SCIENCE CHINA PRESS

SCIENCE CHINA Life Sciences
life.scichina.com link.springer.com
https://doi.org/10.1007/s11427-022-2382-0

LETTER TO THE EDITOR

Development and performance evaluation of a culture-independent nanopore amplicon-based sequencing method for accurate typing and antimicrobial resistance profiling in *Neisseria gonorrhoeae*

Chi Zhang[1,2†], Lulu Zhang[1,2†], Feng Wang[3], Yaling Zeng[3], Liying Sun[1,2], Di Wang[1,2], Yamei Li[1,2], Liqin Wang[1,2] & Junping Peng[1,2*]

[1]NHC Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100176, China;
[2]Key Laboratory of Respiratory Disease Pathogenomics, Chinese Academy of Medical Sciences, Beijing 100730, China;
[3]Shenzhen Center for Chronic Disease Control, Shenzhen 518020, China

†Contributed equally to this work
*Corresponding author (email: pengjp@hotmail.com)

Received 25 November 2022; Accepted 7 June 2023; Published online 1 September 2023

Dear Editor,

Bacterial antimicrobial resistance (AMR) poses a serious threat to global human health (Antimicrobial Resistance Collaborators, 2022). Comprehensive profiling of AMR and accurate molecular typing are important for tracking and controlling the emergence and dissemination of antibiotic-resistant bacteria (Yahara et al., 2021). We previously developed a multiplex amplicon sequencing-based method for directly sequencing AMR-related loci in *N. gonorrhoeae* from clinical samples (Zhang et al., 2021). Furthermore, we proposed a general framework for the method and presented a proof-of-concept evaluation for its feasibility (Potter et al., 2021). However, this prototype method still had two limitations that needed to be resolved. The first limitation was the low accuracy of retrieving homopolymers, which was solved by the development of a novel data processing pipeline in this study. The improved accuracy enabled the data of portable sequencers to perform precise molecular typing of bacteria, which was previously performed using Sanger sequencing or desktop sequencers. The second limitation was the limited batch size per run. The batch size was considerably increased in the current procedure by combining a ligation-based library and optimized PCR assays, resulting in reduced reagent cost per sample. Furthermore, we evaluated the performance of R10.3 flowcell and investigated the relationship between input bacterial genome copies and multiplex PCR amplification efficiency, inferring the limit of detection (LOD) of similar methods based on MinION amplicon sequencing (Cabibbe et al., 2020).

Primers for multiplex PCR were designed and optimized to simultaneously amplify 16 targets, including (i) 7 MLST (multilocus sequence typing) loci (*abcZ*, *adk*, *aroE*, *fumC*, *gdh*, *pdhC*, and *pgm*), (ii) 7 resistance-associated genes used in the NG-STAR typing scheme (*penA*, *mtrR*, *porB*, *ponA*, *gyrA*, *parC*, and 23S rRNA; Demczuk et al., 2017), (iii) NG-MAST loci (*tbpB*, another NG-MAST locus *porB* was included in the NG-STAR typing scheme), and (iv) *porA* for *N. gonorrhoeae* identification. Primer sequences and the optimized volume of each primer required to mix the pool are listed in Table S1 in Supporting Information. WHO *N. gonorrhoeae* reference strains WHO-K, ATCC-49226, 74 *N. gonorrhoeae* isolates, and 51 *N. gonorrhoeae*-positive urethral swabs were used to validate the method. Furthermore, mock specimens were used to determine the LOD of the method. In the bioinformatics pipeline, variants were called with their respective quality scores using Freebayes (Liu et al., 2019). Different thresholds were set based on the types of flowcell (R9.4 or R10.3) and variants (SNP or indel) to filter potential false-positive variants. Thus, the high error of homopolymer calling, which is an intrinsic limitation of nanopore data (Sereika et al., 2022), can be effectively controlled. Figure S1 in Supporting Information depicts an overview of the experimental and data processing workflow used in this study. Sample information, detailed experimental protocol, and data processing pipeline can be found in the materials and methods section of Supporting Information.

