Journal of Experimental Botany
doi:10.1093/jxb/erx289

REVIEW PAPER

A world of opportunities with nanopore sequencing

Richard M. Leggett[1,*] and Matthew D. Clark[1,2]

[1] Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK
[2] University of East Anglia, Norwich NR4 7TJ, UK

* Correspondence: richard.leggett@earlham.ac.uk

Received 12 May 2017; Editorial decision 26 July 2017; Accepted 26 July 2017

Editor: Ian Henderson, University of Cambridge

Abstract

Oxford Nanopore Technologies' MinION sequencer was launched in pre-release form in 2014 and represents an exciting new sequencing paradigm. The device offers multi-kilobase reads and a streamed mode of operation that allows processing of reads as they are generated. Crucially, it is an extremely compact device that is powered from the USB port of a laptop computer, enabling it to be taken out of the lab and facilitating previously impossible in-field sequencing experiments to be undertaken. Many of the initial publications concerning the platform focused on provision of tools to access and analyse the new sequence formats and then demonstrating the assembly of microbial genomes. More recently, as throughput and accuracy have increased, it has been possible to begin work involving more complex genomes and metagenomes. With the release of the high-throughput GridION X5 and PromethION platforms, the sequencing of large genomes will become more cost efficient, and enable the leveraging of extremely long (>100 kb) reads for resolution of complex genomic structures. This review provides a brief overview of nanopore sequencing technology, describes the growing range of nanopore bioinformatics tools, and highlights some of the most influential publications that have emerged over the last 2 years. Finally, we look to the future and the potential the platform has to disrupt work in human, microbiome, and plant genomics.

Key words: Assembly, bioinformatics, diagnostics, nanopore, sequencing, technology, third-generation.

Introduction

> Great things are done by a series of small things brought together
>
> Vincent Van Gogh

Since 2005, when 454 Life Sciences released the first second (or next-) generation sequencing machine, the GS20, few areas of biological science have been left untouched by the potential of genomics technology to deliver insights into the structure and function of living organisms. In the end it was Illumina's technology (having purchased Solexa) of sequencing by synthesis (SBS) combined with bridge amplification of template molecules that won the battle of the next-generation sequencing (NGS) platforms, and they now dominate the sequencer market worldwide. While these next-gen machines still account for the largest share of sequencing output, their need to amplify DNA templates, then extend with single fluorescent nucleotides, and image each step limits them to short 100–400 bp read lengths due to inevitable phasing issues (when templates in a polymerase colony lose synchronicity). These shorter reads make genome, transcriptome, and metagenome assembly more challenging, and leave some areas of even the human genome unresolvable. In contrast, the third generation of sequencing platforms uses single molecule sequencing and, with no phasing issues

limiting read lengths, the challenge instead is to achieve good signal to noise ratios and basecall accuracy. This new wave of technologies is led by the more established player, Pacific Biosciences (PacBio) of Menlo Park, CA, USA and by the relative newcomer, Oxford Nanopore Technologies (ONT), of Oxford, UK. Both technologies analyse individual molecules of DNA with no need for artificial amplification, and generate longer reads than second-generation technologies (typically in the thousands or even tens of thousands of base pairs), but both platforms have a relatively high error rate compared with Illumina's <1% error rates. As the first long read single molecule sequencer, PacBio had to solve problems related to longer reads [new aligners, e.g. BLASR (Chaisson and Tessler, 2012)], lower read accuracy [using adaptor hairpins to create multipass sequencing of both strands and consensus sequence (Travers *et al.*, 2010)], and genome assembly [read correction, and long read capable assemblers (Chin *et al.*, 2013)]. Nanopore research has been built on much of this pioneering work.

Beyond these similarities, there are major differences between the technologies. PacBio technology has been established longer and the company's relatively large machines are typically found in centralized sequencing centres where the large sums required to purchase them can be justified. In contrast, ONT's MinION is a 4-inch long USB-powered device that is provided together with two flowcells and reagents for just US$1000. To a certain extent, ONT's approach has already begun to democratize sequencing, enabling almost any research group anywhere in the world to purchase their own sequencer and to use genomics in their research. The portability of the MinION and the fact that it requires only a laptop's USB port in order to power it has enabled researchers to dream of new experiments that were previously impossible and to consider taking the sequencer to the sample, rather than to bring the sample to the lab, and in so doing overcome a range of regulatory restraints and sample degradation issues. One particularly striking example of this was the recent experiment to deploy a MinION to the International Space Station and carry out sequencing (Castro-Wallace *et al.*, 2016). Finally, and crucially, the MinION operates as a genuine real-time platform, in which individual reads can be analysed as they are sequenced and this information used to determine how long to run the sequencing experiment for—or even whether to eject specific DNA molecules from individual nanopores.

We have now reached a crucial inflection point in the progress of nanopore sequencing. During the first couple of years of the availability of the MinION, researchers were occupied with understanding the technology and its error profile, and imagining where the new paradigm might eventually take them. The samples they worked with were typically bacterial, phage, or amplicon-based samples for which the higher error rate and lower yield of the early versions of the chemistry were better suited. However, the last year has seen large increases in the yield of the platform and significant reductions in the error rate. Members of the community are now beginning to work with much more complex organisms including plants and animals, and the last few months has seen the release of a number of eukaryotic data sets including human whole-genome nanopore data sets (see https://github.com/nanopore-wgs-consortium/NA12878), a human assembly (Jain *et al.*, 2017*a*), and the tomato species *Solanum pennellii* (M.H. Schmidt *et al.*, 2017). The announcement of the higher throughput GridION X5 and PromethION platforms should further facilitate this move to more complex organisms.

Nanopore sequencing

The MinION was first announced at the Advances in Genome Biology and Technology (AGBT) conference in Florida in February 2012, but it was not publically available until an early access programme known as MAP (MinION Access Programme) began in April 2014. However, the idea of using nanopores dates back to the late 1980s and early 1990s and the work of several groups who published key research and patents (for a detailed review, see Deamer *et al.*, 2016). The principle behind the ONT technology is illustrated in Fig. 1. A biological nanopore is built into an electrically resistant artificial membrane and a voltage is applied across the membrane. DNA molecules are prepared according to one of a set of standard library preparation protocols which involve attaching a leader adaptor and motor protein to one strand of DNA. During sequencing, the motor protein unzips dsDNA and passes a single strand through the pore a base at a time. The presence of the DNA molecule in the pore causes a deflection in the current across the pore and this current change can be related to the exact bases present in the pore at that moment—usually based on 5-mer or 6-mer models. On a MinION flowcell is an array of 512 sensors, each connected to four nanopores (one of which is in use at any time), measuring current through the pore thousands of times a second. At any given moment, each of the 512 channels may be in a number of different states depending on whether molecules are being sequenced, captured into the pore, or if pores are empty or blocked. Software called MinKNOW runs on a computer connected to the MinION and records the signal (current) data from all channels, storing it on the local hard drive. Since late 2016, signal data are basecalled to a nucleotide sequence either in real-time by MinKNOW or by a separate offline basecaller. For the first 2 years of the MinION, there was little option other than using the basecalling cloud service provided for free by the ONT subsidiary, Metrichor. This involved uploading the FAST5-format files created by the local control service and downloading basecalled FAST5 reads that needed to be converted to more standard formats such as FASTA and FASTQ. Metrichor characterizes basecalled reads as 'pass' or 'fail' according to whether they pass mean quality criteria. In March 2017, the free Metrichor basecalling service was stopped, though Metrichor offers a number of secondary analysis pipelines through a web service.

