Oxford Nanopore MinION sequencing enables rapid whole-genome assembly of *Rickettsia typhi* in a resource-limited setting

Ivo Elliott[1,2*], Elizabeth M. Batty[2,3,4*], Damien Ming[1], Matthew T. Robinson[1,2], Pruksa Nawtaisong[1], Mariateresa de Cesare[3], Paul N. Newton[1,2†], Rory Bowden[3†]

Affiliations:

1. Laos-Oxford-Mahosot Hospital-Wellcome Trust Research Unit, Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao PDR
2. Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
3. Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
4. Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

Corresponding author: Ivo Elliott ivo@tropmedres.ac

* Contributed equally.

† Co-supervised the project.

Word count: 3075

Abstract

The infrastructure challenges and costs of next-generation sequencing have been largely overcome, for many sequencing applications, by Oxford Nanopore Technologies' portable MinION sequencer. However the question remains open whether MinION-based bacterial whole-genome sequencing (WGS) is by itself sufficient for the accurate assessment of phylogenetic and epidemiological relationships between isolates and whether such tasks can be undertaken in resource-limited settings. To investigate this question, we sequenced the genome of an isolate of *Rickettsia typhi*, an important and neglected cause of fever across

much of the tropics and subtropics, for which only three genomic sequences previously existed. We prepared and sequenced libraries on a MinION in Vientiane, Lao PDR using v9.5 chemistry and in parallel we sequenced the same isolate on the Illumina platform in a genomics laboratory in the UK.
The MinION sequence reads yielded a single contiguous assembly, in which the addition of Illumina data revealed 226 base-substitution and 5,856 in/del errors. The combined assembly represents the first complete genome sequence of a human *R. typhi* isolate collected in the last 50 years and differed from the genomes of existing strains collected over a 90-year time period at very few sites, and with no re-arrangements. Filtering based on the known error profile of MinION data improved the accuracy of the Nanopore-only assembly. However, the frequency of false-positive errors remained greater than true sequence divergence from recorded sequences. While Nanopore-only sequencing cannot yet recover phylogenetic signal in *R. typhi*, such an approach may be applicable for more diverse organisms.

Introduction

Until recently, whole-genome sequencing (WGS) has been the preserve of high-income settings. Although the costs of WGS have dramatically decreased over the past decade[1], the need for initial investment in sequencing platforms and associated equipment and for the technical expertise to run and maintain them have prevented their introduction into lower income settings. Downstream processing of sequencing data is frequently hampered by poor Internet connectivity, a lack of local expertise and the perceived requirement for substantial computational infrastructure.

Oxford Nanopore Technologies (ONT) MinION sequencer is a portable device for DNA and RNA sequencing that generates data for local analysis, in real time. The MinION weighs under 100g and plugs directly into a laptop via a USB port, with no additional computing infrastructure required[2]. Only basic laboratory facilities are needed to extract DNA and prepare sequencing libraries. The MinION has been used to sequence viral[3], bacterial[4,5] and eukaryotic[6-9] genomes. In spite of

its high raw error rate, Nanopore data can in many cases produce highly contiguous assemblies. Illumina short-read data can be added to improve consensus accuracy, in many cases to that of a high-quality draft assembly, without extra finishing steps.

The Lao People's Democratic Republic (Laos) is a landlinked country of ~7 million people in Southeast Asia. Laos is ranked 135th in the world on the Human Development Index (HDI), a composite statistic of life expectancy, education and per capita income[10]. There are few functioning molecular biology laboratories in Laos and no previous published WGS projects.

Rickettsia typhi is an obligate intracellular, Gram-negative bacterium in the family Rickettsiaceae that causes the disease murine typhus and has a worldwide tropical and sub-tropical distribution. The pathogen is transmitted to humans primarily by flea faecal contamination of the bite of the Oriental rat flea *Xenopsylla cheopis.* The disease is an important and grossly under-recognized global cause of febrile illness. Those affected suffer from high fever, headache, myalgia, arthralgia, nausea and vomiting and may have a macular rash. Complications are infrequent but include myocarditis, meningoencephalitis, seizures and renal failure[11]. When diagnosed, recovery is typically rapid after treatment with doxycycline and mortality is low (around 1-2%) with antibiotic treatment[12].

