

Whole genome sequencing reveals transmission of gonococcal antibiotic resistance among men who have sex with men: an observational study

SUPPLEMENTARY APPENDIX

ANTIMICROBIAL SUSCEPTIBILITY TESTING

Antimicrobial susceptibility testing was performed using methods of the Australian Gonococcal Surveillance Programme (AGSP) and in accordance with the National Neisseria Network of Australia.^{1,2} Interpretative criteria are listed in Table S1.

Table S1: AGSP Interpretative Criteria for *N. gonorrhoeae* MIC values

Antibiotic MIC (mg/L)	Sensitive	Less susceptible / Decreased susceptibility	Resistant
Penicillin	<0.06	0.06–0.5	>0.5 #
Ceftriaxone	<0.06	0.06–0.25	Not defined
Azithromycin	<1	Not defined	≥1
Ciprofloxacin	<0.06	0.06–0.5	>0.5
Spectinomycin	≤64	Not defined	>64

All penicillinase-producing *N. gonorrhoeae* (PPNG) isolates are considered resistant to penicillin, irrespective of the MIC value

BIOINFORMATICS METHODS

Raw sequencing reads were processed using a customised pipeline to streamline and automate initial handling of sequence data (<https://github.com/tseemann/nullarbor>). Within this pipeline, Illumina Nextera adapters and low-quality sequence (Phred score <10) were trimmed from raw sequencing reads, with the trimmed sequences subsequently undergoing *de novo* assembly (*SPAdes* v3.7.1),³ annotation (*Prokka* v1.12-beta),⁴ MLST (<https://github.com/tseemann/mlst>), resistance gene detection (<https://github.com/tseemann/abricate>), and pan-genome analysis (*Roary* v3.6.6).⁵ To generate an alignment of core genome single nucleotide polymorphisms (SNPs), sequence reads for each isolate were aligned to the NCCP11945 reference genome (GenBank accession NC_011035.1)⁶ within the *Nullarbor* pipeline using *Snippy* v3.1 (<https://github.com/tseemann/snippy>). Sites that were covered by less than 10 reads were considered absent for that isolate. The remaining non-absent sites were considered to be a SNP if at least 90% of the reads covering that site were different to the reference. The 90% read concordance threshold was set to minimise sequencing error. Any site for which at least one of the isolates had a SNP and none of the isolates were absent, were considered as "core SNP" sites. These were used as input to determine the phylogenomic relatedness of the isolates. SNPs identified between isolates from within an individual or partnership were manually inspected and verified.

The Nullarbor makefile was generated with the following command:

```
$ nullarbor.pl --name partners.20160926 --mlst neisseria --ref NCCP11945.gbk --input samples.tab --outdir partners.20160926 --accurate
```

Within Nullarbor, the following commands were run for each isolate:

```
# Trim reads with Trimmomatic
$ trimmomatic PE -threads $(CPUS) -phred33
/mnt/seq/MDU/READS/sample1/sample1_R1.fastq.gz
/mnt/seq/MDU/READS/sample1/sample1_R2.fastq.gz sample1/clipped_R1.fq.gz
/dev/null sample1/clipped_R2.fq.gz /dev/null
ILLUMINACLIP:conf/trimmomatic.fa:1:30:11 LEADING:10 TRAILING:10
MINLEN:30

# Basic QC estimation of sequencing yield and depth
$ fq --quiet --ref ref.fa sample1/clipped_R1.fq.gz
sample1/clipped_R2.fq.gz > sample1/yield.clean.tab

# QC check species and contamination with Kraken
$ kraken --threads $(CPUS) --preload --paired sample1/clipped_R1.fq.gz
sample1/clipped_R2.fq.gz | kraken-report > sample1/kraken.tab
```

```

# Read alignment with Snippy
$ snippy --cpus $(CPUS) --force --outdir sample1/sample1 --ref
NCCP11945.gbk --R1 sample1/R1.fq.gz --R2 sample1/R2.fq.gz

# De novo draft genome assembly with SPAdes
$ spades.py -o sample1/spades --careful --tmp-dir /tmp --only-assembler
-t $(CPUS) -k 21,33,55,77,97,107,117,127 -1 sample1/clipped_R1.fq.gz -2
sample1/clipped_R2.fq.gz

# Basic assembly QC
$ fa -e -t sample1/contigs.fa

# In silico detection of acquired resistance mechanisms
$ abricate sample1/contigs.fa > abricate.tab

# Draft genome auto-annotation with Prokka
$ prokka --centre X --compliant --force --fast --locustag sample1
--prefix sample1 --outdir sample1/prokka --cpus $(CPUS) --proteins
NCCP11945.gbk sample1/contigs.fa