A number of library preparation protocols are available. Initially, the most popular approach was 2D (double-stranded) sequencing. This involved preparing molecules that incorporate a hairpin adaptor; the motor protein would process the leader adaptor through the pore, followed by the

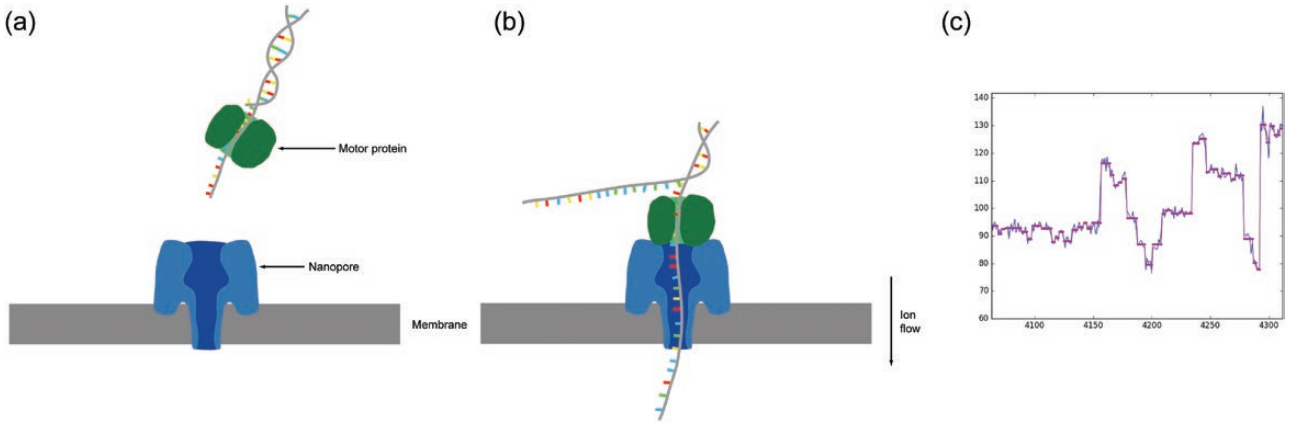

Fig. 1. Nanopore sequencing: (a) A biological nanopore is inserted into an electrically resistant synthetic membrane. A potential is applied across the membrane, resulting in ion flow. Library DNA molecules have adaptors with aliphatic tethers (not shown) which preferentially locate to the membrane for a localized library concentration. (b) The motor protein bound to the other adaptor docks with the pore, and passes the DNA molecule through it. (c) Bases in the nanopore cause disruptions in the current which are characteristic of their sequence (blue line). In some basecallers, the signal is further refined to events (red line) which correspond to distinct pore kmers.

template strand, then the hairpin adaptor and the complement strand. This approach provided the basecaller with two attempts at reading the bases, and generated a combined sequence with higher accuracy basecalls than from a single strand alone. However, as overall accuracy has increased on the platform, 1D library preparations involving sequencing only a single strand of DNA have become more popular. The main advantages of 1D sequencing are much reduced library preparation time (as little as 10 min for the rapid transposase-based kit) and increased yield due to only sequencing one strand of each molecule, and this method is preferred for obtaining the longest reads. As yields have increased, ONT have released a 1D rapid (transposase-based) barcoding kit that supports multiplexing up to 12 samples, for example bacteria to be processed in just 10 min, and a PCR-based barcoding kit that supports up to 96 samples. The early part of 2017 saw the announcement that 2D libraries would be phased out in preference for what is termed $1D^2$ libraries. Here, the two strands of a DNA molecule are delivered to the pore but are not covalently linked. Once the template strand has been sequenced, this approach relies on the complement strand remaining near the pore and being captured by the pore immediately after the template strand, which ONT assert happens in 60% of cases (Clive Brown, March 2017 ONT Technology Update). $1D^2$ also offers accuracy advantages over the existing 2D approach, where complement strands have less accuracy than what is effectively two template strands.

Input DNA requirements for single molecule sequencers tend to be higher than for the second generation of sequencers which use amplification. The standard ONT 1D and 2D ligation kits recommend 1 µg of starting material; good results are possible with half of that, though the best yields—especially for long reads—can benefit from larger amounts. The new rapid transposase kit has a recommended input of 200 ng. ONT's PCR-based protocol needs only 10 ng input, and it is possible to sequence even lower (picogram) amounts either by using increased amplification prior to ONT library creation, or by accepting that pore occupancy, and therefore yields, will be significantly lower. There is growing interest in decreasing input requirements using improved library techniques (lower losses) and improved pore loading, for example using bead-bound library molecules which locate adjacent to pores and load more efficiently. Whilst nanopore and PacBio sequencing require large inputs of DNA, they remain more susceptible to inhibitors of library construction or sequencing, and an additional clean up step may be necessary. Most sample types have now been successfully sequenced, but you might have a more challenging one. ONT maintain a forum (https://nanoporetech.com/community) with extant protocols and rapid advice from MAP colleagues, as well as ONT's staff. Whilst obvious, it bears repeating for emphasis, reads cannot be longer than the input DNA. Thus the more gentle, careful DNA extraction methods often yield superior data. As many of these methods are based on older protocols, it may be worth putting down your kits, reading 'Molecular cloning: a laboratory manual' (Green and Sambrook, 2012), or finding a green-fingered old school molecular biologist and buying them a drink.

Sequencing runs typically last up to 48 h, but the first 24 h tend to produce much higher yields, as flowcell performance gradually declines. At 8 h intervals, mux scans are performed by the system in order to choose the highest performing nanopore in each channel's group of four.

