A Study of DNA and RNA sequencing using the Oxford Nanopore MinION

Nathan E. Houser '21 and C. William Anderson

Department of Chemistry, Hampden-Sydney College, Hampden-Sydney, VA 23943

Introduction

Deoxyribonucleic acid or DNA is a fundamental component of an organism. These molecules are double stranded with each strand being composed of a backbone and interacting nucleobases. Adenine (A), cytosine (C), guanine (G), and Thymine (T) are the four main bases that form specific sequences. These sequences then dictate vital cellular functions and hold vital information to the organism that houses the DNA. The tasks and information that DNA holds makes it vital not only to the functions of an organism but also to understanding that organism [2]. Sequencing the individual bases in DNA allows better understanding of an organism's function. Therefore, methods such as Sanger and illumine sequencing have been developed.

In the 1970s the Sanger sequencing method was developed as an early technique to study DNA [1]. This method begins with spliced fragments of DNA [8]. Each of these fragments is labeled with dideoxynucleotides [8]. Gel electrophoresis then separates shorter and longer fragments [8]. Once separation is complete a laser excites the dideoxynucleotides to create a chromatogram that is in order due to the Gel electrophoresis [8]. In its time this method was commonly employed because it could read around 500 base pairs at the cost of \$3000 [2]. It, however, struggled to read longer sequences such as human chromosome 1 which has 2.47x10⁸ base pairs.

To make longer reads the more modern Illumina sequencing method was developed. Illumina uses purified DNA that has also been fragmented and modified to allow for proper sequencing [3]. Fragments of DNA are then anchored to a chip where modified nucleotides with fluorescent tags are attached [3]. Modified nucleotides attach one at a time and after a round of synthesis a fluorescent camera takes a picture [3]. Using these pictures bases within the fragments of DNA are indicated by the wavelength of the fluorescent tags [3]. This method allows a read length of approximately 100,000,000 base pairs and only required the cost of solutions to run. This technique does carry a large cost to acquire the instrument, which is difficult for smaller laboratories to pay.

The Oxford Nanopore MinION sets its sights on longer and more efficient base pair sequences along with a lower cost. It utilizes a voltage buildup across a membrane with embedded nanopores to determine the sequence of a segment of DNA [1]. Nanopore research begins with voltage patch clamps that require the buildup of voltage to open a pore [1]. Nanopores are small imbedded openings that allow movement into or out of a cell [1]. The MinION instrument utilizes a modified Alpha Hemolysin pore which meets the requirements for DNA movement [1]. During sequencing current passage is measured while bases move through the nanopore [1]. Each base is distinguished by individual size and the length of time it stays in the pore; both factors affect the current that runs through the membrane [1]. This technology has been a promise since the 90s but was not fully released till 2014 [6]. This makes most of the procedures and methods still novel and subject to changes to further efficiency [6].

Despite the novelty of this technique the method promises to advance research in multiple fields. Recent studies into different influenza strains have showed that it was already capable of rapid identification of the Influenza virus [9]. This study also outlined the simple amount and type of equipment necessary for the MinION and conceivable use in the field [9]. Similarly, a study conducted by Dr. Andy Kilianski and others showed that despite a 30% error rate the MinION was able to distinguish samples of E. Coli [4]. The main cause of this high error was flow chemistry which has since been updated [4]. This new technique is also capable of enhancing contiguity when used alongside other sequencing procedures, and as the instrument evolves it becomes a better lone sequencing method for DNA and RNA [5]. Overall the MinION successfully runs around 2,000,000 base pairs and the cost heavy illumina is five hundred times more expensive. On top of these benefits the MinION only weighs 100g and can plug directly into a laptop or computer [7]. These features allow this instrument to be used outside a usual lab setting with ease.

My research is meant to further incorporate the MinION into casual lab use here at Hampden-Sydney college. This began with finding an efficient procedure to sequence DNA. Once a method that could be used was acquired *E. Coli* was tested to ensure the methods success. Hopefully this research allows for better study and understanding of DNA.

Materials and Methods

The MinION is a new instrument with some of the newest sequencing technology. Before setting

up the sequencer I first looked into some of the background on the MinION, Nanopore technology, and how to work a flow cell. I did this using the Oxford Nanopore website which holds the pertinent information regarding the MinION and YouTube which has videos of presentations regarding the MinION and hoe to handle it. It is worth noting that during this time I also set up an account with Oxford Nanopore. This gave me access to the community forums which are a big part in setting up procedures and installing the necessary software.

