



OPEN ACCESS

EDITED BY

Digvijay Verma,
Babasaheb Bhimrao Ambedkar University,
India

REVIEWED BY

Xiaogang Xu,
Fudan University, China
Ahmed M. Moustafa,
University of Pennsylvania, United States

*CORRESPONDENCE

Xin Lu
✉ luxin@icdc.cn
Zhenpeng Li
✉ lizhenpeng@icdc.cn
Biao Kan
✉ kanbiao@icdc.cn

RECEIVED 05 March 2023

ACCEPTED 27 April 2023

PUBLISHED 15 May 2023

CITATION

Zhao W, Zeng W, Pang B, Luo M, Peng Y, Xu J,
Kan B, Li Z and Lu X (2023) Oxford nanopore
long-read sequencing enables the generation
of complete bacterial and plasmid genomes
without short-read sequencing.
Front. Microbiol. 14:1179966.
doi: 10.3389/fmicb.2023.1179966

COPYRIGHT

© 2023 Zhao, Zeng, Pang, Luo, Peng, Xu, Kan,
Li and Lu. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Oxford nanopore long-read sequencing enables the generation of complete bacterial and plasmid genomes without short-read sequencing

Wenxuan Zhao^{1,2}, Wei Zeng^{1,3}, Bo Pang¹, Ming Luo⁴, Yao Peng¹,
Jialiang Xu⁵, Biao Kan^{1,3*}, Zhenpeng Li^{4*} and Xin Lu^{1*}

¹National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, ²Department of Health Statistics, School of Public Health, Shanxi Medical University, Jinzhong, Shanxi, China, ³School of Public Health, Shandong University, Jinan, China, ⁴Yulin Center for Disease Control and Prevention, Yulin, Shanxi, China, ⁵School of Food and Chemical Engineering, Beijing Technology and Business University, Beijing, China

Introduction: Genome-based analysis is crucial in monitoring antibiotic-resistant bacteria (ARB) and antibiotic-resistance genes (ARGs). Short-read sequencing is typically used to obtain incomplete draft genomes, while long-read sequencing can obtain genomes of multidrug resistance (MDR) plasmids and track the transmission of plasmid-borne antimicrobial resistance genes in bacteria. However, long-read sequencing suffers from low-accuracy base calling, and short-read sequencing is often required to improve genome accuracy. This increases costs and turnaround time.

Methods: In this study, a novel ONT sequencing method is described, which uses the latest ONT chemistry with improved accuracy to assemble genomes of MDR strains and plasmids from long-read sequencing data only. Three strains of *Salmonella* carrying MDR plasmids were sequenced using the ONT SQK-LSK114 kit with flow cell R10.4.1, and de novo genome assembly was performed with average read accuracy ($Q > 10$) of 98.9%.

Results and Discussion: For a 5-Mb-long bacterial genome, finished genome sequences with accuracy of $>99.99\%$ could be obtained at 75x sequencing coverage depth using Flye and Medaka software. Thus, this new ONT method greatly improves base-calling accuracy, allowing for the de novo assembly of high-quality finished bacterial or plasmid genomes without the need for short-read sequencing. This saves both money and time and supports the application of ONT data in critical genome-based epidemiological analyses. The novel ONT approach described in this study can take the place of traditional combination genome assembly based on short- and long-read sequencing, enabling pangenomic analyses based on high-quality complete bacterial and plasmid genomes to monitor the spread of antibiotic-resistant bacteria and antibiotic resistance genes.

KEYWORDS

ONT sequencing, accuracy, complete genome, bacteria, plasmid

1. Introduction

Microbial resistance to drugs has become a global issue of widespread concern (Nathan, 2020). The widespread emergence of antibiotic resistance, especially multidrug resistance (MDR), among bacterial strains that cause infections presents difficulties in clinical treatment. Acquired resistance spreads very rapidly compared with chromosomal mutations (Paterson and van Duin, 2017; Bengtsson-Palme et al., 2018). Plasmids are the most common vectors for horizontal gene transfer (San Millan, 2018; Vit et al., 2020). In the concept of One Health, it is crucial to investigate MDR plasmids, because their domain organization is critical to the spread of antimicrobial resistance genes (ARGs) among bacteria (Aslam et al., 2021). Therefore, to track the transmission of ARGs, accurate information on MDR plasmid genomes is essential (Bennett, 2008; Malhotra-Kumar et al., 2016; Jordt et al., 2020).

MDR mostly originates from the accumulation of resistance genes on plasmids (Nikaido, 2009), though resistance genes can also be carried on the chromosome. Identifying these genes and their accurate genomic localization using short-read sequencing data can be difficult (Partridge et al., 2009). To associate independent data with ARGs transmission events, pangenome clustering based on complete plasmid genomes can be applied to surveillance (Li et al., 2022). Short-read sequencing [such as from Illumina and MGI next-generation sequencing (NGS) technologies] has high base-calling accuracy, but the nature of the short reads means that only fragmented draft genomes can be obtained from such data. Instead, scientists would prefer to receive correct, complete genomes as their study advances (Cohen et al., 2019; Kathirvel et al., 2021). Shortly after the introduction of NGS, third-generation sequencing technologies (TGS) emerged, presented by two platforms, Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), giving long and ultra-long sequencing reads, respectively, and enabling coverage of highly repetitive regions and structural variants. PacBio developed the first established single-molecule real-time sequencing technology in 2011 (Athanasopoulou et al., 2021). In 2014, high-throughput, long-read sequencing was made possible by ONT on a portable device MinION (Lu et al., 2016). However, compared with NGS, low base-call accuracy has limited the reliability of ONT data for critical genomic epidemiology tasks (Petersen et al., 2019; Foster-Nyarko et al., 2023). The widely accepted remedy was to use short-read NGS data for error correction of long-read sequence data (Senol Cali et al., 2019; Smith et al., 2020).