MinION performance

Since the MAP programme started in 2014, the MinION platform has seen regular updates to the device, flowcell design, sequencing pore, sequencing kits, and the signal capture and basecalling software. Thus any discussion about MinION performance is a reflection of a particular time point and is quickly out of date. The first sequencing pore used by members of MAP was termed R6 ('R' for 'Reader'), then came R7, R7.3, R9, R9.4, and R9.5, which is the latest iteration of the CsgG-derived pore used from R9 flowcells onwards. An international consortium of MAP members, the MinION Analysis and Reference Consortium (MARC), set out to produce an assessment of MinION quality and consistency by

conducting an experiment in which five different labs across the world sequenced DNA from the same *Escherichia coli* K-12 sample using the R7.3 pore and flowcell (Ip *et al.*, 2015). They found a high variation in yield amongst the 20 flowcells used (ranging from ~20 000 to 140 000 reads per flowcell), but a much more consistent error rate with median accuracy of 89% for 2D pass reads. Recently, the consortium published a Phase 2 analysis which updated results to R9 (Jain *et al.*, 2017*b*), demonstrating an increase of median accuracy to 92% and much increased yield (127 000–217 000 reads per flow cell, four flow cells sequenced). MARC have yet to publish an update for R9.4, but the general consensus within the nanopore community is that accuracy is now up to ~95% for 2D pass reads, and yields of 3–5 Gb are typical (Fig. 2). As with the PacBio platform, insertions and deletions are the predominant error type—R7.3 MARC data found roughly three times the number of indels as substitutions across 2D pass and fail data. This is in contrast to Illumina technology where errors, while rare, tend to be substitutions. The MinION has had particular difficulties with homopolymer sequences, as a stream of identical bases will result in the same current level, and variability in the speed of procession through the pore makes it difficult to discern the exact number of bases from temporal data alone. A new basecaller termed 'scrappie' (https://github.com/nanoporetech/scrappie) has been developed by ONT to address this problem, and early results seem to indicate substantially improved homopolymer resolution (Jain *et al.*, 2017*a*).

ONT have long asserted that nanopore sequencing quality is the same at the beginning and end of the DNA molecule, thus read length is dependent on the DNA extraction and preparation. This means that if you can prepare long DNA molecules for sequencing, the device will produce long sequence data. The standard protocols for library preparation involve a recommended shearing step which tends to produce a mean length of ~7–8 kb. However, even with this approach, long reads are possible and the MARC Phase 2 data includes mapped reads up to 50 kb. Recently, Nick Loman and Josh Quick garnered much attention within the nanopore community for their attempts to sequence extremely long molecules on the MinION. Loman's blog (http://lab.loman.net) details recent success in sequencing accurate *E. coli* and human DNA molecules up to 900 kb long using older phenol–chloroform extraction and minimal use of pipetting in order to avoid unwanted shearing. The same protocols were also used in a recent human sequencing pre-print publication which described generation of reads of up to 882 kb (Jain *et al.*, 2017*a*).

Basecalling

At launch, the only basecalling option for the MinION was using the EPI2ME platform provided by the ONT subsidiary Metrichor (http://www.metrichor.com/). The client software ran on the same computer as the MinION and uploaded raw signal FAST5 files to a remote Metrichor cloud-based server hosted on Amazon Web Services. After a slight delay, basecalled FAST5 files were downloaded back onto the client machine. Initially, ONT adopted an approach to basecalling in which the raw signal was first converted to events (with each event ideally corresponding to a single base movement through the pore; Fig. 2c) and the event space modelled as a Hidden Markov Model (HMM), utilizing the Viterbi algorithm to find the most likely sequence of states consistent with the event data (for more detail, see Schreiber and

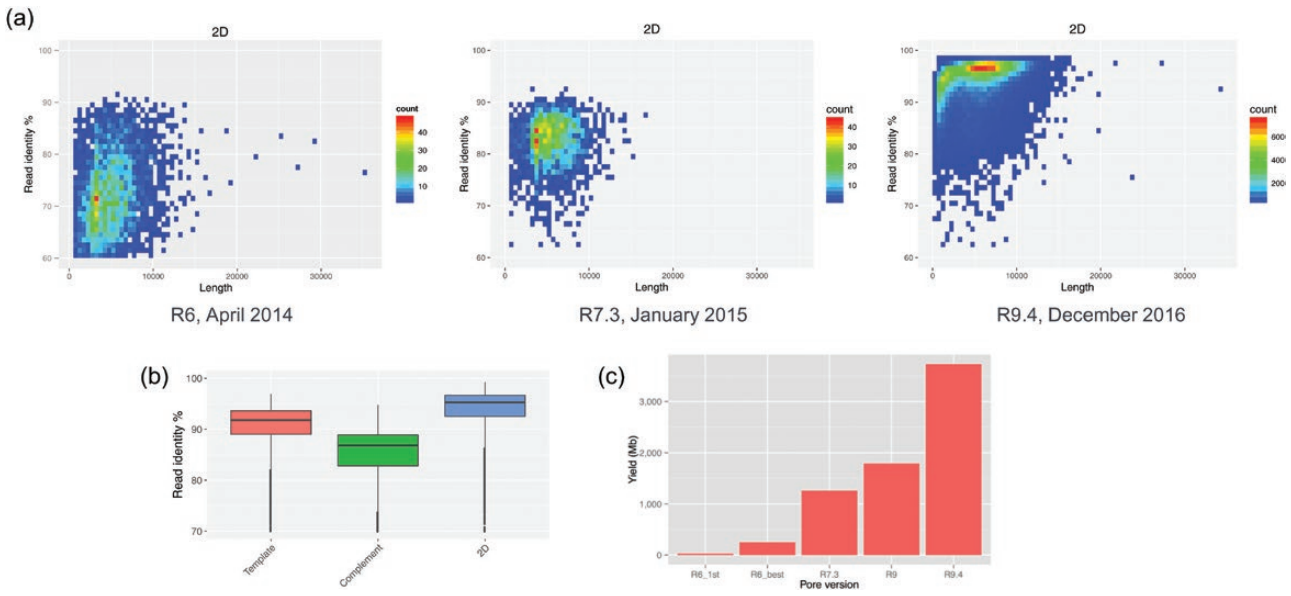

Fig. 2. Improvements in MinION accuracy and yield. (a) Indicative read identity versus length plots for standard 2D library prep on R6, R7.3, and R9.4, based on flowcells run at Earlham Institute using the ONT suggested 2D protocol (with ~8 kb shear). (b) Distribution of read identities for the R9.4 run, showing the difference between template, complement, and 2D consensus read accuracy. (c) Indicative yields. In all cases, results are affected by sample and DNA quality, but the general trend has been upwards. Read length is dependent on DNA preparation. Plots are produced using NanoOK (Leggett *et al.*, 2016).

Karplus; 2015; David *et al.*, 2017). Subsequently, ONT found increased accuracy could be obtained using recurrent neural networks (RNN) to model the event-to-sequence process more effectively.

MinKNOW 1.0.2, released in mid-2016, introduced 1D local basecalling by RNN and removed the reliance on the Metrichor service, though still providing it as an option. Using local basecalling, MinKNOW will attempt to capture signal data and basecall them on the same laptop; depending on the processing power available, the basecalling may lag behind. ONT subsequently released the standalone Albacore basecaller, which can be run locally or on a high performance computing (HPC) cluster. This added the ability to perform 2D basecalling and with MinKNOW 1.5.5 (March 2017), as MinKNOW was brought into parity with Albacore 0.8.4. This coincided with ONT switching off cloud basecalling, though it is unclear at this stage if it will subsequently be offered as a paid for service, something ONT have previously indicated. The period in which there was no local basecalling option provided by ONT saw the release of two third-party basecallers: Nanocall was the first to be published, providing an HMM-based implementation of 1D basecalling and producing comparable results to the Metrichor HMM basecalling of R7.3 data (David *et al.*, 2017); DeepNano emerged at a similar time, adopting an RNN approach and claiming to offer slightly higher accuracy than Metrichor then did (Boža *et al.*, 2017). For a period, both tools provided a viable basecalling approach for those needing to move away from an internet service but have subsequently been left behind by updates to the core platform and ONT's own move to local RNN basecallers. Nevertheless, they represent useful examples for those looking to understand and work with nanopore signal data.

