

FINISHING THE JOB - UTILITY OF LONG-READ SEQUENCING USING THE MINION FOR BACTERIAL GENOMICS

AMADA EL-SABEH^{1*}

Received: 24th of August 2021 / Revised: 27th of August 2021

Accepted: 30th of August 2021 / Published: 22nd of March 2022

Keywords: bacterial genomes; third-generation-sequencing; Oxford Nanopore Technologies MinION; hybrid assembly

Abstract: Sequencing technologies have evolved dramatically since the first two bacterial genomes were published. Currently, due to second generation sequencing, millions of bacterial genomic sequences exist, although a significantly smaller amount represent completely assembled genomes. Third generation sequencing allows the analysis of single molecules, with read-lengths that cover highly complex repetitive regions previously inaccessible by short-read sequencing. However, long-read sequencing is known for producing errors which make long-read-only genome assemblies unreliable or complex if high accuracy is important for further applications. Here, Oxford Nanopore Technology's MinION, the first handheld nanopore sequencing device, is evaluated in comparison with competing sequencing platforms. The MinION's applications, potential and limitations are reviewed, focusing on its utility for bacterial genome *de novo* or hybrid assembly.

FROM FIRST TO THIRD: OVERVIEW OF SEQUENCING TECHNOLOGIES

Over a quarter of a century has passed since the first two bacterial genomes were sequenced in 1995 using what was peak technology at the time (Fleischmann et al., 1995; Fraser et al., 1995). By 2005, the number of sequenced bacteria had increased 150-fold to around 300 and already was 100-fold higher by 2015 (Land et al., 2015). Only six years later, there were almost 46 million bacterial genomic sequences stored on NCBI. This accumulation of sequencing output has risen dramatically over the past decade due to the transition from the “first generation sequencing” technology to the next generation, title which encompasses the second and the most recent third-generation sequencing technologies.

In 1977, Frederick Sanger developed the “chain-termination” technique using radiolabelled analogues of deoxyribonucleotides (Heather and Chain, 2016). Ten years later, Applied Biosystems introduced the first automatic Sanger sequencing machine using capillary based electrophoresis and replaced the radioactive reagents with a fluorometric detection system (Heather and Chain, 2016; Liu et al., 2012). Due to these improvements, Sanger's technique became more efficient and safer, establishing itself as the first generation sequencing technology and was predominantly used until 2005 (Liu et al., 2012). Despite its high accuracy and read-lengths of up to 1 kb (Heather and Chain, 2016), Sanger sequencing was still laborious and could produce only a draft genome, albeit a high-quality one. Although most bacterial genomes were complete when publicly announced, finishing the obtained genome was challenging and required a production line in itself. Moreover, obtaining a draft genome and finishing it could sum up to \$50,000 (Land et al., 2015). Hence, even though Sanger sequencing produced long and highly accurate reads, the data was, at best, limited, as proven by the 15 year-long Human Genome Project (Collins et al., 2003). Furthermore, sequencing projects were handled by the few large institutions with access to the necessary funding and technological infrastructure (Land et al., 2015). Therefore, there was increasing demand for technologies that could provide higher amounts of better-quality data that was needed to uncover the answers to the increasingly complex questions that were being asked.

This eagerly expected breakthrough in sequencing started in 2005 with the launch of several Second-Generation Sequencing (SGS) platforms, beginning with 454 (now Roche 454). A year later were introduced the Genome Analyzer by Solexa (bought by the currently widely-used Illumina) and SOLiD by Agencourt (Sequencing by Oligo Ligation Detection, bought by Applied Biosystems, now part of Invitrogen) (Liu et al., 2012). These platforms and the Sanger method alike use “sequencing by synthesis” techniques, requiring the direct activity of a DNA polymerase to obtain the desired output (comprehensively reviewed by Heather and Chain, 2016). However, these novel SGS platforms were revolutionary due to their ability to sequence, in parallel, numerous PCR-amplified DNA samples (Karlsson et al., 2015). Their high throughput made it so that nowadays, it can cost under \$100 to obtain a bacterial draft genome (Balloux et al., 2018) and above \$1,000 for a human genome (Schwarze et al., 2020), making sequencing a no-brainer for any research team (Land et al., 2015).

Compared to the reads generated by Sanger sequencing, those produced using high throughput SGS platforms are significantly shorter, with reads produced by Illumina having ≤ 150 and maximally 300 bp (Wick et al., 2017). Moreover, SGS platforms yield sequences with higher error rates (~ 0.1 –1%) (Zagordi et al., 2010) than those obtained using Sanger sequencing. Most errors are either systemic ones (Goodwin et al., 2016; Laehnemann et al., 2016), or are a consequence of the amplification step required for sample preparation (Loman et al., 2012; Ross et al., 2013). The major downfall of the short reads is in dealing with genomic repetitive sequences – if a genome contains repeats longer than the maximal read

length supported by the sequencing platform, it is possible that no read will cover the repeat in the genome (Loman et al., 2012). Therefore, in order to complete a genome, more data is required to ensure the necessary coverage and a larger number of contiguous sequences (contigs) has to be closed and assembled (Gurevich et al., 2013).

Prokaryote genomes are much smaller than eukaryotic ones and contain fewer and shorter repetitive sequences, of up to 10 kbp in length. However, prokaryotes may possess plasmids that can have different copy numbers than the chromosome, thus requiring a different read depth for proper assembly, and their replicons are, generally, circular (Wick and Holt, 2021). Despite the less complex genomes of prokaryotes, short-read sequencing technologies are still lacking for finishing genome assemblies. For example, to complete the *de novo* genome assembly of *Francisella* isolates which include insertions sequence elements of varying numbers and longer than Illumina reads, scientists were faced with the impossible task of assembling tens to a couple hundred discrete contigs (Karlsson et al., 2015).

Whereas with Sanger sequencing the costs were driven by sequencing itself, with SGS, the majority of the costs were skewed towards assembling the sequences into a completely finished genome. As a consequence, an increasing number of genomes were published as incomplete drafts and contained multiple contigs of various quality levels (Land et al., 2015). Despite these shortcomings, at least when discussing microorganisms, there is a consensus that most draft genomes are of “good enough quality” for a majority of common applications (Land et al., 2015, 2014). Studies that compared finished and draft versions of microbial genomes found that information was not significantly lost from generating draft genomes using Illumina sequencing (Mavromatis et al., 2012). However, it was established that the location of antimicrobial resistance genes can be of epidemiological significance; considering that repetitive insertion sequences commonly flank antimicrobial resistance genes, an incompletely assembled genome would be useless to researchers trying to assess if the resistance gene of interest is found in the chromosome or on a plasmid (Wick et al., 2017a). Furthermore, the selection of a reference genome and a performant bioinformatic pipeline is the critical factor in accurately calling single nucleotide polymorphisms (SNPs), which is crucial for preventing and tracking the transmission of microorganisms and predicting phenotypic characteristics such as antimicrobial resistance (Bush et al., 2020).

