



## Original Article

## Assessing the performance of the Oxford Nanopore Technologies MinION



T. Laver<sup>a,\*</sup>, J. Harrison<sup>a,1</sup>, P.A. O'Neill<sup>a,b</sup>, K. Moore<sup>a,b</sup>, A. Farbos<sup>a,b</sup>, K. Paszkiewicz<sup>a,b</sup>, D.J. Studholme<sup>a</sup>

<sup>a</sup> Biosciences, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, UK

<sup>b</sup> Wellcome Trust Biomedical Informatics Hub, Geoffrey Pope Building, Stocker Road, University of Exeter, Exeter EX4 4QD, UK

## ARTICLE INFO

## Article history:

Received 4 December 2014

Received in revised form 5 February 2015

Accepted 18 February 2015

## Keywords:

DNA sequencing

MinION

Nanopore

## ABSTRACT

The Oxford Nanopore Technologies (ONT) MinION is a new sequencing technology that potentially offers read lengths of tens of kilobases (kb) limited only by the length of DNA molecules presented to it. The device has a low capital cost, is by far the most portable DNA sequencer available, and can produce data in real-time. It has numerous prospective applications including improving genome sequence assemblies and resolution of repeat-rich regions. Before such a technology is widely adopted, it is important to assess its performance and limitations in respect of throughput and accuracy. In this study we assessed the performance of the MinION by re-sequencing three bacterial genomes, with very different nucleotide compositions ranging from 28.6% to 70.7%; the high G + C strain was underrepresented in the sequencing reads. We estimate the error rate of the MinION (after base calling) to be 38.2%. Mean and median read lengths were 2 kb and 1 kb respectively, while the longest single read was 98 kb. The whole length of a 5 kb rRNA operon was covered by a single read. As the first nanopore-based single molecule sequencer available to researchers, the MinION is an exciting prospect; however, the current error rate limits its ability to compete with existing sequencing technologies, though we do show that MinION sequence reads can enhance contiguity of de novo assembly when used in conjunction with Illumina MiSeq data.

© 2015 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The Oxford Nanopore Technologies (ONT) MinION [20] is a new sequencing technology that is currently available as part of an early access and development scheme: the MinION Access Programme [21]. This programme allowed early access to the MinION for participating sequencing centres. The results produced by this study are based on the first round of the ONT MinION Access Programme, using the company's R6 sequencing chemistry.

The MinION will most likely be the first commercially available sequencer that uses nanopores. Nanopore sequencing has been shown to be able to discriminate individual nucleotides by

measuring the change in electrical conductivity as DNA molecules pass through the pore [23,28]. Nanopore sequencing does not rely on sequencing by synthesis as most current major technologies do. Laszlo et al. [13] sequenced the phi X 174 genome using another nanopore based technology, demonstrating that nanopore sequencing can produce long reads that are accurate enough to enable them to be aligned back to their reference genomes.

The MinION has several attributes that give it the potential to replace or complement existing sequencing technologies for some applications. The technology offers read lengths of tens of kilobases, with theoretically no instrument-imposed limitation on the size of reads that can be generated. The MinION uses nanopores to sequence a single DNA molecule per pore [11]; this has significant potential advantages over the current widely used sequencing technologies (Ion Torrent, Illumina), which rely on sequencing clusters of amplified DNA molecules. Sequencing a single molecule removes the necessity for PCR amplification and its associated biases [1]. The device has a low capital cost, is by far the most portable DNA sequencer available and can produce data in real-time, although at this stage the samples still require library preparation prior to sequencing – a process that has yet to be optimised. It has applications in scaffolding genome sequences

**Abbreviations:** NRPS, non-ribosomal peptide synthase; ONT, Oxford Nanopore Technologies.

\* Corresponding author. Tel.: +44 7706943765.

**E-mail addresses:** [twl207@exeter.ac.uk](mailto:twl207@exeter.ac.uk) (T. Laver), [jh288@exeter.ac.uk](mailto:jh288@exeter.ac.uk) (J. Harrison), [P.A.O'Neill@exeter.ac.uk](mailto:P.A.O'Neill@exeter.ac.uk) (P.A. O'Neill), [K.A.Moore@exeter.ac.uk](mailto:K.A.Moore@exeter.ac.uk) (K. Moore), [A.Farbos@exeter.ac.uk](mailto:A.Farbos@exeter.ac.uk) (A. Farbos), [K.H.Paszkiwicz@exeter.ac.uk](mailto:K.H.Paszkiwicz@exeter.ac.uk) (K. Paszkiewicz), [D.J.Studholme@exeter.ac.uk](mailto:D.J.Studholme@exeter.ac.uk) (D.J. Studholme).

<sup>1</sup> These authors contributed equally to this work.

<http://dx.doi.org/10.1016/j.bdq.2015.02.001>

2214-7535/© 2015 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

assembled from short reads [3,31] and resolving repeat sequences or haplotypes, being able to span ambiguous regions in a single read, as has been demonstrated for PacBio [27,9]. Future developments may include use in real-time medical diagnostics and forensics, as well as prospective applications as an environmental DNA sensor.

As the MinION is still in its testing stage there is very limited data published data on its performance. Mikheyev and Tin [19] sequenced the lambda phage genome, reporting that, when unalignable reads are taken into account, less than 1% of the sequence produced by the MinION is identical to the reference. Quick et al. [24] were able to sequence an *Escherichia coli* genome demonstrating that the MinION is able to sequence entire bacterial genomes. Ashton et al. [2] used the MinION to resolve the structure and chromosomal insertion site of an antibiotic resistance island in *Salmonella typhi*. They estimated the median accuracy of their MinION data to be between 61.6% and 71.5% based on mapping back to the reference. De novo genome assembly using MinION reads has been demonstrated to achieve improved assembly compared to Illumina sequencing alone by [7].