Despite its importance as a pathogen, to date just three *R. typhi* whole-genome sequences, with wide geographic and temporal distribution, have been published. The type-strain, Wilmington, was isolated from a patient in North Carolina in 1928 and the sequence was published in 2004[13,14]. Two further sequences, originating from a patient in Northern Thailand collected in 1965 and the other from a bandicoot rat, *Bandicota* sp., collected in Burma (now Myanmar) in the 1970s [15-17] were published online in 2014 (Table 1).

We report the successful whole genome sequencing of a human isolate of *R. typhi* using only the MinION platform in Laos. We assess the potential for use of ONT

data alone to perform comparative analyses, and discuss the challenges of undertaking WGS in resource-limited settings.

Materials and Methods

R. typhi culture and DNA extraction

Frozen L929 mouse fibroblast cells (ATCC CCL-1), infected with *R. typhi* isolated from the blood of a patient, TM2540, from Pakkading District (~18.33° N 104.0° E), Bolikhamxay Province, Laos with suspected acute murine typhus infection presenting in 2012 to Mahosot Hospital, Vientiane, were reanimated. Frozen aliquots were briefly thawed at room temperature and then transferred to 25 cm^2 cell culture flasks containing a L929 cell monolayer at 80% confluence in RPMI 1640 medium (Gibco, USA), supplemented with 10% foetal calf serum (Sigma-Aldrich, USA). Flasks were incubated at 35 °C in 5% CO_2 atmosphere for seven days, then cells were mechanically detached and transferred to a 75 cm^2 flask and again incubated at 35 °C in 5% CO_2 atmosphere.

For DNA extraction, cells from three 75 cm^2 flasks were mechanically detached, re-suspended and transferred to a 50 ml conical-bottom centrifuge tube and centrifuged at 3,220×*g* for 10 minutes. The pellet was re-suspended in 3ml fresh medium and transferred to 1.5 ml micro-centrifuge tubes. Tubes were vortexed for one minute, then centrifuged at 300×*g* for 3 minutes. The supernatant was passed through a 2 µm filter (Corning, USA), then mixed with 10 µl/ml DNAase (1.4 µg/µl) and incubated at room temperature for 30 minutes. The mixture was centrifuged at 18,188×*g* for 10 minutes and washed twice with 0.3 M sucrose (Sigma-Aldrich, USA). DNA was extracted from the *R. typhi* pellet using the DNeasy Blood & Tissue kit (Qiagen, USA) and eluted products were stored immediately at -20 °C.

DNA was quantified using Qubit dsDNA High Sensitivity assay kit (Thermofisher, USA) following manufacturer's protocols and assayed for *R. typhi* sequences by quantitative polymerase chain reaction (qPCR) targeting the 17kDa outer membrane antigen[18].

Preparation of MinION sequencing libraries

MinION sequencing libraries were produced using the ONT 1D Genomic DNA by ligation (SQK-LSK108) protocol. Briefly, 1.4 µg of DNA was fragmented in a Covaris g-TUBE (Covaris Ltd., Brighton, U.K.) by centrifugation. Sheared DNA was repaired using NEBNext FFPE repair mix (New England Biolabs, MA, USA). End-repair and dA-tailing were performed with the NEBNext Ultra II End Repair/dA-tailing module. Adapter ligation used the NEB Blunt/TA Ligase Master Mix, and the library was purified using Agencourt AMPure XP beads (Beckman Coulter Inc., CA, USA).

MinION sequencing in Laos

MinION libraries were sequenced for 48 hours on an ONT MinION R9.5 flow cell, connected to a Dell Latitude E5470 XCTO laptop with 256GB SATA Class 20 solid state drive. ONT fast5 data files were base-called using the ONT Albacore module.

Illumina sequencing in Oxford

An Illumina sequencing library was generated from the same sample of *R. typhi* DNA using the Nextera (Illumina) library preparation method. The library was sequenced on the Illumina MiSeq with 2×260bp reads. A total of 1,201,068 read pairs were sequenced, giving 312Mbp of total sequence data.

