Applications and Challenges of Real-time Mobile DNA Analysis

Steven Y. Ko
Computer Science and Engineering
University at Buffalo
stevko@buffalo.edu

Lauren Sassoubre
Civil, Structural, and Environmental Engineering
University at Buffalo
lsassoub@buffalo.edu

Jaroslaw Zola
Computer Science and Engineering
Biomedical Informatics
University at Buffalo
jzola@buffalo.edu

ABSTRACT

The DNA sequencing is the process of identifying the exact order of nucleotides within a given DNA molecule. The new portable and relatively inexpensive DNA sequencers, such as Oxford Nanopore MinION, have the potential to move DNA sequencing outside of laboratory, leading to faster and more accessible DNA-based diagnostics. However, portable DNA sequencing and analysis are challenging for mobile systems, owing to high data throughput and computationally intensive processing performed in environments with unreliable connectivity and power.

In this paper, we provide an analysis of the challenges that mobile systems must address to maximize the potential of portable DNA sequencing, and *in situ* DNA analysis. We explain the DNA sequencing process and highlight the main differences between traditional and portable DNA sequencing in the context of the actual and envisioned applications. We look at the identified challenges from the perspective of both algorithms and systems design, showing the need for careful co-design.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing*;

KEYWORDS

Real-time mobile DNA analysis, mobile DNA sequencing, MinION

ACM Reference Format:
Steven Y. Ko, Lauren Sassoubre, and Jaroslaw Zola. 2018. Applications and Challenges of Real-time Mobile DNA Analysis. In *HotMobile '18: 19th International Workshop on Mobile Computing Systems & Applications, February 12–13, 2018, Tempe , AZ, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3177102.3177114

1 INTRODUCTION

DNA, a polymer made from four basic nucleotides (abbreviated by A, C, G, T), is the main carrier of genetic information. The DNA sequencing is the process in which this information is extracted by converting physical DNA molecules into signals that describe the exact order and type of the constituent nucleotides. The ability to sequence DNA has revolutionized molecular biology, biomedicine and life sciences in general. Among many applications, some we review in Section 2, it is recognized as a critical method for diagnosing and improving human health (e.g. dissecting genetic mechanisms of cancer [15]), identifying pathogens and protecting public health (e.g. detecting and tracking spread of infectious diseases [6]) or understanding our environment (e.g. impact of microorganisms on water, air and soil [14]).

The end-to-end DNA sequencing and analysis involves a combination of laboratory and bioinformatics steps (see Section 3). In a traditional setup, these steps are performed at massive scales by highly trained personnel using expensive benchtop DNA sequencers and supporting computational servers. Consequently, the process has been confined to high-end laboratories, limiting the access and extending the time it takes from sample collection to interpretable results.

The recently introduced portable DNA sequencers, specifically Oxford Nanopore Tech. (ONT) MinION [16] (see Fig. 1), are changing this situation. Compared to the traditional DNA sequencers, these devices are relatively inexpensive and truly mobile: smaller than a cell phone, USB powered, and designed to be easily operable "in the field." Moreover, they use biochemical principles (i.e. nanopore-based single molecule sequencing [11]) that enable near real-time streaming of raw signals as soon as DNA molecules are "sensed," usually within minutes from the process initiation. As a result, portable DNA sequencing emerges as a rapid *in situ* diagnostic tool, especially when DNA samples are difficult or impossible to preserve or transport. Examples include the DNA surveillance of Ebola [13] and Zika [5] during the recent outbreaks, deployments in the Arctic [3] and Antarctic [8], in rainforests of Ecuador [12] and even in the International Space Station [2].

Figure 1: Illumina MiniSeq (left), vs. portable MinION (middle) and forthcoming SmidgION (right). Pictures not to scale. The approximate size in centimeters is marked for reference. The current price of MinION is $1,000 compared to $50K for an entry-level benchtop sequencer.