During the MinION DNA library preparation hairpin structures are added to the end of the double stranded fragments, these fragments are then denatured resulting in one length of single stranded DNA consisting of the forward strand followed by the hairpin sequence then the reverse strand [24]. The MinION generates up to three different types of read for each fragment of DNA that passes through a pore: 'Template', 'Complement' and 'Two Direction'. Initially, the forward strand is sequenced generating the Template read then the hairpin structure is read through followed by the reverse strand, generating the Complement read. Finally, the ONT base calling software attempts to call a consensus sequence of the Template and Complement reads; this resulting consensus sequence is referred to as a Two Direction read. Not all fragments that pass through the pore result in generation of all three read types; some only result in the Template read as output, others in Template and Complement, while only a small minority produce Template, Complement and Two Direction reads. One objective of the current study was to assess whether there were differences between the three types of read, such as read G+C content, read length and error rate.

Extreme G+C content is known to affect the performance of DNA sequencers [1]. To investigate whether the MinION was affected by the nucleotide composition of the target DNA this study re-sequenced a mix of three bacteria with a range G+C content *Borrelia burgdorferi* (28.6%), *Streptomyces avermitilis* (70.7%) and *E. coli* (50.8%).

## 2. Methods

### 2.1. Bacterial DNA

Bacterial DNA was obtained from American Type Culture Collection (ATCC) for *S. avermitilis* (ATCC 35210), *B. burgdorferi* (ATCC 31267) and *E. coli* K-12 (ATCC 10798).

### 2.2. MinION sequencing

1 µg DNA was fragmented using Covaris g-tube centrifuged at 5000 × g for 60 s. 5 µl lambda phage spike-in DNA (CS, ONT) was added to each sample. Fragments were end-repaired and adenylated using NEXTflex Rapid DNaseq kit (Newmarket Scientific #5144-02), purified and concentrated using Ampure XP beads (Beckmann Coulter). Size distribution was checked on a Bioanalyser 7500 DNA chip (Agilent Technologies) (Supplementary Fig. 1) and the concentration determined using the Qubit BR assay (Life

Technologies) before pooling DNA from each species in 50 µl: *S. avermitilis* 576 ng, *B. burgdorferi* 560 ng and *E. coli* 530 ng.

The ONT protocol was followed unless indicated and all reactions carried out at room temperature. Adapters were ligated to the adenylated DNA and purified using 0.4 × volume Ampure XP beads (Beckman Coulter); beads were washed with ONT-supplied wash buffer, and eluted in 25 µl ONT supplied elution buffer. Tether was annealed for 10 min and the library conditioned with the HP motor for 30 min. This pre-sequencing mix was stored briefly on ice. Immediately before sequencing, 6 µl pre-sequencing mix, 140 µl EP and 4 µl fuel mix were mixed very gently before loading on to the MinION flowcell. Additional input material was added to the MinION flowcell at 16 h 33 min.

### 2.3. MiSeq sequencing

For each species (*S. avermitilis*, *B. burgdorferi* and *E. coli*) Illumina fragment libraries were prepared and those containing insert sized averaging 550 bp were selected. DNA was sequenced (300 bp Paired End) on a MiSeq using v3 reagents. Supplementary Table 1 details the number of reads produced for each species.

Data available at the SRA: *B. burgdorferi* SRR1772332, *E. coli* SRR1770413, *S. avermitilis* SRR1770414.

### 2.4. Alignment of MinION reads against reference genome sequences

After sequencing and base calling reads were converted to fasta using Poretools [17] then aligned against a database of the closest available reference genomes for those species: *B. burgdorferi* ATCC 31267 (NC.001318) [5], *S. avermitilis* ATCC 35210 (NC.003155) [10] and *E. coli* strain MG1655 (NC.000913) [26], plus the 3.56 kb sequence of the lambda phage spike-in. The alignment of the MinION reads to the reference genomes was carried out using the LAST alignment software [6,12], as in [24]. The best alignment for each read was selected based on alignment score. Using LAST we aligned 12,632 reads (26.8%) and 38280405 (40.7%) bases. LAST was designed to cope well with long error-prone reads, resulting in higher mapping rates than alignment software designed for short high-fidelity reads such as BWA [15] or Bowtie2 (Langmead and Salzberg, 2012). An update for BWA mem [15] has been released designed for ONT reads. While its author suggests its performance will typically still be inferior to LAST [16] our results suggest the alignment rate is comparable, making it another viable option for aligning MinION reads (Supplementary Table 2).

### 2.5. Calculation of error rates from LAST sequence alignments

To calculate the error rate we counted the number of mismatch positions in the gapped alignment of a read to a reference sequence, thus it is a measure of substitution, insertion and deletion errors. The error rates were then expressed as a percentage of the length of reference sequence aligned against. Some recorded errors may in fact be genuine differences between our DNA samples and the published reference genome sequences, either due to real polymorphism or errors in the published reference sequences. To estimate the frequency of such false-positive errors we re-sequenced each of our genomic DNA samples using the Illumina MiSeq and hence ascertained the number of discrepancies between our DNA samples and the published reference sequences. The MiSeq reads were aligned against the reference genome sequences using Bowtie2 (Langmead and Salzberg, 2012) and differences to the references were evaluated using SAMtools and BCftools [14]. Table 1 shows the number of short variations between our data and the published reference genomes, these suggest that approximately 0.009% of the 'errors' in the MinION data are not errors but genuine differences.