Data analysis

The FAST5 format used on the MinION platform is an instance of the HDF5 standard for hierarchical data storage (https://support.hdfgroup.org/HDF5/). Each read has its own FAST5 file and each file is substantially larger than the size of the sequence data alone, containing signal data and large amounts of other metadata. Standard bioinformatics tools typically require FASTA or FASTQ format files, so it is necessary to extract reads in one of these formats for downstream analysis. ONT do not provide toolsets to enable this, but a number have been developed by the community. Poretools (Loman and Quinlan, 2014) was the first to be released, offering a command line interface to extract FASTA/Q data and to report and plot basic statistics such as yield over time or read length distributions. Released slightly later, poRe offers similar functionality for users of the R statistical environment (Watson *et al.*, 2015). NanoOK also provides FASTA/Q extraction, and will align reads against references and generate a comprehensive report analysing accuracy and the error profile of the data (Leggett *et al.*, 2016). minoTour implements a web front-end to interact with nanopore sequencing data or to monitor and control runs on remote machines in real-time (http://minotour.nottingham.ac.uk).

Once converted to FASTA/Q format, it is important to use alignment tools that can handle the particular error profile of nanopore reads. Early data were error prone enough to be challenging to analyse using standard tools, but as MinION accuracies have increased this has become easier. LAST was initially popular within the nanopore community, offering BLAST-like searches but with an adaptive seed approach in which seeds are chosen for their rareness instead of fixed length matches (Kielbasa *et al.*, 2011). After a nanopore-specific option was added by the authors of BWA-MEM (Li and Durbin, 2009), it soon attained ascendancy. Good results can also be obtained from BLAST, especially with tuned penalty parameters (Goodwin *et al.*, 2015). marginAlign is a nanopore-specific alignment tool that uses expectation maximization to train an HMM and estimate Maximum Likelihood Estimation parameters in order to find higher confidence alignments (Jain *et al.*, 2015). Using marginAlign, the MARC Phase 1 researchers were able to decrease their estimate of overall error rate by 1.1% for 2D pass reads, compared with alignments with BWA or LAST (Ip *et al.*, 2015). GraphMap is an alignment tool for nanopore reads which adopts a process of refinement that progressively reduces the candidate alignment sites until a good quality alignment is obtained (Sović *et al.*, 2016*a*).

Table 1 provides a summary of key third-party software tools for working with nanopore data.

A portable lab

Two of the most exciting attributes of the MinION are its compact size and the ability to power it and operate it from a standard laptop computer. This has enabled researchers to design experiments that would previously have been impossible or that are made difficult by restrictions (physical or legal) when transporting biological materials. Instead of collecting samples and couriering them frozen to the sequencing centre, it is now possible to take the sequencer to the samples in the field for faster results and potentially highlighting sites with special genetic diversity. One high-profile example of this was the project to carry out surveillance of the West African Ebola outbreak of 2014–2015 using the MinION (Quick *et al.*, 2016). All instruments and consumables necessary to set up a genome surveillance lab were packed into <50 kg of aircraft baggage and flown to Guinea. Here, hundreds of samples were sequenced, basecalled, and analysed. This took place before local basecalling was available and there were some challenges in obtaining sufficient mobile internet data bandwidth. These were overcome, and the researchers were able to carry out detailed evolutionary analysis of the outbreak. Some of the same team responsible for the Ebola work then became part of the ZiBRA collaboration which is using MinIONs to carry out surveillance of the Zika virus outbreak in Brazil (Faria *et al.*, 2016, 2017). Another group evaluated the use of nanopore sequencing as part of a rapidly deployable laboratory during a 2016 NATO exercise which simulated an outbreak due to a bioterrorism attack (Walter *et al.*, 2016). Using a MinION operated inside an inflatable

Table 1. *Commonly used third-party nanopore software tools*

Tool	Brief description	Website	Reference
BWA-mem	Read alignment	http://bio-bwa.sourceforge.net	Li and Durbin (2009)
Canu	Correction and assembly of long read data. Supersedes Celera.	http://canu.readthedocs.io/en/latest/	Koren *et al.* (2017)
DeepNano	Basecalling	https://bitbucket.org/vboza/deepnano	Boža et al. (2016)
GraphMap	Read alignment	https://github.com/isovic/graphmap	Sović *et al.* (2016a)
LAST	Read alignment	http://last.cbrc.jp	Kielbasa *et al.* (2011)
LINKS	Scaffolding with long reads	http://www.bcgsc.ca/platform/bioinfo/software/links	Warren *et al.* (2015)
marginAlign	Tuning of nanopore read alignments	https://github.com/benedictpaten/marginAlign	Jain *et al.* (2015)
minimap and miniasm	Rapid mapping and assembly	https://github.com/lh3/miniasm	Li (2016)
minoTour	Web-based monitoring and remote control of runs	http://minotour.nottingham.ac.uk	
Nanocall	Basecalling	https://github.com/mateidavid/nanocall	David *et al.* (2017)
NanoCorr	Error correction using Illumina data	https://github.com/jgurtowski/nanocorr	Goodwin *et al.* (2015)
NanoOK	Read extraction and alignment-based analysis	https://documentation.tgac.ac.uk/display/NANOOK/NanoOK	Leggett *et al.* (2016)
nanopolish (and nanocorrect, which is no longer used)	Signal level analysis—including assembly polishing, event alignment, methylation detection, variant calling	https://github.com/jts/nanopolish	Loman *et al.* (2015)
NaS	Hybrid assembly via nanopore synthetic-long reads.	http://www.genoscope.cns.fr/externe/nas/	Madoui *et al.* (2015)
npScarf	Scaffolding with long reads	https://github.com/mdcao/npScarf	Cao *et al.* (2017)
poRe	Read extraction and run metrics via R	https://github.com/mw55309/poRe_docs	Watson *et al.* (2015)
PoreSeq	Variant calling	https://github.com/tszalay/poreseq	Szalay and Golovchenko (2015)
poretools	Read extraction and run metrics	http://poretools.readthedocs.io/en/latest/	Loman and Quinlan (2014)
Racon	Assembly consensus	https://github.com/isovic/racon	Vaser *et al.* (2017)
SignalAlign	Signal mapping and methylation detection	https://github.com/ArtRand/signalAlign	Rand *et al.* (2017)
SPAdes	Hybrid assembly	http://cab.spbu.ru/software/spades/	Bankevich *et al.* (2012)

tent, they were able to prepare samples, sequence, and perform metagenomic analysis to determine the causative agent. Some challenges remain—such as the amount of input DNA required—but improved yields and accuracy, combined with emerging simplified and automated sample preparation, will make outbreak surveillance activities even simpler in the future.

