temporal spread and bridging across sexual networks of particular strains (9).

In this study we aimed to utilise WGS to better understand transmission dynamics within a potentially new outbreak of *N. gonorrhoeae* infection among heterosexuals living within urban areas of NSW, Australia. This was prompted by the above studies and more recent unpublished data showing increasing notifications among older

individuals in urban areas of Australia, including females. Understanding these trends is important, particularly in light of the potential for increasing disease, including pelvic inflammatory disease among females and also because older age is not typically viewed as a key risk factor for acquiring gonorrhoea amongst heterosexuals. Here, we sequenced the genomes of all available isolates from females that belonged to the most common gonococcal strain identified in the previous studies (5, 6).

Methods

N. gonorrhoeae isolates

A total of 94 *N. gonorrhoeae* isolates from 87 female patients from two previous investigations were included in this study. The samples comprised 31 vaginal swabs, 27 cervical swabs, 33 throat swabs, 1 rectal swab, 1 eye swab and 1 genital swab. Age groups for the patients were; 18-24 years (n = 29), 25-34 (n = 31), 35-44 (n = 14), 45-54 (n = 11), 55 and over (n = 9). The isolates were from NSW, Australia and were isolated in January to June of 2012 (n = 59) and January to June of 2014 (n = 35). The isolate metadata is summarised in Supplementary Table 1. All 94 isolates were indistinguishable on the basis of iPLEX massarray genotyping. It should be noted that different names were used to describe these strains in the previous studies; the year 2012 isolates were originally described as belonging to strain-type “S01” and genotype “G122” (Supplementary Table 3 in (5)). The year 2014 isolates were called “NSW-001”. For consistency in this study, all 94 isolates will be referred to as being of genotype G122.

N. gonorrhoeae isolates from the United Kingdom

To provide context for the degree of variation among the NSW genomes, analyses of genome data from 1,870 *N. gonorrhoeae* isolates from a recent investigation (8) based

in Brighton, United Kingdom (UK) were included. The UK genome data was chosen for comparison because it was the largest collection of *N. gonorrhoeae* genome data with unbiased isolate selection and a collection period (2011-2015) that coincided with the NSW isolates (collected in 2012 and 2014).

DNA extraction, sequencing and assembly

A single colony from each isolate was cultured on LB agar and incubated at 37^{o}C, enriched with 5% CO_2 for approximately 18-24 hours. Genomic DNA was extracted from half of a 10µl loop of culture growth using the Ultraclean Microbial DNA Isolate Kit (GeneWorks, Australia), using the 'Alternative Lysis Method' as per manufacturer's instructions. Libraries were prepared using the Illumina Nextera XT protocol, with 125bp paired-end reads generated using the Illumina HiSeq instrument (AGRF, Melbourne, Australia), with HiSeq 2500 V4 chemistry. Raw sequence reads from the 94 samples were assessed in FastQC v0.11.4 (10) and hard trimmed to 100bp using Nesoni v0.132 (11). The 94 NSW genomes were assembled in parallel with the 1,870 *N. gonorrhoeae* genomes from the UK (8). The UK genomes were acquired from the NCBI Short Read Archive under BioProject PRJNA315363 (https://www.ncbi.nlm.nih.gov/bioproject/315363). The quality trimmed reads from both NSW and the UK datasets underwent *de novo* assembly using SPAdes v3.6.2 (12) and MegaHit v1.0.3 (13), respectively. Sequence read and assembly metrics are summarised for all NSW genomes in Supplementary Table 2. Sequence read data for all NSW isolates were submitted to the NCBI Short Read Archive under BioProject PRJNA392203 (https://www.ncbi.nlm.nih.gov/bioproject/392203). The taxonomic sequence classification tool Kraken v0.10.5 (14) was initially used to confirm each genome as *N. gonorrhoeae* species. Exclusion criteria applied to the UK genomes included: (i) a genome size greater than 3Mbp, (ii) a genome that had >350 contigs

and, (iii) a genome percentage identity according to Kraken of <75% for *N. gonorrhoeae*.