After looking into some background of nanopore technology a laptop was set up to handle MinION sequencing. This setup began with changing certain settings that would interfere with the long times required to sequence. First sleep settings were completely turned off to ensure that the laptop would not shut down after a certain amount of time. The laptop was also continuously plugged in and settings that put the computer into rest mode while charging were turned off. Internet also needed to be continuously supplied to the laptop during sequencing. All of this was done before downloading the proper software.

Two specific desktop agents had to be installed in order to allow the laptop to sequence with the MinION. Both agents as well as a step by step manual for sequencing Lambda DNA can be acquired on nanoporetech.com under the community tab in the getting started section. The first application is the MinKNOW desktop agent. This application directly controls the MinION and performs sequencing runs and basecalls data. When imputing the parameters for a sequencing run a folder needs to be prepared and chosen for MinKNOW to send data to. The EPI2ME desktop agent further aids to analyze basecalled data from the MinION. After a Lambda DNA control experiment the EPI2ME agent works on the alignment of the basecalled data. This software is easily downloaded and the same information for the nanoporetch website is used in the EPI2ME agent.

The MinION sequencer is the first piece of hardware to be set up on the computer before a sequencing run. This device was stored in a desk drawer at room temperature. The box that Oxford Nanopore shipped this sequencer in was used to keep this instrument, its plug, and a test flow cell. This device is easily plugged into the computer and, with the test flow cell inserted, a hardware check can be initiated. This check will ensure that the electronics within the MnION are working correctly. This device is then placed through a flow cell check to identify pore count and sequencing once the proper solutions are added to the flow cell. After use this item was simply placed back in its box and into a drawer.

Flow cells house the nanopore membrane and circuitry required to convert the current into an electrical signal necessary for experimental results. These essential pieces of equipment were kept in a freezer with the storage buffer S in order to preserve the nanopore membrane. When in use flow cells were taken from storage and placed in a sealed container to allow them to come to room temperature without excess condensation. Then after taken out of its container the pore count was checked to confirm at least 800 pores were active. All solutions were carefully added into the priming port in order to prevent air bubbles from entering the sensory array and disturbing the nanopores, and likewise solutions were added dropwise into the sample port to also avoid air bubbles. After all solutions were added and sequencing complete the flow cell was washed with 150 microliters of solution A for 10 minutes then either 500 microliters of solution B was added to begin another sequencing run or 500 microliters of solution S was added to store the flow cell.

Solutions from the flow cell priming kit and Rapid Sequencing kit were required to run a control lambda DNA test and to sequence a purified strand of E. Coli. Both kits were purchased form nanopore tech.com and included Fragmentation mix. Rapid Adapter, Sequencing Buffer, Loading Beads, Flush Buffer, Flush Tether, and lambda DNA. All these solutions were stored in a freezer until ready to be used and were kept on ice during solution preparation. To prepare a library for a sequencing run 7.5 microliters od Lambda DNA and 2.5 microliters of fragmentation mix were added to a PCR tube and incubated in a thermocycler at 30 degrees C for 1 minute and 80 degrees C for another. After this 1 microliter of rapid adapter was added and the library was prepared. To create priming mix 30 microliters of flush tether was added directly into a flush buffer tube. A presequencing mix was also prepared by transferring 34 microliters of sequencing buffer, 4.5 microliters of nuclease free water, 25.5 microliters of loading beads, and 11 microliters of Lambda DNA into an Eppendorf tube and mixing. All waste solutions were properly disposed of and pipette tips used during these processes were placed into a waste receptacle. A sample of E. Coli was obtained by growing a liquid bacterial culture in 15mL conical vials with 5 mL of LB broth, incubated at 37 degrees C and allowed to culture for 24 hours. DNA was then extracted and purified using Qiagen ultra clean microbial kit and the protocol within that kit.

Results

The first goal of this project was to obtain and refine a procedure for sequencing with this technology. The novelty of this instrument means that the methodology is continuously changing. I planned to begin my research by both becoming more familiar with the instrument and lay out an efficient procedure for successive researchers.

I was easily able to begin this process through the community tab on nanoporetech.com. The necessary software needed to be downloaded and the hardware checked before beginning experimentation. Once these were accomplished, I began looking into a control Lambda DNA experiment. The procedure for this was included in the same tab on nanoporetech.com as software download. When I began to run this experiment I at first noticed an absence of working pores required for sequencing. I worked to solve this issue by washing the flow cells with storage and sequencing buffers and by trying to remove air bubbles. A lack of results with these methods led me to attempt a lambda DNA experiment with no available pores. I was able to begin this experiment but files relating to this data



Figure 1: Channels panel for active nanopores available in real time during sequencing of E. Coli DNA.

and the displays on MinKNOW showing flow cell activity were vacant. After some time, another flow cell was acquired with pores. Another control



Figure 2: Duty time of pore channel status over the course of *E*. Coli sequencing

experiment did yield results in the form of a data summary and folders with relevant pore information. Because this data was present and in order to further test the procedure that I acquired I began a sequencing experiment with a test sample of *E. Coli*



Figure 3: Read length Histogram of bases called during E. Coli sequencing.