```

In addition, the following were run on the subsequent output:

```

# Generate core genome SNP alignment and pseudo-whole-genome alignment
$ snippy-core sample1/sample1 sample2/sample2 ... sampleN/sampleN

# Generate table of pairwise SNP distances
$ afa-pairwise.pl core.aln > distances.tab

# Pan-genome analysis with Roary
$ roary -f sample1/roary -v -p $(CPUS) */prokka/*.gff

# Approximate maximum likelihood tree estimation with FastTree
$ FastTree -nt -gtr < core.aln > tree.newick

```

A summary report was generated for each set of isolates run through Nullarbor including quality control metrics, MLST, resistome and pan-genome comparison. NG-MAST and MLST results were inferred *in silico* with *NGMASTER*,⁷ and *mlst* (<https://github.com/tseemann/mlst>) respectively:

```

# In silico MLST
$ mlst --scheme neisseria sample1/contigs.fa > mlst.tab

# In silico NG-MAST
$ ngmaster sample1/contigs.fa > ngmast.tab

```

The pan-genome output from *Roary* was formatted (<https://github.com/kwongj/roary2fripan>) for visualisation and inspected in *FriPan* (<http://drpowell.github.io/FriPan/>).

Resistance-determining regions of individual genomes were identified in *de novo* draft assemblies using *BLAST*.⁸ Resistance gene sequences were extracted, aligned with *MUSCLE*,⁹ and manually inspected in *Geneious* v10.0.5 (<https://www.geneious.com/>) for comparison between strains. Additionally, to verify the mutations detected in draft assemblies, sequence reads were aligned to the FA1090 reference genome (GenBank accession AE004969.1) using Snippy v3.2-dev. The resulting output (snps.tab file) was searched for resistance determinants:

```
# Run snippy to align reads to FA1090 reference
$ snippy --outdir sample1 --ref FA1090.gbk --pe1
sample1/clipped_R1.fq.gz --pe2 sample1/clipped_R2.fq.gz

# Search snps.tab file for resistance determinants
$ for f in */snps.tab ; do if grep NGO1812 $f | grep missense | grep
120/348 ; then echo $f ; fi ; done
```

To detect 23S allele mutations, sequencing reads were aligned to a mutated NCCP11945 reference genome with three copies of the 23S region masked to force alignment to a single locus. An in-house script (<https://github.com/kwongj/ng23S-mutations>) was used to examine the relative proportion of reads with the mutation and determine how many 23S copies had the mutation.

```
# Run snippy to align reads to masked NCCP11945 reference
$ snippy --outdir sample1-23S --ref NG_NCCP11945_23S-masked.gbk --pe1
sample1/clipped_R1.fq.gz --pe2 sample1/clipped_R2.fq.gz

# Run ng23S-mutations script to print results to stdout
$ ng23S-mutations.py sample1-23S sample2-23S ... sampleN-23S
```

For phylogenetic analysis, recombination was identified with *ClonalFrameML*,¹⁰ and filtered using a custom python script (<https://github.com/kwongj/cfml-maskrc>) using the following commands:

```
# Resolve polytomies in the tree from Nullarbor
$ nw_multifurcation.py --nosupport tree.newick > core.tree

# Trim the pseudo-whole-genome alignment from Nullarbor
$ trim-aln.py --end 2232025 core.full.aln > core.chr.full.aln

# Run ClonalFrameML
$ ClonalFrameML core.tree core.chr.full.aln cfml -
ignore_incomplete_sites true

# Filter predicted regions of recombination from alignment
$ cfml-maskrc.py --aln core.chr.full.aln --out maskrc.full.aln cfml
```