The utility of the MinION in a range of clinical settings has also been demonstrated. Applications have included real-time sequencing of a hospital outbreak of *Salmonella* (Quick *et al.*, 2015), identification of antibiotic resistance profiles (Ashton *et al.*, 2015; Bradley *et al.*, 2015; Cao *et al.*, 2016; van der Helm *et al.*, 2017), same day diagnostics and surveillance for tuberculosis (Votintseva *et al.*, 2016), prenatal testing (Cheng *et al.*, 2015; Wei and Williams, 2016), detection of viral pathogens in blood samples (Greninger *et al.*, 2015), identification of bacterial pathogens in urine samples (K. Schmidt *et al.*, 2017), and structural variant analysis in cancer (Norris *et al.*, 2016).

Researchers have attempted to test the MinION in what might be termed 'extreme' environments. Edwards *et al.* (2016) took sequencers to a European High Arctic glacier in order to sequence microbial communities, finding profiles coherent with those obtained by Illumina sequencing. Another group sequenced in the McMurdo Dry Valleys, Antarctica (Johnson *et al.*, 2017). However, perhaps the most extreme environment is represented by a recent NASA experiment aboard the International Space Station (Castro-Wallace *et al.*, 2016) in which MinIONs were used to sequence mouse, *E. coli*, and lambda libraries. Zero gravity was no barrier to sequencing, and researchers found no discernible decrease in MinION quality and throughput. However, the successful sequencing runs were carried out using pre-prepared libraries, and preparing these in space may be more of a challenge.

Genome assembly

Due to the similarities in read length and error profiles between PacBio and ONT technologies, it has been natural for researchers working with the MinION to adopt many of the assembly approaches and tools used for PacBio data. This has meant a return to overlap, layout, consensus assembly methods (originally used with Sanger reads), as well as exploration of hybrid assembly approaches that utilize long reads to scaffold Illumina contigs or Illumina reads to correct the long reads.

The first *de novo* assembly of a complete bacterium from only MinION data was Loman, Quick and Simpson's assembly of *E. coli* K-12 MG1655 from R7.3 data (Loman *et al.*, 2015). Their three-stage pipeline consisted of a custom read correction step (nanocorrect), assembly with Celera

(Myers *et al.*, 2000), followed by an assembly polishing step with Simpson's nanopolish tool. This final step involves using the raw pore signal data in order to compute an improved consensus sequence based on iterative evaluation of possible changes with a HMM. The resulting assembly is a single contig with 99.5% nucleotide identity to the already available reference sequence. The polishing step played a crucial part in the final quality of the assembly, decreasing the substitution and indel rate from 80 and 921 per 100 kb, respectively to 26 and 321 per 100 kb. The drawback of the polishing process is that it is relatively processor intensive, though it is easily parallelized for users with a HPC cluster.

Currently, Canu appears to be the most popular assembly tool for nanopore data. Based on Celera, but with new overlap and assembly algorithms, it has been designed to provide improved support for the error profile of third-generation sequencers and improved run times (Koren *et al.*, 2017). Canu uses its own computationally efficient read correction step which negates the need for the slower nanocorrect; however, the assembly will still be improved by finishing with nanopolish. Miniasm is an assembler for PacBio and nanopore data that sacrifices accuracy for speed (Li, 2016). By omitting read correction and consensus sequence generation, it is able to assemble small genomes in minutes (e.g. *Caenorhabditis elegans* in only 9 min on 16 cores). However, the resulting contigs reflect the underlying errors in the input reads and are significantly less accurate than Canu. A popular tool for pairing with miniasm is Racon (for Rapid Consensus) which provides a stand-alone assembler-neutral consensus module based on a Partial Order Alignment graph and brings assembly quality up to similar levels to Canu (Vaser *et al.*, 2017). As with Canu, running nanopolish after Racon can further improve assembly quality.

A number of assemblies utilizing hybrid approaches have been published which combine nanopore data with highly accurate short read Illumina data. Madoui *et al.* (2015) describe what they term Nanopore Synthetic-long, or NaS, reads which are formed by using nanopore reads as a template for local assemblies of Illumina short-reads. LINKS uses nanopore reads to scaffold contigs assembled from Illumina MiSeq data (Warren *et al.*, 2015). Karlsson *et al.* (2015) also used MinION reads to scaffold Illumina data. Another tool, npScarf, takes advantage of the fact that nanopore reads can be used immediately in order to scaffold Illumina data in real-time (Cao *et al.*, 2017). Although it is still necessary to wait for an Illumina run to complete before running npScarf, the authors argue that it favours a reduction in nanopore resources (flowcells, consumables) as sequencing can be stopped when the assembly has reached an acceptable quality. The assembler SPAdes has an option to utilize nanopore or PacBio data for gap closure and repeat resolution in hybrid assemblies with Illumina data (Bankevich *et al.*, 2012). Nanocorr uses alignments of Illumina reads against nanopore reads to correct the longer reads (Goodwin *et al.*, 2015). MiSeq reads are aligned against the nanopore reads, and an algorithm selects an optimal set of alignments spanning the read. A consensus basecall for the read is then determined using the PacBio tool PBDAG-Con (Chin *et al.*, 2013).

In general, as the yield and read accuracy of the MinION have increased, there has been less focus on hybrid approaches for assembly of small genomes and a realization that nanopore data alone are sufficient to produce accurate results. Evaluations of some of the approaches discussed can be found in Sović *et al.* (2016*b*), Deschamps *et al.* (2016), Judge *et al.* (2016), and Lu *et al.* (2016).

MinION reads have contributed in part or fully to assemblies of a wide range of bacteria (Ashton *et al.*, 2015; Millard *et al.*, 2015; Deschamps *et al.*, 2016; Greninger *et al.*, 2016, 2017; Turton *et al.*, 2016), yeast (Goodwin *et al.*, 2015; Istace *et al.*, 2017), and viruses (Wang *et al.*, 2015; Karamitros *et al.*, 2016). In general, nanopore assemblies of bacteria typically produce more contiguous results than Illumina-only assemblies, and the platform is starting to be seen as a competitor to the PacBio RS2 which has become the *de facto* gold standard for microbial assemblies. Thanks to increases in yield and accuracy, researchers are now starting to assemble larger genomes with MinION data. Pre-prints have recently become available outlining assemblies of the 54 Mb fungal genome *Rhizoctonia solani* (Datema *et al.*, 2016), the 100 Mb *C. elegans* genome (Tyson *et al.*, 2017), the 860 Mb endangered European eel (Jansen *et al.*, 2017) and the Gigabase genome of the tomato species *S. pennellii* (M.H. Schmidt *et al.*, 2017). The human assembly (Jain *et al.*, 2017*a*) is particularly impressive as it features ultra-long reads (N50 of 99.7 kb), a data type which is still only possible on MinIONs. Adding only 5× genome coverage of this doubles the contiguity of the assembly, and their models suggest that 30× coverage of such ultra-long data would have led to a 10-fold increase in contiguity (NG50 >30 Mbp).