Considering these limitations of SGS, a third generation of sequencing platforms emerged in the mid-2010s, represented by the Pacific Biosciences (PacBio) RS II system and the Oxford Nanopore Technology (ONT) MinION™. These third-generation sequencing technologies (TGS) are capable of long read (of up to tens of kilobases) real-time sequencing of individual DNA and even RNA molecules. Like Sanger and SGS platforms, PacBio RS II is also based on sequencing by synthesis, but its advantages lie in the capacity of individually monitoring DNA molecules, the real-time incorporation of fluorescently labelled nucleotides and the much lower sequencing bias (compared to SGS, not Sanger) (Eid et al., 2009; Karlsson et al., 2015; Ross et al., 2013). PacBio achieved these improvements using their SMRT approach: an adapter is ligated at either end of each input DNA molecule, which is circularized and can be sequenced several times in order to achieve a higher accuracy consensus read. However, SMRT reads have high error rates, requiring deep coverage or error correction using short-reads generated with SGS, such as Illumina. For projects targeting large genomes, the yield and high cost per base sequenced using PacBio's RS II system are prohibitive factors. Supplementary, two major downfalls hindered the wide implementation of this technology in small independent labs: the instrument's initial cost and the cost of the additional necessary infrastructure (Madoui et al., 2015).

SMALL BUT MIGHTY: THE MINION™

In 2014, Oxford Nanopore Technologies released the MinION™ to over 1000 laboratories through a beta-testing program, i.e. The MinION Access Program. Although what it delivers is very similar to the PacBio system, the MinION is distinct in several important ways. Unlike previous sequencing-by-synthesis technologies, the MinION is the first commercially available single-molecule nanopore-based sequencer. The device weighs 90 g and is just 10 cm in length, making it a handheld, highly portable device that can connect to any laptop through a USB interface. Even more, the MinION is significantly more financially accessible than any other sequencing platform on the market, library construction is simplified, preparatory PCR-amplification is not necessary and the generated high-throughput data can be acquired and analysed in real-time (Feng et al., 2015; Karlsson et al., 2015; Madoui et al., 2015).

How does it work?

The MinION contains a flow cell with an array of 2048 protein nanopores split into four channels, each with four pores and sensors. Each channel can be individually controlled by an application-specific integrated circuit and allows the simultaneous processing of up to 512 nucleic acid molecules (Cherf et al., 2012; Jain et al., 2016; Magi, Semeraro, et al., 2017). When an external voltage is applied to the flow cell, particles (nucleotides) smaller than the pore size translocate through the pore (**Fig. 1, A**). When shifting, the negatively-charged nucleotides pass through the nanopore, the current flowing through the pore is blocked and its signal interrupted. A sensor detects these changes in ionic current as separate

discrete events and depicts them in what is known as a “squiggle plot” (Fig. 1., B.). Using graphical models, the duration, mean amplitude and variance of the discrete events is statistically analysed as a sequence of 3-6 nucleotide long k-mers (Fig. 1., C.) and directly correlated with the physico-chemical properties of the target molecule (Feng et al., 2015; Jain et al., 2016; Karlsson et al., 2015; Madoui et al., 2015). The translation of the current profile into nucleotide sequence information is done in real time by various base calling software such as Metrichor™ (<https://metrichor.com/technology.html>), Albacore, Guppy, Scrappie or Flappie (Wick et al., 2019) .

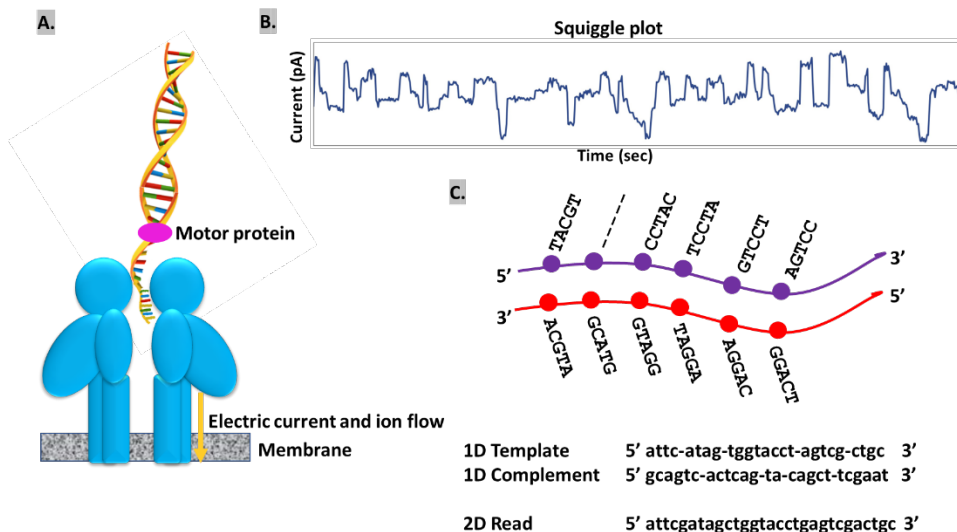


Figure 1. Flow of Oxford Nanopore Technologies MinION sequencing process

A. An adapter preloaded with a motor protein (magenta) unwinds the template strand of a duplex DNA molecule and guides it through the MinION nanopore (blue). As nucleotides pass through the pore, changes in electric current are detected by a sensor as discrete events and are depicted as a **B.** “Squiggle plot” of fluctuating electrical signals, each specific to a nucleotide’s physico-chemical properties; **C.** The k-mers decoded from discrete changes in ionic current and the alignment of 1D and, optionally, 2D base calls which will provide the 2D consensus read.

For sequencing both strands of a duplex DNA molecule, adapters (preloaded with motor proteins, Fig. 1., A.) are ligated at either end of the DNA input (genomic or cDNA) in order to aid in strand capture, ensure the loading of a processive enzyme at the 5'-end of one strand and closely concentrate the DNA material in the nanopore’s vicinity (Ip et al., 2015; Magi, Semeraro, et al., 2017). Therefore, the DNA capture rate is amplified while, on the millisecond timescale, the bound enzyme unwinds the double stranded DNA, guides a single strand through the pore and unidirectionally displaces a single nucleotide along the first (leading) DNA strand, generating the “template read” (Fig 1., C.). As long as the DNA strand is not damaged during sample processing, a hairpin adapter attaches at the opposite end of the DNA molecule and covalently binds together the two strands of a duplex molecule, facilitating the uninterrupted sequencing of the sense and antisense strands. After passing through the hairpin adapter, the enzyme repeats its activity on the remaining complementary strand, yielding the “complement read” which is aligned to the “template read” to generate a consensus 2D read of higher quality than 1D reads (Jain et al., 2016; Karlsson et al., 2015; Madoui et al., 2015). Using the more recent 1D² chemistry which replaced the 2D method, the forward and reverse strands of a DNA duplex are sequenced without the necessity of an adapter to keep the strands in physical contact. This novel method has been proven to yield similar data quality to that obtained using the 2D chemistry, with the advantage of a simpler experimental protocol and higher number of reads produced (Tyler et al., 2018).