A maximum likelihood phylogeny was inferred from the filtered alignment using *RAxML* v8.2.8¹¹ to compare isolates:

```
# Exclude invariant and non-core sites using snp-sites
$ snp-sites -c -o maskrc.core.aln maskrc.full.aln

# Run RAxML
$ raxmlHPC-PTHREADS -s maskrc.core.aln -n raxml.maskrc.tree -m GTRGAMMA
-f a -p 13579 -x 13579 -# 1000 -T 64
```

SUPPLEMENTARY RESULTS

A summary table of isolate and genome characteristics is provided in Supplementary Table S2. For each isolate, over 89% of the reference genome was covered by aligned reads for identifying SNPs. A total of 15,981 core SNP sites were identified, with 1619 genes identified as core genes shared by >99% of isolates from a total of 5229 putative genes.

AUSMDU00005596	3/06/2013	30	B	Pharyngeal	M	33	8602	5093	5	-	59	39	67	78	682	153	65	1	<=64	<=0.008	2	>8	<=0.06	Detected	NonMosaic_II*	penA_294		L421P	blaTEM-1B	tet(M)-12	S91F, D95G	D86N		NEIS0489_81	NEIS0763_327	no	
AUSMDU00005597	3/10/2013	29	B	Pharyngeal	M	31	-	new	563	7363	59	39	67	78	148	153	65	0.5	<=64	0.03	8	<=4	0.125	Not detected	NonMosaic_IX*	penA_645	G120K, A121D	L421P	-A		S91F, D95G	E91G		NEIS0488_635	NEIS0489_808	NEIS0763_332	yes
AUSMDU00005598	3/10/2013	29	B	Rectal	M	31	-	new	563	7363	59	39	67	78	148	153	65	0.5	<=64	0.03	8	<=4	0.125	Not detected	NonMosaic_IX*	penA_645	G120K, A121D	L421P	-A		S91F, D95G	E91G		NEIS0488_635	NEIS0489_808	NEIS0763_332	yes
AUSMDU00005599	3/10/2013	29	A	Rectal	M	33	-	new	563	7363	59	39	67	78	148	153	65	0.5	<=64	0.03	8	<=4	0.125	Not detected	NonMosaic_IX*	penA_645	G120K, A121D	L421P	-A		S91F, D95G	E91G		NEIS0488_635	NEIS0489_808	NEIS0763_332	yes
AUSMDU00005600	30/12/2013	26	A	Urethra	M	20	8030	4786	563	7363	59	39	67	78	148	153	65	0.5	<=64	0.03	8	<=4	0.125	Not detected	NonMosaic_IX*	penA_645	G120K, A121D	L421P	-A		S91F, D95G	E91G		NEIS0488_635	NEIS0489_808	NEIS0763_332	yes
AUSMDU00005601	31/12/2013	31	A	Urethra	M	27	10162	5957	188	8123	59	39	170	78	148	153	65	0.125	<=64	<=0.008	<=0.03	>8	<=0.06	Not detected	NonMosaic_II	penA_166			tet(M)-5					NEIS0488_514	NEIS0489_804	NEIS0763_11	no
AUSMDU00005602	31/12/2013	31	B	Rectal	M	33	10162	5957	188	8123	59	39	170	78	148	153	65	0.125	<=64	<=0.008	<=0.03	>8	<=0.06	Not detected	NonMosaic_II	penA_166			tet(M)-5					NEIS0488_514	NEIS0489_804	NEIS0763_11	no
AUSMDU00005603	30/12/2013	26	B	Rectal	M	24	8030	4786	563	7363	59	39	67	78	148	153	65	0.5	<=64	0.03	4	<=4	0.125	Not detected	NonMosaic_IX*	penA_645	G120K, A121D	L421P	-A		S91F, D95G	E91G		NEIS0488_12	NEIS0489_808	NEIS0763_332	yes
