Spring 2023

Evaluation of an Adaptive Sampling Approach to Characterize Microbes Associated with Pneumonia in White-Tailed Deer

Claire Nowak

Follow this and additional works at: https://digitalcommons.library.umaine.edu/honors

Part of the Bioinformatics Commons, and the Other Genetics and Genomics Commons

This Honors Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Honors College by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.
EVALUATION OF AN ADAPTIVE SAMPLING APPROACH TO CHARACTERIZE
MICROBES ASSOCIATED WITH PNEUMONIA IN WHITE-TAILED DEER

by
Claire Nowak

A Thesis Submitted in Partial Fulfillment
of the Requirements for a Degree with Honors
(Molecular and Cellular Biology)

The Honors College

University of Maine

May 2023

Advisory Committee:
Pauline Kamath, Ph.D., Assistant Professor of Animal Health, Advisor
Benjamin King, Ph.D., Assistant Professor of Bioinformatics
Mark Brewer, Ph.D., Professor of Political Science, Preceptor in the Honors College
Melanie Prentice, Postdoctoral Research Associate, School of Food and Agriculture
ABSTRACT

Infectious diseases have a tremendous global impact, adversely affecting the health and well-being of humans, domestic livestock, and wildlife. Consequently, pathogen surveillance in wild animals is essential for managing the risk of disease transmission to humans and domesticated animals, as well as for understanding host-pathogen interactions. However, pathogen detection methods are often focused on one to a few pathogen species, which limits our understanding of the distributions and effects of multiple co-infecting pathogens on host individuals and populations. In this study, I employed a metagenomic sequencing approach to (1) characterize the microbial community in a white-tailed deer (*Odocoileus virginianus*) that had succumbed to pneumonia, and (2) evaluate an adaptive sampling sequencing approach that more efficiently targets microbial sequences by excluding the host genome during sequencing. I first performed metagenomic sequencing on DNA extracted from deer lung tissue and then sequenced the same sample using the adaptive sampling approach on a MinION platform (Oxford Nanopore). The results from the metagenomic sequencing were compared to that from the adaptive sampling approach. While the expectation was to see an increase in microbes identified using the adaptive sampling approach, I found a similar list of microbial species using both the adaptive and standard approaches. However, we also found that the adaptive sequencing approach was more efficient in identifying a similar set of microbes in a shorter time frame and could be more cost-effective. This study validated the use of the adaptive sampling method as a more rapid tool for sequencing pathogens while also showing limitations in the ability to identify a broader range of microbes.
ACKNOWLEDGMENTS

There are multiple people who helped me cross the finish line in completing my thesis. First, I would like to thank my advisor, Dr. Pauline Kamath, who helped me refocus and reframe my project to accomplish my goal of graduating with honors. I would like to thank my wonderful and supportive family and friends who continued to cheer me on and listen to/coach me throughout the process. Lastly, I would like to thank my ASP staff (Pete, Lauren, and Rose) and fellow Center Directors (Stephen, Kam, and Alexis) who helped me catch up on work while I worked long hours into the night during training and set up to get the document to its final stages.

I would like to acknowledge US Fish and Wildlife Service Pittman-Robertson Funds through the Wisconsin Department of Natural Resources, Daniel Storm with the Wisconsin DNR, Melanie Prentice for help with the lab work as well as the other co-authors on the preliminary deer study.
# TABLE OF CONTENTS

INTRODUCTION ................................................................. 1
Sample collection ............................................................... 5
DNA extraction, quantification, and purification ......................... 5
Metagenomic sequencing ..................................................... 5
Bioinformatics analyses ........................................................ 6
RESULTS ................................................................................. 9
MinION run quality: Standard vs. Adaptive runs ....................... 9
Dataset sequence quality and alignment ................................ 12
Pathogen identification .......................................................... 15
DISCUSSION ......................................................................... 18
FUTURE DIRECTIONS & CONCLUSIONS ................................. 22
LITERATURE CITED ............................................................. 23
AUTHOR’S BIOGRAPHY .......................................................... 25
LIST OF TABLES AND FIGURES