Genotyping

In silico analyses were used to determine multilocus sequence type (MLST) for all NSW and UK genomes. Briefly, the MLST profile was determined using mlst v2.1 (15), providing a seven-locus typing scheme commonly used for Neisseria species (*abcZ, adk, aroE, fumC, gdh, pdhC* and *pgm*). For the 94 NSW genomes, *N. gonorrhoeae* multi-antigen sequence type (NG-MAST) was determined using NGMASTER (16). Sequences with different NG-MAST profiles were compared using BioEdit v7.0.9.0 (17). Novel NG-MAST profiles were uploaded to the NG-MAST database (http://www.ng-mast.net/). The NG-MAST predictions for the UK genomes were kindly provided by Dr David Eyre, University of Oxford, UK (personal communication).

Phylogenetic analyses

All phylogenetic trees were constructed using a core SNP alignment following the removal of recombinant regions. Genomes were aligned using Parsnp v1.2 (18) by randomly selecting a NSW genome (AU2012-573) to act as a reference, followed by the prediction and removal of recombinant regions using Gubbins v2.1.0 (19). Phylogenetic analyses based on core SNP alignments for both the NSW and UK genomes were achieved using RAxML v8.2.9 (20) using a general-time reversible nucleotide substitution model with a GAMMA correction for site variation. Phylogenetic trees were visualised using FigTree v1.4.2, Evolview v2 or iTOL v3 (21-23).

Mutations associated with antimicrobial resistance

All 94 NSW genomes and closely related UK genomes, underwent annotation via Prokka v1.11 (24). A combination of the method outlined below and BIGSdb (25) were used to assess loci associated with antimicrobial resistance (AMR) including; *gyrA, folP, mtrR, parC, parE, penA, ponA, porB, rpoB, rpsE, rspJ* and 23S rRNA. Using BLAST+ and reference sequences derived from characterised WHO-F and WHO-U strains (26), loci were extracted from an annotated multi-fasta file using seqtk (27). Sequences were aligned and inspected using Clustal Omega (28).

Results

Phylogenetic analysis of NSW genotype G122 isolates in the context of a broad collection of UK isolates

The core SNP phylogeny (Figure 1A) included all NSW isolates and 1,796 of the 1,870 UK genomes; 74 UK genomes were removed from further analysis on the basis of the above exclusion criteria. The unrooted maximum likelihood phylogeny in Figure 1A shows a diverse range of strains from the UK. Figure 1B shows all 94 NSW genomes tightly cluster in a single clade with six UK genomes (SRR3350090, SRR3343655, SRR3350214, SRR3350225 and SRR3343534, SRR3350146). All genomes in this clade (hereafter called the G122 clade) were MLST 7359, while all 6 UK genomes and the majority of NSW genomes (n = 86) were NG-MAST 4186. The remaining NSW isolates comprised four previously described NG-MAST types (NG-MAST 6759, n = 2; NG-MAST 6767, n = 2; NG-MAST 15344, n = 2 and NG-MAST 15348, n = 1) as well as one novel NG-MAST profile (NG-MAST 15609, n = 1). All NG-MAST profiles shared the same *tbpB* 241 allele but differed by their *porB* sequences which shared >99% nucleotide identity, indicating they all belonged to the same genogroup (29); which we assigned as “genogroup 4186”. Based on the most

recent common ancestor, isolates in the G122 clade differ by <200 core SNPs from isolates in the nearest sister clade, comprising 9 UK genomes that are all MLST 7826 and NG-MAST 2487 (Figure 1B). Both clades form a distinct lineage approximately 1000 core SNPs distant to the nearest split in the underlying *N. gonorrhoeae* phylogeny (Figure 1A).

Phylogenetic analysis of the G122 clade

To investigate the genetic relationships between G122 isolates at the single nucleotide level we determined a core SNP phylogeny of all 94 NSW genomes and the six related UK genomes (Figure 2). The majority of isolates belong to one of five clusters (C1 to C5) comprising six or more isolates that each share a defining SNP and a common ancestor indicating a transmission network (Figure 2). The maximum core SNP difference between genomes in each cluster, C1 to C5 respectively, were 6, 4, 19, 21 and 14 SNPs. Of these clusters, only three (C3, C4 and C5) comprised isolates from both years 2012 and 2014 whereas C1 and C2 comprised isolates from year 2012 only. Age group for clusters C3 to C5 ranged from 18-24 to 55 and over (median = 25.5, 32 and 25 years respectively), while C1 ranged from 18-24 to 45-54 (median = 35.5) and C2 ranged from 25-34 to 45-54 (median = 45.5). The percentage of cervical/vaginal and throat samples respectively for each cluster were 33% and 67% for both C1 and C2, 57% and 36% for C3, 65% and 35% for C4 and 58% and 42% for C5 (summarised in Supplementary Table 3).