AUSMDU00005604	3/07/2014	32	A	Pharyngeal	M	19	1614	1053	33	7828	126	39	67	158	148	153	133	0.25	<=64	<=0.008	1	<=4	<=0.06	Not detected	NonMosaic_IX	penA_285	G120K, A121D	L421P	-A		S91F, D95G			NEIS0488_70	NEIS0489_842	NEIS0763_332	no
AUSMDU00005605	3/07/2014	32	A	Rectal	M	19	1614	1053	33	7828	126	39	67	158	148	153	133	0.25	<=64	<=0.008	1	<=4	<=0.06	Not detected	NonMosaic_IX	penA_285	G120K, A121D	L421P	-A		S91F, D95G			NEIS0488_70	NEIS0489_842	NEIS0763_332	no
AUSMDU00005606	3/07/2014	32	B	Urethra	M	43	1614	1053	33	7828	126	39	67	158	148	153	133	0.25	<=64	0.016	1	<=4	<=0.06	Not detected	NonMosaic_IX	penA_285	G120K, A121D	L421P	-A		S91F, D95G			NEIS0488_70	NEIS0489_842	NEIS0763_332	no
AUSMDU00005607	14/08/2014	20	A	Rectal	M	50	8709	2656	1579	-	126	39	-	238	148	153	133	0.25	<=64	<=0.008	<=0.03	<=4	0.25	Not detected	NonMosaic_II	penA_166								NEIS0488_12	NEIS0489_287	NEIS0763_8	no
AUSMDU00005608	14/08/2014	20	A	Pharyngeal	M	50	8709	2656	1579	-	126	39	-	238	148	153	133	0.25	<=64	<=0.008	<=0.03	<=4	0.25	Not detected	NonMosaic_II	penA_166								NEIS0488_12	NEIS0489_287	NEIS0763_8	no
AUSMDU00005609	14/08/2014	20	B	Urethra	M	42	8709	2656	1579	-	126	39	-	238	148	153	133	0.25	<=64	<=0.008	<=0.03	<=4	0.25	Not detected	NonMosaic_II	penA_166								NEIS0488_12	NEIS0489_287	NEIS0763_8	no
AUSMDU00005610	8/09/2014	12	A	Urethra	M	26	9654	1808	1744	-	126	39	170	238	-	153	133	0.25	<=64	<=0.008	<=0.03	<=4	0.125	Not detected	NonMosaic_II	penA_166								NEIS0488_12	NEIS0489_287	NEIS0763_8	no
AUSMDU00005611	8/09/2014	12	A	Rectal	M	26	9654	1808	1744	-	126	39	170	238	-	153	133	0.25	<=64	<=0.008	<=0.03	<=4	0.25	Not detected	NonMosaic_II	penA_166								NEIS0488_12	NEIS0489_287	NEIS0763_8	no
AUSMDU00005612	8/09/2014	12	B	Pharyngeal	M	22	9654	1808	1744	-	126	39	170	238	-	153	133	0.25	<=64	<=0.008	<=0.03	<=4	0.25	Not detected	NonMosaic_II	penA_166								NEIS0488_12	NEIS0489_287	NEIS0763_8	no
AUSMDU00005613	8/09/2014	12	B	Rectal	M	22	9654	1808	1744	-	126	39	170	238	-	153	133	0.25	<=64	<=0.008	<=0.03	<=4	0.25	Not detected	NonMosaic_II	penA_166								NEIS0488_12	NEIS0489_287	NEIS0763_8	no
AUSMDU00005614	22/09/2014	33	A	Urethra	M	35	1407	908	110	7360	109	39	170	78	148	153	65	1	<=64	0.03	8	<=4	0.25	Not detected	Mosaic_XXXIV	penA_266	G120K, A121N	L421P	-A		S91F, D95G	S87R		NEIS0488_12	NEIS0489_12	NEIS0763_327	yes
AUSMDU00005615	22/09/2014	33	B	Rectal	M	24	1407	908	110	7360	109	39	170	78	148	153	65	1	<=64	0.06	8	<=4	0.25	Not detected	Mosaic_XXXIV	penA_266	G120K, A121N	L421P	-A		S91F, D95G	S87R		NEIS0488_12	NEIS0489_12	NEIS0763_327	yes

