SHORy REPORT

Epidemiological typing of *Neisseria gonorrhoeae* and detection of markers associated with antimicrobial resistance directly from urine samples using next generation sequencing

R M A Graham, C J Doyle, A V Jennison

ABSTRACT

**Objectives** To investigate the potential for next generation sequencing (NGS) to be used directly on clinical specimens that have tested positive for *Neisseria gonorrhoeae* by nucleic acid amplification testing (NAAT), to generate information on epidemiological genotyping and antimicrobial resistance (AMR) markers.

**Methods** DNA was extracted from 13 *N. gonorrhoeae* NAAT-positive urine specimens, enriched for microbial DNA and sequenced using the Ion Torrent PGM workflow. Sequences that aligned to the human genome were filtered out and the remaining sequences were de novo assembled. The resulting contigs were searched for regions of interest using Ridom SeqSphere. MLST and NG-MAST alleles were assigned according to the schemes at PubMLST.org and NG-MAST.net, respectively.

**Results** In total, 11 of the 13 samples tested generated a sufficient number of *N. gonorrhoeae* sequence reads to provide full coverage of the genome at a depth of 6–130×. Complete MLST and NG-MAST sequence types could be generated for each of these samples. The presence of 10 different AMR markers was investigated, and both previously reported and novel mutations were identified in genes associated with reduced susceptibility to several antimicrobials.

**Conclusions** We found that sequencing the entire genome of *N. gonorrhoeae* directly from clinical samples is possible using NGS, and that multiple levels of *N. gonorrhoeae* typing information can be generated. As NAAT only testing becomes more common, this method could be used to detect both known and novel mutations associated with AMR and to generate genotyping information, supporting AMR and epidemiological surveillance in the absence of culturing.

INTRODUCTION

*Neisseria gonorrhoeae*, the causative agent of gonorrhoea, continues to be a major public health concern, due to a continued rise in the number of infections globally and increasing levels of antimicrobial resistance (AMR). Over time, gonococci have developed resistance to previously used first-line antimicrobials such as penicillin, tetracycline, macrolides and fluoroquinolones, and isolates with decreased susceptibility to extended-spectrum cephalosporins have now been reported in several countries. Antimicrobial sensitivity testing and surveillance of *N. gonorrhoeae* in the community using epidemiological typing methods such as NG-MAST and MLST provide important information on transmission networks, disease clusters, and the emergence and spread of antimicrobial-resistant strains. In many countries, however, the diagnosis of gonorrhoea is now carried out predominately by nucleic acid amplification testing (NAAT), limiting the availability of isolates for epidemiological typing and phenotypic antimicrobial susceptibility testing. A number of molecular markers have been identified in *N. gonorrhoeae* for reduced susceptibility to antimicrobials; however, screening for these markers requires multiple molecular assays and may miss novel mutations. By sequencing the entire genome of *N. gonorrhoeae* directly from a clinical sample using next generation sequencing (NGS), typing information can be generated and both known and novel mutations associated with AMR identified, although the effect of any novel mutations on susceptibility to antimicrobials would require phenotypic analysis. We report on the potential for NGS performed directly on *N. gonorrhoeae*-positive clinical samples to generate data for AMR markers and epidemiological genotyping.

METHODS

A total of 13 deidentified urine-Cobas PCR media specimens were used in this study. These specimens were known to be positive for *N. gonorrhoeae* using the Roche Cobas 4800 CT/NG test, but no further information regarding the load of gonococcal DNA present in the samples was provided. DNA was extracted using the MasterPure DNA extraction kit (Epicentre, Madison Wisconsin, USA), according to the manufacturer’s instructions. Total DNA was enriched for microbial DNA using the NEBNext microbiome kit (New England BioLabs, Ipswitch, Massachusetts, USA). Extracted DNA was sheared to 300 bp using the Covaris S220 ultrasonicator and fragment libraries were prepared using the Ion Plus Fragment library kit and sequenced on an Ion Torrent PGM using the Ion PGM IC 200 Kit and 316v2 chips (Life Technologies, Carlsbad, California, USA) according to the manufacturer’s instructions. Sequences that aligned to the human genome were removed and the remaining sequences were de novo assembled using Genome Assembly, Resistance, Epidemiology (GARE) pipeline. Sequences aligning to the *N. gonorrhoeae* FA1090 genome were uploaded to the European Nucleotide Archive under project
RESULTS AND DISCUSSION