**Figure 1.** Comparison of the efficiency of the Standard (A,C) versus Adaptive (B, D) 10

**Figure 2.** Comparison of the quality of the Standard versus Adaptive sequencing runs 11

**Figure 3.** Comparison of the flow cell maps of the Standard versus Adaptive sequencing runs 12

**Figure 4.** Base quality across sequence position for the Standard (A), Adaptive (B), and Subsample (C) datasets. 14

**Figure 5.** Microbial species identified and their abundance across the Standard, Standard_subset and Adaptive datasets 16

**Table 1.** Output summary comparison of three sequence datasets generated from the Standard (19NX80, 19NX80_subsample) and Adaptive (19NX80a) MinION sequencing runs. 13

**Table 2.** Consistency of microbial species identified across datasets. 17
INTRODUCTION

Infectious diseases are an increasingly important issue in the world today, as more of the emerging diseases we see in humans today are zoonotic, and stem from wild animals (Jones et al. 2008). Using genomics for the surveillance of pathogen infections is becoming increasingly popular and feasible with the development of new genomic technologies; however, studying infectious diseases through a genomics lens has not been something done widely in wildlife (Blanchong et al. 2016). Genomics has been used for incredibly diverse applications, with diagnosing disease, monitoring disease across populations, and tailoring treatments through personalized medicine being prime examples (Clark and Pazdernik 2013; Ghosh et al. 2018). Applying genomics to studying disease can assist in understanding how host parasites interact with other pathogens. This project aims to focus on how genomics can be used for the identification and surveillance of potential pathogenic agents in wildlife hosts. We specifically applied a metagenomic approach using an Oxford Nanopore MinION device, which involved the real-time exclusion of the host genome to enable surveillance of whole microbial communities (including potential pathogens).

Pneumonia, a respiratory disease, poses significant challenges for the white-tailed deer population, which holds economic and ecological importance (Torres et al. 2021). White-tailed deer contribute to the economy through sectors like hunting and wildlife tourism. Disease not only impacts the deer population but also disrupts the ecological balance and poses economic consequences for hunting-related industries and tourism. Wildlife management agencies and researchers in Wisconsin are actively working on monitoring and understanding the causes of pneumonia-related mortality, implementing
measures to reduce deer densities, and promoting sustainable practices to preserve the white-tailed deer population and its contributions to the economy and ecological well-being (WDNR). However, as pneumonia-related mortality in deer has become a pressing issue in Wisconsin, the WDNR was unsure what was causing the outbreak and deaths in deer that could not be linked to Chronic Wasting Disease (CWD) (Gilbertson et al. 2022). Following this, further study was done that found several bacterial species that have been associated with pneumonia and/or other diseases of ruminants, including *Trueperella pyogenes, Pasteurella multocida, and Fusobacterium necrophorum* (Prentice et al, in preparation).

Understanding the environment for a host organism can be incredibly important as many of the parasites and pathogens that impact them can also have implications for the health of humans and domestic animals (Han et al. 2016). While it is important to understand that factors such as overpopulation, high host density, and habitat fragmentation exacerbate disease transmission, so can the host organism itself (Haydon et al. 2002). Due to this, the application of genomics to wildlife can help us have a better understanding of zoonotic diseases. While the study described here applied a genomics approach to wildlife disease, it more specifically identified and characterized the microbial community within a host organism, in this case, a single white-tailed deer whose death was suspected to be from pneumonia. This microbial data associated with a pneumonia-related fatality will be valuable for further examination of disease etiology as well as relationships between hosts and microbial communities.

Microbial communities, consisting of diverse microorganisms, play crucial roles in nutrient cycling and decomposition, interacting with each other and their environment.
Understanding these interactions is vital in ecology, microbiology, and medicine, especially regarding host disease susceptibility. Identifying specific pathogens is essential for accurate diagnosis, treatment, and prevention, enabling a deeper understanding of disease dynamics and the development of effective strategies. In complex diseases like pneumonia that involve multiple pathogens, further research is needed to comprehend the effects of whole microbial communities and co-infection on disease outcomes. Co-infection occurs when two or more pathogens interact within the same organism and with the host, influencing transmission and increasing host vulnerability to infections. Studies have shown that co-infecting pathogens can impact the transmission of additional pathogens (Devi et al. 2021, Viney and Graham 2013). The current study focuses on characterizing the microbial communities associated with disease in deer, utilizing metagenomic sequencing approaches. The accessibility and applicability of metagenomic sequencing have increased due to technological advancements and cost reduction, underscoring the importance of refining these approaches to study diverse microbial communities associated with polymicrobial diseases. This method provides a comprehensive understanding of the entire microbial community present, unlike previous approaches that focused on individual microbes.