**Table 1**  
Differences to published references genomes.

Species	SNPs	Indels
<i>S. avermitilis</i>	722	148
<i>E. coli</i>	402	17
<i>B. burgdorferi</i>	144	16

Clearly this small number of false-positive errors does not substantially affect the overall estimate of sequencing error-rate.

### 2.6. Calculating G+C content versus coverage

To investigate a potential bias against extreme G+C sequences we split the *E. coli* and *B. burgdorferi* genomes into 1000 bp windows using BEDTools [25] then using the LAST alignment of the MinION reads against the reference genome sequences we evaluated the coverage depth of the alignment for those windows using BEDTools.

### 2.7. Assembling *E. coli* using MinION reads

The Illumina MiSeq *E. coli* paired end reads were combined where possible using FLASH [18] resulting in 71456 overlapped reads and 562512 uncombined paired end reads. We extracted the MinION reads that aligned to *E. coli*. We generated an assembly using Spades 3.5.0 [22] (ONT MinION specific setting) with these MinION reads and the Illumina MiSeq data. The assembly was evaluated using QUAST [8].

**Table 2**  
Summary statistics for the MinION reads.

Read type	Read count	Mean length (bp)	Standard deviation of length (bp)	Maximum length (bp)
Template	35,946	1951	3007	98,366
Complement	8270	1827	2549	44,769
Two direction	2877	3088	2958	28,365

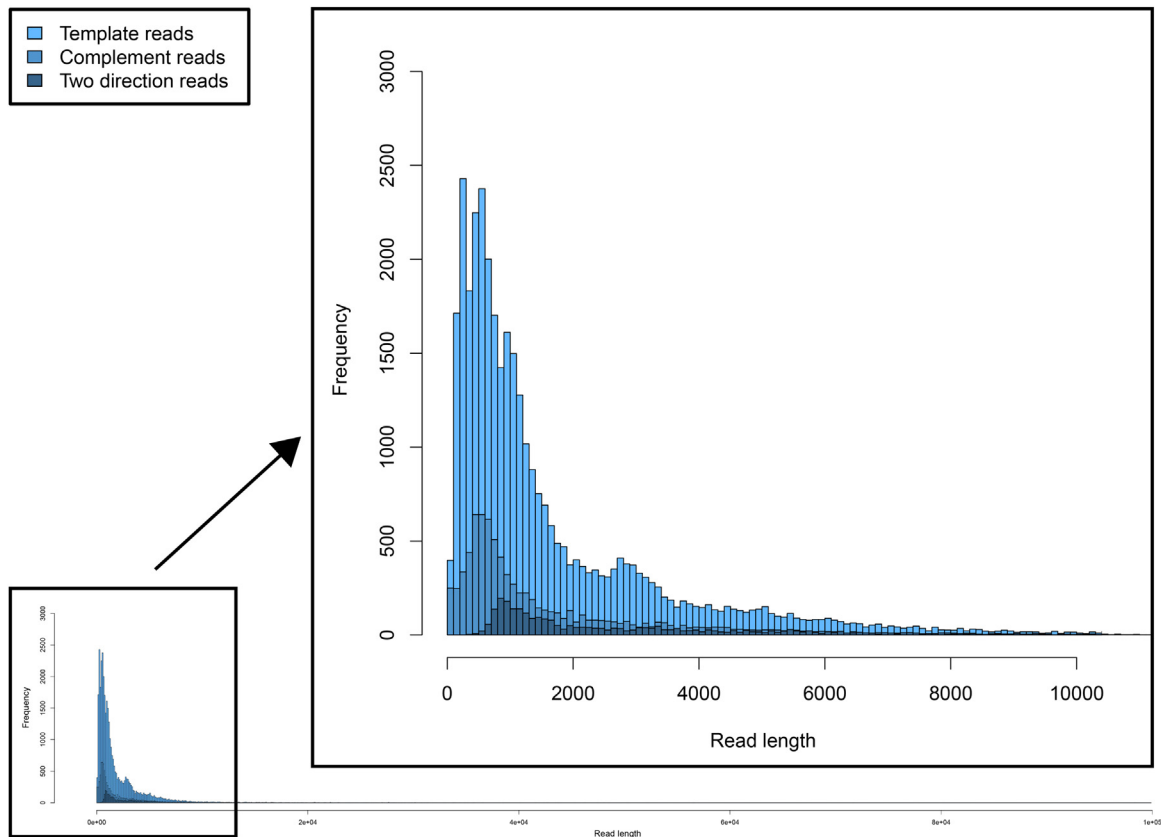
## 3. Results and discussion

### 3.1. Overview of sequence data

We constructed a sequencing library containing genomic DNA from three bacterial strains in equal quantities, as described in Section 2. This single MinION run generated Template sequence reads for 35,946 different DNA fragments, but only 23.0% produced Complement reads and only 8.0% yielded Two Direction reads (Table 2). The longest single read generated was 98,366 bp. As shown in Fig. 1 reads of this extreme length were the exception and not representative of the distribution; the majority of reads for all three read types have read lengths of less than 2000 bp.