Bioinformatic analysis

The species composition of the MinION reads was identified using Centrifuge software[19] testing against the prebuilt non-redundant database and keeping only the best hit per read.
Assembly strategies that used MinION data alone, or in combination with Illumina short reads, were assessed. A draft genome assembly was generated

from the MinION reads using Canu[20] with the suggested parameters for ONT sequencing "correctedErrorRate=0.120 -nanopore-raw" and an estimated genome size of 1.1Mb, and polished using Nanopolish[4]. The Illumina reads were mapped to the Canu+Nanopolish assembly using bwa 0.7.12[21] and the mapped reads were used to error-correct the assembly using Pilon v1.22[22]. To check for enrichment of errors in regions of low coverage, the ONT reads were mapped onto the corrected assembly using bwa mem to determine coverage and the coverage at error positions was compared to the total coverage distribution. To look for enrichment of errors near homopolymers, we compared the distance to the nearest homopolymer of 5bp or longer in length for each corrected error to a random sample of positions in the genome. BUSCO v3 using the OrthoDB version 9 bacterial dataset was used to assess gene content[23]. The Illumina short reads alone were assembled using Unicycler[24]. Prokka v1.14[25] was used to annotate both the new assemblies and the existing *R. typhi* genomes to give consistent data for comparison. The Wilmington reference strain was used to train the gene model for the Prodigal gene prediction used in Prokka, and this training set was used to annotate the other three genomes. The new genome was rotated to begin with the *yqfL* gene for consistency with other *R. typhi* genomes.

Data availability

The three existing *R. typhi* genomes were downloaded from RefSeq under the accession numbers NC_006142.1, NC_017066.1 and NC_017062.1. The sequence data is available at the ENA under project PRJEB2567. The new strain is named TM2540 and the assembly is available at the ENA under accession ERZ497871.

Results

MinION flow cells and reagents were shipped from Oxford, United Kingdom to Laos at +4°C. On dispatch 1,242 active pores were available and on receipt after 72hrs travel, 1,103 active pores were available. Sequencing using the MinKNOW platform was performed for 48 hours and generated approximately 250,000

reads, with a total fast5 file size of 20GB. 222,848 reads passed quality filters and were used in further analysis.

We assessed the species composition of the reads which passed quality filters. Since the *Rickettsia typhi* sample was grown in L929 fibroblast cells, we expected to see some contaminating reads from the mouse genome. 134,802 (60%) of the reads were classified as belonging to the genus *Rickettsia*. 126,447 reads were classified as *Rickettsia typhi* at the species level and 41,428 (19%) were classified as belonging to the genus *Mus* (Figure 1).

The FASTQ files produced by Albacore were assembled using Canu, an assembler designed for noisy single-molecule sequencing such as that produced by the MinION. This produced a single-contig assembly of 1,078,916bp. Nanopolish, a tool to use the signal-level data from Oxford Nanopore sequencing, was used to improve the consensus sequence of the Canu assembly, producing a new assembly of 1,099,322bp. To further correct errors in the assembly produced from long reads alone, we polished the Canu+Nanopolish genome assembly using Pilon, which uses short read sequencing data to correct errors. We repeated rounds of mapping and correction with Pilon until no further errors were corrected. After 4 rounds of correction, Pilon did not correct any further errors, with the exception of two short indels which were removed and then re-inserted by successive rounds of polishing. 94.2% of errors were corrected by the first round of polishing, with a further 4.8% corrected in the second round (Supplementary Table 1).

After this process, a total of 6,082 errors were corrected, with over 95% (5,856) being small insertions and deletions, with the result that 12kb was added to the genome during polishing. In comparison, when Pilon was run on the Canu draft genome before Nanopolish polishing, 19,970 errors were corrected, confirming that Nanopolish improves the draft genome.

We assessed the ONT read coverage and proximity to a homopolymer run (5 bases or longer) for each of the errors we corrected with Pilon. We observed no

difference in the coverage distribution at sites with errors compared to the overall coverage distribution (Supplementary Figure 1). However, the positions around homopolymer runs were enriched for errors (Figure 2) – 15% of the final polished genome is within 5bp of a homopolymer run (165kb), while 43% of errors (2,641) fall into these regions.

The Illumina data alone was assembled using Unicycler to compare with assemblies using ONT only and the combined ONT and Illumina data. The assembly generated by Unicycler had 29 contigs with a total length of 1,161,002bp, and an n50 of 166kb. However, this included 20 short contigs under 6kb, which are closely related to the mouse genome by BLAST search. These contigs are identified by their low coverage depth relative to *Rickettsia typhi* contigs. After removal of contaminating mouse contigs, the Unicycler assembly has 8 contigs with a total length of 1,111,455bp.