In this study, our main objectives were to produce a high-quality finished genome only by *de novo* assembly based on long-read sequencing and to offer solid evidence for resistance gene analysis, supporting its use in the genome-based epidemiological analyses. This approach offered a more efficient and cost-effective replacement for traditional methods that required both long- and short-read sequencing. First, the reference genome of the strains was generated. Then, the sequencing data were obtained using the new ONT SQK-LSK114 kit with flow cell R10.4.1, and use of long-read sequencing data only for *de novo* genome assembly of these strains, combined with the error correction of the data itself. Finally, the accuracy of these genomes was verified against the reference data. To evaluate the accuracy of our method, we assessed the single nucleotide variations (SNVs), insertions (INs), and deletions

(DELs), which are common *de novo* assembly errors (Boostrom et al., 2022). To compare the new sequencing method with the earlier sequencing methods, all the samples were sequenced using the SQK-LSK110 kit and the R9.4 flow cell. In comparison with the previous version (Sereika et al., 2022), sequencing quality was substantially improved in the latest ONT chemistry and has potential implications for monitoring the spread of antibiotic-resistant bacteria and antibiotic-resistant genes.

2. Materials and methods

2.1. Samples

Salmonella strains were collected from the surveillance of healthy people, and we constructed a strain bank from these strains. Antimicrobial susceptibility testing was conducted and interpreted using the broth microdilution method recommended by the Clinical and Laboratory Standards Institute. Three MDR *Salmonella* strains were recovered from the strain bank and isolated with *Salmonella* agar (CHROMagar Company, Paris, France); strains were identified using the Vitek-II system (bioMérieux, Lyon, France). The genomic DNA from each strains was extracted by boiling and freeze-thawing processes, and the resulting supernatant was recovered for use as the PCR template (Doyle et al., 2012; Ding et al., 2020; Fan et al., 2022).

2.2. Sequencing

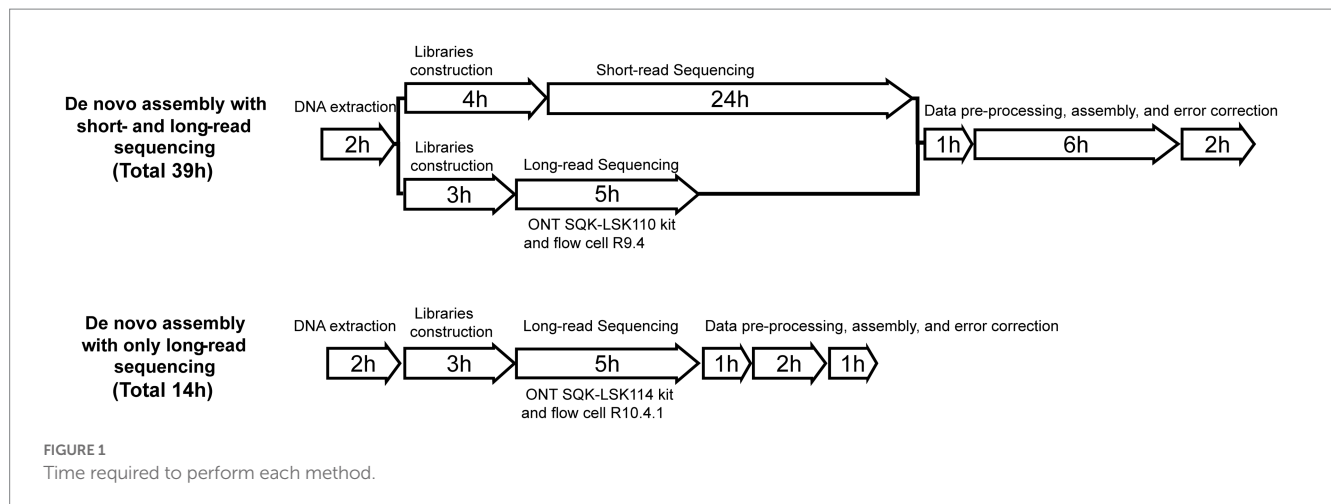
Genomic DNA was extracted using the TIANamp Bacteria DNA Kit (TIANGEN, China) and quantified using a Qubit V4 Fluorometer (Thermo Fisher, United States). Sequencing libraries were prepared using Ligation Sequencing Kit V14 (SQK-LSK114, ONT) and sequenced using R10.4.1 flow cells (FLO-MIN114, ONT) on a GridION device (ONT) with MinKNOW v22.08.9 and super-accuracy base-calling mode selected. Other parameters were kept at their defaults. To compare the new sequencing method used in our study with the earlier sequencing methods, we sequenced all the samples using the SQK-LSK110 kit and the R9.4 flow cell.

NGS libraries were constructed using the MGIEasy FS DNA Library Prep Set (MGI, China) and sequenced on the MGISEQ-200RS sequencing platform (MGI).

2.3. Data analysis

Guppy v6.2.11 (Wick et al., 2019) was used to extract the bases from the downloaded fast5 data and turn them into standard fastq files. NanoPlot v1.20.0 was used to assess the level of sequencing quality (De Coster et al., 2018). NanoFilt v2.8 (De Coster et al., 2018) was used to remove sequences that were <1,000 bp long with quality value < 10. In addition, 50 bp were removed from the front and back ends of each clean data record.