HotMobile '18, February 12–13, 2018, Tempe , AZ, USA

ACM ISBN 978-1-4503-5630-5/18/02…$15.00
https://doi.org/10.1145/3177102.3177114

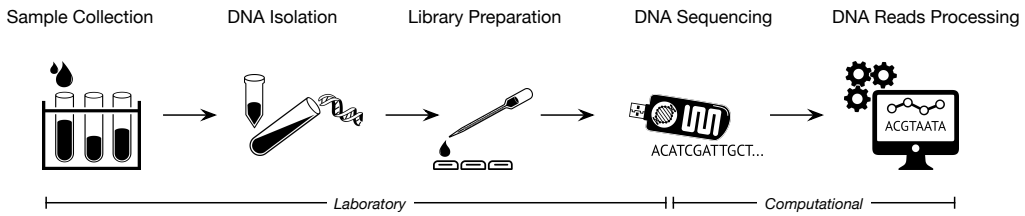

Figure 2: Major steps in a DNA sequencing workflow.

However, portable DNA sequencing and analysis are challenging for mobile systems. This is because the underlying computations, as well as data and communication intensive operations, have to be balanced to ensure the desired quality of the analysis while running in real-time, typically in the energy and bandwidth constrained environments. Currently, mobile DNA sequencing is driven by the existing bioinformatics tools designed primarily for desktop systems, and organized into mobile workflows in an *ad hoc* manner. While these solutions have been successful in the initial trials, they cannot be expected to scale if the underlying problems of processing speed, energy efficiency and resilience are not addressed. As the technology behind portable DNA sequencing matures and becomes more accessible, it is reasonable to assume that it will be adopted by individual consumers. Consequently, the underlying algorithms and software will have to operate in a wide range of conditions and workloads, and under varying resources, while remaining easy to use.

In this paper, we discuss the challenges that mobile systems must address to maximize the potential of portable DNA sequencing, and *in situ* DNA processing. We look at the identified challenges from the perspective of both algorithms and systems design, and argue for careful co-design and functionality separation. The paper is written from the systems perspective, for readers with no prior background in genomics or bioinformatics. Its slightly extended version is available from https://arxiv.org/abs/1711.07370.

2 SEQUENCING APPLICATIONS

We begin with a discussion of general applications that highlight the enabling nature of the portable DNA sequencing.

Sequencing When Time is Critical: Rapid diagnosis of infectious diseases is critical for protecting human health [6]. The DNA sequencing can be used to identify an infectious agent, assess its responsiveness to vaccines or antibiotics, and prescribe the best treatment. In some diseases, starting the right treatment within hours is vital [1, 7], and when responding to epidemics, real-time genomic surveillance increases situational awareness (e.g. by tracking evolution rate and transmission patterns), helping with planning and resource allocation [5, 13]. Importantly, the same principles apply in detection and mitigation of biological threats [18]. However, the current culture-based laboratory methods for identifying pathogens take days, and in fact some pathogens are difficult or impossible to grow in culture.

Because emerging portable DNA sequencers are free from these limitations, they offer a significant advantage when time is critical. One excellent example is the response to the recent Ebola epidemics, where a mobile laboratory using MinIONs, transported in a standard airplane bag, was deployed in Guinea [13]. Despite logistic difficulties, such as lack of continuous electric power and poor Internet connectivity, the laboratory became operational within two days, was generating results within 24 hours from receiving a sample, and provided valuable insights into disease dynamics.

Sequencing When Location is Critical: The samples we know the least about, and would like to study by DNA sequencing, are usually located far from established sequencing facilities. This is especially true for metagenomic studies in which communities of microbial organisms are sampled directly from their native environments [14]. However, many types of samples cannot be transported due to the legal (e.g. export barriers), or practical considerations (e.g. cost effectiveness). Furthermore, sequencing in laboratory is too constrained to study rapidly changing environments.