Additionally, the adaptive sampling technique allows for real-time sequencing of the desired target, enhancing the efficiency of metagenomic analysis (Martin et al. 2022). This is exclusive to MinION sequencing platform, developed by Oxford Nanopore Technologies, which is a portable, real-time DNA sequencing technology. The adaptive method uses a specific program to access genomic libraries to which sequence data is compared to and filtered based on parameters set by the user. It utilizes a protein
nanopore within the MinION device to detect and analyze individual DNA molecules as they pass through (Lu et al. 2016). The portability of MinION allows for its use in various settings, including remote locations and clinical environments. It provides real-time sequencing data, enabling immediate analysis. Moreover, MinION sequencing has the capacity to sequence long DNA fragments, making it valuable for studying complex genetic structures.

I applied bioinformatics to analyze metagenomic sequence data from a white-tailed deer to gain insights into the diversity of microbial communities associated with pneumonia-related mortality, and to evaluate the potential for a more targeted sequencing approach that could be applied to wildlife disease surveillance. The adaptive sampling approach was employed to selectively target non-host sequences, which was expected to optimize the analysis and reduce the number of deer sequences generated. Overall, I aimed to use a metagenomic sequencing approach to (1) characterize the microbial community in a white-tailed deer (*Odocoileus virginianus*) that had succumbed to pneumonia, and (2) evaluate an adaptive sampling sequencing approach that more efficiently targets microbial sequences by excluding the host genome during sequencing. I first performed metagenomic sequencing on DNA extracted from deer lung tissue and then sequenced the same sample using the adaptive sampling approach on a MinION platform (Oxford Nanopore). I hoped to find a method of pathogen surveillance that would lend insight into host susceptibility that would prove to be more efficient and effective in identifying infecting microbes.
MATERIALS AND METHODS

Sample collection

White-tailed deer were collared and monitored in south-central Wisconsin by the Wisconsin Department of Natural Resources (WDNR)(Gilbertson et al. 2022). These deer were being monitored for several diseases. For this study, I specifically used a single sample (19NX80) from a male deer aged 2-4 years, with the body in good condition that was determined to have died of an infection from severe bronchopneumonia (Prentice et al, in preparation).

DNA extraction, quantification, and purification

After the deer was deceased, whole genomic data was extracted from the deer’s lung tissue with a spin-column-based method using the Qiagen DNeasy PowerLyzer PowerSoil Kit, following the manufacturer's protocol (Qiagen, Germantown, MD). A few modifications were applied to the protocol to lyse and homogenize the tissue using a Qiagen Powerlyzer (Qiagen, Germantown, MD) prior to extraction as described in Prentice et al. (in preparation) more effectively. After tissue lysis, the manufacturer's instructions were followed. The DNA was then quantified using a Qubit™ 4 Fluorometer (Invitrogen, Waltham, MA). The quality was then assessed with gel electrophoresis using a standard 100bp ladder (New England BioLabs, Ipswich, MA).

Metagenomic sequencing

I utilized the MinION long-read sequencing platform developed by Oxford Nanopore Technologies (Oxford, UK) for metagenomic sequencing (Jain et al., 2016). Two genomic libraries were prepared following the SQK-LSK 109 ligation sequencing kit and sequenced on the MinION platform: one for a “Standard” sequencing run and the
other for the “Adaptive” sequencing run. The 19NX80 deer sample was then sequenced two times (using either approach) an R9.4.1 flow cell. The two sequencing runs lasted for 72 hours each. The starting number of pores on the flow cells used for the two runs were comparable, with 1,029 and 1,105 pores available for sequencing at the start of the Standard and Adaptive runs, respectively.