Determining the minimum depths is critical for inferring the required data size and thus estimating the suitable sequencing time. The results showed that the hac model (Figure 1A; Figure S2 in Supporting Information) outperformed the fast model (Figures S3 and S4 in Supporting Information; 1.02 versus 11.17 average mismatches, $P<0.001$). The difference in performance between the R10.3 flowcell and R9.4 flowcell in terms of an average number of mismatches was not statistically significant (7.88 versus 4.30 average mismatches, $P=0.126$). For depths of 30×, 45×, and 60×, the results of typing and identification of AMR-related sites were correct in all groups. For a more comprehensive evaluation of the accuracy, samples with different phenotypic and genetic characteristics were tested. Results of nine clinical samples showed the method had good performance in typing and identification of AMR-related sites (Figure 1B). For the R10.3 flowcell group, mismatches were only observed in one isolate (Clinical Sample a04). The average number of mismatches of isolates was lower than that of clinical samples,

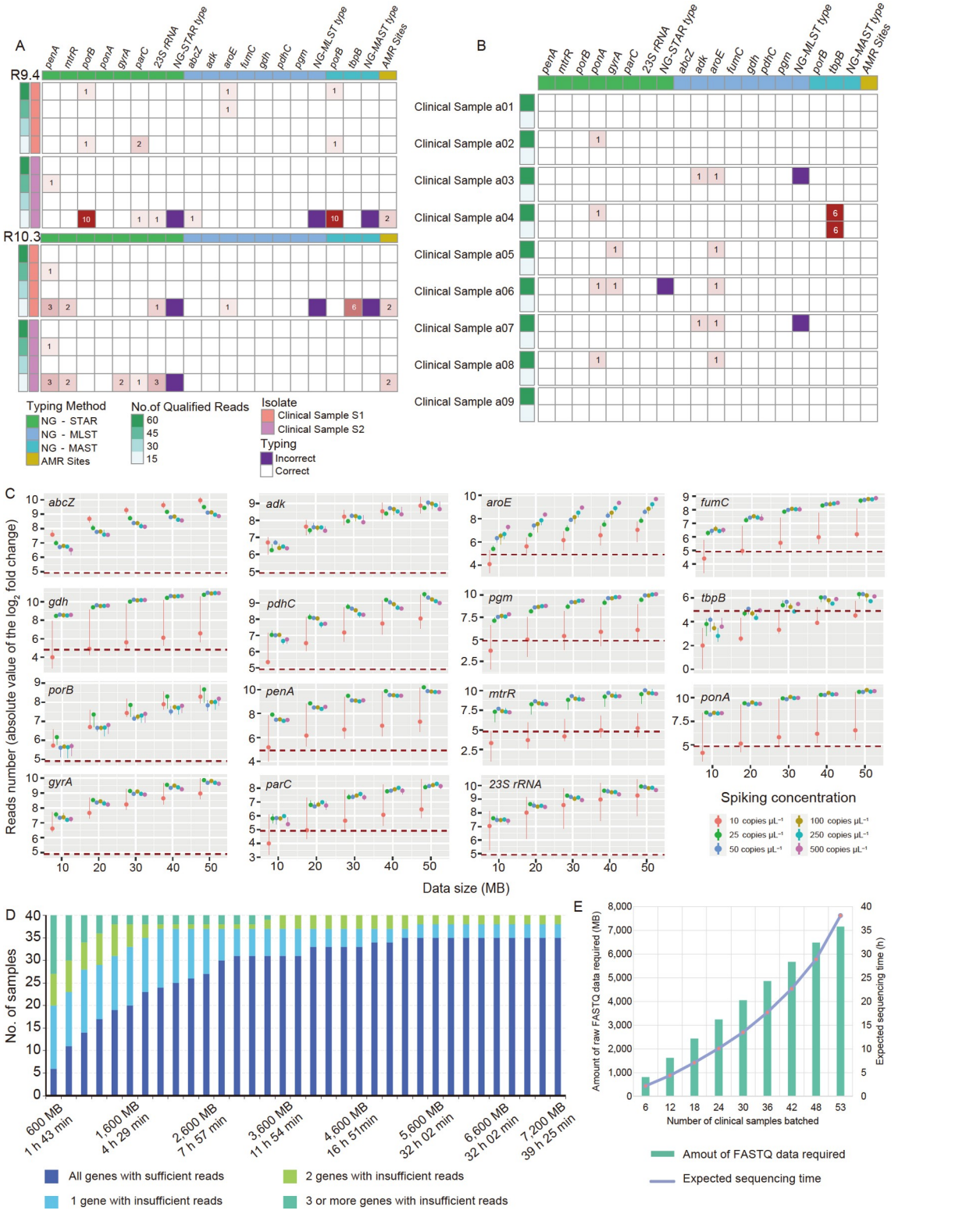

Figure 1. Detection performance and validation of the method. A, Heatmaps depicting the relationship between depths and accuracy of the consensus sequences using analytes of clinical samples. Raw data were base-called using the hac model. Numbers in the cells indicate the number of mismatched bases, with darker cells indicating more mismatches. Cells without numbers indicate full agreement between consensus sequences and confirmed sequences. B, Heatmap of the results of accuracy assessment using nine clinical samples. Raw sequencing data were base-called using the hac model. Consensus sequences were constructed from refined subsets with 60 qualified reads. C, Relationship between the number of qualified reads and spike concentrations. Mock specimens were tested in triplicate with each spike concentration. Circles represent the medians of the three replicates, and error bars indicate the maximum and minimum values of the three replicates. For each replicate, 10, 20, 30, 40, and 50 MB clean FASTQ files were extracted for downstream analyses. The dotted red line indicates 30 qualified reads, ensuring correct typing and identification of AMR-related sites. D, Dynamic relationship between the number of eligible samples and the amount of output data. E, Amount of raw FASTQ format data required and the corresponding suggested sequencing time when a certain number of clinical samples are batched within a run.

https://doi.org/10.1007/s11427-022-2382-0

but the difference was not significant (0.36 versus 1.39 mismatches, P=0.062; Figure S5 in Supporting Information). This distinction is mostly due to the existence of colonizing microbes in the urogenital tract that share genomic similarities with *N. gonorrhoeae*, resulting in non-specific amplification. Susceptibilities to ceftriaxone (CRO), azithromycin (AZM), and ciprofloxacin (CIP) were predicted for each sample. The algorithm and results of prediction are listed in Table S2 in Supporting Information.