Several versions of the MinION chemistry and base-calling software have existed since the device’s launch. The latest R9.4.1 chemistry version is based on the CsgG protein nanopore. Coupled with a computational approach that uses deep learning for base calling and the device’s “fast mode”, the MinION is able to sequence 450 bases/ second and yields up to 10 Gb of data from a 48-hour sequencing run (Magi, Semeraro, et al., 2017). Using circa 40 flow cells per genome, application of the R9.4 chemistry allowed the first nanopore sequencing of human genomes with a coverage of over 30× (Jain et al., 2018).

Benefits of using the MinION™ in lieu of other sequencing platforms

Many of the advantages of using the MinION device stem from its low cost of entry, size, portability and the ultra-long reads it produces. In addition, it can outperform other NGS platforms in several applications.

Variant detection

Called genetic variants, differences between genomes comprise of single nucleotide variants (SNVs) and structural variants (genomic alterations of at least 50 bp) such as copy number variations (mainly gene deletions and duplications), insertions, inversions, and translocations. Structural variants are responsible for more variable bases than single nucleotide variations, being involved in functional changes across populations and species and in the onset of many diseases, Mendelian and cancer alike (Magi, Semeraro, et al., 2017; Mahmoud et al., 2019). Due to the fact that an abundance of diseases is caused by mutations that can affect both genes and the non-coding genome, whole-genome sequencing has been rapidly adopted for clinical purposes because it provides a detailed set of a patient's individually specific mutation profile, allowing the customization of their treatment scheme.

Until recently, because short-read SGS platforms produce mistakes which translate into the assembled genomes, robust methods for the detection of structural variants were lacking. In 2017, using MinION's nanopore technology and the reads of up to hundreds of kilobases it can produce, Stancu et al. sequenced the whole diploid genomes of two patients at more $\geq 11\times$ coverage depth. The team employed a new computational pipeline and resolved the long-range structure of a complex structural variant in the patients. Even more, the long reads obtained from MinION facilitated the identification of significantly more structural variants than those which were detected using short-read Illumina sequencing data of the same two genomes (Stancu et al., 2017).

Results obtained using the MinION for genetic discovery indicated that although the generated data was reliable enough to be useful for detecting small variants at high recall rates, sequencing errors affected the precision of the approach employed, be it genome resequencing or assembly (Magi, Semeraro, et al., 2017). This is especially significant because, contrary to expectations, errors do not occur randomly during MinION's nanopore sequencing. A characteristic of ONT's sequencing technology is that errors generally occur in homopolymers and segmentation of the current profiles are problematic in homopolymer stretches longer than 6 nucleotides, which is the maximal k-mer size detected by the pore at a time (David et al., 2017). Furthermore, errors affect bases differently, as follows: deleted bases are primarily A and T which follow A and T nucleotides, while the most frequent substitutions are of C to G and vice versa (Magi, Giusti, et al., 2017). Therefore, recurrent errors are prone to appear even at sequencing coverages of $\geq 30\times$ and can lead to the discovery of false substitutions once every 10-100 kb and insertion/ deletion events every 1-10 kb. Despite this drawback and depending on the approach used for variant discovery, due to its specificity and sensitivity, MinION nanopore sequencing can be confidently used for the detection of copy number variants with high accuracy, with better results than those obtained using PacBio long-reads or SGS short-reads (Magi, Giusti, et al., 2017; Magi, Semeraro, et al., 2017).

Analysis of RNA expression

For RNA analysis on SGS platforms, several preparatory techniques are required such as fragmentation, conversion of RNA into complementary DNA and PCR amplification. These steps introduce experimental biases and impede the accurate determination of gene expression (Garalde et al., 2018; Oszolak and Milos, 2011). These are critical issues for clinical applications such as determining resistance to certain antibiotics. For example, resistance to aminoglycosides can arise from ribonucleotide modifications such as 7-methylguanoside. When RNA is converted into cDNA, epigenetic modifications can no longer be observed. However, TGS platforms such as ONT's MinION can directly measure RNA in its native molecules without any preparatory shearing and qPCR steps and, even more, there is no limit to the length of the sequenced molecules (Smith et al., 2019). Thus far, ONT's MinION has been used to evaluate the resistome of four extensively drug-resistant *Klebsiella pneumoniae* clinical isolates (Pitt et al., 2020), estimate 3' poly(A) tail length (Krause et al., 2019; Workman et al., 2019), produce an accurate profile of the transcriptome (Bolisetty et al., 2015; Sessegolo et al., 2019), and identify novel isoforms (Byrne et al., 2017; Krizanović et al., 2018) and RNA modifications (extensively referenced in Cozzuto et al., 2020).

Base modification detection

Previous to the advent of single-molecule sequencing technologies, the state-of-the-art method for identifying genome-wide DNA methylations was based on treatment with sodium bisulfite followed by SGS, which did not offer data regarding long-range methylation patterns. Nowadays, using nanopore technology, the methylation of nucleic acids can be directly detected in native molecules at the nucleotide level, in both DNA and RNA. Prior to the launch of the MinION, two groups independently showed that a single-channel nanopore system can distinguish between all five types of C-5 cytosine variants in synthetic DNA (Schreiber et al., 2013; Wescoe et al., 2014), with accuracy ranging from 92-98% for a target cytosine in a known sequence (Wescoe et al., 2014). Since then, two groups have used the MinION and developed software algorithms to identify cytosine methylation in human and/ or bacterial genomic DNA with more than 80% accuracy (Rand et al., 2017;

Simpson et al., 2017). Simpson et al.'s method discriminated among cytosine and 5-methylcytosine, while Rand et al.'s tool also identified 5-hydroxymethylcytosine, i.e. only two or three, respectively, of the five types of C-5 variants known. Additionally, in the case of the method employed by Simpson et al., their training set only focused on fully methylated genomic regions and did not detect those containing heterogenous methylation (Simpson et al., 2017).

However, four years after its launch, the MinION was already sensitive enough to sequence 5 picograms of purified 16S *E. coli* rRNA detected in 4.5 µg of total human RNA and to identify 7-methylguanosine and pseudouridine modifications (Smith et al., 2019). In 2020, Cozzuto et al. made available their open-source workflow for the analysis of direct RNA sequencing data, named MasterOfPores. The pipeline converts raw current intensities into processed data, maps the reads, predicts RNA modifications and estimates poly(A) tail lengths. The MasterOfPores workflow can be easily run on any computer with the Unix OS, does not require the installation of additional software and allows for four direct RNA MinION sequence data sets to be fully processed and analysed in 10 h on 100 CPUs (Cozzuto et al., 2020). Therefore, despite the current limitations, continuous efforts are being made in improving base-call accuracy and developing bioinformatic tools suitable for the accurate and thorough identification of the different base modifications in the genome and transcriptome.