Signal level analysis

One of the unique attributes of the MinION platform is that it provides access not only to the DNA sequence, but also to the raw signal used for basecalling. While signal data are inherently more difficult to interpret, in many ways they are a richer source of data because the basecalling process can cause information, such as DNA modification, to be lost. As we have discussed previously, analysis of raw signal has enabled researchers to implement their own basecallers and to create algorithms for genome polishing going back to the raw signal in a method reminiscent of examining Sanger trace files when reads disagree. Two other exciting uses of signal data are in the analysis of DNA modification and the direct control of individual nanopores.

Before the release of the MinION, researchers had already demonstrated that nanopores can discriminate between 5-methylcytosine, 5-hydroxymethylcytosine, and cytosine bases (Laszlo *et al.*, 2013, Schreiber *et al.*, 2013) through changes in the ionic current levels. Unlike with second-generation sequencing approaches, no special treatment of the DNA is required in order to detect the modifications, but conversely this means that the results are less binary than bisulphite treatment C>T transitions, and so greater coverage of the modified bases increases certainty of the signal analysis. Two groups have since demonstrated modified base analysis

on the MinION, both producing tools that train HMMs to discriminate modified bases. Simpson *et al.* (2017) implemented their model within their existing nanopolish package and demonstrate 95% accuracy in detecting 5-methylcytosine. Rand *et al.* (2017) have produced SignalAlign, a tool for alignment of signal data to a reference with mapping of methylation data, and demonstrated mapping of three cytosine variants and two adenine variants. Both groups anticipate the further development of these models for the detection of other classes of DNA modification. ONT's recent release of a direct RNA sequencing kit also offers the unique possibility to discriminate RNA modifications, though this research is currently immature. PacBio can also call DNA modifications from untreated DNA, due to changes in the kinetics of the sequencing polymerase as it interacts with the base, but the depths needed to call a specific modification accurately vary depending on the modification and can be very high, for example 250-fold for the most common eukaryotic modification, 5-methylcytosine (Davis *et al.*, 2013).

ReadUntil is the name of an Application Programming Interface (API) from Oxford Nanopore that gives software control over individual nanopores. Specifically, it provides a function to reverse the voltage across a pore, causing the ejection of the molecule and offering the possibility of rejecting DNA molecules which the user is not interested in sequencing. This unique capability of nanopore sequencing could enable a researcher to target sequencing to species of interest, or to specific regions of interest in a genome. For the approach to be efficient, a decision must be made as quickly as possible and from the earliest bases. However, standard basecallers do not call the sequence until the whole molecule has been sequenced. Thus, in order to decide if a molecule is required or not, software must make a decision based on analysis of the raw signal data or basecall using partial data. Loose *et al.* (2016) demonstrated the first practical application of ReadUntil and the first instance of real-time target enrichment by using an established signal processing algorithm called Dynamic Time Warping to match nanopore event data to pre-calculated signals. They used 11 amplicons covering 22 kb of the lambda genome and showed they were able to sequence amplicons of choice selectively—generating more even coverage of all amplicons, or high coverage of specific amplicons and low coverage of others. Possible future uses for ReadUntil might include targeting reads from a large insert library to address specific issues such as a duplicated gene cluster, gaps in genome assemblies, or sequencing specific species within a metagenomic community. However, there are significant technical challenges in performing the analysis rapidly enough to be efficient, especially as the number of targeted sequences increases and the speed of nanopore sequencing also increases.

The future for nanopore sequencing

Early 2017 saw the first labs take delivery of the PromethION, ONT's high-throughput sequencing platform. While the MinION takes a single 512-channel flowcell, the PromethION provides for up to forty-eight 3000-channel flowcells, with each flowcell able to be run independently. ONT suggests that the PromethION will exceed the best platform Illumina has, with a theoretical maximum of 11 Tb per PromethION run compared with a published specification of up to 6 Tb for a NovaSeq 6000 with the as yet unreleased S4 flowcell. At the time of writing, PromethION flow cell manufacturing is not fully underway and we are still waiting to see what the throughput is in the hands of typical users. In the first quarter of 2017, ONT also announced the GridION X5 which is effectively a unit consisting of five MinIONs and some built-in computing for basecalling. Both of these platforms enable enormous quantities of long read data to be generated at a price point close to, or matching, Illumina short reads, and will be crucial in the continued expansion of nanopore sequencing into complex plants and animals. At the other extreme, ONT has also announced that they are developing the SmidgION, a tiny 128-channel nanopore sequencer that attaches to, and is powered by, a mobile phone; this is planned for a late 2017 release. To reduce the price of sequencing runs, ONT is internally testing a simplified flowcell (and with fewer pores) which attaches to a reusable flowcell dongle (flongle) housing the signal detection hardware. Devices such as the flongle (MinION) and especially the SmidgION offer improved opportunities for ecological, environmental, and clinical applications.

Library preparation is already a relatively simple process for MinION, but ONT are keen to make it easier and more consistent. VolTRAX is a compact microfluidic device designed to automate nanopore library preparation. It consists of a USB-powered base into which a consumable cartridge is placed with an array of fluid containing pixels on its surface. Software running on a connected laptop controls the movement of drops of fluid around the cartridge surface, from pixel to pixel, including through defined heating zones. ONT provide a protocol for their 1D rapid library prep, but are in the process of developing other protocols and will enable users to develop their own workflows. Original designs for VolTRAX showed it attached to the top of a MinION, offering the tantalizing prospect of putting cell solutions into the device and receiving raw sequencing data out of the bottom. Though this is not realized in the current first release, it remains a target for ONT, possibly via another straw-like device, currently codenamed Zumbador, in which cells or tissues are inputted at the top, and material flows between partitions filled with chemicals and lyophilized enzymes until ready to sequence libraries emerge from the bottom.

New library preparation protocols continue to be developed by ONT, and one of the most eagerly awaited has been the direct RNA sequencing protocol which was recently made available. This offers the prospect of amplification-free (unbiased), full-length transcripts, as well as access to RNA base modification data. Early versions of the RNA basecaller are not yet as accurate as for DNA, but we are likely to see rapid progress as more members of the nanopore community start to exploit the potential of direct RNA sequencing. Two preprints were recently released which demonstrate the potential of direct RNA sequencing. The first, from within ONT, demonstrated successful sequencing of yeast transcripts and

full-length reads of human rhinovirus (7.5 kbp RNA virus) in which they demonstrated that the signal could be used to discriminate m6A base modifications (Garalde *et al.*, 2016). The second paper (Smith *et al.*, 2017) demonstrated the ability to detect two modifications (7mG and pseudouridine) at known positions in full-length 16S rRNA MinION reads from *E. coli*. Access to RNA modification data could provide additional power to discriminate between highly similar sequences, such as 16S rRNA, within a genus, species, or even physiological states.

