

Recovery of small plasmid sequences via Oxford Nanopore sequencing

Ryan R. Wick^{1,*}, Louise M. Judd¹, Kelly L. Wyres¹ and Kathryn E. Holt^{1,2}

Abstract

Oxford Nanopore Technologies (ONT) sequencing platforms currently offer two approaches to whole-genome native-DNA library preparation: ligation and rapid. In this study, we compared these two approaches for bacterial whole-genome sequencing, with a specific aim of assessing their ability to recover small plasmid sequences. To do so, we sequenced DNA from seven plasmid-rich bacterial isolates in three different ways: ONT ligation, ONT rapid and Illumina. Using the Illumina read depths to approximate true plasmid abundance, we found that small plasmids (<20 kbp) were underrepresented in ONT ligation read sets (by a mean factor of ~4) but were not underrepresented in ONT rapid read sets. This effect correlated with plasmid size, with the smallest plasmids being the most underrepresented in ONT ligation read sets. We also found lower rates of chimaeric reads in the rapid read sets relative to ligation read sets. These results show that when small plasmid recovery is important, ONT rapid library preparations are preferable to ligation-based protocols.

DATA SUMMARY

Supplementary figures, tables, data and code can be found at: github.com/rrwick/Small-plasmid-Nanopore and bridges.monash.edu/articles/dataset/Small_plasmid_Nanopore_data/13543754. The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

INTRODUCTION

Plasmids are extra-chromosomal pieces of DNA present in many bacterial genomes [1, 2]. While smaller than the chromosome, they are important genomic components that can confer key phenotypic traits and contribute to gene flow within/between species. Most plasmids are circular, though linear plasmids also exist [3], and for some bacterial species it is common to find multiple plasmids in a single genome [4]. Plasmids come in a broad range of sizes, from <1 kbp to >300 kbp [5, 6]. For simplicity, here we categorise plasmids as 'small' or 'large' using a threshold of 20 kbp [7]. Both small and large plasmids can carry clinically relevant genes, such as virulence or antimicrobial resistance (AMR) genes [8–10].

Unlike many large plasmids, small plasmids do not encode their own conjugative transfer systems, but many do appear to be readily mobilizable between host cells, and are therefore similarly important for the transfer of genetic material and the spread of AMR [11].

Short-read sequencing (e.g. Illumina platforms) of bacterial genomes can typically only produce fragmented draft assemblies [12]. It is difficult to reconstruct plasmid sequences from short-read assemblies [13], and this impedes the ability to accurately monitor the spread of key genes in bacterial populations [14, 15]. In contrast, long-read sequencing (e.g. Pacific Biosciences and Oxford Nanopore Technologies platforms) allows for complete bacterial genome assembly, with each chromosome or plasmid assembled into a single contig [16, 17]. Oxford Nanopore Technologies (ONT) long-read platforms are especially well suited to bacterial genomics, as they are cost-effective and allow for easy multiplexing of samples, enabling sufficient data generation for simultaneous completion of 12 or more genomes at a cost of <100 USD each [18–20]. There are many genome assembly tools appropriate for use with ONT reads, some of which also use short reads for what is known as hybrid assembly [12, 21]. Other tools

Received 21 February 2021; Accepted 11 June 2021; Published 25 August 2021

Author affiliations: ¹Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, VIC, 3004, Australia; ²Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK.

***Correspondence:** Ryan R. Wick, rrwick@gmail.com

Keywords: Plasmids; whole-genome sequencing; Oxford Nanopore sequencing; long-read sequencing.

Abbreviations: AMR, antimicrobial resistance; CARD, Comprehensive Antibiotic Resistance Database; GTDB, Genome Taxonomy Database; IG, Integrative Genomics Viewer; IS, insertion sequence; ONT, Oxford Nanopore Technologies; QC, quality control; VFDB, Virulence Factor Database.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. One supplementary table and eight supplementary figures are available with the online version of this article.

000631 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

operate on ONT reads alone, an approach we will call long-read-only assembly [22–24].

While ONT sequencing offers many advantages in bacterial genomics, during a number of sequencing projects [18, 25–30], we have anecdotally observed that small plasmids are frequently absent from ONT long-read-only assemblies. This can have serious downstream consequences, e.g. if a small plasmid containing an AMR determinant is missing from the assembly, this can result in an incorrect antimicrobial susceptibility prediction (a failure to detect resistance), known as a ‘very major error’ as it can lead to prescribing an ineffective antimicrobial [31]. We hypothesised that this problem was due in part to our use of the ONT ligation-based library preparation (SQK-LSK109 kit). During extraction, DNA is incidentally fragmented, typically resulting in fragments ~10 kbp in size [32] onto which barcode and adapter sequences are attached via blunt-end ligation during library preparation. Large plasmids are likely to be fragmented into one or more linear pieces, but small plasmids may avoid fragmentation and remain circular. As such, the small plasmids will have no blunt ends, will not have any adapter ligated and will thus be unavailable for sequencing (Fig. 1). Upon sequencing of the library, this leads to underrepresentation of small plasmids in the resulting ONT read set, which may in turn cause the assembler to fail to produce contigs for them [24]. Deliberately increasing fragmentation before DNA preparation could mitigate this effect by creating smaller fragments, however this would result in shorter read lengths, which negates the benefits of long-read sequencing and can negatively impact the assembly contiguity [24, 33].