In one of the runs, I applied an adaptive sampling approach, excluding host genome sequencing in real-time during sequencing by aligning sequence reads to a white-tailed deer reference genome database (Martin et al., 2022). During sequencing, the adaptive sampling algorithm analyzed the current signal in real-time and dynamically adjusted the flow rate of DNA molecules through the nanopore based on the signal-to-noise ratio of the current signal. If the signal was clear, the algorithm increased the flow rate of DNA molecules, resulting in faster sequencing throughput. If the signal was noisy or unclear, the algorithm slowed down the flow rate to improve the accuracy of the sequencing. By excluding sequence reads aligning to the deer genome, we aimed to target non-host (microbial) sequences and reduce the number of deer sequences generated.

**Bioinformatics analyses**

Three sequence datasets were compared within the bioinformatics analysis: the 19NX80 and 19NX80_subsample datasets were generated from the “Standard” MinION sequencing run, and the 19NX80a dataset was generated from the “Adaptive” MinION sequencing run. The subsample was used as it subset the sequencing data for the original sample to a more comparable number of reads generated across all other samples that were run in the previous study. All analyses were conducted using the high-performance computing cluster through the Ohio Supercomputer Center. The bioinformatics analyses
involved crucial steps such as data cleaning, sequence alignment, read depth analysis, and quality score assessment. These steps ensure accurate and reliable results by removing errors, assessing coverage, and identifying low-quality regions.

The bioinformatics approach included a four-step sequential process of (1) base-calling, (2) quality assessment, (3) alignment to the reference genome, and (4) pathogen identification. First, raw sequence data from the MinION output was base-called using Guppy v.6.0.0 (Oxford Nanopore Technologies, Oxford, UK). Second, the quality of the two sequencing runs (Standard, Adaptive) was assessed using MinIONQC (Lanfear et al. 2018) and the quality of the three sequence datasets (19NX80, 19NX80_subsample, 19NX80a) used were assessed using FastQC v0.11.7 (Andrews 2010), summarizing results across datasets using MultiQC v1.10.1 (Ewels et al. 2016). Quality control is a crucial step in bioinformatics analysis, involving data cleaning, alignment, read depth analysis, quality score analysis, and assessment of replicates and controls. The process ensures accurate and reliable results by identifying and removing errors, assessing coverage, and evaluating data reproducibility. Third, I aligned the sequence datasets to the host (deer) genome, and from this, the unmapped reads were filtered out (keeping anything greater than 200bp). Finally, unmapped reads were blasted against NCBI BLASTN database 2.13 (Basic Local Alignment Search Tool) to identify microbial sequences. High-quality reads were aligned to a reference genome or database of known pathogen sequences, and taxonomic classification algorithms assign them to a specific species or strain, potentially detecting genetic variants. The potential pathogen identification data were then compiled and a list of species with 100 or more hits was
compiled and any non-microbial (e.g., mammalian) species were filtered out. The remaining species were compared across the three datasets.

Finally, I compared the run and sequence quality between the Standard and Adaptive runs. To evaluate the use of adaptive sample sequencing on the MinION, I then compared the microbial species identified from both the standard and subsample datasets to the output of the adaptive run (standard vs. adaptive; subsample vs. adaptive). The sequence read counts in the subsample were similar, making this comparison control for variation in the quantity of read count data generated among runs.
RESULTS

MinION run quality: Standard vs. Adaptive runs

I observed that the Standard and Adaptive MinION sequencing runs differed in the time to maximal data yield. In particular, the adaptive run reached the maximum total yield in approximately half the time (~20-25 hours) of time required in the standard run (~40-45 hours) (Figure 1A, 1B). Simultaneously, the number of reads generated dropped to zero once the runs reached maximum yield, with the Standard run reaching zero reads after more time than Adaptive run (Figure 1C, 1D). This is also supported by a decrease in mean read length between roughly 12-48 hours into the adaptive sampling run, possibly due to there being minimal quality microbial data left to sequence (Figure 2B). In comparison, the standard run maintained a higher mean read length throughout the run (Figure 2A). The quality (mean Q score) of the adaptive versus standard runs was also different, with a more pronounced decrease in high-quality (Q>7) reads throughout the adaptive run, in comparison to a more consistent output of high-quality reads during the standard run (Figure 2C, 2D). Lastly, I found there were differences in the flow cell usage between runs, such that there appeared to be little to no active sequencing of reads in large sections of the flow cell used for the adaptive run (Figure 3), likely due to the use of an older flow cell.
Figure 1. *Comparison of the efficiency of the Standard (A, C) versus Adaptive (B, D) sequencing runs.* The graphs are arranged with the Standard on top and the Adaptive below. Plots A and B depict the yield produced per hour the sample was running on the flow cell. Plots C and D depict the number of reads produced per hour while the sample was running on the flow cell.
Figure 2. Comparison of the quality of the Standard versus Adaptive sequencing runs. The graphs are arranged with the Standard on top and the Adaptive below. Plots A and B depict the mean read length produced over the hours the sample was run on the flow cell. Plots C and D depict the quality of the sequences produced over hours of the run.
Figure 3. *Comparison of the flow cell maps of the Standard versus Adaptive sequencing runs.* The image shows pores on the flow cell that were utilized, including the quantity and the quality (Q) of sequenced material per pore, during the Standard (A) versus Adaptive (B) sequencing runs. The original flowcell started with roughly 1,029 pores, while the flow cell used on the adaptive run started with roughly 1,105 pores.