### 3.2. *S. avermitilis* sequences were under-represented

By aligning MinION sequence reads against published reference genomes, we tried to assign each read to its most likely genome of origin (i.e. *B. burgdorferi*, *S. avermitilis* or *E. coli*). Reads from *S. avermitilis* were clearly under-represented (Table 3), as there were equal abundances (by mass) of each bacterial genome in the



**Fig. 1.** Distribution of MinION read lengths. Frequency distributions of lengths of reads obtained from the MinION run. Data shown for each of the three read types Template, Complement and Two direction, superimposed.

**Table 3**  
The number of reads of each type which aligned to each species.

Read type	<i>S. avermitilis</i>	<i>E. coli</i>	<i>B. burgdorferi</i>	Lambda	Unaligned
Template	226	2703	6752	1246	25,018
Complement	44	203	773	28	7222
Two direction	0	268	317	71	2221

sequencing library. Given the high G+C content of *S. avermitilis* compared to *B. burgdorferi* or *E. coli*, this suggests that G+C content may be the explanatory factor. However, this analysis does not exclude the possibility that some other property of the *S. avermitilis* DNA was responsible (e.g. methylation or other modification of the DNA). It is also not clear whether the under-representation arises from fewer *S. avermitilis* DNA molecules being sequenced (e.g. because they are out-competed for pores) or if the DNA was sequenced with a higher error rate resulting in lower alignment rates. The overall error rate of the aligned reads is 38.2% but is higher for the *S. avermitilis* reads (Table 4) (4.3 and 5.2 percentage points higher for the template and complement reads respectively). However when the aligned portions of all the reads are examined there is no clear correlation between G+C content and error rate (correlation coefficient of 0.198) (see also supplementary Figs. 2 and 3).

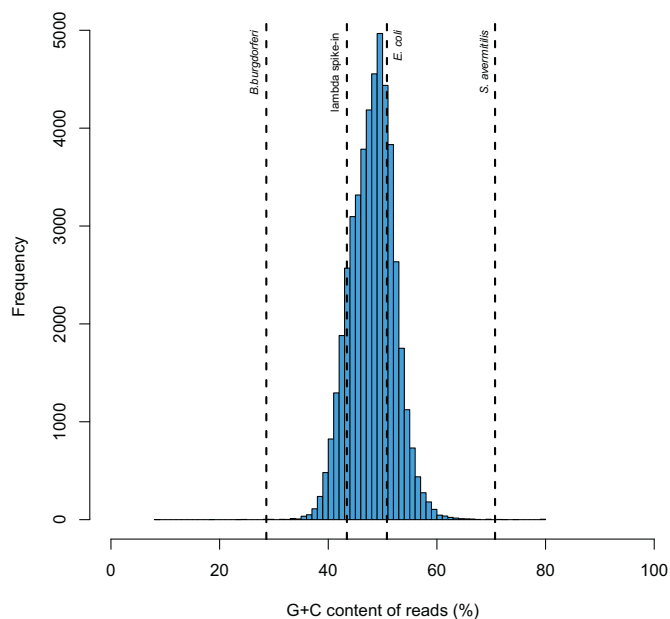
To further explore whether there was a bias against high G+C sequences we split the *E. coli* and *B. burgdorferi* genomes into windows and evaluated the relationship between coverage depth and G+C content. The correlation between high G+C and lower coverage was very weak for *E. coli* (correlation coefficient of  $-0.0171$ ) while for *B. burgdorferi* it was in the opposite direction (correlation coefficient of 0.444), suggesting that if there is any trend at all, it is that extreme G+C results in lower coverage (Supplementary Figs. 4 and 5). The lack of windows in *E. coli* and *B. burgdorferi* with G+C content as high as *S. avermitilis* prevents a true examination of the effect of extreme G+C using this method.

### 3.3. G+C content of reads is not the same as the sequence aligned to

The mean G+C content of the MinION reads is 47.2%. As shown in Fig. 2 the GC content of the reads does not correspond to the G+C content of all of the input genomes; extreme G+C sequences which would be expected to be generated from *B. burgdorferi* and *S. avermitilis* are not present in the reads. However as shown in Table, the most likely genome of origin for many reads is *B. burgdorferi*, suggesting that reads were in fact generated from this genome. The lack of extreme G+C reads appears to be due, at least in part, to the fact that the G+C content of the aligned portion of a read is different to that of the section of the reference to which it aligns (Fig. 3). As shown in Fig. 4 the distribution of G+C content for the aligned sections of reads (Fig. 4A) is different to that of the sections of reference sequence to which they align (Fig. 4B); the extremes of G+C content found in the reference seem to be shifted towards intermediate G+C in the reads. This could be caused by substitution errors in the sequencing effectively inserting random bases in the reads which will result in reads with more intermediate G+C content than the sequenced DNA fragment.