Real-time targeted sequencing

For clinical applications especially, it is crucial to obtain and analyse genomic and transcriptomic data as quickly as possible. Neither Sanger nor SGS can match ONT's MinION in terms of on-site sequencing due to its portability and accessibility. The MinION can be used immediately upon arrival in an outbreak area without the need for calibration procedures or an actual laboratory set-up, which could be problematic in certain regions because of logistical (sample transportation and storage) or political issues (Lu et al., 2016). Even more, the MinION has a "Read Until" function: a mix of DNA fragments can be applied to the flow cell and when a DNA strand translocates through the nanopore, the current intensity profiles are compared to the expected pattern for a target sequence. If there is no match, that DNA strand is rejected by the nanopore, which will continue its analysis of a different DNA strand; otherwise, sequencing continues. Thus, for clinical applications such as in-field and point-of-care, the "Read Until" function dramatically reduces the time elapsed from sample acquisition to result analysis (Jain et al., 2016). Targeted reverse transcription PCR coupled with MinION sequencing have already been used to obtain rapid data turnaround for the management of disease outbreaks such as those of the Severe Acute Respiratory Syndrome Coronavirus 2, Ebola and Zika viruses (Paden et al., 2020; Quick et al., 2017, 2016).

De novo assembly

A major feature of the MinION is the read lengths it yields, which dramatically surpass those produced by the best performing SGS platforms. In its launch year, the MinION had already been used by Ip et al. for sequencing *E. coli* genomic DNA; the team obtained 1D and 2D read lengths of over 300 kb and up to 60 kb, respectively (Ip et al., 2015). In the same year, Loman and collaborators used only MinION sequence reads to *de novo* assemble the *E. coli* K-12 MG 655 chromosome into a single contig of 4.6 Mb, which had correctly ordered genes and 99.5% nucleotide identity. Instead of relying on the assemblers available at the time, which were unable to handle MinION sequencing errors, they corrected the long reads in two phases, by using a multiple-alignment process followed by polishing via a probabilistic model of the signal-level data (Loman et al., 2015). Although their approach yielded satisfactory results, the MinION produces data with high error rates compared with Sanger or SGS platforms, which have made its use problematic for *de novo* assembly algorithms designed for short reads and fewer errors (Goodwin et al., 2015). To overcome this impediment, continuous efforts have been made since the launch of the MinION Access Programme to develop computational approaches which are capable of accurately processing the high-throughput and error-prone MinION long-read sequences (Magi, Semeraro, et al., 2017). To drive this point forward, a comprehensive review of the performance of long-read assemblers for prokaryote whole genome sequencing was first published in 2019 and has been updated yearly since then (Wick and Holt, 2021).

Thus far, although ONT's nanopore technology has been primarily used for microbial sequencing, efforts have already been made to sequence and assemble several significantly more demanding eukaryotic genomes. For example, Istace et al. used the MinION to perform *de novo* sequencing and assembly of 21 genetically diverse *Saccharomyces cerevisiae* isolates and obtained assembly contiguities 14 times higher compared to Illumina-only assemblies. 65% of the chromosomes were covered by only one or two contigs, which enabled the accurate discovery and inspection of long structural variants present across the 21 sequenced genomes, variations which were generally missed using only short-read sequencing (Istace et al., 2017). Schimdt et al. used the MinION to sequence and assemble the genome of an even more complex eukaryote, namely that of *Solanum pennellii*, a wild tomato species. They obtained an assembly with an N50 value of 2.5 MB, but although the genome assembled using raw nanopore sequences was structurally highly similar to their reference genome, it was rich in homopolymer deletions and had a high error rate. Finally, the team used a hybrid assembly approach (see below) by applying Illumina short-reads to finesse the nanopore-reads assembly to an error rate lower than 0.02 when compared against the Illumina data set. Therefore, although nanopore technology and its dedicated assemblers are being continuously

improved, *de novo* genome assembly requires careful data analysis and polishing to obtain accurate and relevant error rates, followed by checking the genome quality and gene content (Schmidt et al., 2017).

Hybrid assembly

Because the majority of publicly-available genomes are incomplete, especially in the case of thousands of sequenced bacteria, significant interest and effort have been applied towards combining the complementary advantages of the accurate but short reads produced by Illumina sequencing and long but error prone reads resulted from TGS platforms. This method is known as “hybrid assembly” and can be achieved using one of two distinct approaches, namely short-read-first or long-read-first methods. The former method employs a scaffolding tool which uses long reads to join together Illumina contigs. The disadvantages of this approach are the misassembled sequences which arise from quite common scaffolding mistakes (Hunt et al., 2014). In long-read-first approaches either the uncorrected long reads are assembled first and short reads are used to correct errors in the assembly (Koren et al., 2017), or short reads are firstly used to error-correct long reads, the final assembly using the already-corrected long reads (Koren et al., 2012; Salmela and Rivals, 2014). Regardless if the error-correction is performed before or after the assembly, a higher long-read depth is required when the long-read-first approach is employed (Wick et al., 2017b). However, due to the read lengths produced by PacBio and ONT exceeding those of the length of repeats found in most bacterial genomes, complete hybrid genome assemblies have been achieved with even only one contig per replicon (Conlan et al., 2014; Koren et al., 2017).

THE MINION™: FOR THE LAST BRICK IN THE WALL

As previously discussed, genome assembly using only short-read sequences generally results in a number of unordered contigs. For example, an attempt by Karlsson et al. to *de novo* assemble the *Francisella* FSC996 chromosome (32% GC content) using Illumina short-reads with a 1000× coverage resulted in 40 contigs. The team changed tactics and performed hybrid genome assemblies on different *Francisella* strains using either MinION or PacBio long-reads coupled with Illumina-generated short-reads. Only a fifth of the long-reads from a single MinION run were sufficient to obtain a correct genome scaffold and easily assemble the contigs into a complete genome. Despite the older R7.3 chemistry available at that time and the higher error rates associated with it compared to the current R9.4.1, a single MinION run yielded a genome of ~99.8% sequence accuracy (Karlsson et al., 2015).

In 2017, Wick et. al used barcoded ONT libraries sequenced in multiplex on a single MinION flowcell and a hybrid-assembly approach to resolve the large genomes of 12 *Klebsiella pneumoniae* isolates (Wick et al., 2017a). The data was assembled either using ONT-only reads or a hybrid approach using complementary Illumina data, which had been previously found by the team to be insufficient for identifying the location of antimicrobial resistance genes (Gorrie et al., 2017). The group observed that ONT-only assemblies were prone to high error rates and were not substantially improved by read depth, although this parameter did positively influence sequence accuracy. The most accurate ONT-only assembly they obtained had an error rate equivalent to one error per 287 bp, implying that more often than not, a 1 kbp gene could contain an error. They concluded that such an assembly would not facilitate resistance allele or multi-locus sequence typing or be appropriate for studying phylogenomics or drug resistance transmission. Even more, they highlighted that the obtained MinION data did not accurately represent the small plasmids present in the bacterial isolates. Consequently, Unicycler, the assembly pipeline the group used, was unable to automatically complete the genome because of the missing plasmid sequence data. This issue is a great example of the importance of sample preparation depending on the scientific interest. The team hypothesized that either the DNA extraction method they used was improper for small DNA fragments such as plasmids, or the compromised plasmid recovery was a consequence of omitting the DNA shearing step during library preparation. Hence, because any existing small plasmids remained circular, no DNA strand ends were free to bind to the ONT adapters and the small plasmids evaded sequencing. Despite these caveats, the use of MinION-Illumina hybrid read sets resulted in 12 finished genomes, with a cost of around 150 USD per strain (Wick et al., 2017a).

Lemon et al. focused on testing the performance of the MinION in sequencing already isolated plasmid DNA. They resequenced three plasmids from a reference *K. pneumoniae* isolate; the accuracy of the draft genome was 99% when assembled using only MinION reads, the value increasing to 99.9% when the draft assembly was polished using Illumina MiSeq short-reads. The group also sequenced plasmid DNA from previously uncharacterized antibiotic resistant *E. coli* and *K. pneumonia* clinical isolates. The MinION reads enabled the facile detection of drug resistance genes in the draft genome assembly. Interestingly, by using isolated plasmid DNA instead of whole genomic DNA, full annotation of antimicrobial resistance genes was possible with quite low read depth, using only 2000-5000 reads, which can be produced within 20 minutes of sequencing (Lemon et al., 2017).

Even though Wick and colleagues were able to completely assemble the large *K. pneumoniae* genomes with MinION read lengths of N50>20 kb and over 14× coverage, their previous experience indicated that bacterial strains whose genomes have more frequent and larger repetitive sequences may require more distinct approaches (Wick et al., 2017b). For example,

some *Shigella* (cause of bacillary dysentery) genomes contain a couple hundred pseudogenes, many high-copy-number repeats of ~1 kbp associated with hundreds of copies of insertion sequence elements and numerous indels, translocations and inversions. Their findings demonstrated that compared to *K. pneumoniae*, *Shigella* genomes require around twice the nanopore sequencing depth to obtain complete hybrid assemblies (Wick et al., 2017b; Yang et al., 2005). Another problematic species is *Acinetobacter*, with one of the earliest multiple-antibiotic-resistant isolates *A. baumannii* strain A1 containing a highly repetitive biofilm-associated gene variable in length but which can reach over 25 kbp (Holt et al., 2016). Holt et al. did assemble the genome using PacBio and Illumina reads, but this required significant manual intervention in both the assembly and annotation steps. In contrast, Wick et al. used simulated reads from the reference genome produced by Holt et al. and the hybrid assembly tool Unicycler and obtained an exclusively automatic complete genome assembly of *A. baumannii* strain A1 (Wick et al., 2017b).

Similar to Wick et al., Todd and colleagues also used sample multiplexing for short-read Illumina and long-read MinION platforms and assembled the obtained data using Unicycler. They succeeded in assembling seven genomes of *Fusobacterium* into highly accurate and singular complete chromosomes, compared to the previously available draft assemblies containing 24-67 fragmented contigs. Even more, they revealed the presence of a genomic inversion of over 450 kb in the previously existing *F. nucleatum* subsp. *nucleatum* ATCC 25586 genome assembly, which they managed to correct using the hybrid assembly approach (Todd et al., 2018).

In 2018, two independent groups at the National Microbiology Laboratory (Public Health Agency of Canada) sequenced four well-characterized isolates in replicate, using the latest flow cells, sequencing chemistries and software available at the time. When designing an experiment, consistent workflows between replicate runs are ideal. However, in this study, because of the accelerated evolution of MinION sequencing (reaction chemistries, software), the researchers were unable to ensure technical consistency throughout their experiments. Furthermore, the inter-run variability regarding the obtained yields made it challenging to estimate yield per flow cell based solely on the number of samples and input DNA. Despite these caveats, the group observed that both sequencing yield and quality had improved throughout the experiment, sequence alignment accuracies being over 94% for 1D and 2D chemistries alike, the resulting data being equally suitable for genome assembly. Overall, the high error rate (which has since been improved and is easier to overcome by the available software), the inconsistencies observed between runs caused by the rapidly changing kits, reagents and software were considered limiting factors for adopting MinION sequencing for wide-scale use. Despite this, the study recognized the advantages the MinION brings for whole genome sequencing of bacteria and its capacity for pathogen identification even in samples with DNA concentrations lower than those recommended by ONT (Tyler et al., 2018).

In a recent publication, two reference strains and two field isolates of *Campylobacter jejuni* were sequenced using Illumina MiSeq and MinION. The sequences were assembled using either designated assemblers for short-reads (SPAdes) and long-reads (Canu), respectively, or Unicycler, which performs hybrid genome assemblies using both read types. The Illumina raw data assembled using SPAdes had the most nucleotide identity and genes correctly annotated when compared to the PacBio-generated reference genomes, but the short-read lengths yielded fragmented contigs and a greater, misrepresenting number of coding sequences. On the other hand, MinION-only assemblies using Canu were contiguous and enabled the easy identification of plasmids, but had the least accuracy and contained numerous errors such as substitutions and indels, which lead to inaccurate gene annotations and sequence typing. Finally, the MiSeq and MinION data were combined to obtain hybrid genome assemblies. The number of mismatches was slightly higher than in the Illumina-only assembly, possibly because the assembly pipeline (Unicycler) heavily relies on MinION data in regions of low Illumina read coverage. The assembly accuracies were improved when the amount of MinION data used in the assembly was increased from 40× to 200× and the resulted hybrid-assembled genomes were contiguous and completely circularized. The bacterial genomes constructed using the hybrid approach were the most useful for identifying plasmids, large genomic rearrangements and repetitive elements such as genes coding for ribosomal and transport RNA (Neal-McKinney et al., 2021).

Similar findings were reported by Goldstein et al., who sequenced nine bacterial genomes with GC contents varying from low to high, using either Illumina MiSeq or the MinION. They tested short-read-first and long-read-first assembly approaches. Regarding bacterial strains with extreme GC contents, the researchers mentioned that because of the bias from the Illumina libraries, polishing the long-read data using MiSeq sequences was a challenge. In spite of this, they concluded that using MinION reads for initial assembly followed by Illumina short-reads for error-correction provided the most contiguous genomes, which were accurate enough for annotating challenging regions to sequence, such as secondary metabolite biosynthetic gene clusters and insertion sequence elements (Goldstein et al., 2019).

CONCLUSION: THE ONLY WAY IS FORWARD

The MinION is a paradigm-shifting device due to its nanopore sequencing method, portability and commercial diffusion to research and clinical applications. The use of nanopore sequencing greatly improves *de novo* genome assemblies when

considering N50, contig numbers and its robustness with extreme GC content organisms, allowing the detection and exploration of structural and sequence variants. Even though assembling MinION nanopore reads alone is feasible, issues with sequence accuracy and small plasmid recovery were reported. Therefore, scientists have called for more suitable approaches to library preparation and base-calling algorithms to tackle the caveats of *de novo* assemblies relying exclusively on nanopore sequences (Wick et al., 2017a). Because long-read sequencing is becoming more common especially in microbial genomics, long-read assembly is also on a continuous rise. Therefore, the development and refinement of designated assemblers is paramount for the scientific community to be able to actually use the full potential of these rapidly and dynamically-evolving sequencing technologies (Wick and Holt, 2021).

Despite the drawbacks that nanopore sequencing has for *de novo* genome assembly, it has become a robust and highly-appreciated method for succeeding where other platforms have not. Although at first glance, the MinION seems to be highly similar to PacBio in that they produce similarly sized reads, the two systems have been proven to have quite distinct applications and target users. Whereas the MinION stands out due to its convenient size, PacBio platforms are dramatically large and heavy and require a substantial initial investment, being more suitable for sequencing centres where space and infrastructure is not an issue. The MinION is the complete opposite (Karlsson et al., 2015). ONT's device has thus far been shown to be easily operated in the field for rapid real-time pathogen identification, which is advantageous for more general sample analysis than what can be achieved using real-time PCR. The device performed very well even with lower DNA input than what the manufacturer deemed necessary, as indicated by Tyler et al. Even more, the MinION has already gone where no other sequencing device has. As part of a 6-month long NASA experiment on the International Space Station, it was used to successfully sequence bacteriophage, bacterial and eukaryote DNA in microgravity. In parallel, on Earth, *de novo* assembly of the MinION data showed over 96% consensus pairwise identity. The MinION's performance was benchmarked against Illumina's MiSeq and PacBio's RS II platforms, the results being promising for MinION applications for in-space monitorization of human health and response to spaceflight and the identification of DNA-based extra-terrestrial lifeforms (Castro-Wallace et al., 2017).

Already a jack of many trades, the MinION has proven itself a robust and reliable master for microbial sequencing and assembly, especially when combined with complementary sequencing methods such as those provided by Illumina. As was discussed above, hybrid genome assembly using MinION long reads enables closing the gap regions where the reach of short-read sequencers was insufficient. Complete and accurate genomes are essential for a variety of scientific and clinical applications among which phylogenetic studies and infectious disease epidemiology, for which a hybrid assembly approach is, momentarily, the golden standard for obtaining high-quality, detailed and accurate data sets. However, a hybrid assembly approach is not always necessary and the additional labour and costs may be misplaced for bulk sequencing of bacterial isolates. On the other hand, demonstrating genetic relatedness or identity is commonly a necessity when managing pathogen outbreaks and establishing regulatory action. The most accurate comparison between isolates would be ensured by using hybrid-assembled genomes due to their completeness and contiguity, which enables the distinction of potential drug resistance genes and virulence factors belonging either to plasmid or to chromosomal DNA. If interested in the characterization of the complete genetic content of a bacterium or if wanting to compare highly related isolates, then hybrid assembly is the best method for obtaining the additional detail which may lack from incomplete short-read-only or error-prone long-read-only assemblies (Neal-McKinney et al., 2021).

To conclude, whereas other sequencing platforms require high capital investment which restricts the location of sequencing infrastructure to high-brow sequencing centres (Stancu et al., 2017), the MinION is currently unmatched in terms of range of applications, up-front costs, ease of use, initial set-up, space and infrastructure requirements. Furthermore, the MinION is an extremely flexible device and opens the genomics, transcriptomics and epigenomics field to virtually anyone with sufficient funds and knowledge; as long as a Unix-compatible computer is in sight, the MinION can be used. Although similar technologies in terms of long-read sequencing do exist, the MinION stands out because it can reach smaller, independent laboratories which may be financially unable or just undesiring of a high investment such as that required for a PacBio system. Currently, the biggest caveats of MinION sequencing continue to be the error rate, which has been declining since its launch, and the dynamic changes to the flow cells, reagents and kits. The higher error rates can be handled via a few different approaches such as bioinformatic work-arounds, increasing sequencing depth, error-correcting using already available short-read datasets or sequencing in parallel using a complementary SGS platform. Regarding the common changes to the technology, these seem to be problematic mainly for wide-scale use. Considering the timeline of sequencing technologies and although fast evolving, nanopore sequencing has just surpassed its infancy and, by all accounts, its future is bright and shining and will open many doors for us (or rather, in the case of genomes, close them).

REFERENCES

- Balloux, F., Brønstad Brynildsrud, O., van Dorp, L., Shaw, L. P., Chen, H., Harris, K. A., ... Eldholm, V. (2018, December 1). From theory to practice: translating whole-genome sequencing (WGS) into the clinic. *Trends in Microbiology*, Vol. 26, pp. 1035–1048. Elsevier. <https://doi.org/10.1016/j.tim.2018.08.004>
- Bolisetty, M. T., Rajadinakaran, G., and Graveley, B. R. (2015). Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biology*, 16(1), 1–12. <https://doi.org/10.1186/s13059-015-0777-z>
- Bush, S. J., Foster, D., Eyre, D. W., Clark, E. L., De Maio, N., Shaw, L. P., ... Walker, A. S. (2020). Genomic diversity affects the accuracy of bacterial single-nucleotide polymorphism-calling pipelines. *GigaScience*, 9(2), 1–21. <https://doi.org/10.1093/GIGASCIENCE/GIAA007>
- Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., ... Vollmers, C. (2017). Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nature Communications*, 8(1), 1–11. <https://doi.org/10.1038/ncomms16027>
- Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., ... Burton, A. S. (2017). Nanopore DNA sequencing and genome assembly on the International Space Station. *Scientific Reports*, 7(1), 1–12. <https://doi.org/10.1038/s41598-017-18364-0>
- Cherf, G. M., Lieberman, K. R., Rashid, H., Lam, C. E., Karplus, K., and Akeson, M. (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nature Biotechnology*, 30(4), 344–348. <https://doi.org/10.1038/nbt.2147>
- Collins, F. S., Morgan, M., and Patrinos, A. (2003). The Human Genome Project: Lessons from large-scale biology. *Science*, Vol. 300, pp. 286–290. Science. <https://doi.org/10.1126/science.1084564>
- Conlan, S., Thomas, P. J., Deming, C., Park, M., Lau, A. F., Dekker, J. P., ... Segre, J. A. (2014). Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing *Enterobacteriaceae*. *Science Translational Medicine*, 6(254). <https://doi.org/10.1126/scitranslmed.3009845>
- Cozzuto, L., Liu, H., Prysacz, L. P., Pulido, T. H., Delgado-Tejedor, A., Ponomarenko, J., and Novoa, E. M. (2020). MasterOfPores: A workflow for the analysis of oxford nanopore direct RNA sequencing datasets. *Frontiers in Genetics*, 11. <https://doi.org/10.3389/fgene.2020.00211>
- David, M., Dursi, L. J., Yao, D., Boutros, P. C., and Simpson, J. T. (2017). Nanocall: An open source basecaller for Oxford Nanopore sequencing data. *Bioinformatics*, 33(1), 49–55. <https://doi.org/10.1093/bioinformatics/btw569>
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., ... Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, 323(5910), 133–138. <https://doi.org/10.1126/science.1162986>
- Feng, Y., Zhang, Y., Ying, C., Wang, D., and Du, C. (2015, February 1). Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics and Bioinformatics*, Vol. 13, pp. 4–16. Elsevier. <https://doi.org/10.1016/j.gpb.2015.01.009>
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... Venter, J. C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223), 496–512. <https://doi.org/10.1126/science.7542800>
- Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., ... Venter, J. C. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science*, 270(5235), 397–403. <https://doi.org/10.1126/science.270.5235.397>
- Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., ... Turner, D. J. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods* 2018 15:3, 15(3), 201–206. <https://doi.org/10.1038/nmeth.4577>
- Goldstein, S., Beka, L., Graf, J., and Klassen, J. L. (2019). Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genomics*, 20(1), 1–17. <https://doi.org/10.1186/s12864-018-5381-7>
- Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., and McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Research*, 25(11), 1750–1756. <https://doi.org/10.1101/gr.191395.115>
- Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, Vol. 17, pp. 333–351. Nature Publishing Group. <https://doi.org/10.1038/nrg.2016.49>
- Gorrie, C. L., Mirc Eta, M., Wick, R. R., Edwards, D. J., Thomson, N. R., Strugnell, R. A., ... Holt, K. E. (2017). Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clinical Infectious Diseases*, 65(2), 208–215. <https://doi.org/10.1093/cid/cix270>
- Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., A. G., V. S., ... G, T. (2013). QUAST: quality assessment tool for genome assemblies. 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Heather, J. M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, Vol. 107, pp. 1–8. Elsevier. <https://doi.org/10.1016/j.ygeno.2015.11.003>

- Holt, K. E., Hamidian, M., Kenyon, J. J., Wynn, M. T., Hawkey, J., Pickard, D., and Hall, R. M. (2016). Genome sequence of *Acinetobacter baumannii* strain A1, an early example of antibiotic-resistant global clone 1. *Genome Announcements*, 3(2). <https://doi.org/10.1128/genomeA.00032-15>
- Hunt, M., Newbold, C., Berriman, M., and Otto, T. D. (2014). A comprehensive evaluation of assembly scaffolding tools. *Genome Biology* 2014 15:3, 15(3), 1–15. <https://doi.org/10.1186/GB-2014-15-3-R42>
- Ip, C. L. C., Loose, M., Tyson, J. R., de Cesare, M., Brown, B. L., Jain, M., ... Olsen, H. E. (2015). MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research*, 4. <https://doi.org/10.12688/f1000research.7201.1>
- Istace, B., Friedrich, A., D'Agata, L., Faye, S., Payen, E., Beluche, O., ... Aury, J. M. (2017). De novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *GigaScience*, 6(2). <https://doi.org/10.1093/gigascience/giw018>
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., ... Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* 2018 36:4, 36(4), 338–345. <https://doi.org/10.1038/nbt.4060>
- Jain, M., Olsen, H. E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1), 1–11. <https://doi.org/10.1186/s13059-016-1103-0>
- Karlsson, E., Lärkeryd, A., Sjödin, A., Forsman, M., and Stenberg, P. (2015). Scaffolding of a bacterial genome using MinION nanopore sequencing. *Scientific Reports* 2015 5:1, 5(1), 1–8. <https://doi.org/10.1038/srep11996>
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., ... Phillippy, A. M. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology* 2012 30:7, 30(7), 693–700. <https://doi.org/10.1038/nbt.2280>
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 27(5), 722–736. <https://doi.org/10.1101/gr.215087.116>
- Krause, M., Niazi, A. M., Labun, K., Torres Cleuren, Y. N., Müller, F. S., and Valen, E. (2019). TailFindR: Alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA*, 25(10), 1229–1241. <https://doi.org/10.1261/rna.071332.119>
- Križanović, K., Echchiki, A., Roux, J., and Šikić, M. (2018). Evaluation of tools for long read RNA-seq splice-aware alignment. *Bioinformatics*, 34(5), 748–754. <https://doi.org/10.1093/bioinformatics/btx668>
- Laehnemann, D., Borkhardt, A., and McHardy, A. C. (2016). Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Briefings in Bioinformatics*, 17(1), 154–179. <https://doi.org/10.1093/bib/bbv029>
- Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T. H., ... Ussery, D. W. (2015, February 27). Insights from 20 years of bacterial genome sequencing. *Functional and Integrative Genomics*, Vol. 15, pp. 141–161. Springer. <https://doi.org/10.1007/s10142-015-0433-4>
- Land, M., Hyatt, D., Jun, S. R., Kora, G. H., Hauser, L. J., Lukjancenko, O., and Ussery, D. W. (2014). Quality scores for 32,000 genomes. *Standards in Genomic Sciences*, 9(1), 1–10. <https://doi.org/10.1186/1944-3277-9-20>
- Lemon, J. K., Khil, P. P., Frank, K. M., and Dekker, J. P. (2017). Rapid nanopore sequencing of plasmids and resistance gene detection in clinical isolates. *Journal of Clinical Microbiology*, 55(12), 3530–3543. <https://doi.org/10.1128/JCM.01069-17>
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., ... Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, Vol. 2012. J Biomed Biotechnol. <https://doi.org/10.1155/2012/251364>
- Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., and Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, 30(5), 434–439. <https://doi.org/10.1038/nbt.2198>
- Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, 12(8), 733–735. <https://doi.org/10.1038/nmeth.3444>
- Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics and Bioinformatics*, Vol. 14, pp. 265–279. Elsevier. <https://doi.org/10.1016/j.gpb.2016.05.004>
- Madoui, M. A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., ... Aury, J. M. (2015). Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics*, 16(1), 1–11. <https://doi.org/10.1186/s12864-015-1519-z>
- Magi, A., Giusti, B., and Tattini, L. (2017). Characterization of MinION nanopore data for resequencing analyses. *Briefings in Bioinformatics*, 18(6), 940–953. <https://doi.org/10.1093/bib/bbw077>
- Magi, A., Semeraro, R., Mingrino, A., Giusti, B., and D'Aurizio, R. (2017). Nanopore sequencing data analysis: State of the art, applications and challenges. *Briefings in Bioinformatics*, 19(6), 1256–1272. <https://doi.org/10.1093/bib/bbx062>
- Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., and Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biology*, Vol. 20, pp. 1–14. BioMed Central. <https://doi.org/10.1186/s13059-019-1828-7>