ONT are not the only company working on nanopore technology. Hitachi have recently presented work on solid state nanopores (Goto *et al.*, 2016) and Genia (purchased by Roche) have a technology based on biological nanopores and highly modified DNA copies. However, no other company has yet reached market with a product. ONT's main competitors then are PacBio's long-read technology and Illumina's short-read, high-throughput technologies. Both companies have tended to downplay the impact of ONT's technology publicly, but it is clear that ONT's continuing improvements in throughput and accuracy have moved nanopore sequencing to a point where it has become a contender to both of the longer standing platforms. As ONT's business model includes options to buy platforms for little or no cost (paying only for consumables), many labs may be attracted to nanopore sequencers without writing an equipment grant. Assuming the continuation of improvements in nanopore sequencing (by no means guaranteed), PacBio are probably the most threatened, followed by Illumina's assembly and then resequencing markets. Illumina SBS and cluster chemistry is strongest in the 'tag-counting' space of functional genomics (e.g. RNA-seq, ChIP-seq, etc.) but one can envisage how a long read SAGE-type method (Velculescu *et al.*, 1995) might threaten even this. As users of all three technologies, we hope that competition drives innovation, increased performance, and lower prices that enable new biological discoveries.

Acknowledgements

In our MinION work, we are grateful for the contributions of other members of Earlham Institute's MinION taskforce, especially Darren Heavens, Yuxuan Lan, and Robert Davey, and also to the Platforms & Pipelines team, Tim Stitt, and the NBI Computing infrastructure for Science (CiS) group. We thank Ned Peel for commenting on a manuscript draft. We have also appreciated the expertise of the wider MinION community, particularly through our interactions with members of the MARC consortium. Our MinION work is supported by BBSRC TRDF award BB/N023196/1, BBSRC National Capability Grants BB/J010375/1 and BB/CCG1720/1, BBSRC Institute Strategic Programme Grant BB/J004669/1, and BBSRC Core Strategic Programme Grant BB/J004669/1. The authors have not received direct financial contributions from ONT, but have received a small number of free flowcells as part of the MAP and MARC programmes. RML is in receipt of travel and accommodation expenses to speak at an ONT conference and is on a PhD student advisory team with a member of ONT staff.

References

Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, Wain J, O'Grady J. 2015. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nature Biotechnology **33,** 296–300.

Bankevich A, Nurk S, Antipov D, *et al*. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology **19,** 455–477.

Boža V, Brejová B, Vinař T. 2017. DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads. PLoS One **12,** e0178751.

Bradley P, Gordon NC, Walker TM, *et al*. 2015. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. Nature Communications **6,** 10063.

Cao MD, Ganesamoorthy D, Elliott AG, Zhang H, Cooper MA, Coin LJ. 2016. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION™ sequencing. GigaScience **5,** 32.

Cao MD, Nguyen SH, Ganesamoorthy D, Elliott AG, Cooper MA, Coin LJ. 2017. Scaffolding and completing genome assemblies in real-time with nanopore sequencing. Nature Communications **8,** 14515.

Castro-Wallace SL, Chiu CY, John KK, *et al*. 2016. Nanopore DNA sequencing and genome assembly on the International Space Station. Preprint bioRxiv doi:10.1101/077651.

Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics **13,** 238.

Cheng SH, Jiang P, Sun K, Cheng YK, Chan KC, Leung TY, Chiu RW, Lo YM. 2015. Noninvasive prenatal testing by nanopore sequencing of maternal plasma DNA: feasibility assessment. Clinical Chemistry **61,** 1305–1306.

Chin CS, Alexander DH, Marks P, *et al*. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods **10,** 563–569.

Datema E, Hulzink RJM, Blommers L, Valle-Inclan JE, Van Orsouw N, Wittenberg AHJ, De Vos M. 2016. The megabase-sized fungal genome of *Rhizoctonia solani* assembled from nanopore reads only. Preprint bioRxiv doi:10.1101/084772.

David M, Dursi LJ, Yao D, Boutros PC, Simpson JT. 2017. Nanocall: an open source basecaller for Oxford Nanopore sequencing data. Bioinformatics **33,** 49–55.

Davis BM, Chao MC, Waldor MK. 2013. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. Current Opinion in Microbiology **16,** 192–198.

Deamer D, Akeson M, Branton D. 2016. Three decades of nanopore sequencing. Nature Biotechnology **34,** 518–524.

Deschamps S, Mudge J, Cameron C, Ramaraj T, Anand A, Fengler K, Hayes K, Llaca V, Jones TJ, May G. 2016. Characterization, correction and de novo assembly of an Oxford Nanopore genomic dataset from *Agrobacterium tumefaciens*. Scientific Reports **6,** 28625.

Edwards A, Debbonaire AR, Sattler B, Mur LAJ, Hodson AJ. 2016. Extreme metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78 N. Preprint bioRxiv doi:10.1101/073965.

Faria NR, Quick J, Claro IM, *et al*. 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature **546,** 406–410.

Faria NR, Sabino EC, Nunes MR, Alcantara LC, Loman NJ, Pybus OG. 2016. Mobile real-time surveillance of Zika virus in Brazil. Genome Medicine **8,** 97.

Garalde DR, Snell EA, Jachimowicz D, *et al*. 2016. Highly parallel direct RNA sequencing on an array of nanopores. Preprint bioRxiv doi: 10.1101/068809.

Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Research **25,** 1750–1756.

Goto Y, Yanagi I, Matsui K, Yokoi T, Takeda K. 2016. Integrated solid-state nanopore platform for nanopore fabrication via dielectric breakdown, DNA-speed deceleration and noise reduction. Scientific Reports **6,** 31324.

Green MR, Sambrook J. 2012. Molecular cloning: a laboratory manual, 4th edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Greninger AL, Streithorst J, Chiu CY, Miller S. 2016. First draft genome sequences of *Neisseria* sp. strain 83E34 and *Neisseria* sp.

strain 74A18, previously identified as CDC eugonic fermenter 4b species. Genome Announcements **4,** e01277-16.

Greninger AL, Nacache SN, Federman S, *et al*. 2015. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Medicine **7,** 99.

Greninger AL, Streithorst J, Golden JA, Chiu CY, Miller S. 2017. Complete genome sequence of sequential *Pandoraea apista* isolates from the same cystic fibrosis patient supports a model of chronic colonization with in vivo strain evolution over time. Diagnostic Microbiology and Infectious Disease **87,** 1–6.

Ip CLC, Loose M, Tyson JR, *et al*. 2015. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. F1000Research **4,** 1075.

Istace B, Friedrich A, d'Agata L, *et al*. 2017. *De novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. GigaScience **6,** 1–13.

Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. 2015. Improved data analysis for the MinION nanopore sequencer. Nature Methods **12,** 351–356.