**Table 4**  
Error rate of reads split by type and species aligned to.

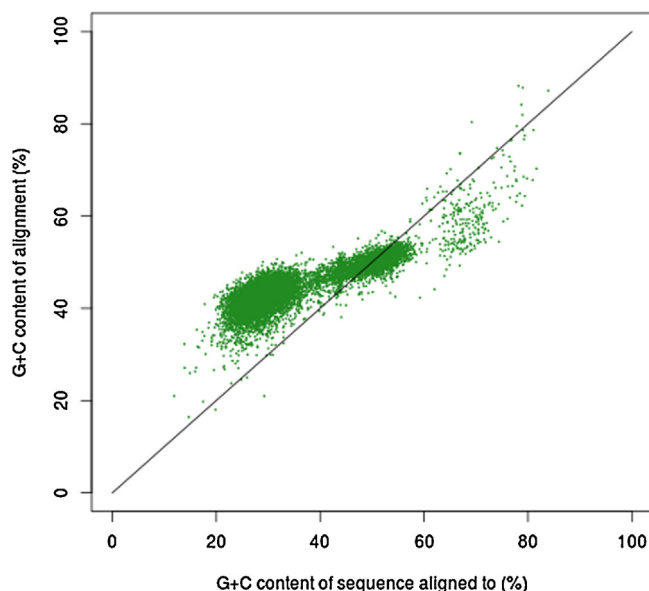
Read type	<i>S. avermitilis</i> (%)	<i>E. coli</i> (%)	<i>B. burgdorferi</i> (%)	Lambda (%)
Template	42.5	38.2	38.4	36.9
Complement	43.4	38.2	38.2	38.0
Two direction	NA	37.3	40.8	38.4



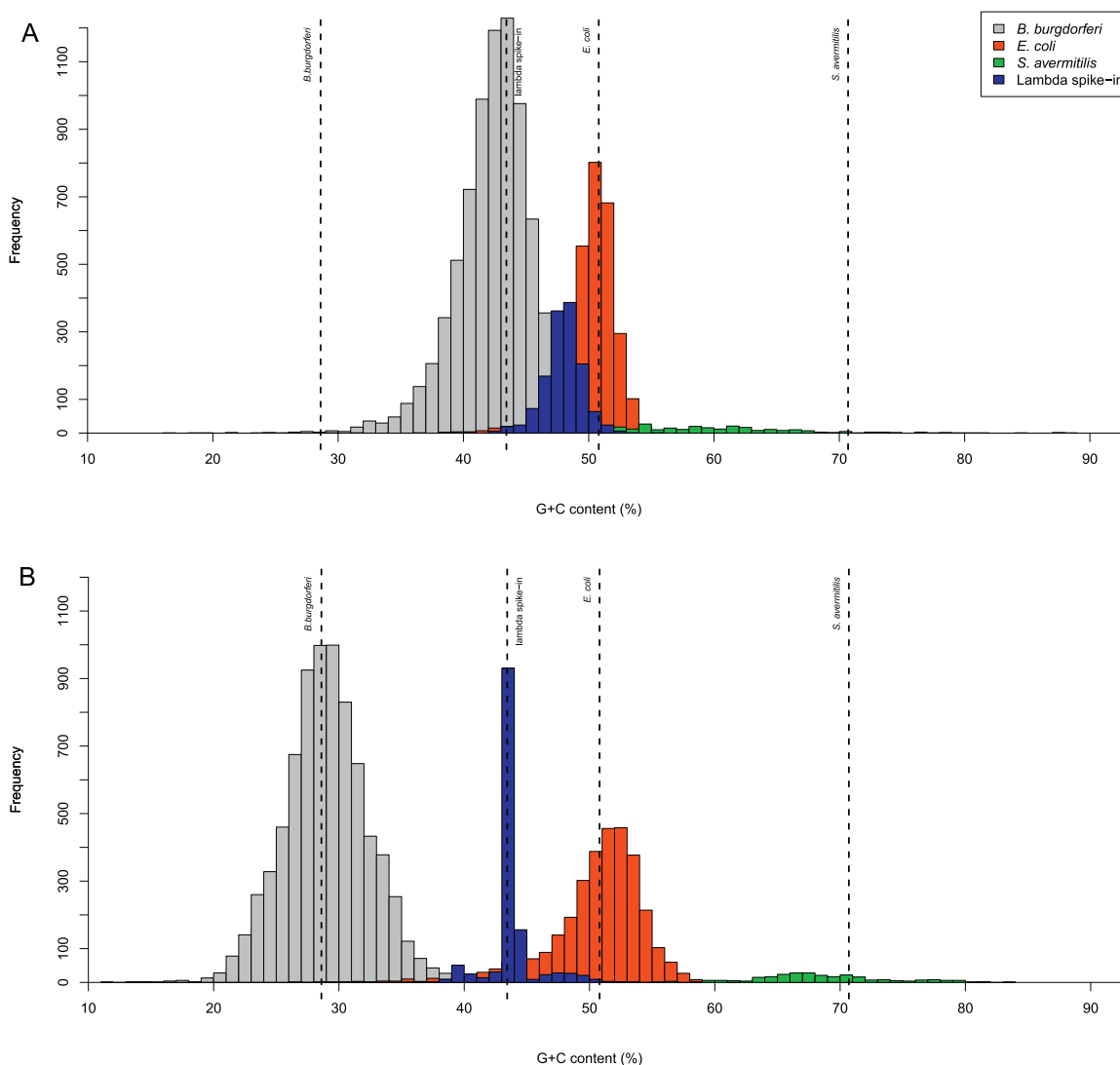
**Fig. 2.** G+C content of MinION reads. A frequency distribution of the G+C content of reads generated by the MinION run. Mean G+C content of each reference genome is included for comparison.

### 3.4. 25 genes covered by single MinION read

The long read lengths generated by the MinION have important possible applications not available to traditional short reads sequencing technologies. These reads (up to 98 kb in this study) are more than enough to span important genomic features such as secondary metabolite clusters, repeat rich regions and operons. Several interesting classes of bacterial genes are long and modular, made up of multiple partially repeated segments; these genes include non-ribosomal peptide synthase (NRPS) and TAL effectors. Because of the repetitive nature of these gene sequences, they are notoriously difficult to assemble using short-read sequencing



**Fig. 3.** G+C content of aligned portions of MinION reads against corresponding reference sequence. Plot of G+C content of the aligned portion of a read versus the G+C content of the section of the reference to which it aligns. Included is a line to demonstrate the relationship if the two were equal.