DNA. This had a low pore availability but did show initial results in the form of flow cell activity shown in figures 1,2, and 3.

These pieces of information confirmed that during the hours of sequencing data collection on DNA had taken place.

Raw data on individual bases was then collected via the designated file on the laptop

Multiple data files were created as a result of sequencing including a sequence summary that gave

an overview of the results. The Fast5_pass file held multiple sequences of bases that MinKNOW had identified as shown in figure 4.

11 file_version 0 0 1 11 asic_temp 0 0 34.007441 • 0 11 asic_wrsion 0 0 1 11 flow_celid 0 0 1 11 flow_celid 0 1 0 33.005560 • 0 0 81.00455 10.0 • 0 11 sample_id 0 0 1	1.0 8 1.402C 0 [0] [0] [0] [0] [0] [0] [0] [0] [0] [update I I 1 e-type I 0 minion Tguppy_version I 0 I I + 0 E-Col1_7_11 unfig I 1 I minon	00 U (1 A [[[]] auto_update_sour + 0 000] distribution_tus - 0.3707D706 + 8 [][[]] [[[]] installation_type + 0 000 protocol_run_id fus_1.1.1_currentinton_fpga_1.1.0	(1111 asic_1d 1 1 (1111 asic_1d 1 1 (1 0 1 1) tasic_1d 1 1) tasic_1d 1) 1 1 nc + 0 1 X 0 4264686-582-4c3 wctrlewuto + (1111	0 • 0 1111 as hirror.confordnanoportal.com/soft 8 1111 distribution_versic 1111 local_firmare_file n=27fc-f732ee2243e3 • 8 1 version 1 0 5.5.2 0
1 0 11 L4 11 000 cd 11 pore_type 0 0 1 11 digitisation 071 0.4 channel_id#(00 11 11 0.8 1 0.1 start_max 0 0 1 11 01 11 11 11 11 11 11 11 11 11 11 11 1	1 11 14 not_set 111 1 11	P + P 000 run_id + 0 TREE 9999999 000 offset 0 71 44 49 ext_tagsN 000 00 1 vead_id 0 X 0 26678922 38 X TREE 9999999) af2be888706ff8 99999999 Af2be888706ff8 0 1 Af2be88706ff8 1 Af2be88706ff8 Af2be88706ff8 0 1 Af2be88706ff8 0 1 Af2be88706ff8 0 1 Af2be88706ff8 0 1 Af2be8706ff8 0 <	2ae6a7640185b16645cc23020 range ? (β.4) 49 999999999999 •β median_before	•0 1 1 •0-8•8 111 121 8-41491 0
1 0 Illingumstation 019-07-11191/43/062 0 0 10 01 0 0 Illingumstry A 1 0 Illingumstry A	ian_200 6 [11]	1 % 0 TATE 0 1 % Million Million 0 1 1 % Million Million 0 1 1 % Million Million Million 0 1 1 % Million Million<	3777777777777777777777777777777777777	II version I 0 1 3-3.2 sector II 1 1 1 1 1 sector II 1 0 1 4 sector II 1 0 1 1 sector II 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	

Figure 4: data of *E*. coli sequencing in file FAH60112_4f2be808706ff82ae4a7449185b16c445cc23029_0 under fast5-pass with sequenced bases highlighted in blue.

This file was analyzed using BLAST to determine the sequenced DNA. Using this method, it was confirmed with 94% accuracy and a 0.0 EE value that the sequenced DNA was from a strand of *E. Coli* as shown in figure 5. These results indicated that I had