Dataset sequence quality and alignment

The quality of the sequence read data was fairly consistent across all three datasets (19NX80, 19NX80_subsample, 19NX80a) analyzed. The 19NX80 dataset had a significantly higher output of sequence reads (7,641,637 reads), in contrast to the 19NX80_subsample (1,700,000 reads) and 19NX80a (2,742,001 reads) datasets (Table 1). The 19NX80_subsample was smaller than the 19NX80 dataset as the predetermined reads were selected from the standard dataset. However, all three datasets were similar in the average quality scores of the bases called along the read positions of the sequences produced (Figure 4). This implies that the Adaptive method produces sequences that are still of reputable quality and can be used to analyze the microbial communities present in the host.
Table 1. Output summary comparison of three sequence datasets generated from the Standard (19NX80, 19NX80_subsample) and Adaptive (19NX80a) MinION sequencing runs. The runs were compared using FastQC output methods, with the Subsample included as it better matched the Adaptive run, in terms of the total number of read sequences. The outputs demonstrate the sequence quantity and quality for the three datasets examined and the proportion of identified reads.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Percent Duplicates (%)</th>
<th>Percent GC (%)</th>
<th>Mean Sequence Length (bp)</th>
<th>Total Read Sequences</th>
<th>Proportion of unmapped reads (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19NX80</td>
<td>3.26</td>
<td>43</td>
<td>1007</td>
<td>7641637</td>
<td>1.1</td>
</tr>
<tr>
<td>19NX80 Subsample</td>
<td>2.19</td>
<td>43</td>
<td>1006</td>
<td>1700000</td>
<td>1.1</td>
</tr>
<tr>
<td>19NX80a</td>
<td>1.60</td>
<td>42</td>
<td>782</td>
<td>2742001</td>
<td>.7</td>
</tr>
</tbody>
</table>
Figure 4. Base quality across sequence position for the Standard (A), Adaptive (B), and Subsample (C) datasets. The graphs depict the consistency of the quality scores across all three of the runs, suggesting comparable base quality scores across read positions.
Pathogen identification

Across the three datasets, there were 20 microbial species identified, including both bacteria and DNA phage viruses, with many species present in all three datasets (Table 2). When comparing the 19NX80 and the 19NX80_subsample dataset, I found that some species with lower frequency detections (<300 sequences) were no longer detected, most of species that were relevant to the study remained present in both datasets, with the exception of *Klebsiella quasipneumoniae*. This was similarly true in comparing the Adaptive dataset (19NX80a) to the Standard dataset (19NX80 vs. 19NX80a), likely due to differences in the quantity of read data. When comparing the Adaptive dataset results (19NX80a) to a subset of the Standard dataset (19NX80_subsample), which had more comparable read numbers, I found only 1 species was not identified in the Adaptive dataset, *Fusobacterium nucleatum* (Table 2). In summary, the Standard method was able to detect a greater number of microbes present with more blast results (Figure 4), whereas the subsample and the adaptive sampling did not yield as diverse results given a lower quantity of read count data. Nonetheless, the Adaptive sampling approach identified similar potential pathogens to the Standard approach when controlling for the number of reads generated by a run. Overall, the results of the three datasets were mostly consistent with regards to the species type across the three runs, particularly with regards to the top 3 species identified (*Escherichia coli, Fusobacterium necrophorum, Escherichia phage*).
**Table 2. Consistency of microbial species identified across datasets.** In all three of the datasets, there were 14 consistent microbial species with the most common across the board being *Escherichia coli.*

<table>
<thead>
<tr>
<th>Species</th>
<th>Standard</th>
<th>Adaptive</th>
<th>Subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>Fusobacterium necrophorum</em></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>Escherichia phage</em></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>Klebsiella quasipneumoniae</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clostridium botulinum</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Salmonella enterica</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Enterobacteria phage</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Paenibacillus sonchi</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Clostridium novyi</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Stenotrophomonas maltophilia</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Bacteriophage sp.</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Campylobacter phage</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Enterobacteriaceae bacterium</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Fusobacterium nucleatum</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clostridium septicum</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Fusobacterium gonidiaformans</em></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><em>Fusobacterium canifelinum</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chrysodeixis includens</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fusobacterium pseudoperiodonticum</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clostridium haemolyticum</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia albertii</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Streptococcus equi</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Streptococcus pyogenes</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clostridium phage</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5. Microbial species identified and their abundance across the Standard, Standard_subsample and Adaptive datasets. The top microbial species were identified using BLAST (NCBI) and after applying a filter to exclude mammalian host sequences. From the species count data, frequency histograms were created from pivot tables, with sorting applied from highest to lowest microbes present in the sample.
DISCUSSION

Metagenomic sequencing is a valuable approach for pathogen identification, enabling us to shed light on key microbes potentially linked to pneumonia-related mortality in deer, and that warrant further investigation. Moreover, our evaluation of adaptive sampling revealed its efficiency in producing comparable results with reduced resource utilization and faster turnaround time. Although adaptive sampling during sequencing did not outperform in detecting low-frequency microbial species, its effectiveness in achieving similar outcomes more efficiently holds promise for future investigations and resource optimization. The utilization of the adaptive sampling sequencing method for pathogen surveillance is something that has not been done before.

Metagenomic sequencing provides valuable insights into the microbial species present in a sample, but its effectiveness is influenced by the current sequence data available in the utilized online databases. Therefore, it is crucial to conduct further research such as comparisons to traditional diagnostic approaches like culturing as well as analyses to enhance the reliability and accuracy of the results. Additionally, the accuracy of the available databases for the host genome and the submitted sequences from other researchers play a significant role in linking the identified potential pathogens as causative agents of disease.

This study has demonstrated that the adaptive sequencing method can be used to create a similar species list and provide similar results if used as a microbial surveillance system, however it does not demonstrate significant improvement in this regard. This is seen through the overall top species identified being similar to other studies within the same system, such as *Escherichia coli* (Gilbertson et al. 2022). This study also
demonstrated that the metagenomic approach is useful in determining additional pathogens in system than just few that were noted in previous studies. By finding the twenty-five species in the sample we have a better understanding of the host and how it is becoming more susceptible to pneumonia.

Adaptive sampling offers a valuable approach to enriching target materials within a sample, proving especially advantageous when the target material is scarce or present in low abundance (Martin et al. 2022). This technique not only enhances the concentration of the desired material but can also be employed to simultaneously enrich multiple targets, thus maximizing the efficiency and cost-effectiveness of the process. However, in this study, I focused on the depletion or the removal of host sequence data from a sample. This seemed to present challenges as it may have resulted in the unintended loss of the target material or the inadvertent elimination of other pertinent components. This was a novel approach and initially, I expected this option would increase the efficiency and ability to detect a broader range of species present. The adaptive sampling was unable to deplete the host as well as I had predicted and overloaded the flow cell. This meant that it did not blast or read as many results, thus the depletion method may not be the best fit for adaptive sampling. The adaptive sampling method may be better suited for the enrichment method where certain pathogens are selected for. This can be attributed to the large size of the host genome and could be the reason why the proportion of mapped to unmapped reads remind similar across the results. So, while we did find it to be more efficient, the latter was not true, and there could be multiple reasons for this, due to the age of the flow cell, the pores on the flow cell being overwhelmed, and/or the large size of the host genome presenting a computational problem. Nevertheless, adaptive
sampling-based enrichment (or depletion) provides crucial advantages, including improved control over the concentration and purity of the target material. Moreover, this targeted strategy enhances the specificity and sensitivity of subsequent analyses, thereby facilitating more accurate and reliable downstream investigations. It also leads to a lower requirement for complex and time-consuming sample preparation methods, which proves particularly valuable when working with samples that possess limited amounts of starting material. Thus, adaptive sampling streamlines workflows and enhances the feasibility of various scientific investigations and applications through reaching yield capacity and read counts in a much shorter period of time.

Flow cells, an integral component of the MinION sequencing technology, can undergo degradation over time as a consequence of repeated use. This is something we perhaps experienced in this study, with the use of an older flow cell for the adaptive sequencing method, for which some of the pores appeared to be either degraded or impacted by human error (Figure 3). This degradation can have significant implications for the accuracy and quality of sequencing data. One consequence of degraded flow cells is the increased likelihood of errors in base calling, which subsequently affects the overall sequencing accuracy. Additionally, with the aging of flow cells, a decrease in pore occupancy can occur, meaning that fewer pores remain active during sequencing. Consequently, this reduction in pore occupancy leads to diminished sequencing throughput, resulting in lower coverage and potentially compromising the sensitivity to detect low-abundance microbial species. Another issue associated with older flow cells is the increased susceptibility of the pores to clogging or physical damage. These conditions can further compromise sequencing quality, leading to higher error rates and reduced
overall data reliability. On the other hand, human error, such as introducing an air bubble to the flow cell during loading could have similar results and cause a decrease in data produced. Moreover, the presence of long reads of the genome may overload the pores, exacerbating these issues. Therefore, it is important to consider the state and age of the flow cells in order to ensure optimal sequencing performance and data quality.

The microbial world is a vast and diverse realm, estimated to harbor up to one trillion distinct species of microorganisms, of which only a fraction have been identified and thoroughly characterized. However, our ability to study these microorganisms is often impeded by their inherent difficulty to be cultured in laboratory settings, thereby limiting the application of conventional microbiological techniques. To overcome this challenge, the identification and characterization of microorganisms frequently rely on genome or metagenome sequencing approaches. Yet, the complexity and diversity of microbial communities present significant challenges in accurately sequencing and analyzing their genetic material. Compounding the issue, the lack of comprehensive databases and reference genomes for many microorganisms further complicates their identification and taxonomic classification, particularly for those that are rare or previously unknown. Addressing these limitations is crucial for advancing our understanding of the microbial world and unraveling its numerous ecological, medical, and biotechnological implications.
FUTURE DIRECTIONS & CONCLUSIONS

The limited knowledge and understanding of the functions and interactions of many microorganisms poses challenges in interpreting the ecological and clinical significance of their presence or absence in various environments. This knowledge gap hinders our ability to grasp the broader implications of microbial communities. However, the field of microbiology is continuously advancing, with the emergence of new technologies and methods specifically designed to study the intricate microbial world. To address these challenges, ongoing research and exploration are essential to expand our understanding of microbial diversity, unravel their ecological roles, and elucidate their potential impacts on human health. Through persistent scientific investigation, we can uncover the hidden complexities of microbial interactions, paving the way for more comprehensive interpretations of microbial communities and their significance in diverse contexts.
LITERATURE CITED


AUTHOR’S BIOGRAPHY

Claire Nowak was born and raised in Geneva, Illinois, attending Geneva Community High School. While pursuing a Bachelor of Science molecular and cellular biology, Claire served as the Vice President of Recruitment and Operations for Delta Phi Epsilon, as well as Secretary, New Member Educator, and Ritualist. She also served as the Vice President of Recruitment for the Panhellenic Council, Programming and PR for Order of Omega, a Senator for the University of Maine Student Government, and Secretary of Operation H.E.A.R.T.S. She also received the Honors Thesis Fellowship funding from the Maine INBRE grant. She was nominated for Sorority Woman of the Year and The Andrew David Gerke Award and was a part of the Golden Key Honor Society. She worked as an Office Manager in the Office of Fraternity and Sorority Affairs throughout her senior year. Upon graduation, Claire plans to pursue a career in Women’s Public Health.