**Fig. 4.** G + C content of aligned portions of MinION reads and the reference sequence aligned to. Frequency distribution of the G + C content of aligned portions of the reads (A) and the G + C content of the sections of the reference genome they align to (B). Mean G + C content of each reference genome is included for comparison.

technologies. For example, Fig. 5A shows a section of the alignment generated from the MinION sequencing data. Highlighted is a single MinION read aligned to a 20,016 bp region of the reference genome spanning the entire length of one copy of the *E. coli* rDNA operon (5088 bp in length). Fig. 5B shows a NRPS gene cluster in the *E. coli* genome which is 53,661 bp in length and contains 49 genes. This MinION run has generated reads which span large portions of the cluster, one of which covers 28,134 bp of this NRPS cluster including 25 of its constituent genes. Repetitive regions are problematic when trying to assemble genomic data using short read sequencing technologies as it is not possible for one “short read” to span an entire region of interest [30].

### 3.5. MinION reads improved *E. coli* de novo assembly

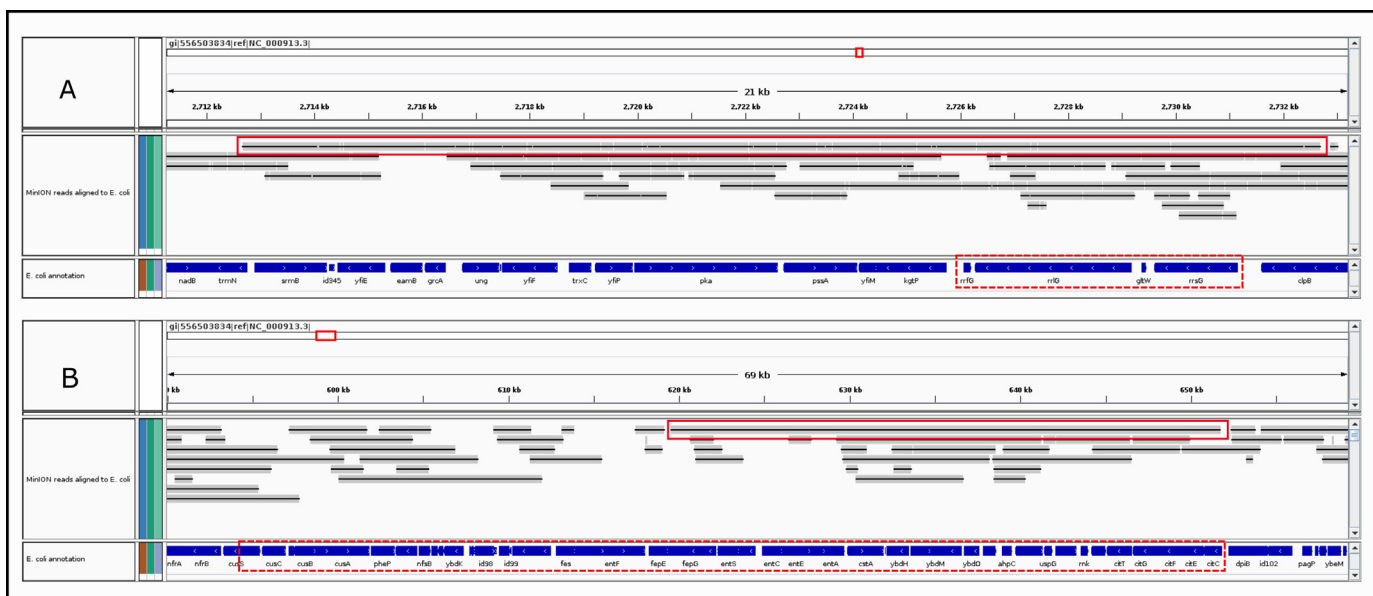
To demonstrate how the long reads produced by the MinION can be used to improve genome assemblies we extracted the MinION reads which aligned to the *E. coli* genome and used these in a combined assembly with Illumina MiSeq data. The resulting assembly had 84 contigs of at least 200 bp, a longest contig of 442595 bp and an N50 of 199079 bp compared to the assembly using only MiSeq data which contained 116 contigs, whose longest contig was 299472 bp with an N50 of 159445 bp. However when

the assemblies were evaluated using QUAST [8] the results show seven more misassemblies in the MinION aided assembly. These findings show that even at this early stage in the development of this technology, the MinION can offer substantial improvement in assembly length.

### 3.6. The error rate of the aligned reads remains constant over a MinION run

In order to evaluate the performance of the MinION over the duration of a run and whether there are characteristics of the data which vary over run time, a time series was generated. Higher mean read lengths were observed during the first 8 hours of operation (Fig. 6), perhaps suggesting that, if read length is your primary concern the initial stages of a run are optimal for this purpose. The alignment rate varies across the run time (Fig. 7) while the number of reads generated falls off towards the end of the run. However the error rate of the aligned reads remains relatively consistent throughout the run, suggesting that the quality of the data at the end of the run will not necessarily be any worse than at the beginning, so running the machine for as long as convenient will be beneficial rather than detrimental.