- Mavromatis, K., Land, M. L., Brettin, T. S., Quest, D. J., Copeland, A., Clum, A., ... Kyripides, N. C.** (2012). The Fast Changing Landscape of Sequencing Technologies and Their Impact on Microbial Genome Assemblies and Annotation. *PLoS ONE*, 7(12), e48837. <https://doi.org/10.1371/journal.pone.0048837>
- Neal-McKinney, J. M., Liu, K. C., Lock, C. M., Wu, W. H., and Hu, J.** (2021). Comparison of MiSeq, MinION, and hybrid genome sequencing for analysis of *Campylobacter jejuni*. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-84956-6>
- Ozsolak, F., and Milos, P. M.** (2011). RNA sequencing: Advances, challenges and opportunities. *Nature Reviews Genetics*, Vol. 12, pp. 87–98. Nature Publishing Group. <https://doi.org/10.1038/nrg2934>
- Paden, C. R., Tao, Y., Queen, K., Zhang, J., Li, Y., Uehara, A., and Tong, S.** (2020). Rapid, Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases*, 26(10), 2401–2405. <https://doi.org/10.3201/eid2610.201800>
- Pitt, M. E., Nguyen, S. H., Duarte, T. P. S., Teng, H., Blaskovich, M. A. T., Cooper, M. A., and Coin, L. J. M.** (2020). Evaluating the genome and resistome of extensively drug-resistant *Klebsiella pneumoniae* using native DNA and RNA Nanopore sequencing. *GigaScience*, 9(2), 1–14. <https://doi.org/10.1093/gigascience/giaA002>
- Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., ... Loman, N. J.** (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols*, 12(6), 1261–1266. <https://doi.org/10.1038/nprot.2017.066>
- Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., ... Carroll, M. W.** (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), 228–232. <https://doi.org/10.1038/nature16996>
- Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., and Paten, B.** (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*, 14(4), 411–413. <https://doi.org/10.1038/nmeth.4189>
- Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., ... Jaffe, D. B.** (2013). Characterizing and measuring bias in sequence data. *Genome Biology*, 14(5), 1–20. <https://doi.org/10.1186/gb-2013-14-5-r51>
- Salmela, L., and Rivals, E.** (2014). LoRDEC: Accurate and efficient long read error correction. *Bioinformatics*, 30(24), 3506–3514. <https://doi.org/10.1093/bioinformatics/btu538>
- Schmidt, M. H.-W., Vogel, A., Denton, A. K., Istace, B., Wormit, A., van de Geest, H., ... Usadel, B.** (2017). De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell*, 29(10), 2336–2348. <https://doi.org/10.1105/tpc.17.00521>
- Schreiber, J., Wescoe, Z. L., Abu-Shumays, R., Vivian, J. T., Baatar, B., Karplus, K., and Akeson, M.** (2013). Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, 110(47), 18910–18915. <https://doi.org/10.1073/pnas.1310615110>
- Schwarze, K., Buchanan, J., Fermont, J. M., Dreau, H., Tilley, M. W., Taylor, J. M., ... Wordsworth, S.** (2020). The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genetics in Medicine*, 22(1), 85–94. <https://doi.org/10.1038/s41436-019-0618-7>
- Sessegolo, C., Cruaud, C., Da Silva, C., Cologne, A., Dubarry, M., Derrien, T., ... Aury, J. M.** (2019). Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules. *Scientific Reports*, 9(1), 1–12. <https://doi.org/10.1038/s41598-019-51470-9>
- Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., and Timp, W.** (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, 14(4), 407–410. <https://doi.org/10.1038/nmeth.4184>
- Smith, A. M., Jain, M., Mulrone, L., Garalde, D. R., and Akeson, M.** (2019). Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS ONE*, 14(5). <https://doi.org/10.1371/journal.pone.0216709>
- Stancu, C. M., Van Roosmalen, M. J., Renkens, I., Nieboer, M. M., Middelkamp, S., De Lig, J., ... Kloosterman, W. P.** (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-01343-4>
- Todd, S. M., Settlege, R. E., Lahmers, K. K., and Slade, D. J.** (2018). *Fusobacterium* genomics using MinION and Illumina sequencing enables genome completion and correction. *MSphere*, 3(4). <https://doi.org/10.1128/msphere.00269-18>
- Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., and Corbett, C. R.** (2018). Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Scientific Reports*, 8(1), 1–12. <https://doi.org/10.1038/s41598-018-29334-5>
- Wescoe, Z. L., Schreiber, J., and Akeson, M.** (2014). Nanopores discriminate among five C5-cytosine variants in DNA. *Journal of the American Chemical Society*, 136(47), 16582–16587. <https://doi.org/10.1021/ja508527b>
- Wick, R. R., and Holt, K. E.** (2021). Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research*, 8, 2138. <https://doi.org/10.12688/f1000research.21782.4>
- Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E.** (2017a). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics*, 3(10), e000132. <https://doi.org/10.1099/MGEN.0.000132>

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017b). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, 13(6), e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>

Wick, R. R., Judd, L. M., and Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20(1), 1–10. <https://doi.org/10.1186/s13059-019-1727-y>

Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., ... Timp, W. (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods*, 16(12), 1297–1305. <https://doi.org/10.1038/s41592-019-0617-2>

Yang, F., Yang, J., Zhang, X., Chen, L., Jiang, Y., Yan, Y., ... Jin, Q. (2005). Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Research*, 33(19), 6445–6458. <https://doi.org/10.1093/nar/gki954>

Zagordi, O., Klein, R., Däumer, M., and Beerenwinkel, N. (2010). Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Research*, 38(21), 7400–7409. <https://doi.org/10.1093/nar/gkq655>

ACKNOWLEDGEMENTS AND AUTHOR AFFILIATION

The author was supported by a grant from the Romanian National Authority for Scientific Research and Innovation, CNCS-UEFISCDI, project no. PN-III-P4-ID-PCE-2020-0656: Sequencing the genome of a useful bacteria: *Paenarthrobacter nicotinovorans* – next step in extending its biotechnological applications.

1 Biochemistry and Molecular Biology Laboratory, Department of Biology, Alexandru Ioan Cuza University of Iasi, Iasi 700506, Romania

* amada.elsabeh@uaic.ro