To investigate the quality and abundance of gonococcal DNA present in the samples, sequences were mapped to the
*N. gonorrhoeae* FA1090 reference genome (GenBank accession number: NC_002946) using CLC Genomics workbench 8. Enrichment of microbial DNA was found to be an essential step, as when DNA extracted from the samples was not enriched, fewer than 1% of sequences aligned to the *N. gonorrhoeae* genome. This number of reads was sufficient to identify the presence of *N. gonorrhoeae* in a sample using the K-mer Finder tool at the Center for Genomic Epidemiology (CGE) (https://cge.cbs.dtu.dk/services/KmerFinder-2.0/), but gave insufficient coverage of the genome to allow characterisation. Following enrichment for microbial DNA, in 11 of the 13 samples the percentage of sequences that aligned to the *N. gonorrhoeae* genome ranged from 2% to 43% of the total reads, which was sufficient to provide an average depth of coverage of the FA1090 genome of 6–130×. With the exception of the sample with only 6× coverage, sequences from all other samples provided greater than 10× coverage, and although 6× coverage is lower than ideal, the quality of the sequences generated from this sample was high, and robust typing information was still able to be generated. Enrichment for microbial DNA did not sufficiently improve the number of gonococcal sequence reads for two of the samples, so no further analysis was done. The presence of other *Neisseria* species in the specimens was excluded by *rplF* gene and K-mer analysis, which are methods used to accurately identify and discriminate between different *Neisseria* species. To determine NG-MAST and MLST types, sequences from each sample were de novo assembled and the contigs were screened for the target loci using Ridom SeqSphere, which also assigned alleles and sequence types (STs) according to the schemes at NG-MAST.net and PubMLST.org, respectively. Each of the 11 samples had a unique NG-MAST ST, and none of the samples had more than one NG-MAST ST detected, which would have been indicative of a same site mixed infection. Novel *tbpB* sequences were identified for two samples, which were assigned new allele types and STs. The sample with 6× coverage across the genome had regions in the *porB* and *tbpB* genes with only 4× coverage, which is lower than the minimum coverage recommended by Ridom SeqSphere. However, there were no ambiguities in these regions and all four reads were identical at all positions so it is likely that the consensus sequences generated and thus the ST assigned were accurate. All seven MLST loci were found in each sample at greater than 5× coverage, and nine different MLST STs were present, as described in table 1. In addition to the standard MLST, the 20 locus eMLST (pubMLST.net) profiles could be determined for all samples except S7, which had insufficient coverage of the *pip* gene, demonstrating the potential for higher resolution genotyping to be performed using the data generated.

The assembled sequence data for each sample was investigated for the presence of mutations in eight genes associated with reduced gonococcal antimicrobial susceptibility (table 1) using Ridom SeqSphere. Additionally, the presence of acquired resistance genes was investigated using the ResFinder tool at CGE (https://cge.cbs.dtu.dk/services/ResFinder-2.1/). A number of AMR markers were identified in the samples, the most common (9/11) being the A39T amino acid substitution in MtrR. This mutation has been associated with low-level resistance to

<table>
<thead>
<tr>
<th>Sample</th>
<th>NG-MAST</th>
<th>MLST</th>
<th>mtrR††</th>
<th>penA§</th>
<th>penA genotype*†</th>
<th>rpsL§</th>
<th>gyrA¶</th>
<th>parC¶</th>
<th>ponA*</th>
<th>23S rRNA gene#</th>
<th>Presence of Penicillinase-encoding plasmid*</th>
<th>TetM-encoding plasmid$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>9920</td>
<td>7363</td>
<td>A39T</td>
<td>A121G</td>
<td>II</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>L421P</td>
<td>WT</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>S2</td>
<td>8709</td>
<td>9363</td>
<td>A39T</td>
<td>WT</td>
<td></td>
<td>V57M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>S3</td>
<td>21</td>
<td>1579</td>
<td>Deletion of A**</td>
<td>G120K</td>
<td>A121G</td>
<td>IX (D345a, A501V, P551L)</td>
<td>V57M</td>
<td>WT</td>
<td>WT</td>
<td>L421P</td>
<td>C2611T (4)††</td>
<td>No</td>
</tr>
<tr>
<td>S4</td>
<td>7138</td>
<td>11 430††</td>
<td>A39T</td>
<td>WT</td>
<td>II</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>S91F</td>
<td>D95A</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>S5</td>
<td>12 424††</td>
<td>11 431††</td>
<td>A39T</td>
<td>WT</td>
<td>II</td>
<td>V57M</td>
<td>S91F</td>
<td>E91G</td>
<td>WT</td>
<td>WT</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>S6</td>
<td>9654</td>
<td>11 428</td>
<td>A39T</td>
<td>WT</td>
<td>II</td>
<td>V57M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>S7</td>
<td>12 425††</td>
<td>8134</td>
<td>A39T</td>
<td>WT</td>
<td>II</td>
<td>V57M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>S8</td>
<td>4822</td>
<td>1901</td>
<td>Deletion of A</td>
<td>G120K</td>
<td>A121N</td>
<td>XXXIV (Mosaic allele)</td>
<td>V57M</td>
<td>S91F</td>
<td>S87R</td>
<td>L421P</td>
<td>C2611T (2)††</td>
<td>No</td>
</tr>
<tr>
<td>S9</td>
<td>10 558</td>
<td>9363</td>
<td>A39T</td>
<td>WT</td>
<td>II</td>
<td>V57M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>S10</td>
<td>9909</td>
<td>7359</td>
<td>A39T</td>
<td>WT</td>
<td>II</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>S11</td>
<td>4186</td>
<td>7359</td>
<td>A39T</td>
<td>WT</td>
<td>II</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