The ONT rapid preparation kits (e.g. the SQK-RBK004 rapid barcoding kit) offer a potential solution. Unlike the ligation approach, the rapid approach uses a transposase enzyme to simultaneously cleave DNA and attach barcode/adaptor sequences (Fig. 1). Since rapid preparations do not rely on blunt-end ligation, they should be active on sequences that are circular such as unfragmented small plasmids. Hence, in this study we compared the performance of ONT ligation and rapid library preparations for whole-genome sequencing of bacterial genomes containing small plasmids. Specifically, we aimed to quantify plasmid read depths and determine whether ONT rapid preparations give a more accurate representation of small plasmid abundance than ONT ligation preparations. To assess this, we used short-read Illumina sequencing, whose libraries are unbiased with respect to the starting length of the molecules, as the gold standard for quantifying plasmid abundance.

METHODS

Bacterial isolates, DNA extraction and sequencing

We included seven bacterial isolates in this study (Table 1, Fig. S1a, available in the online version of this article), each containing small plasmids (identified from previous analyses) [34–37] and belonging to different bacterial species: *Acinetobacter baumannii*, *Citrobacter koseri*, *Enterobacter kobei*, an unnamed *Haemophilus* species (given the placeholder name

Impact Statement

Researchers who use Oxford Nanopore Technologies (ONT) platforms to sequence bacterial genomes can currently choose from two library preparation methods. The first is a ligation-based approach, which uses ligase to attach sequencing adapters to the ends of DNA molecules. The second is a rapid approach, which uses a transposase enzyme to cleave DNA and attach adapters in a single step. There are advantages to each preparation, for example ligation can produce better yields but rapid is a simpler procedure. Our study reveals another advantage of rapid preparations: they are more effective at sequencing small plasmids. We show that sequencing of ligation-based libraries yields fewer reads derived from small plasmids, making such plasmids harder to detect in bacterial genomes. Since small plasmids can contain clinically relevant genes, including antimicrobial resistance (AMR) or virulence determinants, their exclusion could lead to unreliable conclusions that have serious consequences for AMR surveillance and prediction. We therefore recommend that researchers performing ONT-only sequencing of bacterial genomes should consider using rapid preparations whenever small plasmid recovery is important.

Haemophilus sp002998595 in GTDB R202) [38, 39], *Klebsiella oxytoca*, *Klebsiella variicola* and *Serratia marcescens*. These isolates were cultured overnight at 37 °C in Luria-Bertani broth and DNA was extracted using GenFind V3 according to the manufacturer’s instructions (Beckman Coulter) (Fig. S1b). The same DNA extract was used to sequence each isolate using three different approaches: ONT ligation, ONT rapid and Illumina (Fig. S1c). For ONT ligation, we followed the protocol for the SQK-LSK109 ligation sequencing kit and EXP-NBD104 native barcoding expansion (Oxford Nanopore Technologies). For ONT rapid, we followed the protocol for the SQK-RBK004 rapid barcoding kit (Oxford Nanopore Technologies). All ONT libraries were sequenced on MinION R9.4.1 flow cells. For Illumina, we followed a modified Illumina DNA Prep protocol (catalogue number 20018705), whereby the reaction volumes were quartered to conserve reagents. Illumina libraries were sequenced on the NovaSeq 6000 using SP reagent kits v1.0 (300 cycles, Illumina Inc.), producing 150 bp paired-end reads with a mean insert size of 331 bp. We repeated this process (from culture to sequencing) to generate a set of technical replicates. For the first technical replicate, a refuel (with the EXP-FLP002 flow cell priming kit) was performed at the 18h point of the ONT runs to boost yield. However, no refuelling step was required for the second replicate. All ONT read sets were basecalled and demultiplexed using Guppy v3.6.1 (Fig. S1d, e). Basecalled reads (FASTQ format) are available in the supplementary data repository, github.com/rrwick/Small-plasmid-Nanopore.