It has been widely used in earlier studies and proved to be the most accurate method to assemble reference genomes utilizing short- and long-read sequencing (Petersen et al., 2019; Boostrom et al., 2022; Sereika et al., 2022). In our study, we utilized this



method to generate the reference genomes. Long-read data with 500× depth and short read data with 500× depth were used, and were assembled using the Unicycler hybrid assembler v0.4.8 (Wick et al., 2017). Pilon v1.24 (Walker et al., 2014) was used to polish these genomes. The obtained genomes were used as references for the following analyses.

Raw data packets were generated periodically during nanopore sequencing, typically at 6-min intervals (approximately 4,000 reads). The depth of coverage was based on packet size. After the clean data were generated, it was divided into depths: 1× (6 min), 5× (12 min), 10× (25 min), 20× (40 min), 30× (75 min), 50× (105 min), 75× (175 min), 100× (265 min), 150× (360 min), 200× (460 min), 250× (560 min), 300× (660 min), 350× (760 min), 400× (850 min), 450× (960 min), and 500× (1,050 min). The final depth of coverage was estimated based on the actual data size. The goal was to determine the saturation sequencing time point and depth.

Flye has proven to be the most effective tool for *de novo* genetic assembly (Boostrom et al., 2022). Therefore, the default parameters of Flye v2.8.2 (Kolmogorov et al., 2019) were used for *de novo* assembly, and QUASt v5.2.0 (Gurevich et al., 2013) was used to evaluate the quality of genome assembly. Finally, errors were corrected by applying Medaka v1.2.2 three times.¹ BLAST v2.11.0 was utilized to determine the identity of the fastq file in comparison to the reference genome. Snippy v4.4.5 was used to compare the assembled fasta file to the reference genome and obtain the number of SNPs, insertions, and deletions.² ARG genes were aligned using the ResFinder database. FastANI v1.33 (Jain et al., 2018) was used to calculate genome-wide average nucleotide identity between genomes. R v4.1.0 and BRIG v0.95 (Alikhan et al., 2011) were used to visualize the outcomes (Figure 1). For detailed usage instructions, please refer to the Supplementary material.

The complete sequences of all three strains have been deposited in the GenBank database, Their BioProject ID is PRJNA937772.

3. Results

3.1. Characteristics of sequences

In total, we used three MDR *Salmonella* strains (SA15303, SA14318, and SA17155) that have been shown by PCR to carry the *mcr*, *ndm*, and *tet* genes, respectively. After hybrid genome assembly from long and short reads, we used short reads for multiple rounds of polishing. The aim was to obtain a high-quality reference genome for each strain. The complete genomes for strains SA15303, SA14318, and SA17155 were 5.1, 4.7, and 5.2 Mb long, respectively. The plasmids of these strains were 29, 199, and 287 kb in size, and contained 4, 9, and 12 ARGs, respectively. Thus, the plasmid size as well as the resistance genes varied widely between the strains. There were more homopolymer regions on the larger plasmids (Table 1).

In experiments to test genome generation using only long-read sequencing data, we used a combination of the latest ONT kit v14 (SQK-LSK114) and flow cell R10.4.1 to obtain the raw data. After sequencing for 18 h, each flow cell generated about 500× data, with 39%–42% of the reads having lengths > 5 kb. Multiple reads of >60,000 bp were obtained (Figure 2A). Reads with quality value Q10 exceeded 85%, while those with Q20 exceeded 40% (Table 2).

We used an identical procedure to sequence the genomes of the three strains. The read lengths and Q-score distribution of the three sets of sequence data were similar (Figure 2A). The average length was 5 kb, and the average Q value was 16.70. We considered a Q value < 10 to represent low-quality reads and removed these (approximately 14% of the whole sequencing data) before assembly. When compared with the reference genome sequences, the average read accuracy was 98.9% (Figure 2B), with the lower quartile > 95%.