Portable DNA sequencers reached the level where they can be operated in some of the most demanding environments. For example, in one recent study, a battery-only powered laboratory, consisting of MinIONs and an *ad hoc* cluster of two laptops without Internet connection, was harnessed for *in situ* analysis of microorganisms found in the depths of South Wales Coalfield [4]. Although the entire process was far from simple, the study demonstrated that DNA sequencing in remote locations is currently feasible. Other studies [2, 8] serve as a further proof of principle.

Future: With the continuing improvements to the sequencing technology, and simplification and automation of DNA extraction and preparation protocols (see next section), we may expect that portable DNA analysis will become a ubiquitous tool. Rapid medical diagnostics, forensics, agriculture, and general exploration of microbial diversity on Earth and in the outer space, are just some domains that will benefit. However, the most exciting opportunities are in consumer genomics. ONT has been promoting the idea of Internet of Living Things (IoLT) [19], where anyone will be able to sequence anything anywhere, opening endless possibilities for DNA-driven discoveries. Yet, because even short DNA sequencing runs can easily deliver gigabytes of data, which may require hours to analyze, the success of IoLT will depend on the ability of mobile and cloud computing to provide adequate support.

3 PORTABLE VS. BENCHTOP

In order to understand computational challenges in portable DNA sequencing, it helps to first look at the end-to-end DNA sequencing process, and how this process differs between the current benchtop sequencing platforms and the emerging portable sequencers.

Table 1: General characteristics of different steps in DNA sequencing.

	Library Preparation		DNA Sequencing		DNA Reads Processing	
	MinION	Illumina	MinION	Illumina	MinION	Illumina
Time Scale	Minutes-hours	Hours-days	Real-time, Minutes-hours	Days-weeks	Real-time, Minutes-hours	Batch-mode, Minutes-days
Equipment	Basic, portable laboratory equipment (e.g. pipettes, centrifuge)		Small portable USB device	Benchtop machine	Laptop to data center	
Energy	1-2W		1W	Up to 10KW	No data available, sustained high load	
Software	N/A		Proprietary firmware, drivers and control software		Usually open source, complex string and statistical algorithms	
Advantages	Fast and easy protocol	Protocols for very low DNA mass	Inexpensive, portable, real-time, long reads	Low cost per base, low error rate	Streaming and interactive sequencing	Many tested workflows available
Challenges	High mass of input DNA	Time and labor intensive	High error rate, high cost per base	Short reads, expensive device	High error rate	Short read length, high data volume

3.1 How DNA is Studied

A typical DNA sequencing workflow involves both laboratory and computational steps (see Fig. 2 and Tab. 1). While the specifics of the protocols executed in every step may vary, the main steps remain the same irrespective of the DNA sequencing platform.

Sample Collection: The first step is to obtain the material from which DNA will be extracted. The choice and quantity of material to sample are dictated by the particular application. Sample types include patient's blood (e.g. in epidemiology), feces (e.g. when studying gut microbial flora), water or soil (e.g. when tracking biological contaminants), etc. The samples are preserved, e.g. by freezing or adding a chemical buffer, to minimize degradation or contamination until they can be further processed. As we mentioned earlier, sometimes samples may be impossible to adequately preserve, necessitating immediate processing.

DNA Isolation: In this step, the collected samples are subjected to chemical, mechanical or thermal processing to extract and purify the DNA molecules. DNA extractions performed in a laboratory with the standard commercially available kits take from minutes to hours, and in addition to the basic tools like pipettes, involve heating/cooling and specialized equipment, e.g. a centrifuge. Thus, this step requires access to power supply, or the use of improvised substitute solutions.

Library Preparation: The purified DNA is further processed, to make it compatible with the sensing machinery of the sequencer. This usually involves basic biochemical processing that nevertheless may require complex protocols. The step becomes nuanced when DNA has to be biochemically barcoded or amplified as required in some applications. Overall, the entire library preparation takes from several minutes to several days, and usually does not require additional equipment beyond the portable laboratory tools.

DNA Sequencing: Once the DNA library is ready, the actual sequencing can be performed. Currently, several sequencing platforms are available, e.g. Illumina, PacBio, Ion Torrent or Oxford Nanopore. They differ in how DNA is detected and read, which translates into differences in: sequencing speed and throughput (i.e. the number of nucleotides detected per unit of time), length of the output reads (i.e. how long DNA fragments sequencer can sense), and error rate (i.e. how many incorrectly detected nucleotides one may expect in the output). These differences are crucial, since they directly affect the downstream processing and analysis. The current sequencers are controlled by computers, which also receive and store output data. With the exception of the MinION, they are not portable, taking days to complete a single sequencing run in laboratory. Finally, the sequencing process involves additional consumable resources, such as biochemical reagents and flow cells – devices in which the actual DNA sequencing happens.

DNA Reads Processing: This final step is purely computational. Its goal is to first convert signal produced by a sequencer into DNA reads, and then analyze these reads for insights. Because of the volatility of DNA, and the technical limitations of the sequencing platforms, DNA is hard to sequence as a single large molecule (e.g. a chromosome). Instead, it is sequenced in fragments that the sequencer is able to sense. The raw signal produced for each detected DNA fragment is run through base calling algorithms to generate DNA reads – the actual strings where detected nucleotides are represented by corresponding letters, commonly referred to as bases (**A**denine, **C**ytosine, **G**uanine, **T**thymine).

The collected DNA reads are the input to bioinformatics analysis. Here different workflows can be applied, depending on how samples had been prepared for sequencing, and what are the questions of interest. For example, *de novo* DNA assembly aims at reconstructing genome from input reads, while metagenomic analysis uses DNA reads to detect, classify, and functionally annotate microorganisms present in the sequenced samples. However, irrespective of the applied analysis, the common denominator is the reliance on compute and memory intensive string, combinatorial and statistical algorithms, ranging from massive graphs construction and traversal, through clustering, to large databases querying. Consequently, this step requires access to non-trivial computational resources, often exceeding capabilities of a single laptop or even a desktop computer.

3.2 Comparison

To compare portable and benchtop sequencing we concentrate on the MinION, currently the only portable sequencer, and Illumina platform, the dominating benchtop solution. We make our comparison in the context of *Library Preparation*, *DNA Sequencing*, and *DNA Reads Processing*, since these steps vary between platforms. Table 1 summarizes our comparison.

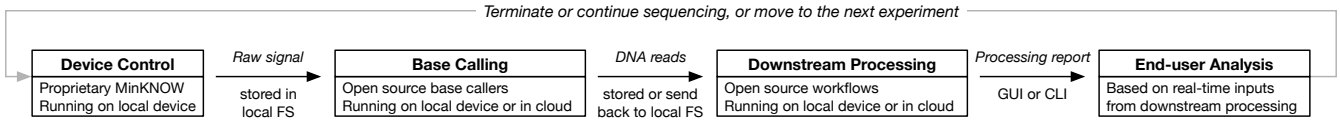

Figure 3: Current MinION software workflow.

Library Preparation: The attractiveness of MinION is in rapid sequencing protocol that can be executed in the field. The protocol takes roughly 10 min, and can be automated by using portable hands-off sample preparation hardware [17]. However, compared to standard protocols, it has limitations: it usually leads to lower quality sequencing output (i.e. with higher error rates), and it requires significant amount of input DNA (~240 ng). In comparison, protocols for the Illumina platform, while much more complex and labor intensive (most take hours), can be performed with an order of magnitude less DNA, without impacting sequencing quality.

DNA Sequencing: The MinION is based on the idea of nanopore sequencing, where DNA molecules pass through organic nanoscale sensors in a flow cell [11]. This approach has three main advantages: first, it permits portable and easy to use design with a minimal power consumption, second, it enables real-time sequencing – the signal gathered by hundreds of nanopores in a sequencer becomes immediately available for downstream processing, third, it can produce long reads (current average is ~7K bases compared to ~250 bases for Illumina). All this comes at the price of high error rate (10%-30% compared to ~0.1% for Illumina), which may lead to poor sequencing yield and complicates downstream analysis. Finally, because of the lower throughput, the cost of sequencing a single base is relatively higher (if we exclude upfront investments). This is in part because the MinION flow cells last for at most 48h of continuous sequencing.

DNA Reads Processing: The DNA reads produced in real-time by MinION allow for flexible approach to bioinformatics analysis, with the emphasis on streaming and one-pass strategies run outside of data centers. This makes "interactive" sequencing possible: the process can be terminated as soon as enough reads have been collected to answer question at hand. Long reads simplify tasks such as querying of databases or DNA assembly, but the high error rate makes other tasks (e.g. variants detection) challenging, or requires more input reads to resolve uncertainties. In contrast, the Illumina platform is high-throughput and high-volume oriented. The data from a single batch run is typically processed by well established, and to some extent standardized, workflows executed in a data center close to the sequencing facility. Low error rate combined with high DNA reads volume, makes tasks such as variants detection possible. However, short reads force complex algorithms on other tasks, e.g. assembly.

4 MOBILE COMPUTING PERSPECTIVE

We are now ready to highlight some of the open problems in portable DNA analysis as pertaining to mobile computing. To help understand the problems, we first provide the overview of the state of the art in mobile DNA analysis software. Then we discuss the limitations and open problems. While we base our discussion on the current MinION platform, we believe that the points we make are equally applicable to the future generation of portable sequencers.

4.1 Current Software Overview

Figure 3 shows the MinION software architecture. The first component is MinKNOW – the proprietary control and signal acquisition software. MinKNOW is responsible for the configuration and supervision, including initiation and termination, of sequencing runs. It also receives and stores DNA signals generated by the ASIC in the sequencer's flow cell.

The second component is base calling software. Here multiple options are available [20], including execution on the host device or delegation to a specialized cloud service (e.g. Metrichor [10]). The current basecallers are very compute intensive, e.g. they typically involve Recurrent Neural Nets (RNNs), and are often recommended to run in the cloud.

The final software component consists of the application specific bioinformatics tools selected by end-users. Here, bioinformatics and computational biology keep delivering many open source solutions to choose from. Usually, these tools are organized into a pipeline of their own, and may involve multiple processing stages (e.g. preprocessing to remove low-quality data, querying a database to find sequences matching given DNA reads, building a tree of genetic relationships, etc.). Depending on the complexity, this DNA analytics may be deployed on the host machine, but more frequently it will be running in the cloud.

4.2 Limitations

The software architecture discussed above has been already used with success to showcase the promise of portable DNA sequencing. However, the approaches thus far focused on manual and *ad hoc* organization of the existing tools to run in a mobile setup, without addressing many important concerns, including a systematic approach to energy, data and network management. While this is a great first step, this strategy has too many limitations to be sustainable in a long term.

Energy Management: Energy is one of the most critical resources in mobile environments. It is especially important for mobile DNA analysis, because a sequencer is directly attached to, and draws energy from, a host device. Then, computational tasks at different stages of the sequencing workflow may run for tens of minutes to hours. Yet, the current software tools do not have any mechanisms to consider energy as a managed resource. In fact, bioinformatics software tools are routinely designed under the assumption that they will be executing either on computational servers or in data centers, with abundant main memory and storage, parallel execution capabilities, and stable power supply.

Data Management: DNA sequencer generates large volumes of data, which flows through various processing stages. For example, a 48h continuous run may produce over 200 GB of output, scattered across millions of files. Furthermore, if a cloud backend is used for the analysis, data is moved back and forth between a mobile device and the backend.

Currently, data is managed independently by each software component. This puts an unnecessary burden on software designers since they need to implement not only core functionality (e.g. a base calling algorithm) but also data management logic (e.g. data transfers, ensuring interoperability with other software, etc.). It also hinders rapid deployment of new data management mechanisms, as they have to be integrated into multiple and disjoint software elements.