Jain M, Koren S, Quick J, *et al*. 2017*a*. Nanopore sequencing and assembly of a human genome with ultra-long reads. Preprint bioRxiv doi:10.1101/128835.

Jain M, Tyson J, Loose M, *et al*. 2017*b*. MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. F1000Research **6,** 760.

Jansen HJ, Liem M, Jong-Raadsen SA, *et al*. 2017. Rapid de novo assembly of the European eel genome from nanopore sequencing reads. Preprint bioRxiv doi:10.1101/101907.

Johnson SS, Zaikova E, Goerlitz DS, Bai Y, Tighe SW. 2017. Real-time DNA sequencing in the antarctic dry valleys using the oxford nanopore sequencer. Journal of Biomolecular Techniques **28,** 2–7.

Judge K, Hunt M, Reuter S, Tracey A, Quail MA, Parkhill J, Peacock SJ. 2016. Comparison of bacterial genome assembly software for MinION data and their applicability to medical microbiology. Microbial Genomics **2,** e000085.

Karamitros T, Harrison I, Piorkowska R, Katzourakis A, Magiorkinis G, Mbisa JL. 2016. *De novo* assembly of human herpes virus type 1 (HHV-1) genome, mining of non-canonical structures and detection of novel drug-resistance mutations using short- and long-read next generation sequencing technologies. PLoS One **11,** e0157600.

Karlsson E, Lärkeryd A, Sjödin A, Forsman M, Stenberg P. 2015. Scaffolding of a bacterial genome using MinION nanopore sequencing. Scientific Reports **5,** 11996.

Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. Genome Research **21,** 487–493.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research **27,** 722–736.

Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM, Bartlett JJ, Pavlenok M, Gundlach JH. 2013. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proceedings of the National Academy of Sciences, USA **110,** 18904–18909.

Leggett RM, Heavens D, Caccamo M, Clark MD, Davey RP. 2016. NanoOK: multi-reference alignment analysis of nanopore sequencing data, quality and error profiles. Bioinformatics **32,** 142–144.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25,** 1754–1760.

Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics **32,** 2103–2110.

Lu H, Giordano F, Ning Z. 2016. Oxford nanopore MinION sequencing and genome assembly. Genomics, Proteomics and Bioinformatics **14,** 265–279.

Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nature Methods **12,** 733–735.

Loman NJ, Quinlan AR. 2014. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics **30,** 3399–3401.

Loose M, Malla S, Stout M. 2016. Real-time selective sequencing using nanopore technology. Nature Methods **13,** 751–754.

Madoui MA, Engelen S, Cruaud C, Belser C, Bertrand L, Alberti A, Lemainque A, Wincker P, Aury JM. 2015. Genome assembly using nanopore-guided long and error-free DNA reads. BMC Genomics **16,** 327.

Millard AD, Westblade LF, LiPuma JJ, Vavikolanu K, Read TD, Pallen MJ, Burd EM, Constantinidou CI. 2015. Draft genome sequence of the *Pandoraea apista* LMG 16407 type strain. Genome Announcements **3,** e01300-15.

Myers EW, Sutton GG, Delcher AL, *et al*. 2000. A whole-genome assembly of *Drosophila*. Science **287,** 2196–2204.

Norris AL, Workman RE, Fan Y, Eshleman JR, Timp W. 2016. Nanopore sequencing detects structural variants in cancer. Cancer Biology and Therapy **17,** 246–253.

Quick J, Ashton P, Calus S, *et al*. 2015. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. Genome Biology **16,** 114.

Quick J, Loman NJ, Duraffour S, *et al*. 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature **530,** 228–232.

Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B. 2017. Mapping DNA methylation with high-throughput nanopore sequencing. Nature Methods **14,** 411–413.

Schreiber J, Karplus K. 2015. Analysis of nanopore data using hidden Markov models. Bioinformatics **31,** 1897–1903.

Schmidt K, Mwaigwisya S, Crossman LC, *et al*. 2017. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. Journal of Antimicrobial Chemotherapy **72,** 104–114.

Schmidt MH, Vogel A, Denton A, *et al*. 2017. Reconstructing the gigabase plant genome of *Solanum pennellii* using nanopore sequencing. Preprint bioRxiv doi:10.1101/129148.

Schreiber J, Wescoe ZL, Abu-Shumays R, Vivian JT, Baatar B, Karplus K, Akeson M. 2013. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. Proceedings of the National Academy of Sciences, USA **110,** 18910–18915.

Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. Nature Methods **14,** 407–410.

Smith AM, Jain M, Mulroney L, Garalde DR, Akeson M. 2017. Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. Preprint bioRxiv doi:10.1101/132274.

Sović I, Križanović K, Skala K, Šikić M. 2016*b*. Evaluation of hybrid and non-hybrid methods for de novo assembly of nanopore reads. Bioinformatics **32,** 2582–2589.

Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. 2016*a*. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nature Communications **7,** 11307.

Szalay T, Golovchenko JA. 2015. De novo sequencing and variant calling with nanopores using PoreSeq. Nature Biotechnology **33,** 1087–1091.

Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Research **38,** e159.

Turton JF, Doumith M, Hopkins KL, Perry C, Hopkins KL, Perry C, Meunier D, Woodford N. 2016. Clonal expansion of *Escherichia coli* ST38 carrying a chromosomally integrated OXA-48 carbapenemase gene. Journal of Medical Microbiology **65,** 538–546.

Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP. 2017. Whole genome sequencing and assembly of a *Caenorhabditis elegans* genome with complex genomic rearrangements using the MinION sequencing device. Preprint bioRxiv doi:10.1101/099143.

van der Helm E, Imamovic L, Hashim Ellabaan MM, van Schaik W, Koza A, Sommer MOA. 2017. Rapid resistome mapping using nanopore sequencing. Nucleic Acids Research **45,** e61.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Research **27,** 737–746.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. Science **270,** 484–487.

Votintseva AA, Bradley P, Pankhurst L, *et al*. 2016. Same-day diagnostic and surveillance data for tuberculosis via whole genome sequencing of direct respiratory samples. Journal of Clinical Microbiology **55,** 1285–1298.

Walter MC, Zwirglmaier K, Vette P, Holowachuk SA, Stoecker K, Genzel GH, Antwerpen MH. 2017. MinION as part of a biomedical rapidly deployable laboratory. Journal of Biotechnology **250,** 16–22.

Wang J, Moore NE, Deng YM, Eccles DA, Hall RJ. 2015. MinION nanopore sequencing of an influenza genome. Frontiers in Microbiology **6,** 766.

Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJ, Birol I. 2015. LINKS: scalable, alignment-free scaffolding of draft genomes with long reads. GigaScience **4,** 35.

Watson M, Thomson M, Risse J, Talbot R, Santoyo-Lopez J, Gharbi K, Blaxter M. 2015. poRe: an R package for the visualization and analysis of nanopore sequencing data. Bioinformatics **31,** 114–115.

Wei S, Williams Z. 2016. Rapid short-read sequencing and aneuploidy detection using MinION Nanopore technology. Genetics **202,** 37–44.