BUSCO was used to quantitatively measure the genome completeness of our assemblies based on a set of 148 conserved single-copy genes (Table 2). None of the genes were found in the initial Canu assembly. Inspection of the TBLASTN results shows that while there were hits to many of the conserved genes, they fall below the threshold where BUSCO will call them as present. After polishing with Nanopolish using ONT data only, 14.2% of the genes could be found as a complete copy, with a further 32.4% found as fragmented genes. Two rounds of Pilon polishing improves this to 88.5% of genes found as complete copies. Running BUSCO on the Wilmington reference strain gives the same number of complete genes, suggesting this is the complete set of the orthologous reference genes that exists in *R. typhi.* The Unicycler assembly also contains this complete set of genes.

The final polished genome assembly is 1,111,939bp, extremely similar to the three available genomes, which range from 1,111,496-1,112,957bp in length (Table 1). Alignment of the four whole genomes in Mauve[26] shows that the genomes are co-linear and that no rearrangements have taken place between strains (Supplementary Figure 1).

The TM2540 genome contained 866 predicted genes, very close to the 865 genes predicted in all other strains. Roary was used to define the core genome (genes present in all strains at 90% identity) of the four *R. typhi* strains, which included 863 genes. Of the 5 accessory genes not found in all strains, two were found in three of the four strains, and three were unique to a single strain. Two of the genes not found in all samples were the same gene, but clustered separately due to a 150bp insertion in the B991CWPP strain that caused the sequence to fall below the 90% identity threshold to cluster with the other three strains. In two of the five accessory genes, a single base insertion in one strain has created a premature stop codon, splitting a gene into two sequences in that strain. In the final case, the gene is present in three strains next to a duplicate pseudogene which has lost the start and end of the gene, while in the fourth strain a deletion has removed the end of the pseudogene and the start of the gene, leaving two incomplete and presumably inactive copies of the gene. All of the 5 variably present genes are annotated as hypothetical genes.

Table 3 shows the SNP differences between strains determined from a whole genome alignment, with 15-23 SNPs separating each strain.

We compared the results of the final polished assembly to the best assemblies generated using Illumina data alone (filtered Unicycler assembly) or ONT data alone (Canu + Nanopolish assembly). The Illumina assembly has 866 predicted genes, and the core genome using this assembly has 862 core genes and 7 accessory genes. This includes one gene unique to the Illumina assembly, which is part of a gene split into two pieces at a contig boundary. The ONT assembly has 1745 predicted genes, with many genes duplicated in tandem, likely due to sequence errors, and the core genome using this assembly has only 774 genes.

Discussion

WGS of pathogens is an increasingly important tool in both research and clinical settings, whose use has increased over the last decade as its cost has decreased

and availability improved. However, the initial investment in sequencing and computing equipment and the technical expertise needed to produce and analyse data remain a barrier to its introduction into resource-poor settings. Laos has very few clinical or research laboratories and relatively poor Internet connectivity[27]. Although our research laboratory can perform molecular diagnostics and cell culture, to date there has been no sequencing capacity, necessitating the slow and costly shipment of samples to other countries to undertake sequencing projects. The portable nature of the MinION sequencer has allowed the first bacterial WGS to be performed in Laos. We were able to run all steps of our bioinformatic analysis (except Centrifuge) on a laptop computer (MacBook Pro 2017) without continuous Internet access, demonstrating that sequencing and analysis can take place in relatively remote settings.

Murine typhus is an important and severely neglected tropical and subtropical disease of worldwide distribution. In Vientiane an estimated 10% of non-malarial fevers in adult inpatients is caused by *R. typhi*[28]. Surprisingly, for a disease to which many millions of people are potentially exposed, there are only three published whole genome sequences, of which the last human isolate dates back more than 50 years.

The *R. typhi* genome is 1.1Mb and contains almost no repeats, which makes genome assembly for this species relatively simple. Running Canu with the default ONT parameters was sufficient to assemble the genome into a single contig and we polished the assemblies using both Nanopolish, to utilize the signal-level data from ONT sequencing, and Pilon, to use the extra information available from Illumina sequencing with its lower error rate. While its synteny with other genomes suggests that there are no large misassemblies in our genome, the two methods of polishing corrected many small SNP and indel errors in the ONT only draft genome. Thus even for relatively simple bacterial genomes, a combination of ONT with data from another technology with a low error rate is currently necessary to produce an accurate sequence. In the case of *R. typhi*, the strains are so closely related that to distinguish between them requires a highly accurate sequence, and using ONT data alone the sequencing

errors are more numerous than the true differences between strains. Removing regions of the genome close to homopolymer runs would remove a large number of positions with errors by filtering out a comparatively small amount of the genome, but would still leave over 2,000 more errors than true differences. ONT data alone may suffice for applications which can work with sequences which still retain some base-pair level errors, such as determining large-scale genome rearrangements[29], determining the species of an unknown sample[30,31], or detecting antimicrobial resistance genes[32]. Increases in sequencing accuracy, combined with bioinformatics advances in assembly and polishing software, may allow for the future use of ONT data alone to give complete and accurate sequences, increasing the applications for this device.

Comparative analysis of the four *R. typhi* genomes shows very little variation between strains. Very few genes are variably present between strains, and gene order is completely conserved. Although these four samples were collected from humans and a rodent over a long timeframe, we see only 23 SNPs between the two most divergent strains, no large-scale genome rearrangements, and very few indels. The low number of differences suggests that whole-genome sequencing may not be a useful tool to analyse the relationships between strains of *R. typhi*, as there may be few differences between strains and no recent shared ancestry.

Some limitations remain on the use of ONT's MinION platform. Currently individual flow cells remain costly at USD500-USD900 each[33], depending on the number purchased. However, the cost per sample can be reduced since multiplex sequencing now allows for multiple bacterial strains to be sequenced on the same flow cell[34]. Ongoing improvements in the technology continue to increase overall output and per-run capacity. Novel technologies are likely to improve access to sequencing when computational resources and support are limited and allow more cost-efficient sequencing of single samples[35].

MinION instruments continue to be used in an ever-widening set of challenging circumstances[3,36-41]. Here we demonstrate the current applicability of the MinION to resource-poor settings where some laboratory infrastructure exists,

but WGS capacity is unavailable. With concerns in some countries about the export of biological samples for WGS in other countries, MinION systems could facilitate countries without current WGS facilities to undertake such work.

Acknowledgments

We are very grateful to Associate Professor Bounthaphany Bounxouei, Director of Mahosot Hospital, the staff of the Microbiology Laboratory, LOMWRU and wards, Assistant Professor Chanphomma Vongsamphan, Director of Department of Health Care, Ministry of Health, and H.E. Professor Bounkong Syhavong, Minister of Health, Laos, for their help and support; Narongchai Tongyoo and Phonepasith Panyanouvong for growing the isolate in cell culture; Duncan Parkes for assistance with performing base-calling.

Financial Support

This study was supported by Ivo Elliott's Wellcome Trust Research Training Fellowship (105731/Z/14/Z and in part by Core Awards from the Wellcome Trust (090532/Z/09/Z and 203141/Z/16/Z).

Author contributions

Conflict of interests:

IE, PNN, EMB, MTR, PN None

References

1. Institute NHGR. The Cost of Sequencing a Human Genome - National Human Genome Research Institute (NHGRI).
2. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 2016; **17**(1): 239.

3. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016; **530**(7589): 228-32.
4. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* 2015; **12**(8): 733-5.
5. Risse J, Thomson M, Patrick S, et al. A single chromosome assembly of Bacteroides fragilis strain BE1 from Illumina and MinION nanopore sequencing data. *GigaScience* 2015; **4**(4): 60.
6. Istace B, Friedrich A, d'Agata L, et al. de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *GigaScience* 2017; **6**(2): 1-13.
7. Jansen HJ, Liem M, Jong-Raadsen SA, et al. Rapid de novo assembly of the European eel genome from nanopore sequencing reads. *Sci Rep* 2017; **7**(1): 7213.
8. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 2018.
9. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res* 2015; **25**(11): 1750-6.
10. Hdr.undp.org. Human Development Index (HDI). 2015. http://hdr.undp.org/en/content/human-development-index-hdi (accessed 14 November 2017).
11. Dittrich S, Rattanavong S, Lee SJ, et al. Orientia, rickettsia, and leptospira pathogens as causes of CNS infections in Laos: a prospective study. *Lancet Glob Health* 2015; **3**(2): e104-12.
12. Dumler JS, Walker DH. Rickettsia typhi (Murine Typhus). In: Mandell GL, Bennett JE, Dolin R, eds. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. 7th ed. Philadelphia, USA: Churchill Livingstone; 2010: 2525-8.
13. McLeod MP, Qin X, Karpathy SE, et al. Complete genome sequence of Rickettsia typhi and comparison with sequences of other rickettsiae. *J Bacteriol* 2004; **186**(17): 5842-55.
14. Maxcy KF. Endemic typhus fever of the southeastern United States: reaction of the Guinea Pig. *Public Health Reports* 1929; **44**(11): 589-600.
15. Institute USDoEJG. Rickettsia typhi TH1527 2014. https://gold.jgi.doe.gov/project?id=11636 (accessed 14 November 2017).
16. Sankasuwan V, Pongpradit P, Bodhidatta P, Thonglongya K, Winter PE. Murine typhus in Thailand. *Trans R Soc Trop Med Hyg* 1969; **63**(5): 639-43.
17. Institute USDoEJG. Rickettsia typhi B9991CWPP. 2014. https://gold.jgi.doe.gov/project?id=11634 (accessed 14 November 2017).
18. Jiang J, Chan TC, Temenak JJ, Dasch GA, Ching WM, Richards AL. Development of a quantitative real-time polymerase chain reaction assay specific for Orientia tsutsugamushi. *Am J Trop Med Hyg* 2004; **70**(4): 351-6.
19. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res* 2016; **26**(12): 1721-9.
20. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome Res* 2017; **27**(5): 722-36.

21. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. https://arxiv.org/abs/1303.3997 (accessed 30/01/2018 2018).
22. Walker BJ, Abeel T, Shea T, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* 2014; **9**(11): e112963.
23. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015; **31**(19): 3210-2.
24. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017; **13**(6): e1005595.
25. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**(14): 2068-9.
26. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010; **5**(6): e11147.
27. Speedtest. Speedtest Global Index. 2018. http://www.speedtest.net/global-index (accessed 13 August 2017).
28. Phongmany S, Rolain JM, Phetsouvanh R, et al. Rickettsial infections and fever, Vientiane, Laos. *Emerg Infect Dis* 2006; **12**(2): 256-62.
29. McGinty RJ, Rubinstein RG, Neil AJ, et al. Nanopore sequencing of complex genomic rearrangements in yeast reveals mechanisms of repeat-mediated double-strand break repair. *Genome Res* 2017; **27**(12): 2072-82.
30. Parker J, Helmstetter AJ, Devey D, Wilkinson T, Papadopulos AST. Field-based species identification of closely-related plants using real-time nanopore sequencing. *Sci Rep* 2017; **7**(1): 8345.
31. Batovska J, Lynch SE, Rodoni BC, Sawbridge TI, Cogan NO. Metagenomic arbovirus detection using MinION nanopore sequencing. *J Virol Methods* 2017; **249**: 79-84.
32. Schmidt K, Mwaigwisya S, Crossman LC, et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* 2017; **72**(1): 104-14.
33. . https://nanoporetech.com/products - modal=comparison (accessed 23 March 2018.
34. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* 2017; **3**(10): e000132.
35. . https://nanoporetech.com/products/flongle (accessed 09/05/2018 2018).
36. Bates M, Polepole P, Kapata N, Loose M, O'Grady J. Application of highly portable MinION nanopore sequencing technology for the monitoring of nosocomial tuberculosis infection. *Int J Mycobacteriol* 2016; **5 Suppl 1**: S24.
37. Menegon M, Cantaloni C, Rodriguez-Prieto A, et al. On site DNA barcoding by nanopore sequencing. *PLoS One* 2017; **12**(10): e0184741.
38. Faria NR, Sabino EC, Nunes MR, Alcantara LC, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. *Genome Med* 2016; **8**(1): 97.
39. Goordial J, Altshuler I, Hindson K, Chan-Yam K, Marcolefas E, Whyte LG. In Situ Field Sequencing and Life Detection in Remote (79 degrees 26'N) Canadian

High Arctic Permafrost Ice Wedge Microbial Communities. *Front Microbiol* 2017; **8**: 2594.
40. Johnson SS, Zaikova E, Goerlitz DS, Bai Y, Tighe SW. Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. *J Biomol Tech* 2017; **28**(1): 2-7.
41. Castro-Wallace SL, Chiu CY, John KK, et al. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Sci Rep* 2017; **7**(1): 18022.

Strain	Source	Accession	Genome length (bp)	Predicted genes	GC percentage
Wilmington	Human, North Carolina, USA. 1928.	NC_006142.1	1111496	865	0.289
B9991CWPP	Bandicoot rat, Myanmar. 1970s	NC_017062.1	1112957	865	0.289
TH1527	Human, Chiang Rai, Thailand. 1965	NC_017066.1	1112372	865	0.289
TM2540	Human, Bolikhamxay, Laos. 2012	ERZ497871	1111939	866	0.289

Table 1. Strain information. The Wilmington, B991CWPP, and TH1527 strains are previously assembled strains of R. typhi*; TM2540 is the assembly produced in this study.*

	Complete	Fragmented	Missing
Canu only	0.00%	3.40%	96.60%
Canu + Nanopolish	14.20%	32.40%	53.40%
Canu + Nanopolish + Pilon (1 round)	85.10%	2.70%	12.20%
Canu + Nanopolish + Pilon (2 rounds)	88.50%	0.70%	10.80%
Canu + Nanopolish + Pilon (3 rounds)	88.50%	0.70%	10.80%
Canu + Nanopolish + Pilon (4 rounds)	88.50%	0.70%	10.80%
Unicycler (Illumina only)	88.50%	0.70%	10.80%
Reference genome (Wilmington)	88.50%	0.70%	10.80%

Table 2. Proportion of BUSCO bacterial gene set found as complete genes, fragmented genes, or missing for different assemblies of the TM2540 strain, and the reference Wilmington assembly. All complete genes were single-copy.

	Wilmington	B9991CWPP	TH1527	TM2540
Wilmington	0	23	17	22
B9991CWPP	23	0	16	21
TH1527	17	16	0	15
TM2540	22	21	15	0

Table 3. Pairwise differences between strains based on the whole genome alignment.