**Fig. 5.** Single MinION reads able to span important genes. Images generated using IGV [29] showing the alignment of MinION reads to the *E. coli* reference genome. (A) A rDNA operon. (B) A NRPS gene cluster. Highlighted with continuous red lines are the reads spanning the relevant sections and the dashed lines highlight the genomic regions of interest.

### 3.7. MinION quality scores do not follow the Phred scale

The per base quality scores of other sequencing technologies correspond with the Phred scale [4] where scores indicate a specific likelihood of error for that base; for example a Phred score of 20 indicates there will be 1 error for every 100 bases with that score. The MinION quality scores do not follow Phred expected error rates; the same quality score for the MinION does not equate to the same error rate as Phred (see supplementary Fig. 8).

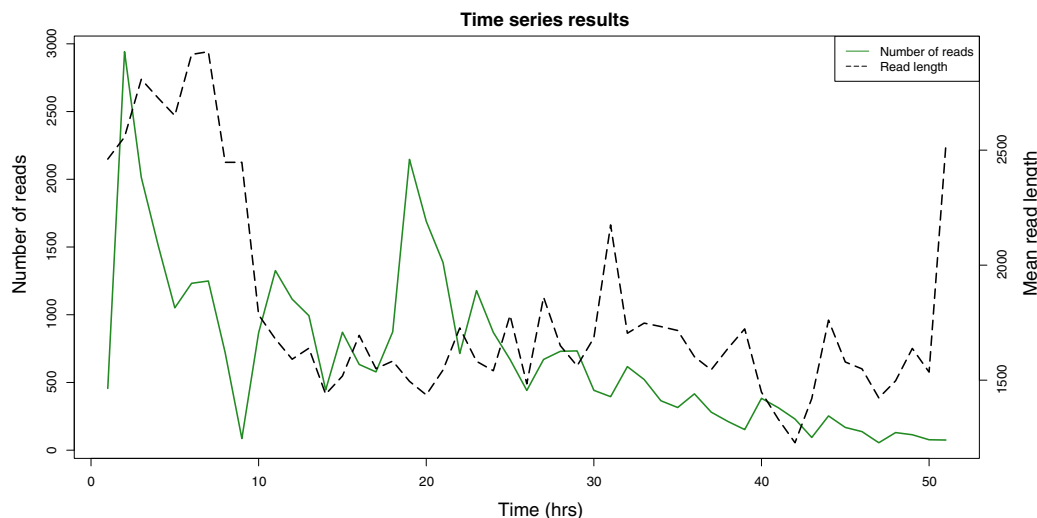
### 3.8. Comparisons to publically available MinION data

The error rates measured on our MinION data are similar to those for other public data on the MinION. Using our methods on the data published by Mikheyev and Tin [19] we calculated their error rate for single direction reads as 40.2% based on 35.1% read aligned (25.4% bases aligned), while 32.7% of their Two Direction reads aligned (11.4% bases aligned) with 40.1% error. This data was

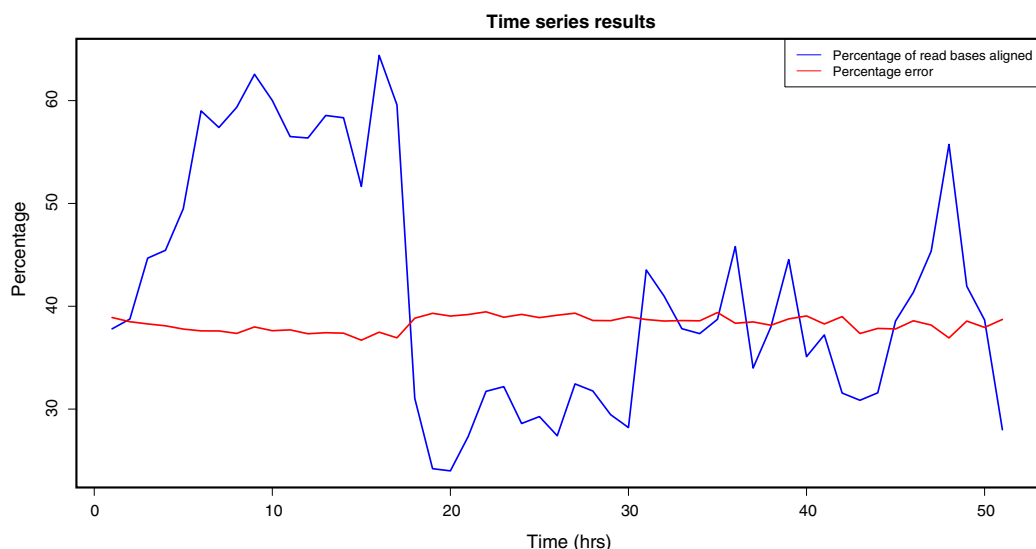
generated using the same R6 MinION chemistry as the data published in this study.

Due to the experimental nature of the MinION, the sequencing chemistry is rapidly evolving. The data presented in this study was generated using R6 sequencing chemistry; to explore if our results for error rate and the effect of G+C content held true for the R7 chemistry we evaluated data from [24]. Re-analysing this data with our methods resulted in 57.8% of template reads aligned (55.4% of bases aligned), with 37.5% error, but more promisingly their High Quality Two Direction reads resulted in 82.5% of reads aligned (82.3% of bases aligned), with 26.6% error. This suggests that the error rate for the high quality reads is improving as the technology evolves. As we have already been able to demonstrate that MinION reads can both cover biologically important genes and be used to generate improved genome assemblies the technology will only have more applications as it improves.