* Associated with decreased susceptibility/resistance to penicillin.
† Associated with decreased susceptibility to cephalosporins.
‡ Associated with decreased susceptibility to ceftriaxone.
§ Associated with decreased susceptibility to azithromycin.
¶ Associated with decreased susceptibility/resistance to tetracycline.
†† Contains an A deletion in the 13 bp inverted repeat region of the mtrF promoter.
†‡ Parentheses indicate number of alleles with this mutation.
§§ STs not previously identified.
STs, sequence types; WT, wild type.
macrolides and β-lactam antibiotics, and decreased susceptibility to quinolones when present in conjunction with the mtrR G43D mutation, but reportedly has limited effect on susceptibility when present on its own. More strongly linked to antibiotic resistance is an adenine deletion in the promoter of mtrR, which was present in two specimens, S3 and S8. These specimens also had several other AMR markers and characteristics suggestive of multiple drug resistance. S3 had the type IX penA allele, which is not mosaic in structure but has D345A, A501V and P551L mutations, which are associated with reduced susceptibility to ceftriaxone when present in conjunction with the G120K, A121D mutations in penB that were also present in this sample. S8 was determined to be MLST ST1901 and NG-MAST ST4822. N. gonorrhoeae with MLST ST1901 form a globally distributed multidrug-resistant gonococcal clone, and NG-MAST ST4822 has been associated with resistance to multiple drugs in South Africa, a phenotype also seen in many Queensland NG-MAST ST4822 isolates (unpublished data). Specimen S8 had the type XXXIV penA mosaic allele, the L421P mutation in penA, and the G120K and A121N mutations in penB, AMR markers that are identical to those described in the South African NG-MAST ST4822 isolates. Both S3 and S8 also had mutations in gyrA and parC, rpsJ and the 23SrRNA genes, which are associated with reduced susceptibility to quinolones, tetracycline and azithromycin, respectively. Penicillinase-producing plasmid sequences were present in two specimens, and TetM-encoding plasmid sequences were identified in one specimen. Given these specimens were from a non-sterile site and organisms other than N. gonorrhoeae will be present, it is possible that the AMR markers detected do not belong to N. gonorrhoeae, and care should be taken in interpreting these results. However, the chromosomal markers identified in this study have been used in previous studies as targets for real-time PCR performed directly on clinical samples, and the single-nucleotide polymorphisms present have a high specificity for N. gonorrhoeae. As N. gonorrhoeae was not cultured from these specimens, phenotypic antimicrobial sensitivity testing data were not available, so we are unable to confirm if the genotypic markers detected contribute to the AMR profile of N. gonorrhoeae present in the samples; however, these markers have been shown to be good predictors of reduced susceptibility to antimicrobials in several studies (reviewed in ref. 2). Furthermore, the relationship between genotype and phenotype with regard to antimicrobial sensitivity is complex and involves interactions between the products of several genes. In many cases, multiple mutations are necessary to produce reduced susceptibility to a given drug. In this study, sequencing all gonococcal DNA present in N. gonorrhoeae NAAT-positive specimens by NGS could, in the majority of samples, generate extensive information on AMR markers and genotypes. This approach, while not yet cost effective for routine public health surveillance, does demonstrate a means to produce valuable information on potential AMR where a culture is not available and replaces the need for multiple molecular tests targeting individual markers. The ability to identify novel mutations that may have an effect on antimicrobial sensitivity is also highly beneficial. In several of the specimens tested in this study, previously undescribed mutations in AMR marker genes were identified. These include a R44H mutation in mtrR and A59T mutation in gyrA. Whether these mutations have any effect on antimicrobial sensitivity cannot be inferred from this study, but they may warrant further investigation by phenotypic testing. The specimens provided in this study were identified as N. gonorrhoeae positive by NAAT only, meaning that there were no matched bacterial isolates available to confirm the results generated by NGS. This is a limitation of the study to date, with further validation of the method using N. gonorrhoeae isolates intended.

This study has demonstrated that NGs can be used to generate extensive N. gonorrhoeae typing information directly from clinical samples using simple bioinformatics tools. Using this technology, we were able to identify multiple AMR markers and determine NG-MAG and MLST ST in N. gonorrhoeae-positive samples diagnosed by NAAT only. As NAAT only testing for N. gonorrhoeae becomes more common, the ability to perform epidemiological typing and antimicrobial sensitivity testing directly on clinical samples will be vital to ensure this valuable information can continue to be collected.

Handling editor Jackie A Cassell

Acknowledgements The authors acknowledge Sullivan Nicolaides Pathology for providing the specimens used in this study.

Contributors RMAG designed the study, performed testing and data analysis, and wrote the paper. CID performed data collection and analysis. AVJ advised on study design, data analysis and manuscript preparation.

Competing interests None declared.

Ethics approval This study was approved by the Queensland Health Forensic and Scientific Services Human Ethics Committee, approval number HEC15-010.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES