On-Site MinION Sequencing

Lucky R. Runtuwene, Josef S. B. Tuda, Arthur E. Mongan, and Yutaka Suzuki

Abstract DNA sequencing has reached an unprecedented level with the advent of Oxford Nanopore Technologies' MinION. The low equipment investment, ease of library preparation, small size, and powered only by a laptop computer enable the portability for on-site sequencing. MinION has had its role in clinical, biosecurity, and environmental fields. Here, we describe the many facets of on-site sequencing with MinION. First, we will present some field works using MinION. We will discuss the requirements for targeted or whole genome sequencing and the challenges faced by each technique. We will also elaborate the bioinformatics procedures available for data analysis in the field. MinION has greatly changed the way we do sequencing by bringing the sequencer closer to the biodiversity. Although numerous limitations exist for MinION to be truly portable, improvements of procedures and equipment will enhance MinION's role in the field.

The arrival of MinION, a portable long read DNA sequencer, has changed many aspects of DNA sequencing. The low overhead cost, ease of library preparation, small size, and use of a USB port of a laptop computer as its power supply have placed DNA sequencing at the frontier of research in the field. Researchers now have the freedom to perform DNA sequencing in the field where they collect samples. MinION has proven to be very beneficial in the clinical, biosecurity, and environmental fields (Hardegen et al. 2018).

Since its availability on the market through the MinION Access Program (MAP) in 2015, MinION has significantly evolved. From the first iteration that produced only up to 1 GB of data, the most recent flow cell is able to produce up to10–20 GB of data. The accuracy of MinION has improved dramatically from 65 to 88% (Lu et al. 2016) to greater than 90% (Oxford Nanopore Technologies

L. R. Runtuwene (✉) · Y. Suzuki (✉)
Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan
e-mail: luckyruntuwene@edu.k.u-tokyo.ac.jp; ysuzuki@hgc.jp

J. S. B. Tuda · A. E. Mongan
Faculty of Medicine, Sam Ratulangi University, Manado, Indonesia

Y. Suzuki (ed.), *Single Molecule and Single Cell Sequencing*, Advances in Experimental Medicine and Biology 1129,
https://doi.org/10.1007/978-981-13-6037-4_10

2017). Many options are available in proprietary kits, ranging from a PCR-free quick preparation to high-accuracy PCR-based kits. Special purposes kits are also available, e.g., barcoding, low input, and ultralow input. Furthermore, many aspects of DNA sequencing have also been simplified with the prospective release of equipment, such as Zumbador (simplifying DNA extraction and library preparation), Voltrax (simplifying library preparation), MinIT (a portable CPU to substitute the whole laptop), and even further miniaturization of MinION itself through sequencing on a smartphone.

The MinION principle relies on the recognition of an electrical current disturbance in a nanopore protein. When a strand of DNA is inside the nanopore, the nucleotides induce a change in the basal electrical current. This electric current disturbance pattern is specific to kmers and is recognized by sensors and translated into a sequence of nucleotides by an algorithm. The exclusion of lasers and fluorescence, which are commonly used in traditional and high-throughput DNA sequencers, permits the sequencing technology to be portable and energy efficient. Base-calling can be performed on-the-go or after sequencing. An online connection is also optional; thus, sequencing can occur in seclusion.

One of the most powerful aspects of MinION is the ability to sample biodiversity that is not easily sampled or requires an unmanageable amount of time. Many of the planet's biological resources are locked in the countries inaccessible to DNA sequencers. Consequently, instead of taking the sample to a laboratory, a group of researchers decided to take DNA sequencing to a remote rainforest of Tanzania (Menegon et al. 2017). These researchers successfully identified a frog species based on its DNA barcode, i.e., the small subunit of ribosomal RNA (16s rRNA). Despite the high error rate of the first generation of flow cell, a generation of consensus sequence exhibited 97% similarity to the reference genome of the frog. Another group of researchers conducted on-site sequencing in the Ecuadorian Choco rainforest to sequence the DNA barcodes of a toad and three species of snakes (Pomerantz et al. 2018), two of which were previously undescribed. These researchers appreciated how mobile laboratories and portable sequencing can accelerate species identification in remote areas. Extreme conditions also complicate the transport of samples to a laboratory, e.g., in the Canadian high Arctic. In this setting, MinION has been utilized to identify microorganisms through 16s rRNA sequencing. These researchers were able to identify bacteria, archaea, and eukaryotes in the permafrost layer (Goordial et al. 2017). Despite the high error rate, the results were still comparable to Illumina sequencing. The region resembles the planet Mars and acts as a test area should the need to identify Martian microorganisms arise. Similarly, DNA sequencing has been performed in space at the International Space Station (ISS) (Castro-Wallace et al. 2017). Here, three types of DNA from *E. coli* K-12, bacteria phage, and BALB/C mice were sequenced over four MinION runs, and the results did not differ between the space and Earth, suggesting MinION makes sequencing in extreme environments possible.

The advantage of MinION's low overhead cost is welcomed by a group of researchers working with metagenomics in waste sludge (Hardegen et al. 2018). Sequencing is rarely used for this purpose partially due to the required investment.

Here, MinION has aided in the analysis of real-time changes in microorganisms in the biomass industry, increasing the efficiency of the production process. MinION has been implemented in food security by identifying begomoviruses causing cassava mosaic disease (Boykin et al. 2018), further protecting the loss of productivity for 800 million people worldwide. The researchers performed the experiments in Tanzania, Uganda and Kenya, and all of these location lack sequencing facilities. The benefit of real-time analysis allowed the identification of begomoviruses as early as 11 s after sequencing. MinION has been applied for genomic surveillance, e.g., yellow fever outbreak in Brazil. By sequencing 52 whole genome sequences of yellow fever virus (YFV), the researchers identified an early case of sylvatic transmission. Seventy percent of the new sequences were generated using MinION (N. Faria et al. 2018), further demonstrating MinION's ability to complement and potentially replace established techniques.

When working on-site to identify the causative agent of a disease, researchers are puzzled by the available methods to enrich pathogenic nucleic acids. Several methods might be employed in this setting. Direct metagenomics sequencing is preferable when the causative agent is not known. However, this method has a direct correlation with the concentration of pathogens in the sample (Houldcroft et al. 2017). A project in Brazil to improve the molecular surveillance and sequencing of Zika virus (ZIKV) reported that direct metagenomics is difficult to perform with MinION when the viral concentration in the sample is low (N. R. Faria et al. 2016). The sequencing projects were performed in a mobile laboratory across five federal states in Northern region of Brazil. A total of 1349 samples were tested in 16 days during this field trip. Due to the low virus concentration, the genomes generated were fragmented and exhibit less than 50% coverage. A low sample concentration must undergo repeated sequencing using several flow cells to reach the adequate depth, which is not feasible in the field where resources are limited. To overcome this problem, a PCR-amplified tiling sequencing has been developed specifically for ZIKV (Quick et al. 2017). This technique requires segmental amplification using multiple sets of primers to amplify the entire genome of a virus. This methodology can be applied to samples containing as few as 50 genome copies per reaction, and the result are available 1–2 days following clinical sample acquisition. This technique can also be applied in other virus genomes. For example, the same technique was employed in the Ebola outbreak in Africa on 2016. MinION was used to sequence the entire genome of 146 EBOV samples (Quick et al. 2016). Using reverse transcription and tiling amplification with PCR, the fastest confirmation of EBOV infection was attainable within 1 h. Further, by combining the results obtained in the field with a database, the rate of EBOV mutation is determined. Again, successful detection depends on the viral titer. While this technique is feasible for small-genome viruses, sequencing viruses with 20–50-kB genomes using this technique is impractical due to the number of primers needed. The third technique commonly used to enrich viral genomes is oligonucleotide hybridization. Specific or degenerate probes bound to a solid phase (e.g., streptavidin bead) bind to the virus nucleic acids and are pulled down. The captured virus is subsequently adapter ligated and low-cycle amplified prior to sequencing. However, this technique

Fig. 1 Portable PCR

has not been tested in an on-site sequencing context; this technique most consistently yields full genome results in the laboratory.

When full-genome sequencing of a pathogen is not likely, targeted sequencing is the method of choice. Parts of the genome can be amplified by multiple methods and subsequently sequenced. The use of this technique offers the advantage of cost-effectiveness by mixing amplicons together in one sequence run. The mixture can be either multiple target genes of a sample or multiple samples of the same gene. Due to the high output of the most recent flow cell, a researcher can sequence up to 96 amplicons and decrease the cost of a sequencing run to $5.2–9.4 per amplicon. Multiplexing can be achieved by using the proprietary kit (only available in 1D rapid sequencing mode) or the self-produced barcodes attached to the forward primer of target gene (for compatibility with 1D^2 sequencing mode). As a proof-of-concept, we have performed a multiplex sequencing of amplicons of genes related to drug resistance in *Plasmodium falciparum* in the field (Runtuwene et al. 2018). DNA was extracted using spin-columns, and the 1D^2 sequencing protocol for genomic DNA was employed. Amplification was conducted using the mobile thermal cycler (Fig. 1), which only requires a USB connection to a laptop computer. The flow cell contained a mixture of 11 amplicons of nine genes, and the sequencing was run for 48 h. Consensus sequence was created for each amplicon per sample, yielding an 84.56% accuracy for sequencing. This seemingly low accuracy despite the use of flow cell R9.4 (i.e., the recent version) is due to the abundance of AT-rich and homopolymer tracts in the parasite genome. Nonetheless, we determined the parasites' drug-resistance status in Manado (Indonesia) (Fig. 2 presents the remote laboratory where we performed the sequencing), Thailand, and Vietnam from blood samples that were either frozen or preserved in an FTA card.

An interesting application for the use of MinION to assist in an outbreak is the development of a system to detect all of the viruses causing hemorrhagic diseases (Brinkmann et al. 2017). This technique involves targeted sequencing using two pools of 285 and 256 primer pairs for the identification of 46 virus species. Target

Fig. 2 MinION in the remote laboratory

genes are amplified with PCR. Using clinical specimens, the panel enables characterization of the causative agent within 10 min of sequencing, and a definitive diagnosis can be procured in less than 3.5 h.

Currently, the required starting DNA concentration is approximately 1 μg for 1D^2 and 10–100 ng for 1D sequencing. Clinical samples rarely yield high concentrations for pathogen sequencing. In the event of the unavailability of a thermal cycler, PCR can be substituted with isothermal amplifications to enrich the input DNA. Many techniques ranging from targeted [e.g., loom-amplified isothermal amplification (LAMP)] to whole genome isothermal amplifications [e.g., multiple displacement amplification (MDA)] are possible. LAMP relies on a denaturing-DNA-polymerase, such as Bst polymerase, to amplify a genome segment flanked by two to three pairs of primers (Notomi et al. 2000). The denaturing nature of the polymerase (i.e., simultaneously denatures the double strand of DNA and elongates the primers) allows the LAMP reaction to be run in isothermal conditions, thus completely eliminating the need for a thermal cycler. Further, LAMP reagents can be dried to assure their stability upon transportation to the field (Hayashida et al. 2015). Applying the LAMP technique as the amplification method, we performed a genomic epidemiology study of dengue virus in Indonesia and Vietnam (Yamagishi et al. 2017). Up to 141 and 80 DENV-positive samples were amplified isothermally and sequenced with MinION. We were able to determine the infecting virus serotype and reported a successful detection rate of 79%. Serotype can also be determined despite the 74–80% identity. We also developed LAMP combined with MinION sequencing to detect and differentiate among five species of malaria parasites by sequencing 18 s rRNA genes.

One of the main bottlenecks of MinION's analysis (and all next-generation sequencing) is the requirement of a bioinfomatician to handle the magnitude of data generated. At least a basic knowledge in computer science is necessary to start navigating the information encoded in the output files. MinION's raw data are provided in an incomprehensible binary language that first must be converted to human

readable files. Conversion to such files called FASTQ files is performed through ONT's proprietary Metrichor, which has been replaced with Nanonet, Albacore, Guppy, and Scrappie, which are also from ONT. These software programs identify DNA sequences directly from raw data and subsequently enhance accuracy. The output FASTQ file is a set of four lines of readable characters that contains a read identifier, the read sequences, a plus sign, and its quality. At this point, data analysis depends on the computer skill of the researcher. Realizing that not all researchers are adept in computer science, ONT has released their suite software to assist with analysis using a graphical user interface. This program is accessible through EPI2MEAgent software and is currently a paid service to the members of the MinION community. *What's in My Pot* (WIMP) is an example of one of these software programs that takes the FASTQ file as input and maps sequence fragments to a database to provide a set of possible answers to the question: "What organisms are likely be sequenced?" A more specific approach is to map the 16s rRNA amplicon sequences to the 16 s rRNA database through *16s* workflow to know the possible bacteria genus (sometimes species) in the sequenced sample. These software programs are helpful for the crude identification of an organism, which is one of the strongest advantages of applying MinION in the workflow. However, the downside is that an internet connection is required to execute these cloud-based software programs. A more experienced researcher in bioinformatics will typically use offline software in a Linux environment.

Acclimatization to a Linux environment represents the other half of the on-site sequencing pipeline, especially in the field where it is difficult to access the internet. Linux provides many open-source software programs for biologists. Using these software programs, endless possibilities for MinION analysis are available. A decent and powerful laptop computer is required for most of these programs. Mapping software programs (or mappers) are the Swiss army knife in on-site sequencing. These programs map the input FASTQ files to a set of reference genomes. Burrows-Wheeler Aligner (BWA) is probably the most well-known mapper as it has served as a staple in complementing Illumina sequencers since the dawn of the next-generation sequencing. It is designed to be compatible with short read sequencing; however, it currently has a MinION-mode to use with MinION's long reads and somewhat noisy sequencing. LAST is also another mapper that is comparable to BWA in terms of processing speed. Minimap is a mapper with superior speed. The second iteration (Minimap2) is threefold and tenfold faster than BWA-MEM for mapping >100 bp short reads and >10 kb long reads, respectively (Li 2018).

Although organism identification is sufficiently achieved using a mapper alone, different tasks require different software tools. For example, variant detection in real-time Ebola surveillance in Africa used Nanopolish. The software directly detects variants using the event-level ('squiggle') data generated by the MinION to evaluate candidate variants found in the aligned reads. However, this technique is not compatible with non-standard genomes (e.g., *P. falciparum*) given that Nanopolish tends to yield false positive results. Noisy MinION sequencing can gain benefits from consensus sequence generation. When we employed this technique to

detect the SNPs in genes conferring drug resistance in *P. falciparum*, it improved accuracy in the obsolete flow cell R7 and the most recent flow cell R9 to 73.46% and 84.56%, respectively (Runtuwene et al. 2018). Nevertheless, these processes were not straightforward and required knowledge in programming language.

In summary, on-site sequencing has propelled the advance of genome research. In this field, MinION has greatly reduced the equipment investment cost, simplified library preparation, and improved the accessibility to biodiversity. Although numerous limitations prevent the complete adoption of MinION as a true portable device, accessories currently being developed for MinION. They will simplify everything from DNA extraction to laptop computers and even the sequencer itself, causing these devices more portable and cheaper in the foreseeable future.

References

Boykin LM, Ghalab A, Rossitto De Marchi B, Savill A, Wainaina JM, Kinene T, et al. Real time portable genome sequencing for global food security. bioRxiv. 2018:1–10.

Brinkmann A, Ergu K, Radoni A, Tufan ZK, Domingo C, Nitsche A. Development and preliminary evaluation of a multiplexed amplification and next generation sequencing method for viral hemorrhagic fever diagnostics. PLoS Negl Trop Dis. 2017;11(11):1–13.

Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins H, Mcintyre ABR, et al. Nanopore DNA sequencing and genome assembly on the international space station. Sci Rep. 2017;7(18022):1–12. https://doi.org/10.1038/s41598-017-18364-0.

Faria NR, Sabino EC, Nunes MRT, Carlos L, Alcantara J, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. Genome Med. 2016;8(97):1–4. https://doi.org/10.1186/s13073-016-0356-2.

Faria N, Kraemer N, Hill S, Goes de Jesus J, de Aguiar R, Iani F, et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. bioRxiv. 2018:1–35.

Goordial J, Altshuler I, Hindson K, Chan-yam K. In situ field sequencing and life detection in remote (79°26'N) Canadian high Arctic permafrost ice wedge microbial communities. Front Microbiol. 2017;8(2594):1–14. https://doi.org/10.3389/fmicb.2017.02594.

Hardegen J, Latorre-Perez A, Vilanova C, Thomas G, Simeonov C, Porcar M. Liquid co-substrates repower sewage microbiomes. bioRxiv. 2018.

Hayashida K, Kajino K, Hachaambwa L, Namangala B, Sugimoto C. Direct blood dry LAMP: a rapid, stable, and easy diagnostic tool for human African trypanosomiasis. PLoS Negl Trop Dis. 2015;9(3):e0003578. https://doi.org/10.1371/journal.pntd.0003578.

Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol. 2017;15(3):183–92. https://doi.org/10.1038/nrmicro.2016.182.

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:1–7. https://doi.org/10.1093/bioinformatics/bty191.

Lu H, Giordano F, Ning Z. Oxford Nanopore MinION sequencing and genome assembly. Genomics, Proteomics Bioinformatics. 2016;14(5):265–79. https://doi.org/10.1016/j.gpb.2016.05.004.

Menegon M, Cantaloni C, Rodriguez-prieto A, Centomo C, Abdelfattah A, Rossato M, et al. On site DNA barcoding by nanopore sequencing. PLoS One. 2017;12(10):1–18.

Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, Hase T. Loop-mediated isothermal amplification of DNA. Nucleic Acids Res. 2000;28(12):E63. https://doi.org/10.1093/nar/28.12.e63.

Oxford Nanopore Technologies. New basecaller now performs "raw basecalling", for improved sequencing accuracy. 2017. Retrieved from https://nanoporetech.com/about-us/news/new-basecaller-now-performs-raw-basecalling-improved-sequencing-accuracy

Pomerantz A, Arteaga A, Bustamante L, Pichardo F, Coloma LA, Salazar-valenzuela D, et al. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. GigaScience. 2018;7:1–14. https://doi.org/10.1093/gigascience/giy033.

Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228–32. https://doi.org/10.1038/nature16996.

Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12(6):1261–76. https://doi.org/10.1038/nprot.2017.066.

Runtuwene LR, Tuda JSB, Mongan AE, Makalowski W, Frith MC, Imwong M, et al. Nanopore sequencing of drug- resistance-associated genes in malaria parasites, Plasmodium falciparum. Sci Rep. 2018;8(8286):1–13. https://doi.org/10.1038/s41598-018-26334-3.

Yamagishi J, Runtu LR, Hayashida K, Mongan AE, Thi AN, Thuy LN, et al. Serotyping dengue virus with isothermal amplification and a portable sequencer, (January), 1–10. 2017. https://doi.org/10.1038/s41598-017-03734-5.