To explore if the issues with extreme G+C content sequences that were suggested by our data were still present for the updated



**Fig. 6.** Read data over time during the MinION run. Plot of number of reads and their mean length generated per hour during the MinION. Additional input material was added to the MinION flowcell at 16 h 33 min.



**Fig. 7.** Fluctuation in alignment and error rates over time during a MinION run. Plot showing percentage of read bases aligned per hour during the MinION run based on the alignment by LAST and their error rate. Additional input material was added to the MinION flowcell at 16h 33 min.

chemistry we repeated our evaluation of coverage versus G + C content across windows of the *E. coli* genome for the R7 data. The results suggest a weak correlation between G + C content and depth of coverage (correlation of coefficient of  $-0.141$  for Template reads and  $-0.0816$  for High Quality Two Direction) a similar finding to our results gained from the R6 chemistry (Supplementary Figs. 6 and 7).

#### 4. Conclusions

Our results demonstrate that in spite of its high error rate the MinION is able to generate extremely long reads, is able to span regions of interest in a single read and is able to improve the contiguity of genome assemblies. As well as the high error rate, the MinION's possible difficulties with high G + C content sequences, demonstrated in this study, will also need to be addressed before the device is put into widespread use.

Our analysis of data generated by [24] on the R7 MinION chemistry suggests that the error rate for the High Quality Two Direction reads is improving as the technology evolves, although we suggest the potential issues with sequencing extreme G + C sequences is still present. The lower error rate generated from the Two Direction reads produced with the updated MinION chemistry gives cause for optimism that future version of the MinION might be able to generate reads with a greatly reduced error rate while still retaining the long read length and low per unit costs that make this such an exciting technological prospect.

#### Acknowledgments

This work was funded by the University of Exeter Sequencing Service as part of the Oxford Nanopore MAP programme. The University of Exeter Sequencing Service is supported by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1). TL was supported by the BBSRC Industrial Case Studentship award BB/H016120/1. JH was supported by a BBSRC PhD studentship (BB/F017367/1).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.bdq.2015.02.001](https://doi.org/10.1016/j.bdq.2015.02.001).

#### References

- [1] Aird D, Ross MG, Chen W, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 2011;12(2):R18.
- [2] Ashton PM, Nair S, Dallman T, Rubinio S, Rabsch W, Mwaigwisya S, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat Biotechnol* 2015;33:296–300.
- [3] Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinform* 2014;15(211):1–9.
- [4] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using Phred I accuracy assessment. *Genome Res* 1998;8(3):175–85.
- [5] Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 1997;390:580–6.
- [6] Frith MC, Hamada M, Horton P. Parameters for accurate genome alignment. *BMC Bioinform* 2010;11(80):1–14.
- [7] Goodwin SS, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz M, McCombie R. Oxford nanopore sequencing and de novo assembly of a eukaryotic genome. *bioRxiv* 2015.
- [8] Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29(8):1072–5.
- [9] Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, Hon L, et al. Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res* 2014;24:688–96.
- [10] Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, et al. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* 2003;21(5):526–31.
- [11] Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA* 1996;93(24):13770–3.
- [12] Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res* 2011;21(3):487–93.
- [13] Laszlo AAH, Derrington IM, Ross BC, Brinkerhoff H, Adey A, Nova IC, et al. Nanopore sequencing of the phi X 174 genome. *Quant Biol* 2014:1–39.
- [14] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25(16):2078–9.
- [15] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics* 2013:1–3.
- [16] Li H. BWA-MEM for long error-prone reads; 2014 [Online] Available from: <http://lh3.github.io/2014/12/10/bwa-mem-for-long-error-prone-reads/> [accessed 09.01.15].
- [17] Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* 2014;30(23):3399–401.
- [18] Magoç T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011;27(21):2957–63.
- [19] Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour* 2014;14(6):1097–102.
- [20] Nanoporetech.com [1]. The MinION device: a miniaturized sensing system; 2014 [Online] Available from: <http://tinyurl.com/m6uboj> [accessed 26.11.14].
- [21] Nanoporetech.com [2]. A guide to MAP; 2014 [Online] Available from: <http://tinyurl.com/q86a72v> [accessed 26.11.14].
- [22] Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 2013;20(10):714–37.

- [23] Olasagasti F, Lieberman KR, Benner S, Cherf GM, Dahl JM, Deamer DW, et al. Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat Nanotechnol* 2013;5(11):798–806.
- [24] Quick J, Quinlan AR, Loman NJ. A reference bacterial genome dataset generated on the MinION portable single-molecule nanopore sequencer. *GigaScience* 2014;3(22):1–6.
- [25] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26(6):841–2.
- [26] Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, et al. *Escherichia coli* K-12: a cooperatively developed annotation snapshot-2005. *Nucleic Acids Res* 2006;34(1):1–9.
- [27] Satou K, Shiroma A, Teruya K, Shimoji M, Nakano K, Juan A, et al. Complete genome sequences of eight *Helicobacter pylori* strains with different virulence factor genotypes and methylation profiles, isolated from patients with diverse gastrointestinal diseases on Okinawa Island, Japan, determined using PacBio single-molecule real-time technology. *Genome Announc* 2014;2(2):1–2.
- [28] Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci USA* 2009;106(19):7702–7.
- [29] Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14(2):178–92.
- [30] Todd J, Saltzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2012;13:36–46.
- [31] Utturkar SM, Klingeman DM, Land ML, Schadt CW, Doktycz MJ, Pelletier DA, et al. Evaluation and validation of *de novo* and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics* 2014;30(19):2709–16.