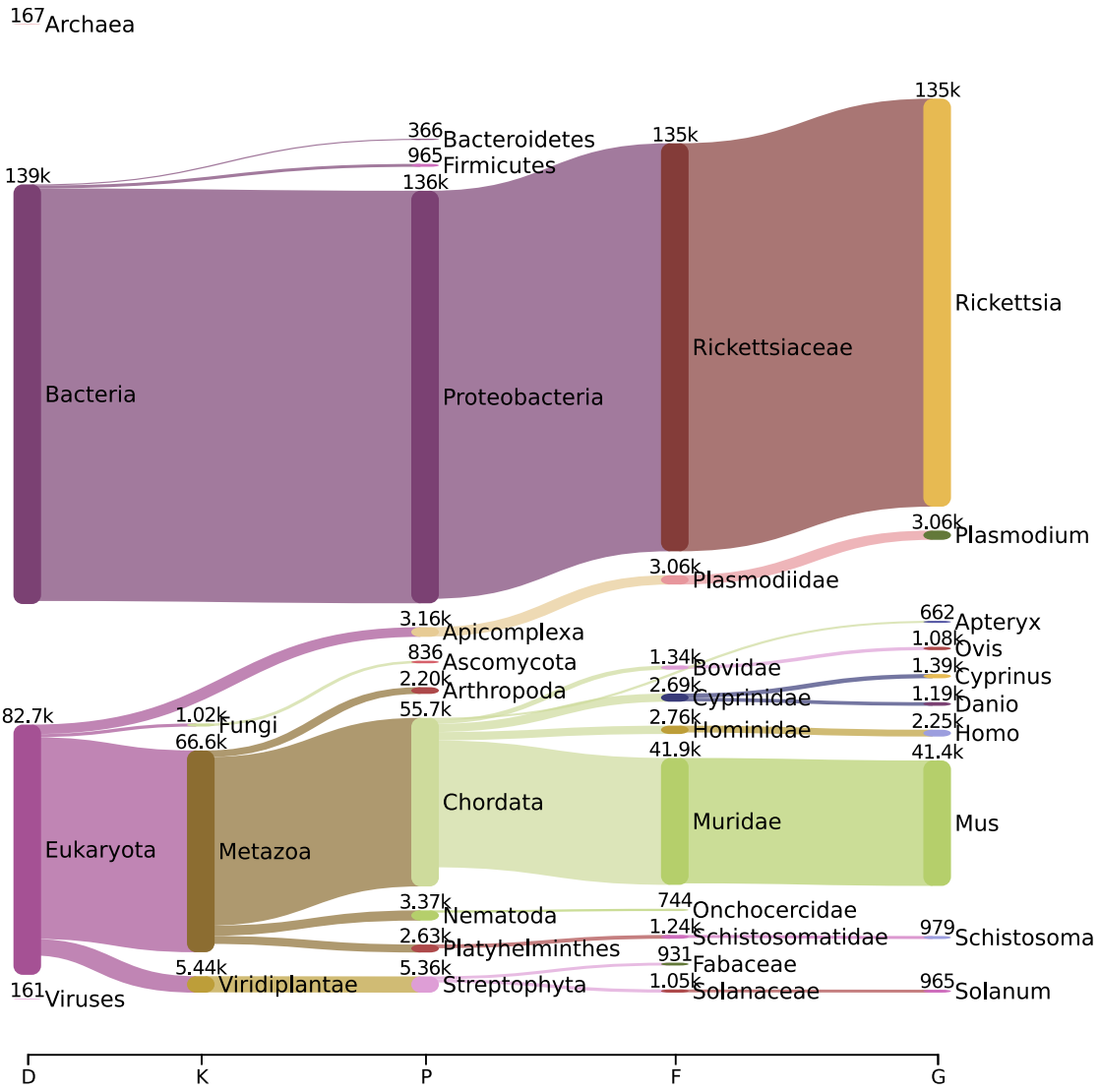

Figure 1. Pavian visualization showing the proportions of ONT reads assigned up to the genus level[42].

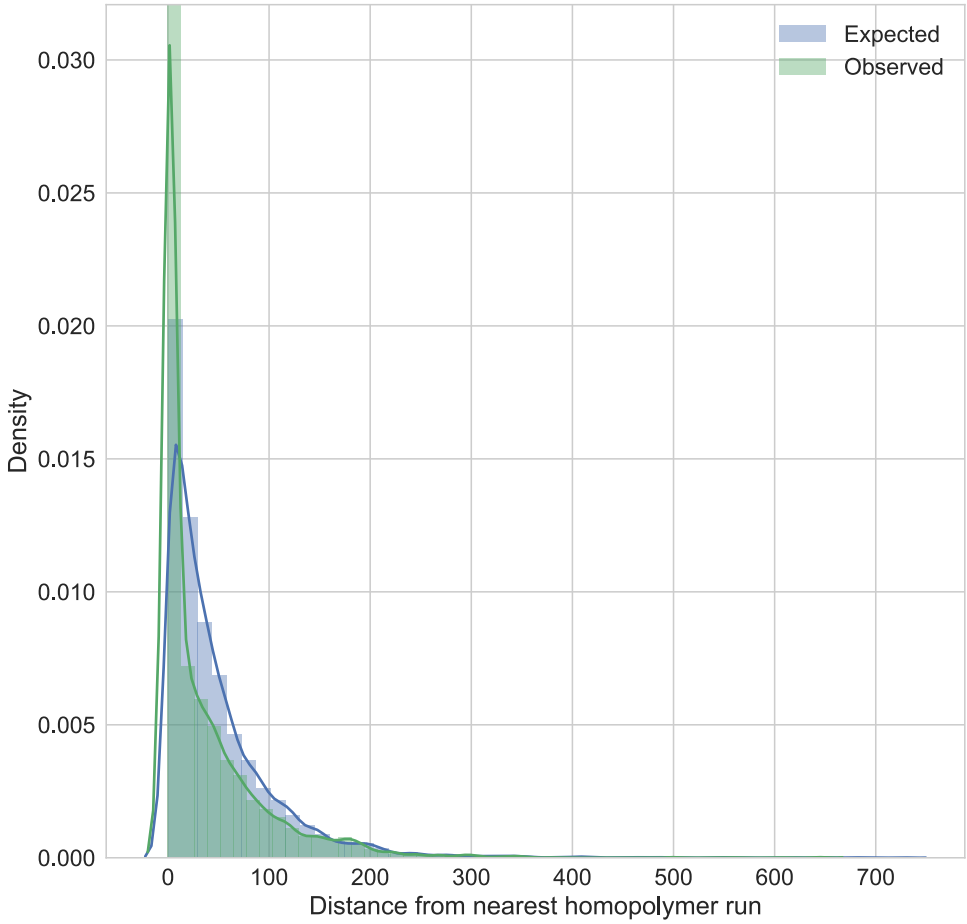

Figure 2. The observed distance from a homopolymer run for the errors corrected by Pilon in the draft genome, compared to the expected distribution if errors were randomly distributed across the genome.