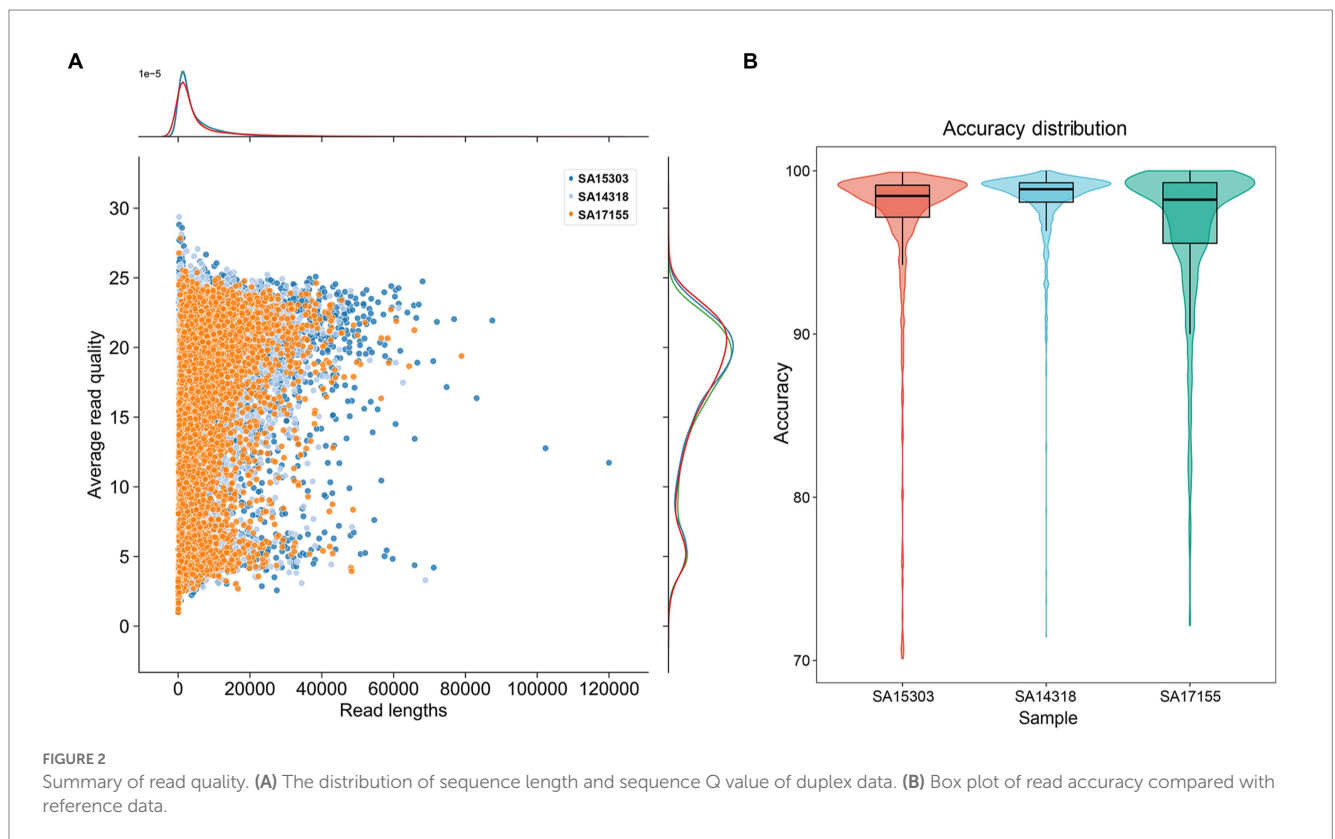
To compare the new sequencing method with the earlier sequencing methods, all the samples were sequenced using the SQK-LSK110 kit and the R9.4 flow cell. It was important to note that the new sequencing technique shown considerable increases in both quality and accuracy. More specifically, accuracy increased from 92% to 98% and Q20 increased from 13% to 42% (Table 2).

1 <https://github.com/nanoporetech/medaka>

2 <https://github.com/tseemann/snippy>

TABLE 1 Details of the strains used in this study and their plasmids.

	SA15303	SA14318	SA17155
Total size	5,144,932 bp	4,742,137 bp	5,240,120 bp
Plasmid size	286,565 bp	29,276 bp	199,024 bp
Plasmid type	IncHI2A	IncX1	IncHI2
Number of resistance genes on the plasmid	9	4	12
Number of homopolymer regions on the plasmid	131	7	94
MDR profiles	Chloramphenicol-Gentamicin-Tetracycline-Trimethoprim-Sulfamethoxazole	Aztreonam-Cefazolin-Ceftazidime-Ceftriaxone-Chloramphenicol-Ciprofloxacin-rtapenem-Meropenem-Imipenem-Tetracycline-Sulfamethoxazole	Cefazolin-Ceftazidime-Ceftriaxone-Gentamicin Chloramphenicol-Ciprofloxacin-rtapenem-Meropenem-Imipenem-Tetracycline-Sulfamethoxazole



3.2. Genome assembly and error correction

After removing the low-quality reads, *de novo* assembly was performed using Flye to obtain preliminary results. Medaka was then run three times to correct errors. About 10× coverage (100 Megabyte, 25 min) was able to obtain complete plasmid sequences for all three strains. However, to obtain the complete sequence of both chromosomes and plasmids, coverage needs to be increased to 30× (300 Megabyte, approximately 75 min; Table 3).

Compared with the reference genome sequences, the errors generated can be classified as DELs, INS, and SNVs. SNVs represented

the largest number of errors (48%), followed by DELs (26%). When the depth of the sequencing data reached 75× (750 Megabyte, 175 min), the assembly error rate was stable, and we obtained satisfactory sequences (Table 3). After assembly, the error-correction capabilities of Medaka were clearly visible; with DEL errors being the majority. Among them, 83.33% of DEL errors were eliminated, and, subsequently, 68.42% of SNV errors were corrected (Table 4). Medaka tends to fix DEL errors, which were more likely to occur in *de novo* assembly. The initial error correction impact of Medaka was notable, whereas the second and third error correction effects had no appreciable improvement (Supplementary Table 1).

TABLE 2 Summary of sequencing data.

		SA15303	SA14318	SA17155
Sample concentration (ng/μL)		97.2	95.2	96.2
SQK-LSK114 kit and flow cell R10.4.1	Total file size (gigabytes)	5.03	4.73	4.95
	Depth	508×	497×	519×
	Total reads	361,153	309,825	318,461
	Reads > 5 kbp (%)	42.06	39.34	46.15
	Average length of reads (kbp)	5.83	4.98	5.18
	Percentage of Q10(%)	88.27	85.58	87.2
	Percentage of Q20(%)	52.67	42.31	46.25
	Average accuracy of reads (%)	98.94	98.88	98.83
	Alignment error rate (%)	2.19	2.11	1.99
SQK-LSK110 kit and flow cell R9.4	Total file size (gigabytes)	5.05	4.87	4.89
	Depth	512×	512×	511×
	Total reads	656,556	457,263	504,270
	Reads > 5kbp (%)	28.53	33.15	37.22
	Average length of reads (kbp)	3.96	4.59	4.14
	Percentage of Q10(%)	89.74	85.26	82.52
	Percentage of Q20(%)	13.65	14.22	18.09
	Average accuracy of reads (%)	94.02	92.53	93.84
	Alignment error rate (%)	5.79	7.18	5.11

TABLE 3 The number of errors in chromosome and plasmid sequences.