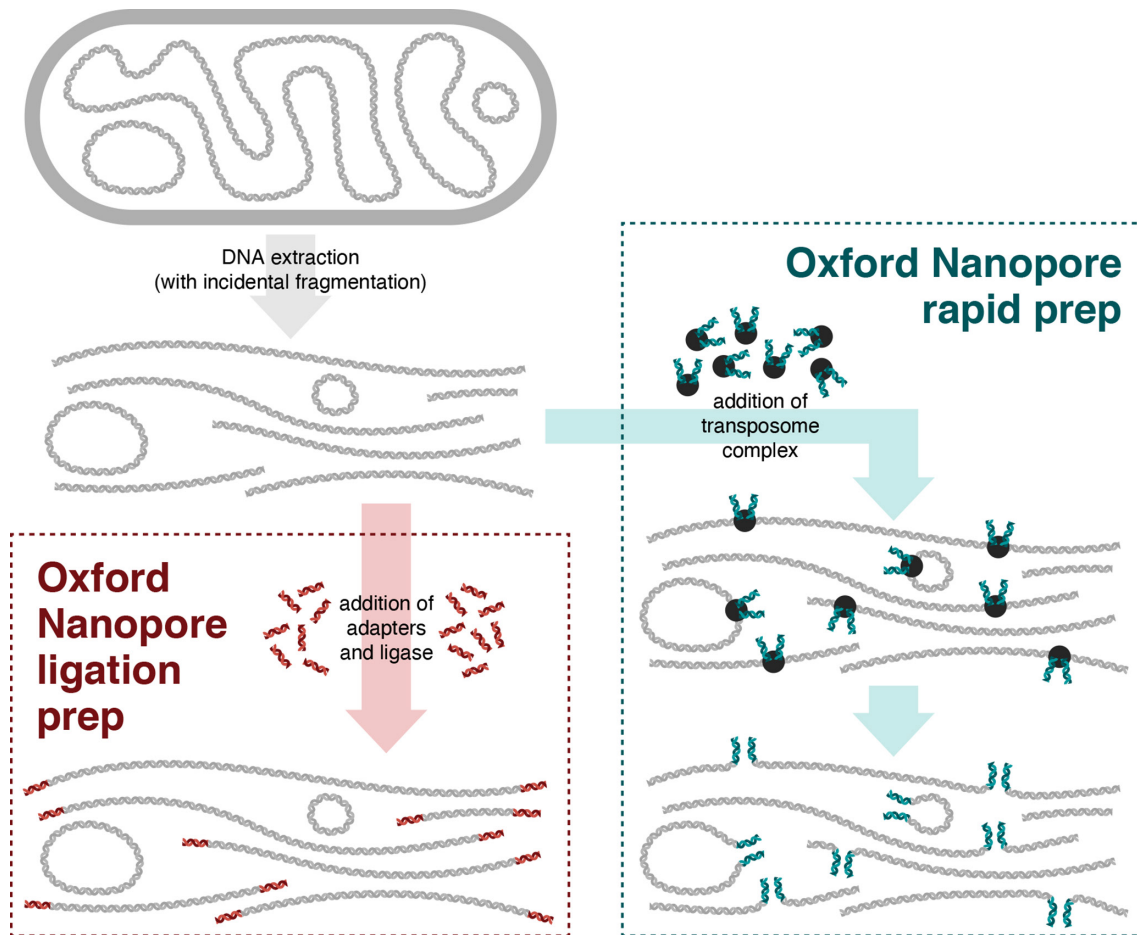


Fig. 1. Conceptual illustration of Oxford Nanopore ligation and rapid sample preparation methods. When circular DNA is extracted from a bacterial cell (top-left), incidental fragmentation of the DNA occurs. The ligation preparation (bottom-left) comprises blunt-end ligation of barcodes/adapters onto DNA molecules, so circular pieces of DNA will not receive adapters and thus remain unavailable for sequencing. The rapid preparation (right) uses a transposome enzyme to add barcodes/adapters into the middle of DNA molecules, making both linear and circular DNA available for sequencing.

Reference genome assembly

To assign reads to their replicon of origin, we required a reference assembly for each of the seven genomes in the study. We first produced separate assemblies for each genome and technical replicate (14 assemblies in total) using pooled reads from each sequencing run (ONT ligation, ONT rapid and Illumina) (Fig. S1f). We then merged the two replicate assemblies for each genome to produce a single final assembly (Fig. S1g). Illumina read QC and trimming was performed by fastp v0.20.1 [40] using default parameters. ONT read QC was performed by Filtlong v0.2.0 using no external reference, a minimum read length of 1 kbp, a minimum mean quality of 80 and a minimum window quality of 60 (quality values refer to Filtlong percent identity estimates based on Phred scores). We used Tricycler v0.3.3 [41] to produce a consensus long-read assembly for each isolate from 15 input assemblies (five Flye v2.8 [23] assemblies, five miniasm/Minipolish v0.3/v0.1.3 [42] assemblies and five Raven v1.1.10 [43] assemblies), each made using independent sets of randomly subsampled ONT