Genome AU2012-768 contained 33 distinct core SNPs, recombination analysis identified three highly SNP-dense regions that were filtered prior to tree-building.

Manual inspection of these regions revealed multiple SNPs were shared between AU2012-768 and the outlying UK genome (SRR3360636) used to root the phylogeny.

Diversity of G122 isolates within patients

Six individuals with more than one isolate were included in our study (Patients 'PA' to 'PF'). The respective genomes from these six patients differed by less than seven core SNPs. Patient 'PA' had a throat and cervical swab collected on the same day, and was the only individual with more than one isolate that had indistinguishable core SNPs. Patients 'PB', 'PC' and 'PD' also had their respective isolates collected on the same day from different sample sites, but showed a difference of 2-7 core SNPs. The remaining patients, PE and PF, each had different isolates collected between one and eleven days respectively and both differed by 2 core SNPs.

Genetic links between patients

There were 6 instances where isolates had indistinguishable core SNPs and have been labelled as groups G1 to G6; these comprised either 3 isolates (G3) or 2 isolates each (G1, G2, G4, G5 and G6). Group G4 involved a single patient (patient 'PA' as described above) whereas the remaining 5 groups all involved different patients. The three isolates from group G3 were collected within 6 weeks of each other (data not shown) even though the collection dates provided in this study indicate a 3 month period (Supplementary Table 1). The isolates for groups G1, G2, G5 and G6 were all collected within one month of each other.

Antimicrobial resistance determinants of the G122 clade

Consistent with antibiotic susceptibility data for G122 isolates (5), the 94 NSW isolates had wildtype copies of the assessed AMR genes with the exception of harbouring the PBP2 345A insertion (Supplementary Table 4). Although the 345A insertion has been associated with penicillin-resistance (30) all NSW isolates were phenotypically susceptible to penicillin.

Discussion

Overall the sequencing results show that isolates from the previously identified SNP-type are indeed closely related strains and, given they were all from females, provides further evidence of sustained transmission of this strain amongst heterosexuals. Moreover, the enhanced resolution provided by the phylogenetic analyses has provided new information regarding transmission of this strain within local sexual networks.

This study included six patients (PA to PF) for which more than one isolate was collected. The observed SNP variation within each patient was found to be minimal, ranging from 0-7 core SNPs. This within-patient genomic stability is not surprising and is consistent with a previous study (8). However, it provides an interesting context when considering the results for the five groups of patients (G1, G2, G3, G5 and G6; Figure 2) all of which involved different individuals having identical core SNPs. Such results are suggestive of tight transmission networks, perhaps even a common infection source for these patients. Of note were three females belonging to group G3 that shared identical core SNPs over a six week period. This observation would be consistent with a common infection source, possibly a male having multiple sexual partners and not actively seeking treatment. Behavioural data were not

available to confirm this. Nevertheless, it should be noted that in the absence of these genomic data it would otherwise be very difficult to identify such clusters, particularly when patients are often unwilling or unable to provide details regarding sexual partners (31). Hence, these data highlight a potential role that WGS could have (assuming it could be achieved within a timely manner) in enhancing contact tracing, whereby the data are used to help pinpoint key groups of individuals for intervention.

The data also provide some evidence that CSW and certainly unprotected fellatio are not key drivers perpetuating this G122 genotype in the population. Previous reports have shown high rates of condom use for vaginal sex, but lower rates for both oral sex and non-paying partners among CSW (32, 33). Thus, if CSW were responsible then one would have expected clusters of pharyngeal infections among these individuals. In contrast, the phylogeny depicts clusters comprising a more even distribution of both pharyngeal and vaginal infections, suggesting that CSW alone is unlikely to be responsible for sustaining transmission of this genotype.

Of further interest in this study was the number of females in older age groups, with 21% of patients being 45 years or older. While there was some limited evidence of transmission within older ages (for example, cluster C2 predominately comprised individuals 45 years or older), the overall phylogeny suggested that gonorrhoea was not being sustained within distinct networks of older ($\geq$ 45 years) individuals. Rather, the observed networks comprised a broad range of age groups and suggested that older individuals were acquiring infections that were otherwise being sustained in younger, more sexually active age groups. Further studies are needed to investigate this. However, current Australian federal policies promote sexual health to younger

individuals and those of reproductive age as they are considered to be the most at risk for further complications (34). Given our data indicate networks of individuals comprising of varying ages, education on STI prevention may be more suited towards a broad range of age groups, including older ages and especially those engaging in new sexual relationships (35).