		Running time (min)	25	75	175	1,050
SA15303	Chromosome	Coverage depth	9×	31×	75×	477×
		Whether cyclic	N	Y	Y	Y
		Size (bp)	1,199,135	4,848,505	4,848,501	4,848,495
		Errors	532	10	9	5
	Plasmid	Coverage depth	11×	41×	94×	594×
		Whether cyclic	Y	Y	Y	Y
		Size (bp)	286,566	286,565	286,567	286,564
		Errors	15	2	2	3
SA14318	Chromosome	Coverage depth	13×	39×	81×	450×
		Whether cyclic	N	Y	Y	Y
		Size (bp)	3,517,587	4,712,839	4,712,855	4,712,849
		Errors	123	11	8	9
	Plasmid	Coverage depth	6×	22×	55×	306×
		Whether cyclic	Y	Y	Y	Y
		Size (bp)	29,281	29,266	29,276	29,276
		Errors	19	0	0	0
SA17155	Chromosome	Coverage depth	11×	31×	71×	466×
		Whether cyclic	N	Y	Y	Y
		Size (bp)	3,256,928	5,031,756	5,030,613	5,030,601
		Errors	535	12	9	8
	Plasmid	Coverage depth	9×	23×	51×	358×
		Whether cyclic	Y	Y	Y	Y
		Size (bp)	199,056	199,024	199,022	199,023
		Errors	50	4	2	3

At a depth of sequencing data of 500×, the genomes of the three strains still had, respectively, 8, 9, and 11 errors that could not be corrected when compared with the reference genome sequences. Among SNV errors, G was frequently misidentified as T (8/9) and C was frequently misidentified as A (4/5). Duplication of the same base was the cause of insertion and deletion errors that could not be corrected. On chromosomes, 88.89% of the errors were found in homopolymer regions (Supplementary Table 2). Likewise, on plasmids, all INS and DEL errors were found in homopolymer regions (Figure 3). The reason for the above faults may be that Medaka is not adapted to the latest ONT model; thus, sequencing correction may improve with future software updates.

When the depth of sequencing data reached 75× coverage, the quality of the assembled sequences was near optimal. The average nucleotide identity was >99.9975% between each assembled sequence and the reference genome (Figure 4).

TABLE 4 Errors before and after correction.

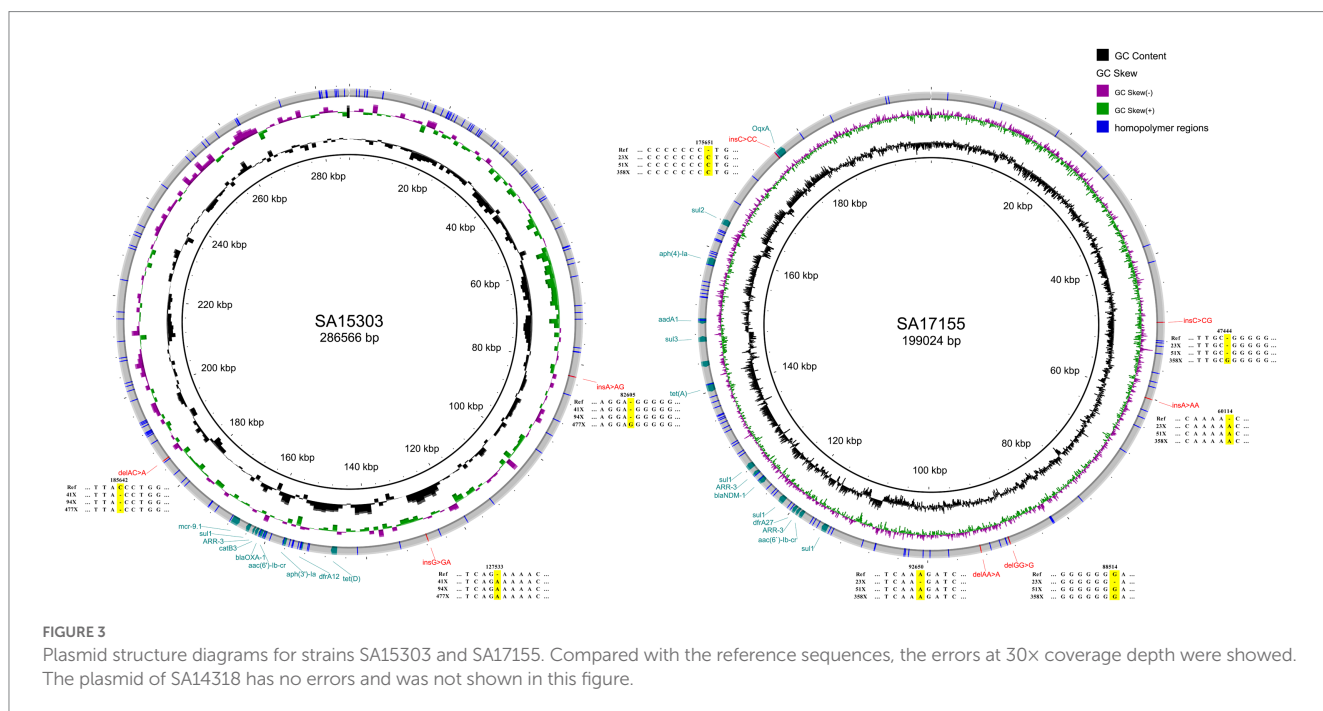
	Type of variant	Before correction	After correction
SA15303	Variant-DEL	10	2
	Variant-INS	3	7
	Variant-SNP	3	2
SA14318	Variant-DEL	8	0
	Variant-INS	2	2
	Variant-SNP	7	5
SA17155	Variant-DEL	6	1
	Variant-INS	1	4
	Variant-SNP	9	6

Compared with sequences at 75× coverage depth, chromosomal sequences at 500× coverage depth had an average error reduction of 1.67, whereas plasmids had an average error increase of 0.67. Thus, although the 500× sequencing required a further 875 min and generated 6.5-fold more data than the 75× sequencing, the effects were not noticeably improved.

The latest ONT kit SQK-LSK114 and flow cell R10.4.1 have dramatically reduced errors in the assembled genomes. For data coverage that exceeded 75× (750 Megabyte, 160 min), using only Nanopore sequencing data in *de novo* genome packing, complete sequences were achieved that nearly matched the accuracy of NGS without requiring short-read data.

4. Discussion

The spread of ARGs carried by plasmids is a major public health issue (San Millan, 2018). As mobile genetic elements that can carry ARGs and be transferred easily between different bacterial species, plasmids enable the quick and effective dissemination of ARGs (Vit et al., 2020). The emergence of MDR and the spread of drug resistance between bacterial strains can both be aided by plasmids (Paterson and van Duin, 2017). It is critical to comprehend how bacteria acquire and spread resistance to antibiotics, in addition to the molecular mechanisms of this phenomenon (Jordt et al., 2020). High-quality finished genomes of drug-resistant bacteria are required to monitor the transmission of antimicrobial resistance. The current tools for ARG detection are known to be highly accurate when used with short-read sequencing data. However, short-read sequencing cannot provide accurate localization of ARGs, which can be carried by both bacterial plasmids and chromosomes. More frequent cross-host resistance epidemic events initiated by plasmids carrying ARGs have occurred (Li et al., 2022), suggesting the greater potential for ARG transmission through plasmids in humans, food, animals, and the environment. To



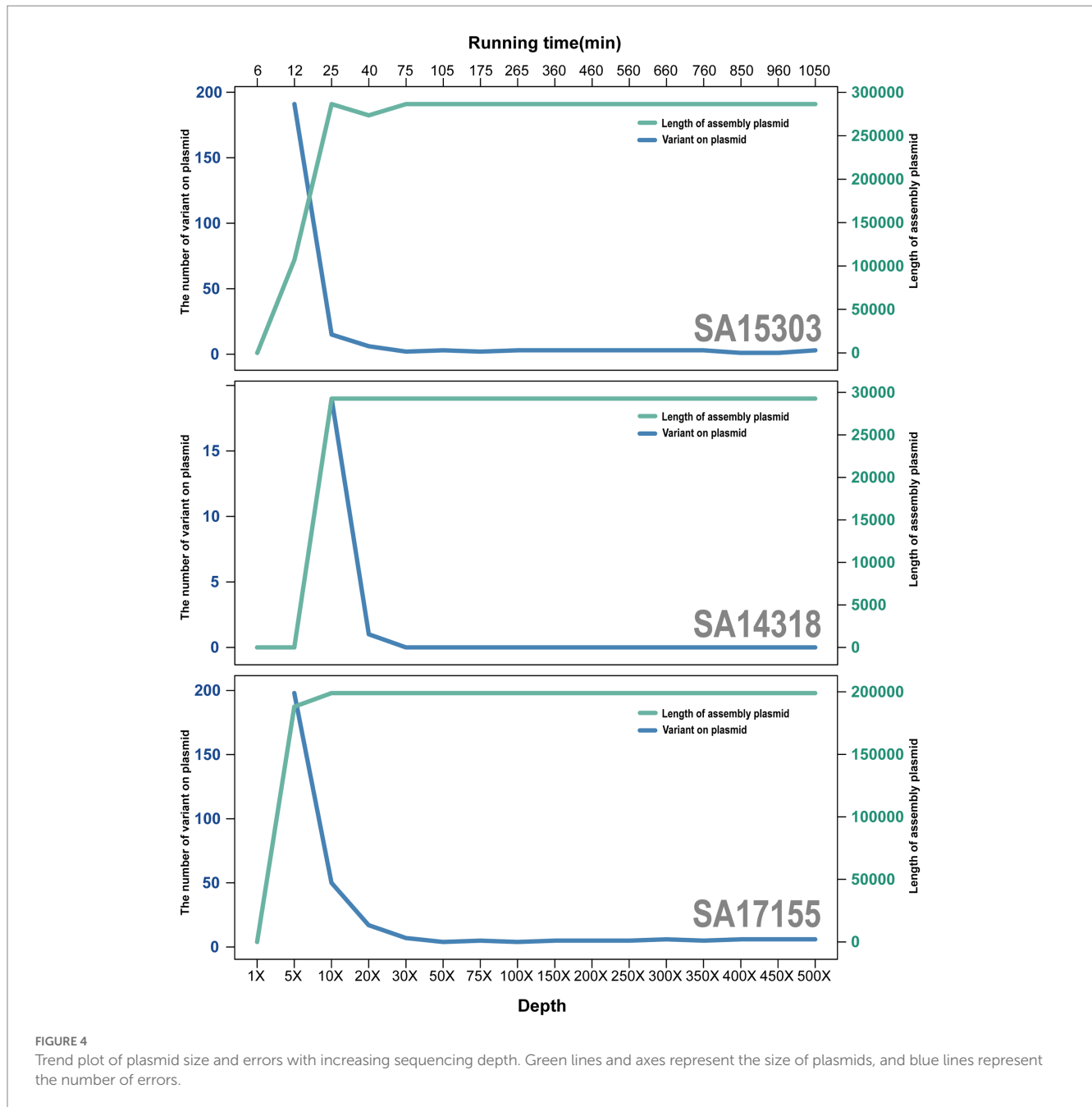


FIGURE 4
Trend plot of plasmid size and errors with increasing sequencing depth. Green lines and axes represent the size of plasmids, and blue lines represent the number of errors.

obtain high-quality finished genomes, it has been necessary to include short-read polishing of long-read sequencing data to correct errors (Senol Cali et al., 2019). This combined approach required two sequencing libraries and was time-consuming, difficult to perform, and can entail high costs. The optimal solution is to improve the accuracy of long-read sequencing. Based on long-read sequencing, one can locate ARGs on chromosomes or plasmids. Horizontal transfer of plasmids plays an important role in the spread of multidrug-resistant bacteria, identifying plasmid-borne resistance genes is necessary to estimate the spread of resistance among bacteria. The coexistence of ARGs is very common, and poses significant public health and food safety threats. Obtaining complete plasmid sequences is the most effective means of detecting ARGs and the coexistence of ARGs on the same plasmid.

In this study, the latest ONT SQK-LSK114 kit with flow cell R10.4.1 was used, and the results demonstrated that a high-quality finished genome could be obtained using *de novo* assembly. Compared to the previous process, the Q-score of reads was significantly improved (Smith et al., 2020). In previous studies, the limitation of ONT sequencing was the relatively high error rate, which in some cases can reach 10% (Khrenova et al., 2022). Low base-call accuracy has limited the reliability of ONT data for critical genomic epidemiology tasks such as ARG and virulence gene detection and typing, serotype prediction, and cluster identification (Foster-Nyarko et al., 2023). In this study, the average read accuracy was 98.9%, indicating that the novel approach has the potential to greatly improve read accuracy. When using the latest ONT SQK-LSK114 kit and R10.4.1 flow cell for long-read sequencing, error correction using

short-read sequence data is not required. The novel ONT sequencing method saves money and time. Both the preparation of short- and long-read libraries as well as sequencing, which take at least 1 day, are unnecessary.

Regarding the new ONT sequencing method, once low-quality reads had been removed, the accuracy of the assembled sequences was nearly identical to that based on NGS. However, homopolymer regions increased the possibility of INS and DEL errors during assembly, and biased SNV errors still needed to be improved. In the future, the analysis pipeline may become more streamlined and effective.

In conclusion, use of the ONT LSK114 kit combined with flow cell R10.4.1 improved the accuracy of *de novo* genome assembly. When sequencing bacteria and plasmid, high-quality complete genomes with ideal coverage and identity were obtained without short-read or reference polishing. To obtain saturated raw data, we recommend acquiring a depth of 100× raw data (1 Gigabytes, 265 min), and performing error correction three times after assembly. This method will save time in obtaining high-quality complete genomes of bacteria for antimicrobial resistance surveillance, and is likely to become a valuable tool for monitoring the transmission of plasmid-borne drug resistance genes (Peter et al., 2020).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

XL, ZL, and BK conceived and designed this study. BP, ML, YP, and JX contributed to the experiment. WZh and WZe contributed to

writing the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program of China (2020YFE0205700 and 2022YFC2303900), the Major Projects of the National Natural Science Foundation of China (22193064), and the Science Foundation (2022SKLID303) of the State Key Laboratory of Infectious Disease Prevention and Control, China.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1179966/full#supplementary-material>

References

- Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402
- Aslam, B., Khurshid, M., Arshad, M. I., Muzammil, S., Rasool, M., Yasmeen, N., et al. (2021). Antibiotic resistance: one health one world outlook. *Front. Cell. Infect. Microbiol.* 11:771510. doi: 10.3389/fcimb.2021.771510
- Athanasopoulou, K., Boti, M. A., Adamopoulos, P. G., Skourou, P. C., and Scorilas, A. (2021). Third-generation sequencing: the spearhead towards the radical transformation of modern genomics. *Life* 12:30. doi: 10.3390/life12010030
- Bengtsson-Palme, J., Kristiansson, E., and Larsson, D. G. J. (2018). Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiol. Rev.* 42:fux053. doi: 10.1093/femsre/fux053
- Bennett, P. M. (2008). Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria: plasmid-encoded antibiotic resistance. *Br. J. Pharmacol.* 153, S347–S357. doi: 10.1038/sj.bjp.0707607
- Boostrom, I., Portal, E. A. R., Spiller, O. B., Walsh, T. R., and Sands, K. (2022). Comparing long-read assemblers to explore the potential of a sustainable low-cost, low-infrastructure approach to sequence antimicrobial resistant Bacteria with Oxford Nanopore sequencing. *Front. Microbiol.* 13:796465. doi: 10.3389/fmicb.2022.796465
- Cohen, K. A., Manson, A. L., Desjardins, C. A., Abeel, T., and Earl, A. M. (2019). Deciphering drug resistance in *Mycobacterium tuberculosis* using whole-genome sequencing: progress, promise, and challenges. *Genome Med.* 11:45. doi: 10.1186/s13073-019-0660-8
- de Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. doi: 10.1093/bioinformatics/bty149
- Ding, Y., Saw, W. Y., Tan, L. W. L., Moong, D. K. N., Nagarajan, N., Teo, Y. Y., et al. (2020). Emergence of tigecycline- and eravacycline-resistant Tet(X4)-producing Enterobacteriaceae in the gut microbiota of healthy Singaporeans. *J. Antimicrob. Chemother.* 75, 3480–3484. doi: 10.1093/jac/dkaa372
- Doyle, D., Peirano, G., Lascols, C., Lloyd, T., Church, D. L., and Pitout, J. D. D. (2012). Laboratory detection of Enterobacteriaceae that produce carbapenemases. *J. Clin. Microbiol.* 50, 3877–3880. doi: 10.1128/JCM.02117-12
- Fan, J., Cai, H., Fang, Y., He, J., Zhang, L., Xu, Q., et al. (2022). Molecular genetic characteristics of plasmid-borne *mcr-9* in *Salmonella enterica* serotype Typhimurium and Thompson in Zhejiang, China. *Front. Microbiol.* 13:852434. doi: 10.3389/fmicb.2022.852434
- Foster-Nyarko, E., Cottingham, H., Wick, R. R., Judd, L. M., Lam, M. M. C., Wyres, K. L., et al. (2023). Nanopore-only assemblies for genomic surveillance of the global priority drug-resistant pathogen, *Klebsiella pneumoniae*. *Microb Genomics* 9:mgen000936. doi: 10.1099/mgen.0.000936
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9
- Jordt, H., Stalder, T., Kosterlitz, O., Ponciano, J. M., Top, E. M., and Kerr, B. (2020). Coevolution of host–plasmid pairs facilitates the emergence of novel multidrug resistance. *Nat Ecol Evol* 4, 863–869. doi: 10.1038/s41559-020-1170-1
- Kathirvel, K., Rudhra, O., Rajapandian, S. G. K., Venkatesh Prajna, N., Lalitha, P., and Devarajan, B. (2021). Characterization of antibiotic resistance and virulence genes of