Low bacterial loads will aggravate uneven amplification, which requires more data to ensure that each target gene has sufficient reads for downstream analysis. We explored the relationship between input genome copies and the number of qualified reads of each target gene, thereby estimating the appropriate input genome copies and the minimum amount of data needed. For each gene, a minimum of 30 qualified reads were deemed sufficient, as this amount can enable correct typing and identification of AMR-related mutations. The results demonstrated for input genome copies larger than 200, 30–40 MB clean FASTQ files were sufficient. Compared with other target genes, the amplification of *tbpB* was suboptimal (Figure 1C; Figure S6 in Supporting Information). To provide strong and trustworthy results, Deeplex Myc-TB, a commercial kit based on amplicon sequencing for *in vitro* diagnosis of *Mycobacterium tuberculosis*, demands a minimum of 1,000 input genomes. Our technique outperformed Deeplex Myc-TB in terms of analytical sensitivity. Furthermore, according to studies on gonococcal bacterial load, our method was sensitive enough to be applied to most samples of gonococcal infections (Bissessor et al., 2011; Priest et al., 2017).

A panel of *Neisseria* spp. strains was tested to assess the cross-reactivity. For each strain, 100 MB clean FASTQ files were analyzed, and the numbers of reads mapped to *porA* were used to evaluate the specificity. For *N. gonorrhoeae* strains, the numbers of reads mapped to *porA* were close to other target genes, with the variances within the same order of magnitude. For nongonococcal *Neisseria* spp. species, such variances were considerable, and the number of reads mapped to *porA* was less than ten (Figure S7 in Supporting Information).

By describing the dynamic relationship between the number of eligible samples and the amount of output data, we evaluated the appropriate batch size per run and the expected sequencing time. If a sample had more than 30 qualified reads of each target gene, it was defined as eligible. The run testing 40 clinical samples yielded 7,200 MB raw FASTQ data in 39 h 25 min, and 35 samples were eligible with 5,400 MB raw data (an average of 135 MB per clinical sample; Figure 1D). Subsequently, the 1,800 MB yielded raw data that can enable the sequencing of ~13 more clinical samples (Figure 1E). For isolate samples, data output was considerably quick, and an isolate can be sequenced with less raw data (an average of 90 MB per isolate). Similarly, up to 80 isolates can be sequenced in one flowcell to make complete use of the data output (Figure S8 in Supporting Information). The sequencing data for each eligible sample was examined using the abovementioned bioinformatics pipeline, and the molecular types of MLST, NG-STAR, and NG-MAST were successfully obtained (Tables S3 and S4 in Supporting Information).