reads of 50× depth. Tricycler consensus assemblies were then polished with Medaka v1.0.3 [44] and Pilon v1.23 [45]. For each genome, the two independent assemblies (one from each technical replicate) were compared using edlib [46]. Wherever differences were found, we used IGV [47] to visually inspect Illumina read alignments (generated with Bowtie2) [48] and ONT read alignments (generated with minimap2) [49] for the region in question to assess whether the difference indicated an assembly error (indicated by a drop in Illumina read depth), and we manually repaired such errors as appropriate. Only one genuine sequence difference was found between the two technical replicates: a 1 bp indel in the smallest plasmid of *E. kobei* MSB1_1B. Two plasmids in *Haemophilus* M1C132_1 only appeared in assemblies for the second technical replicate, as they were missing from all ONT and Illumina read sets in the first replicate, suggesting that they were lost during culturing. A plasmid in *S. marcescens* 17-147-1671 occurred in two variants (a 17.4 kbp version with one copy of IS4321 and an 18.7 kbp version with two copies of IS4321) [50], both

Table 1. Bacterial isolates used in this study. Each genome contained at least one large (≥ 20 kbp) and one small (< 20 kbp) plasmid. Seven plasmids (indicated with *) contained one or more antimicrobial resistance determinants, and two plasmids (indicated with †) contained one or more virulence determinants.

Isolate species and name	Large plasmids (bp)	Small plasmids (bp)
<i>Acinetobacter baumannii</i> J9	145059*	6078*
<i>Citrobacter koseri</i> MINF_9D	64962*	9294
<i>Enterobacter kobei</i> MSB1_1B	136482, 108411	4665, 3715, 2369
<i>Haemophilus</i> M1C132_1	39398	10719, 9975*, 7392, 5675
<i>Klebsiella oxytoca</i> MSB1_2C	118161†, 58472	4574
<i>Klebsiella variicola</i> INF345	250980, 243620*, 31780	5783, 3514
<i>Serratia marcescens</i> 17-147-1671	184477*†, 161385*	17406, 1934

of which occurred in both technical replicates. After curation, we merged the assemblies of the two technical replicates to make a single reference assembly for each of the seven isolates, including the two plasmids that only appeared in the second technical replicate and all versions of plasmids that contained variation. The resultant genomes contained a total of seven chromosomes and 26 plasmids (Table 1, Fig. S2). Using a size threshold of 20 kbp, 14 of the plasmids were classified as ‘small’ and 12 as ‘large’. We screened plasmids for AMR genes using Kleborate v2.0.1 (uses the CARD database) [51] and for virulence genes using the VFDB [52]. Seven of the plasmids (five large and two small) contained one or more antimicrobial resistance genes and two of the plasmids (both large) contained one or more virulence genes (Table S1). The reference assembly sequences (FASTA format) are available in the supplementary data repository, github.com/rrwick/Small-plasmid-Nanopore.

Comparison of ONT library preparation methods

To aid our read-based analyses, we developed custom Python scripts (available in the supplementary data repository, github.com/rrwick/Small-plasmid-Nanopore). We aligned each of the four ONT read sets (ligation run 1, rapid run 1, ligation run 2 and rapid run 2) to each sequence in the merged reference assemblies using minimap2 v2.17 (using the map-ont preset) and the align_reads.py script, which enabled alignment over the start-end junction of circular replicons. These alignments were processed with the assign_reads.py script, which assigned each read to a reference sequence by labelling each position of the read with the reference to which it best aligned (using the following filters: alignment identity $\geq 75\%$, alignment length ≥ 100 bp, mean read identity $\geq 80\%$, mean read coverage $\geq 50\%$). This script also gave a demultiplexing

status to each read: correct (Guppy demultiplexing agreed with reference alignment), incorrect (Guppy demultiplexing disagreed with reference alignment), unclassified (the read was not demultiplexed by Guppy), unaligned (the read did not align to the reference sequences) or chimaera (the read aligned to multiple different reference sequences). The get_depths.py script was used to calculate per-replicon ONT read depths by taking the mean depth across each replicon, excluding repetitive regions common to multiple replicons (identified by cross-replicon minimap2 alignments) as such repeats could make for unreliable alignments. Finally, we aligned each Illumina read set (after QC and trimming with fastp) to its respective reference genome using Bowtie2 v2.3.4.1 and used the get_depths.py script to calculate per-replicon Illumina read depths, again excluding repetitive regions common to multiple replicons.

Plasmid read depths were normalised to their corresponding chromosomal read depth (e.g. a plasmid with a depth twice that of the chromosomal depth had a normalised read depth of two), providing a quantification of plasmid abundance that could be compared between read sets. Read depths were calculated separately for the two technical replicates, as plasmid copy number could differ between DNA extractions. This resulted in three normalised read depths for each plasmid in each technical replicate: ONT ligation, ONT rapid and Illumina (Table S1).

The on-bead tagmentation process in the Illumina DNA Prep produces libraries with a fragment size of ~ 300 – 400 bp, which is considerably smaller than the smallest replicon in our genomes. Therefore, we assumed that unlike long-read ONT sequencing, Illumina sequencing is not significantly biased by replicon size. Since Illumina sequencing is known to have biases regarding GC content [53], we examined this effect in our data by calculating read depth and GC content for each one kbp sequence window in the chromosomes of the seven study genomes (using the depth_and_gc.py script). By plotting the read depth (normalised to the mean depth for 50% GC windows) vs GC content, we estimated that fluctuations in GC content resulted in $< 10\%$ variation of Illumina read depth (Fig. S3). We therefore assume that our normalised Illumina read depth values are a good approximation of the true copy number of each replicon, without adjusting for GC content.

RESULTS AND DISCUSSION

Sequencing data characteristics

The four ONT sequencing runs (ligation run 1, rapid run 1, ligation run 2 and rapid run 2) yielded a total of 15.2 Gbp, 4.3 Gbp, 8.0 Gbp and 9.6 Gbp data, respectively. Per-barcode ONT yields ranged from 64 Mbp to 2.8 Gbp, equating to mean read depths of $10\times$ to $1214\times$ (Table S1). Per-barcode ONT N50 read lengths ranged from 1.9 kbp to 25.8 kbp. Per-barcode Illumina yields ranged from 313 Mbp to 1.06 Gbp ($40\times$ to $232\times$ depth) (Table S1). Since we used pooled ONT read sets (both ligation and rapid) to perform assemblies, all isolates had sufficient data to produce reliable reference genomes. The

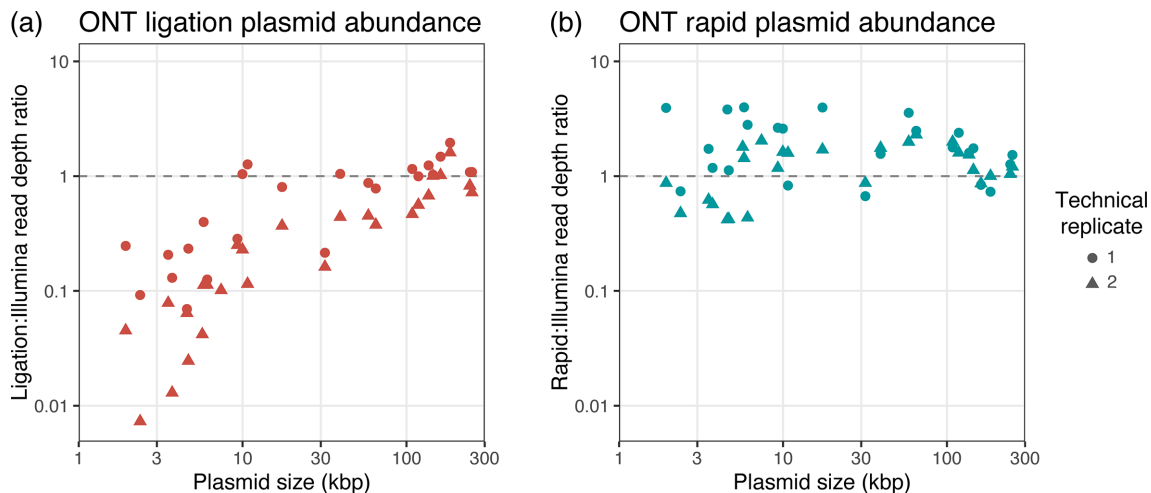


Fig. 2. Plasmid abundance resulting from (a) ligation and (b) rapid ONT library preparation methods. Each point in the plots represents one plasmid, with circles for plasmids in the first technical replicate and triangles for plasmids in the second technical replicate. The read depth ratio is the normalised ONT read depth divided by the normalised Illumina read depth. The dashed lines at ratio=1 indicate perfect agreement of plasmid depths between ONT and Illumina data. Points above the dashed lines indicate plasmids that are overrepresented in ONT reads, while points below the dashed lines indicate plasmids that are underrepresented in ONT reads. For ONT ligation reads (a), small plasmids are systematically underrepresented relative to Illumina reads. For ONT rapid reads (b), plasmid size has no clear effect, and depths for both small (<20 kbp) and large plasmids (≥ 20 kbp) are in good agreement with Illumina reads.

variation in ONT read depth across each replicon was low and did not depend on the library preparation: the mean within-replicon read depth coefficient of variation was 0.10 for both ONT rapid and ONT ligation read sets (0.18 for Illumina read sets) (Table S1). The densities of read-start sites across each replicon were largely compatible with Poisson distributions, showing that most read positions are random (Fig. S4). However, ligation read sets contained more regions where read-start densities deviated from Poisson distributions, suggesting that some sequence regions can be more prone to fragmentation than others during the ligation preparation. This was particularly evident for the *Haemophilus* M1C132_1 and *S. marcescens* 17-147-1671 genomes.

While previous studies have shown that ligation preparations favour read count and rapid preparations favour read length [54], we observed no clear trends (Table S1, Fig. S5). Many factors influence these metrics, including the quality and quantity of extracted DNA, the number of available pores on the flow cell, and the incubation times during library preparation steps. Our results show that for multiplexed bacterial whole-genome sequencing, good yields and read lengths ($>100\times$ depth and >15 kbp N50) are possible from either type of library preparation.

Library preparation type also did not seem to affect the sequence accuracy of reads. All runs had a maximum read identity of $\sim 98\%$, but two of the runs (ligation run 1 and rapid run 2) had a larger proportion of low-identity reads (22 and 23% of reads at $<90\%$ identity vs 15 and 16% for ligation run 2 and rapid run 1, respectively) (Fig. S6). These runs suffered from degradation in translocation speed (the rate at which DNA moves through the nanopore) over the course of the run

(Fig. S7), which may partly explain their lower accuracy [55]. When this problem occurs, it can be mitigated by refuelling an in-progress run using ONT's EXP-FLP002 kit, as we did for both ligation run 1 and rapid run 1. The beneficial effect of refuelling on read accuracy was particularly notable for ligation run 1 but was negligible for rapid run 1 (Fig. S8).

Demultiplexing accuracy was also inconsistent between preparation types, with both the best (0.39%) and worst (3.87%) demultiplexing error rates occurring in rapid runs, compared to error rates of 2.22 and 2.92% for the ligation runs (Table S1). During ligation preparations, barcode sequences are attached on both ends of the reads, while rapid preparations result in barcode attachment to the start of the reads only. This gives users of ligation preparations the option of running Guppy with the `--require_barcodes_both_ends` option (not used in this study) to increase demultiplexing accuracy when needed but with a larger proportion of unclassified reads [56].

Chimaeric read rates

The rate of chimaeras (reads originating from two or more discontinuous pieces of DNA) was notably different between the two preparations: 1.41 and 0.88% chimaeric reads for the two ligation runs vs 0.03 and 0.14% for the two rapid runs (Table S1). There are two broad categories of chimaeric reads: *in silico* chimaeras and ligated chimaeras [57]. *In silico* chimaeras occur when two separate pieces of DNA pass through a pore in quick succession such that the sequencing software mistakes them as a single read. Thus *in silico* chimaeras can potentially happen for either type of preparation. Ligated chimaeras occur when two separate pieces of

DNA are physically joined before sequencing. Since only the ligation preparation involves ligase, ligated chimaeras should be comparatively rare in rapid preparations, which may explain our observation that rapid preparations have fewer chimaeric reads overall. Our results suggest that rapid preparations are preferable when chimaeric reads need to be minimised, but we note that this conclusion is derived from small sample sizes ($n=2$ for each preparation). Additionally, in the context of bacterial whole-genome assembly, we have previously shown that chimaeric read rates of up to ~5% (i.e. exceeding those of our sequencing runs) do not impact assembly quality [24].

Small plasmid abundance

For each assembled plasmid, we calculated a normalised depth ratio: its normalised ONT read depth (i.e. ONT plasmid depth relative to the chromosome) divided by its normalised Illumina read depth (i.e. Illumina plasmid depth relative to the chromosome). Ratios greater than one indicate the plasmid was overrepresented in ONT reads relative to Illumina reads, and ratios less than one indicate the plasmid was underrepresented. Fig. 2 shows the relationship between normalised depth ratios and plasmid size for ligation and rapid ONT preparations.

For ONT ligation reads, there was a clear relationship between the normalised depth ratio and plasmid size ($P=1.1 \times 10^{-10}$, $\tau=0.64$, from a Kendall rank correlation test) (Fig. 2a). ONT rapid read depths showed no such relationship ($P=0.98$, $\tau=0.0025$) (Fig. 2b). Specifically, ONT ligation reads tended to underrepresent small plasmids (<20 kbp), for which the mean normalised depth ratio was 25% (i.e. small plasmids produced ~4× fewer ONT ligation reads than one would expect based on Illumina read depths). The most extreme case was for the 2.4 kbp plasmid in the second technical replicate of *E. kobei* MSB1_1B, which had a normalised depth ratio of <1%, i.e. it was underrepresented by a factor of more than 100. Very large plasmids (>100 kbp) did not suffer from underrepresentation in ONT ligation read sets, presumably because they were likely to be linearised via fragmentation.

The effect of plasmid size on normalised depth ratio in the ligation reads was stronger for the second technical replicate, i.e. plasmids in the second technical replicate were especially underrepresented in ONT ligation read sets. This is concordant with the fact that the read N50 was shorter in ligation run 1 and longer in ligation run 2 (8.2 kbp vs 20.9 kbp) (Table S1, Fig. S5), as shorter reads imply more DNA fragmentation, thus increasing the chance that a circular plasmid will be linearised and sequenced.

These results suggest that both ONT ligation and ONT rapid preparations provide an accurate representation of abundance for very large plasmids (>100 kbp), but only the ONT rapid preparation accurately represents abundance across the full spectrum of plasmid sizes. While these conclusions assume that Illumina reads accurately represent plasmid abundance, this is supported by the fact that Illumina and ONT rapid

plasmid read depths are in good agreement. As our main goal was to understand the effect of library preparation on plasmid results, we kept the DNA extraction method constant for all experiments (Beckman Coulter's GenFind V3). However, it is possible that different DNA extraction methods could affect the capture of small plasmids also, which could be explored in future studies.

The bias against small plasmids during ligation-based library preparations could result in the exclusion of small plasmids from genome assemblies, particularly when performing long-read-only assembly (as opposed to hybrid assembly where Illumina reads are also available) and assembling read sets of modest depth. For example, if a genome containing a four-copies-per-cell small plasmid was sequenced with a chromosomal read depth of 25×, the plasmid sequence should be present at ~100× depth. However, if that small plasmid was underrepresented by a factor of 30 (a realistic possibility based on our data, see Fig. 2a), it might only be sequenced at ~3× depth, making it unlikely to appear in an assembly for that genome.

Conclusions

When faced with the choice of ligation or rapid ONT preparations, researchers must weigh their respective advantages. Ligation kits are versatile and can give greater yields. Rapid kits are faster, require fewer additional resources and can provide longer read lengths (if optimised for in DNA preparation) [54]. Our study reveals two more advantages of the rapid kits: lower chimaeric read rates and better recovery of small plasmids. Rapid preparations are therefore more likely to produce long-read-only assemblies which include all replicons, small plasmids included.

Funding information

This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation [OPP1175797]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. This work was also supported by an Australian Government Research Training Program Scholarship, and KEH is supported by a Senior Medical Research Fellowship from the Viertel Foundation of Victoria. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Conceptualization: R. R. W., K. L. W., K. E. H. Methodology: R. R. W. Software: R. R. W. Formal analysis: R. R. W., K. L. W. Investigation: R. R. W., L. M. J. Resources: R. R. W., L. M. J. Data curation: R. R. W., L. M. J. Writing – original draft: R. R. W. Writing – review and editing: R. R. W., L. M. J., K. L. W., K. E. H. Visualization: R. R. W. Supervision: K. E. H., K. L. W. Project administration: K. E. H., K. L. W. Funding acquisition: K. E. H.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, et al. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 2015;15:141–161.
- Bobay LM, Ochman H. The evolution of bacterial genome architecture. *Front Genet* 2017;8:1–6.

3. Baker S, Hardy J, Sanderson KE, Quail M, Goodhead I, et al. A novel linear plasmid mediates flagellar variation in *Salmonella* Typhi. *PLoS Pathog* 2007;3:0605–0610.
4. Arredondo-Alonso S, Top J, McNally A, Puranen S, Pesonen M, et al. Plasmids shaped the recent emergence of the major nosocomial pathogen *Enterococcus faecium*. *mBio* 2020;11:1–17.
5. Ciok A, Dziewit L, Grzesiak J, Budzik K, Gorniak D, et al. Identification of miniature plasmids in psychrophilic Arctic bacteria of the genus *Variovorax*. *FEMS Microbiology Ecology* 2016;92:1–9.
6. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother* 2018;73:1121–1137.
7. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. *Front Microbiol* 2015;6:1–16.
8. Smalla K, Jechalke S, Top EM. Plasmid detection, characterization, and ecology. *Plasmids* 2015;3:445–458.
9. San Millan A, Escudero JA, Catalan A, Nieto S, Farelo F, et al. β -Lactam resistance in *Haemophilus parasuis* is mediated by plasmid pB1000 bearing bla_{ROB}-1. *Antimicrob Agents Chemother* 2007;51:2260–2264.
10. Anantham S, Hall RM. pCERC1, a Small, Globally Disseminated Plasmid Carrying the dfrA14 Cassette in the strA-strB Gene Cluster. *Microb Drug Resist* 2012;18:364–371.
11. Lanza VF, de Toro M, Garcillán-Barcia MP, Mora A, Blanco J, et al. Plasmid Flux in *Escherichia coli* ST131 Sublineages, Analyzed by Plasmid Constellation Network (PLACNET), a New Method for Plasmid Reconstruction from Whole Genome Sequences. *PLoS Genet* 2014;10:12.
12. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
13. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb Genom* 2017;3:10.
14. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, et al. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Sci Transl Med* 2014;6:254ra126.
15. Weingarten RA, Johnson RC, Conlan S, Ramsburg AM, Dekker JP, et al. Genomic analysis of hospital plumbing reveals diverse reservoir of bacterial plasmids conferring carbapenem resistance. *mBio* 2018;9:1–16.
16. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* 2015;12:733–735.
17. Koren S, Phillippy AM. One chromosome, one contig: Complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol* 2015;23:110–120.
18. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* 2017;3:1–7.
19. Taylor TL, Volkening JD, DeJesus E, Simmons M, Dimitrov KM, et al. Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology. *Sci Rep* 2019;9:1–11.
20. Elliott I, Batty EM, Ming D, Robinson MT, Nawtaisong P, et al. Oxford nanopore MinION sequencing enables rapid whole genome assembly of rickettsia typhi in a resource-limited setting. *Am J Trop Med Hyg* 2020;102:408–414.
21. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. HybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics* 2016;32:1009–1015.
22. Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 2017;27:722–736.
23. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
24. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res* 2019;8:2138.
25. Lam MMC, Wick RR, Wyres KL, Gorrie CL, Judd LM, et al. Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICE_{Kp} in *Klebsiella pneumoniae* populations. *Microb Genom* 2018;4.
26. Lam MMC, Wyres KL, Judd LM, Wick RR, Jenney A, et al. Tracking key virulence loci encoding aerobactin and salmochelin siderophore synthesis in *Klebsiella pneumoniae*. *Genome Med* 2018;10:1–15.
27. Wyres KL, Hawkey J, Hetland MAK, Fostervold A, Wick RR, et al. Emergence and rapid global dissemination of CTX-M-15-associated *Klebsiella pneumoniae* strain ST307. *J Antimicrob Chemother* 2019;74:577–581.
28. Lam MMC, Wyres KL, Wick RR, Judd LM, Fostervold A, et al. Convergence of virulence and MDR in a single plasmid vector in MDR *Klebsiella pneumoniae* ST15. *J Antimicrob Chemother* 2019;74:1218–1222.
29. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, et al. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. *PLoS Genet* 2019;15:1–25.
30. Wyres KL, Nguyen TNT, Lam MMC, Judd LM, van Vinh Chau N, et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive *Klebsiella pneumoniae* from South and Southeast Asia. *Genome Med* 2020;12:11.
31. Jorgensen JH, Ferraro MJ. Antimicrobial susceptibility testing: A review of general principles and contemporary practices. *Clin Infect Dis* 2009;49:1749–1755.
32. Klingström T, Bongcam-Rudloff E, Pettersson OV. A comprehensive model of DNA fragmentation for the preservation of High Molecular Weight DNA. *bioRxiv* 2018.
33. Ou S, Liu J, Chougule KM, Functammasan A, Seetharam AS, et al. Effect of sequence depth and length in long-read assembly of the maize inbred NC358. *Nat Commun* 2020;11:1–10.
34. Hamidian M, Ambrose SJ, Blackwell GA, Nigro SJ, Hall RM. An outbreak of multiply antibiotic-resistant ST49:ST128:KL11:OCL8 *Acinetobacter baumannii* isolates at a Sydney hospital. *J Antimicrob Chemother* 2021;76:893–900.
35. Watts SC, Judd LM, Carzino R, Ranganathan S, Holt KE. Genomic diversity and antimicrobial resistance of *Haemophilus* colonising the airways of young children with cystic fibrosis. *bioRxiv* 2020;2020.11.23.388074.
36. Wyres KL, Hawkey J, Mirčeta M, Judd LM, Wick RR, et al. Genomic surveillance of antimicrobial resistant bacterial colonisation and infection in intensive care patients. *medRxiv* 2020;2020.11.03.20224881.
37. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 2019;20:129.
38. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2020;36:1925–1927.
39. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 2020;38:1098.
40. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–90.
41. Wick RR, Holt KE. Trycycler [Internet]. *GitHub* 2020.
42. Li H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 2016;32:2103–2110.
43. Vaser R, Š M, Šikić M. Raven: a de novo genome assembler for long reads. *bioRxiv* 2020.
44. Wright C, Wykes M. Medaka [Internet]. *GitHub* 2020.

45. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* 2014;9:11.
46. Šošić M, Šikić M. Edlib: A C/C++ library for fast, exact sequence alignment using edit distance. *Bioinformatics* 2017;33:1394–1395.
47. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–26.
48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
49. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
50. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;34:D32–6.
51. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, et al. CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;48:D517–25.
52. Chen L, Yang J, Yu J, Yao Z, Sun L, et al. VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res* 2005;33:325–328.
53. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, et al. Characterizing and measuring bias in sequence data. *Genome Biol* 2012;02:1.
54. Jain M, Koren S, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 2018;36:338–345.
55. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 2020;21:1–16.
56. Wick RR, Judd LM, Holt KE. Deepbiner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Comput Biol* 2018;14:11.
57. White R, Pellefigues C, Ronchese F, Lamiable O, Eccles D. Investigation of chimeric reads using the MinION. *F1000Res* 2017;6:631.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.