										_				
Dea	criptions	Graphic Summary	Alignments	Taxonomy										
Se	Sequences producing significant alignments Download 🗸 Manage Columns 👋 Show 100 🗸 🕴													
Select all 100 sequences selected GenBank Graphics Distance tree of											ree of results			
			D	escription		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession			
	Escherichia d	oli strain BH212 chromosome.	complete genome			1840	3719	94%	0.0	85.94%	CP029239.1			
~	Enterobacter	Enterobacteria phage 0276, complete genome					1840	94%	0.0	85.94%	MH547045.1			
~	Escherichia d	Escherichia coli K-12 strain C3028, complete genome					5546	94%	0.0	85.94%	CP014272.1			
•	Escherichia c	Escherichia coli B strain C3029, complete genome					3705	94%	0.0	85.94%	CP014269.1			
~	Enterobacter	Enterobacteria phage HK630, complete genome					1840	94%	0.0	85.94%	JQ086376.1			
~	Enterobacter	Enterobacteria phage lambda, complete genome					1840	94%	0.0	85.94%	J02459.1			
~	Cloning vector TLE97-3, phage lambda lacZ translational fusion vector, complete sequence					1834	1834	94%	0.0	85.88%	U39286.1			
~	Cloning vector TLF97-2, phage lambda lacZ translational fusion vector, complete sequence					1834	1834	94%	0.0	85.88%	U39285.1			
•	Cloning vecto	r TLF97-1, lambda phage lacZ	translational fusion ver	tor, complete sequence		1834	1834	94%	0.0	85.88%	U39284.1			
~	Cloning vecto	r lambda TXF97, lacZ transcrip	tional fusion vector, co	mplete sequence		1834	1834	94%	0.0	85.88%	U37692.1			
~	Enterobacter	a phage HK629, complete gene	ame			1818	1818	94%	0.0	85.72%	JQ182735.1			
~	Cloning vector lambdaS2775, complete sequence					1483	1483	73%	0.0	86.61%	EU421722.1			
	Synthetic cor	struct clone HO-HIS phytochels	atin synthetase (SPAC:	3H1.10) and histidinol-ph	sphate aminotransferase imidazole acetol phosph	1482	1482	73%	0.0	86.60%	JE340119.2			

Figure 5: BLAST of file

FÄH60112_4f2be808706ff82ae4a7449185b16c445cc23029_0 shown above under the ECOLI file designated for E. Coli sequencing data.

obtained an effective procedure to sequence with the MinION

Conclusions

The MinION is a capable and impressive instrument. The crux of my research was familiarizing myself with its methods in order to incorporate it into more casual use. My results show that initially making this kind of technology cooperate can be a struggle. I attribute this to the novelty of the technology itself and the methods used to operate it. Overtime, however, this is a promising endeavor. Putting some work into producing a procedure will make the MinION an effective sequencing option at Hampden-Sydney College. I would advise moving forward with similar research with the MinION to make it more viable for casual lab use. I attribute most of my struggle to the breaking down of storage solution over time and flow cell issues.

Moving forward the next research that I would recommend would be RNA sequencing using the MinION. If this were able to be done here, then it would give the MinION a broader range of applications.

Acknowledgements

I would like to thank The Office of Undergraduate Research, first and foremost, for giving me the opportunity to work on such a great project with so many amazing people. I would also like to thank Ms. Hines and Mrs. Jenkins for helping me with any materials I needed, Dr. Anderson for his guidance throughout this project, all the chemistry professors who helped me and my peers, especially Christopher Parish for his *E. Coli* sample, whose help was instrumental throughout this project.

REFERENCES

• [1] Goodwin, Sara, director. Advanced Sequencing Technologies 2015 – *Introduction to Nanopore Sequencing (Part 1) - Sara Goodwin. YouTube*, 16 May 2018, youtu.be/Gy1cSIUjr1M.

- [2] Griffiths, Anthony J.F. "DNA Sequencing." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., www.britannica.com/science/DNA-sequencing.
- [3] "Illumina Dye Sequencing." *Wikipedia*, Wikimedia Foundation, 10 May 2019, en.wikipedia.org/wiki/Illumina_dye_sequenci ng.
- [4] Kilianski, Andy, et al. "Bacterial and Viral Identification and Differentiation by Amplicon Sequencing on the MinION Nanopore Sequencer." *GigaScience*, vol. 4, no. 1, 2015, doi:10.1186/s13742-015-0051-z.
- [5] Laver, T., et al. "Assessing the Performance of the Oxford Nanopore Technologies MinION." *Biomolecular Detection and Quantification*, vol. 3, 2015, pp. 1–8., doi:10.1016/j.bdq.2015.02.001.
- [6] Magi, Alberto, et al. "Nanopore Sequencing Data Analysis: State of the Art, Applications and Challenges." *OUP Academic*, Oxford University Press, 16 June 2017, academic.oup.com/bib/articleabstract/19/6/1256/3869205.
- [7] "MinION." Oxford Nanopore Technologies, 26 Feb. 2019, nanoporetech.com/products/minion.
- [8] "Sanger Sequencing." *Sigma*, 2019, www.sigmaaldrich.com/technicaldocuments/ articles/biology/sanger-sequencing.html.
- [9] Wang, Jing, et al. "MinION Nanopore Sequencing of an Influenza Genome." *Frontiers in Microbiology*, vol. 6, 2015, doi:10.3389/fmicb.2015.00766.