- ocular methicillin-resistant *Staphylococcus aureus* strains through complete genome analysis. *Exp. Eye Res.* 212:108764. doi: 10.1016/j.exer.2021.108764
- Khrenova, M. G., Panova, T. V., Rodin, V. A., Kryakvin, M. A., Lukyanov, D. A., Osterman, I. A., et al. (2022). Nanopore sequencing for De novo bacterial genome assembly and search for single-nucleotide polymorphism. *Int. J. Mol. Sci.* 23:8569. doi: 10.3390/ijms23158569
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8
- Li, Z., Li, Z., Peng, Y., Lu, X., and Kan, B. State Key Laboratory of Infectious Disease Prevention and Control; National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China (2022). Trans-regional and cross-host spread of mcr-carrying plasmids revealed by complete plasmid sequences — 44 countries, 1998–2020. *China CDC Weekly* 4, 242–248. doi: 10.46234/ccdcw2022.058
- Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14, 265–279. doi: 10.1016/j.gpb.2016.05.004
- Malhotra-Kumar, S., Xavier, B. B., das, A. J., Lammens, C., Butaye, P., and Goossens, H. (2016). Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. *Lancet Infect. Dis.* 16, 283–284. doi: 10.1016/S1473-3099(16)00012-8
- Nathan, C. (2020). Resisting antimicrobial resistance. *Nat. Rev. Microbiol.* 18, 259–260. doi: 10.1038/s41579-020-0348-5
- Nikaido, H. (2009). Multidrug Resistance in Bacteria. *Annu. Rev. Biochem.* 78, 119–146. doi: 10.1146/annurev.biochem.78.082907.145923
- Partridge, S. R., Tsafnat, G., Coiera, E., and Iredell, J. R. (2009). Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol. Rev.* 33, 757–784. doi: 10.1111/j.1574-6976.2009.00175.x
- Paterson, D. L., and van Duin, D. (2017). ‘China’s antibiotic resistance problems’, the lancet. *Infect. Dis.* 17, 351–352. doi: 10.1016/S1473-3099(17)30053-1
- Peter, S., Bosio, M., Gross, C., Bezdan, D., Gutierrez, J., Oberhettinger, P., et al. (2020). Tracking of antibiotic resistance transfer and rapid plasmid evolution in a hospital setting by Nanopore sequencing. *mSphere* 5, e00525–e00520. doi: 10.1128/mSphere.00525-20
- Petersen, L. M., Martin, I. W., Moschetti, W. E., Kershaw, C. M., and Tsongalis, G. J. (2019). Third-generation sequencing in the clinical laboratory: exploring the advantages and challenges of Nanopore sequencing. *J. Clin. Microbiol.* 58, e01315–e01319. doi: 10.1128/JCM.01315-19
- San Millan, A. (2018). Evolution of plasmid-mediated antibiotic resistance in the clinical context. *Trends Microbiol.* 26, 978–985. doi: 10.1016/j.tim.2018.06.007
- Senol Cali, D., Kim, J. S., Ghose, S., Alkan, C., and Mutlu, O. (2019). Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. *Brief. Bioinform.* 20, 1542–1559. doi: 10.1093/bib/bby017
- Sereika, M., Kirkegaard, R. H., Karst, S. M., Michaelsen, T. Y., Sørensen, E. A., Wollenberg, R. D., et al. (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. doi: 10.1038/s41592-022-01539-7
- Smith, C., Halse, T. A., Shea, J., Modestil, H., Fowler, R. C., Musser, K. A., et al. (2020). Assessing Nanopore sequencing for clinical diagnostics: a comparison of next-generation sequencing (NGS) methods for *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 59, e00583–e00520. doi: 10.1128/JCM.00583-20
- Vit, C., Loot, C., Escudero, J. A., Nivina, A., and Mazel, D. (2020). Integron identification in bacterial genomes and cassette recombination assays. *Methods Mol. Biol.* 2075, 189–208. doi: 10.1007/978-1-4939-9877-7_14
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Bouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595
- Wick, R. R., Judd, L. M., and Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 20:129. doi: 10.1186/s13059-019-1727-y