A recent comprehensive investigation applying WGS to *N. gonorrhoeae* isolates within Brighton, UK, allowed us to provide additional context for our genomes among a broader population. Interestingly, our common Australian strain only comprised 0.3% of isolates assessed from the Brighton study. It is difficult to speculate on the significance of these differences, however it may have only recently been introduced into the UK, perhaps via individuals travelling between the two regions. However, we note that NG-MAST 4186, which is associated with genotype G122, has been previously documented elsewhere, including Japan (36) and New Zealand (37) at relatively high prevalence (5.2% and 10.8% of strains respectively). Little is known as to why some *N. gonorrhoeae* strains may be more successful in a given population than others. Of interest was that our G122 genotype lacked almost all commonly reported *N. gonorrhoeae* AMR mechanisms, however current thinking is that these AMR mutations do not confer a fitness disadvantage (38). Therefore it is likely that other factors, possibly unrecognised virulence mechanisms, may be important in sustaining this genotype.

There were several limitations associated with this study, firstly that we did not access behavioural data; and secondly males were not assessed. However, the decision to exclude males was deliberate, and was done so as to focus on heterosexuals and limit

any potential signal from MSM; Thirdly we only included isolates from two six-month time periods that were two years apart.

In summary, we have used whole genome sequencing to confirm transmission of a particular strain of *N. gonorrhoeae* amongst females in urban areas of Australia from 2012 and 2014. The use of WGS and its enhanced resolution has revealed features of local sexual networks that would not otherwise be apparent through routine surveillance activities. These data provide evidence of the additional value that WGS could provide in *N. gonorrhoeae* outbreak investigations. The information may also provide additional benefit for further studies aimed at identifying virulence markers that are important in maintaining such strains in the population.

Funding

This work was supported by the Australian Infectious Diseases Research Centre and the National Health and Medical Research Council (GNT1090456 and GNT1105128 to S.A.B and D.M.W, respectively).

Acknowledgments

We thank Dr Tiffany Hogan from Prince of Wales Hospital, New South Wales for her assistance with this study. This study was conducted as part of the reference work of the National Neisseria Network (NNN), Australia, which is funded by the Australian Government Department of Health.

References

1. Communicable Diseases Network Australia. National Notifiable Diseases Surveillance System. Department of Health, Australian Government; 2016. Available at: http://www9.health.gov.au/cda/source/cda-index.cfm. Accessed 04 May 2017

2. Roberts-Witteveen A, Pennington K, Higgins N, Lang C, Lahra M, Waddell R, et al. Epidemiology of gonorrhoea notifications in Australia, 2007–12. Sexual Health. 2014;11(4):324-31.

3. Lusk MJ, Uddin RN, Lahra MM, Garden FL, Kundu RL, Konecny P. Pharyngeal Gonorrhoea in Women: An Important Reservoir for Increasing Neisseria gonorrhoea Prevalence in Urban Australian Heterosexuals? J Sex Transm Dis. 2013;2013:967471.

4. McDonagh P, Ryder N, McNulty AM, Freedman E. Neisseria gonorrhoeae infection in urban Sydney women: prevalence and predictors. Sex Health. 2009;6(3):241-4.

5. Trembizki E, Wand H, Donovan B, Chen M, Fairley CK, Freeman K, et al. The Molecular Epidemiology and Antimicrobial Resistance of Neisseria gonorrhoeae in Australia: A Nationwide Cross-Sectional Study, 2012. Clin Infect Dis. 2016;63(12):1591-8.

6. Lahra MM, Trembizki E, Buckley C, Donovan B, Chen M, Guy R, et al. Changes in the rates of Neisseria gonorrhoeae antimicrobial resistance are primarily driven by dynamic fluctuations in common gonococcal genotypes. J Antimicrob Chemother. 2017;72(3):705-11.

7. Didelot X, Dordel J, Whittles LK, Collins C, Bilek N, Bishop CJ, et al. Genomic Analysis and Comparison of Two Gonorrhea Outbreaks. MBio. 2016;7(3).

8. De Silva D, Peters J, Cole K, Cole MJ, Cresswell F, Dean G, et al. Whole-genome sequencing to determine transmission of Neisseria gonorrhoeae: an observational study. Lancet Infect Dis. 2016;16(11):1295-303.

9. Grad YH, Kirkcaldy RD, Trees D, Dordel J, Harris SR, Goldstein E, et al. Genomic epidemiology of Neisseria gonorrhoeae with reduced susceptibility to cefixime in the USA: a retrospective observational study. Lancet Infect Dis. 2014;14(3):220-6.

10. Andrews S. FastQC: A Quality Control tool for High Throughput Sequence Data. 2010. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 10 January 2017.

11. Victorian-Bioinformatics-Consortium. Nesoni. 2013. Available at: https://github.com/Victorian-Bioinformatics-Consortium/nesoni. Accessed 12 January 2017.

12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77.

13. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6.

14. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46.

15. Seemann T. mlst. Available at: https://github.com/tseemann/mlst. Accessed 13 January 2017.

16. Kwong JC, Goncalves da Silva A, Dyet K, Williamson DA, Stinear TP, Howden BP, et al. NGMASTER:in silico multi-antigen sequence typing for Neisseria gonorrhoeae. Microb Genom. 2016;2(8):e000076.

17. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series. 1999; 41:95-8.

18. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15(11):524.

19. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43(3):e15.

20. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.

21. Rambaut A. FigTree. 2007. Available at: http://tree.bio.ed.ac.uk/software/figtree/. Accessed 16 February 2017.

22. He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res. 2016;44(W1):W236-41.

23. 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. 2016;44(W1):W242-5.

24. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.

391 25. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome
392 variation at the population level. BMC Bioinformatics. 2010;11:595.
393 26. Unemo M, Golparian D, Sanchez-Buso L, Grad Y, Jacobsson S, Ohnishi M, et
394 al. The novel 2016 WHO Neisseria gonorrhoeae reference strains for global quality
395 assurance of laboratory investigations: phenotypic, genetic and reference genome
396 characterization. J Antimicrob Chemother. 2016.
397 27. Li H. seqtk: Toolkit for processing sequences in FASTA/Q formats. Available
398 at: https://github.com/lh3/seqtk. Accessed 08 February 2017.
399 28. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast,
400 scalable generation of high-quality protein multiple sequence alignments using
401 Clustal Omega. Mol Syst Biol. 2011;7:539.
402 29. European Centre for Disease Prevention and Control. Molecular typing of
403 *Neisseria gonorrhoeae* - results from a pilot study 2010-2011. European Centre for
404 Disease Prevention and Control, Stockholm **2012**. Available at:
405 https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/201211109
406 -Molecular-typing-gonorrhea.pdf. Accessed 11 April 2017.
407 30. Dowson CG, Jephcott AE, Gough KR, Spratt BG. Penicillin-binding protein 2
408 genes of non-beta-lactamase-producing, penicillin-resistant strains of Neisseria
409 gonorrhoeae. Mol Microbiol. 1989;3(1):35-41.
410 31. Doherty IA, Padian NS, Marlow C, Aral SO. Determinants and consequences
411 of sexual networks as they affect the spread of sexually transmitted infections. J Infect
412 Dis. 2005;191 Suppl 1:S42-54.
413 32. Donovan B, Harcourt, C., Egger, S., Watchirs Smith, L., Schneider, K.,
414 Kaldor, J.M., Chen, M.Y., Fairley, C.K.,Tabrizi, S. The Sex Industry in New South
415 Wales: a Report to the NSW Ministry of Health. Kirby Institute University of New

South Wales, Sydney 2012. Available at: http://www.acon.org.au/wp-content/uploads/2015/04/NSW-Sex-Industry-Report-CSRH-2012.pdf. Accessed 10 April 2017.

33. Seib C, Debattista J, Fischer J, Dunne M, Najman JM. Sexually transmissible infections among sex workers and their clients: variation in prevalence between sectors of the industry. Sex Health. 2009;6(1):45-50.

34. Kirkman L, Kenny A, Fox C. Evidence of Absence: Midlife and Older Adult Sexual Health Policy in Australia. Sexuality Research and Social Policy. 2013;10(2):135-48.

35. Heywood W, Lyons A, Fileborn B, Minichiello V, Barrett C, Brown G, et al. Self-reported testing and treatment histories among older Australian men and women who may be at risk of a sexually transmissible infection. Sex Health. 2016.

36. Shimuta K, Unemo M, Nakayama S, Morita-Ishihara T, Dorin M, Kawahata T, et al. Antimicrobial resistance and molecular typing of Neisseria gonorrhoeae isolates in Kyoto and Osaka, Japan, 2010 to 2012: intensified surveillance after identification of the first strain (H041) with high-level ceftriaxone resistance. Antimicrob Agents Chemother. 2013;57(11):5225-32.

37. Heffernan H WR, Williamson D. Institute of Environmental Science and Research Limited 2015. Available at: https://surv.esr.cri.nz/PDF_surveillance/Antimicrobial/Gono/Ngonosurveyfinalreport2015.pdf. Accessed 10 April 2017.

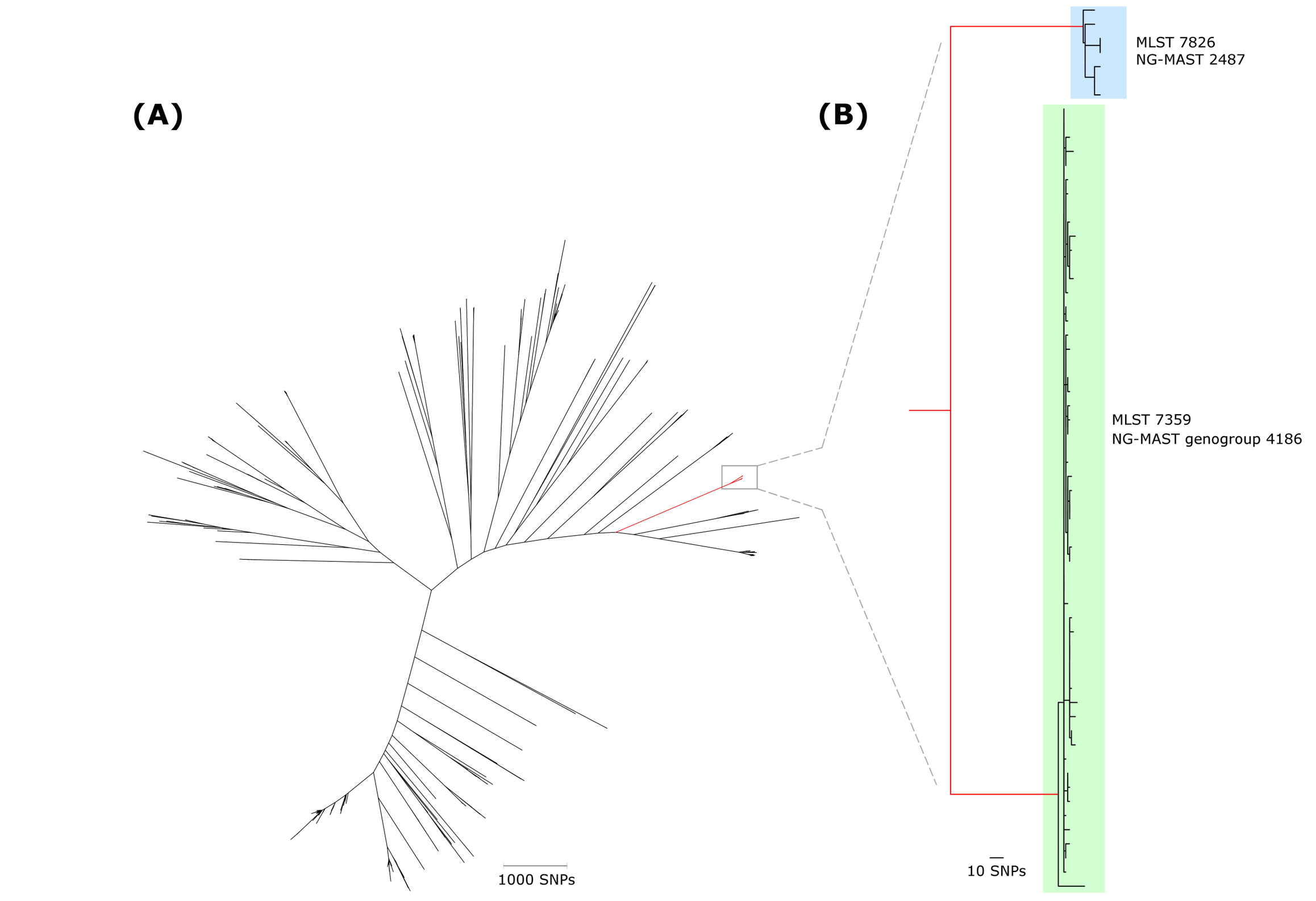
38. Unemo M, Shafer WM. Antimicrobial resistance in Neisseria gonorrhoeae in the 21st century: past, evolution, and future. Clin Microbiol Rev. 2014;27(3):587-613.

Figure 1.

(A) An unrooted maximum likelihood phylogeny constructed with 14,802 core single nucleotide polymorphisms (SNPs) from both New South Wales (NSW; n = 94) and the United Kingdom (UK; n = 1,796) genomes based on non-recombinant core regions. The scale bar represents SNPs. **(B)** An enhanced view of the branch containing two clades highlighted in green and blue. The green clade includes all NSW genomes and 6 UK genomes which comprised MLST 7359 and NG-MAST genogroup 4186. The next closest sister clade (in blue) comprised 9 unrelated MLST (7826) and NG-MAST (2487) UK genomes. The scale bar represents SNPs.

Figure 2.

A rooted maximum likelihood phylogeny constructed with 473 core single nucleotide polymorphisms (SNPs). It comprises the G122 clade containing all 94 New South Wales (NSW) genomes and 6 United Kingdom (UK) genomes. A UK genome (SRR3360636) from the next closest sister clade was selected as the outlying group to root the phylogeny. Five larger clusters of isolates (C1-C5) have been labelled based on the phylogeny. Coloured circles represent isolates from the same patient and coloured squares represent isolates from the UK (see key). Rectangles and their respective colour schemes correspond with date of collection, age group, sample site and NG-MAST profile (see key). The red isolate names with G1-G6 appended to the end, represent isolates that share identical core SNPs. Genome AU2012-768 has 33 core SNPs however its branch is truncated for easier visualisation purposes. The scale bar represents SNPs.


(A)
(B)
MLST 7826
NG-MAST 2487
MLST 7359
NG-MAST genogroup 4186
1000 SNPs
10 SNPs

Patients with >1 Sample
PA
PB
PC
PD
PE
PF

UK Samples
UK: May-2013
UK:Aug-2014
UK: Sep-2014

Year of collection
Jan-Feb 2012
Mar-Apr 2012
May-Jun 2012
Jan-Feb 2014
Mar-Apr 2014
May-Jun 2014

Age group
18-24
25-34
35-44
45-54
55-64+

Sample site
Throat
Cervical/Vaginal
Rectal
Eye

NG-MAST
ST4186
ST6759
ST6767
ST15344
ST15348
ST15609

AU2012-573
AU2014-654
AU2012-646
AU2012-768
AU2012-574
AU2012-511
AU2012-336
AU2012-249
AU2012-63 [G1]
AU2012-57 [G1]
AU2012-65
AU2012-341
C1
AU2014-637
AU2012-161
AU2014-661
AU2012-70
AU2012-62
AU2012-167
AU2012-237 [G2]
AU2012-425 [G2]
AU2012-250
AU2012-338
AU2012-416
AU2012-510
C2
AU2012-248
AU2014-622
AU2014-595
AU2014-596
AU2014-594
AU2014-665
AU2014-634
AU2014-607
AU2012-758
AU2014-647
AU2014-579
AU2012-500
AU2014-612
AU2014-602
AU2014-606
C3
SRR3350090
SRR3350214
AU2012-155
AU2012-331
AU2012-168
AU2012-698
AU2012-701
AU2012-648
AU2012-497
AU2014-686
AU2014-694
AU2014-653
AU2014-682
AU2012-166
AU2012-325
AU2012-340
AU2012-319
AU2012-241
AU2012-323 [G6]
AU2012-332 [G6]
AU2012-329
AU2012-73
SRR3343534
SRR3343655
AU2014-645
AU2014-651
AU2012-649 [G3]
AU2012-420 [G3]
AU2012-807 [G3]
AU2012-75 [G4]
AU2012-74 [G4]
AU2012-61
AU2012-162
AU2012-165
AU2012-240
AU2012-767
AU2014-574 [G5]
AU2014-593 [G5]
AU2014-580
AU2012-572
C4
AU2012-502
AU2012-418
AU2014-589
SRR3350146
SRR3350225
AU2012-68
AU2012-700
AU2012-60
AU2012-501
AU2012-417
AU2014-613
AU2014-635
AU2014-663
AU2014-586
AU2014-690
AU2014-693
AU2014-680
AU2014-620
AU2014-697
AU2014-684
C5
AU2012-238

1 SNP