Table S3: Comparison between distinguishable partner isolates

Partner Site	4A Urethra	4B Rectum	25A Urethra	25B Rectum	26A Urethra	26B Rectum
Susceptibility						
Penicillin MIC (mg/L)	2.0	0.25	0.5	1.0	0.5	0.5
PPNG	ND	ND	ND	ND	ND	ND
Ceftriaxone MIC (mg/L)	0.03	0.03	0.016	0.016	0.03	0.03
Azithromycin MIC (mg/L)			0.125	0.125	0.125	0.125
Ciprofloxacin MIC (mg/L)	≤ 0.03	4	16	16	8	4
Spectinomycin MIC (mg/L)	≤ 64	≤ 64	≤ 64	≤ 64	≤ 64	≤ 64
Typing						
NG-MAST	1419	New	New	New	8030	8030
MLST	1579	8123	9364	9364	7363	7363
WGS comparison						
Number of predicted CDS	2055	2088	2050	2035	2084	2107
SNPs unfiltered	4063	4063	163	163	73	73
SNPs filtered	1498	1498	1	1	25	25

MIC = minimum inhibitory concentration

PPNG = penicillinase-producing *N. gonorrhoeae*

ND = not detected

NG-MAST = *N. gonorrhoeae* multi-antigen sequence typing

MLST = multi-locus sequence typing

WGS = whole-genome sequencing

CDS = coding DNA sequences

SNP = single nucleotide polymorphism

Figure S1: Comparison of four instances where couples had isolates with the same NG-MAST and MLST as another couple. The portions of the tree on the left illustrate the phylogenetic relationships between the isolates. The panel on the right is a heat map of recombinant regions represented by grey vertical lines at the corresponding position across the *N. gonorrhoeae* genome. Red arrows and lines indicate regions where recombination patterns were different between the couples in each instance.



REFERENCES

1. Tapsall J, and members of the National Neisseria Network of Australia. Antimicrobial testing and applications in the Pathogenic Neisseria. In: Merlino J, ed. Antimicrobial Susceptibility Testing: Methods and Practices with an Australian Perspective. Sydney: Australian Society for Microbiology 2004:173-86.
2. Tapsall JW. Use of a quality assurance scheme in a long-term multicentric study of antibiotic susceptibility of Neisseria gonorrhoeae. *Genitourin Med* 1990;66(1):8-13.
3. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19(5):455-77. doi: 10.1089/cmb.2012.0021
4. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30(14):2068-9. doi: 10.1093/bioinformatics/btu153
5. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31(22):3691-3. doi: 10.1093/bioinformatics/btv421
6. Chung GT, Yoo JS, Oh HB, et al. Complete genome sequence of Neisseria gonorrhoeae NCCP11945. *J Bacteriol* 2008;190(17):6035-6. doi: 10.1128/JB.00566-08
7. Kwong JC, Gonçalves da Silva A, Dyet K, et al. NGMASTER: in silico multi-antigen sequence typing for Neisseria gonorrhoeae. *Microbial Genomics* 2016 doi: 10.1099/mgen.0.000076 [published Online First: 25 August 2016]
8. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403-10. doi: 10.1016/S0022-2836(05)80360-2
9. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32(5):1792-7. doi: 10.1093/nar/gkh340
10. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 2015;11(2):e1004041. doi: 10.1371/journal.pcbi.1004041
11. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30(9):1312-3. doi: 10.1093/bioinformatics/btu033