In conclusion, this method outperformed similar methods based on targeted-amplicon sequencing in terms of analytical sensitivity. The results of the validation demonstrated its robustness, outstanding batching capability, and ideal detection performance. The bioinformatics pipeline can reduce the large error of homopolymer calling, allowing exact typing loci recovery not only for MLST with conserved housekeeping gene loci but also for NG-STAR and NG-MAST with highly polymorphic loci. The continued reduction in sequencing costs of this method shows promise in tackling the issues associated with resistance detection.

Compliance and ethics

The authors declare that they have no conflict of interest. This study was performed in accordance with the recommendations of the national ethics regulations and approved by the Institutional Review Board of the Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College (IRB reference number: IPB-2020-12). Clinical samples were collected from the Shenzhen Centre for Chronic Disease Control, and written informed consent was obtained from all participants.

Acknowledgement

This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS) (2021-I2M-1-037, 2021-I2M-1-038), the National Natural Science Foundation of China (82002193), the National Science and Technology Infrastructure of China (National Pathogen Resource Center-NPRC-32), the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2019-PT310-029, 2021-PT310-004) and the Fundamental Research Funds for the Central Universities (3332021092). We thank Siyu Zhou for her support with the bioinformatics analysis and script design.

Supporting information

The supporting information is available online at https://doi.org/10.1007/s11427-022-2382-0. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

Antimicrobial Resistance Collaborators (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399, 629–655.

Bissessor, M., Tabrizi, S.N., Fairley, C.K., Danielewski, J., Whitton, B., Bird, S., Garland, S., and Chen, M.Y. (2011). Differing *Neisseria gonorrhoeae* bacterial loads in the pharynx and rectum in men who have sex with men: implications for gonococcal detection, transmission, and control. J Clin Microbiol 49, 4304–4306.

Cabibbe, A.M., Spitaleri, A., Battaglia, S., Colman, R.E., Suresh, A., Uplekar, S., Rodwell, T.C., and Cirillo, D. M. (2020). Application of targeted next-generation sequencing assay on a portable sequencing platform for culture-free detection of drug-resistant tuberculosis from clinical samples. J Clin Microbiol 58, e00632-20.

Demczuk, W., Sidhu, S., Unemo, M., Whiley, D.M., Allen, V.G., Dillon, J.R., Cole, M., Seah, C., Trembizki, E., Trees, D.L., et al. (2017). *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.

Liu, F., Zhang, Y., Zhang, L., Li, Z., Fang, Q., Gao, R., and Zhang, Z. (2019). Systematic comparative analysis of single-nucleotide variant detection methods from single-cell RNA sequencing data. Genome Biol 20, 242.

Potter, R.F., Ransom, E.M., and Burnham, C.A.D. (2021). The next-generation of *Neisseria gonorrhoeae* antimicrobial resistance testing. Clin Chem 67, 573–575.

Priest, D., Ong, J.J., Chow, E.P.F., Tabrizi, S., Phillips, S., Bissessor, M., Fairley, C.K., Bradshaw, C.S., Read, T. R.H., Garland, S., et al. (2017). *Neisseria gonorrhoeae* DNA bacterial load in men with symptomatic and asymptomatic gonococcal urethritis. Sex Transm Infect 93, 478–481.

Sereika, M., Kirkegaard, R.H., Karst, S.M., Michaelsen, T.Y., Sørensen, E.A., Wollenberg, R.D., and Albertsen, M. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. Nat Methods 19, 823–826.

Yahara, K., Ma, K.C., Mortimer, T.D., Shimuta, K., Nakayama, S., Hirabayashi, A., Suzuki, M., Jinnai, M., Ohya, H., Kuroki, T., et al. (2021). Emergence and evolution of antimicrobial resistance genes and mutations in *Neisseria gonorrhoeae*. Genome Med 13, 51.

Zhang, C., Xiu, L., Li, Y., Sun, L., Li, Y., Zeng, Y., Wang, F., and Peng, J. (2021). Multiplex PCR and Nanopore sequencing of genes associated with antimicrobial resistance in *Neisseria gonorrhoeae* directly from clinical samples. Clin Chem 67, 610–620.