Network Management: Some of the most promising and anticipated applications of mobile sequencing are *in situ* scenarios, where DNA has to be completely handled at a remote location. Here, network connectivity could be sporadic and bandwidth could highly fluctuate. However, the current software is not designed to be adaptive to changing network conditions. It assumes either no network connectivity, and hence runs locally, or depends on full network connectivity, and hence the always-on availability of a cloud service. Moreover, the decision to run locally or remotely is left to the end user, who must decide before executing the experiment.

Consumables Management: A flow cell is the workhorse of a sequencer. It is a consumable that can be used to analyze only a limited number of DNA samples, and within a limited time. Moreover, a flow cell degrades over time, and that translates into progressively lower sequencing throughput and growing error rate. Hence, in truly mobile setups it is necessary to manage flow cells as a scarce resource. While this problem has been recognized, currently no systematic solution exists that would offer the necessary functionality.

4.3 Open Problems and Our Approach

The majority of the limitations we identified above, are cross-cutting issues that involve multiple software components. For example, all software components should have some form of data, energy, and network management – in order to implement a new solution that addresses a limitation for any one of these, we need to work with multiple components and apply the solution across all of them. This is time consuming and error prone, and thus hinders rapid innovation.

To address this challenge, we envision a new software architecture that identifies all necessary functions and separates them into different software components with clean interfaces to ensure interoperability. Specifically, we propose the architecture based on three elements: the data management layer, the DNA analytics layer, and the workflow manager. This architectural separation has the benefit of allowing different components to innovate independently from other components. It also has an advantage of simplifying software development, by allowing each component to focus on its core functionality. It also accommodates the existing software, especially growing set of bioinformatics tools.

Data Management Layer: The goal of having a separate data management layer is to free other software components from the burden of managing data on their own. Thus, the data management layer should provide all functionality related to DNA analysis data. This includes 1) an interface for other components to read and store data, including the backward compatibility support for the POSIX interface that the existing tools use for flat files, 2) efficient algorithms and mechanisms for data management, including discovery, monitoring and delivery, and 3) integration with cloud services for processing delegation and data backup.

Interesting questions arise for the design of a data management layer in all three aspects. First, for the clean slate interface design, the primary question is what kind of abstractions make the most sense for DNA analysis. As mentioned earlier, there are mainly three types of data – raw signals generated by a sequencer, DNA reads (strings), and analysis results (e.g. stored as data tables). Thus, perhaps the most natural interface design is to have an abstraction for each data type. That would allow other components to easily search, access, and when needed join data without dealing with low-level details such as file management.

Second, once we design the interface, the underlying implementation can freely employ various data management mechanisms. For example, it is well-known that DNA has much inherent redundancy due to its small alphabet (four letters) and because it is repetitive. Thus, compression strategies can help reduce the volume of data to manage. However, it is an open question as to how best to compress this data in a mobile setting, considering the trade-offs between computing cost, computing precision and constrained storage. The existing general as well as DNA-specific strategies, which take into account DNA quality and properties of known genomes, may work well, but it remains to be seen which strategy is practical in a mobile setup.

Lastly, DNA sequencing and analysis data needs to be moved back and forth between a mobile device and a cloud service. As discussed earlier, base calling requires extensive RNNs and it is recommended to run it in the cloud. Similarly, some DNA analyses are computationally intensive and require access to large reference databases. Thus, the data management layer needs mechanisms to optimize data transfer. Here again, the existing techniques, such as similarity detection and dynamic chunking, may or may not work well.

DNA Analytics Layer: Ideally, the DNA analytics layer should have multiple sub-components, each implementing one algorithm relevant to DNA reads processing. The goal of this layer is to allow algorithm designers to focus on algorithms and their implementations without worrying about other orthogonal issues, such as data or energy management.

Interesting questions arise if we consider that DNA processing can leverage both mobile devices and cloud services at the same time. This provides an opportunity to revisit current solutions and redesign them such that they become amenable to running in both domains, or in either one of the domains. In fact, we can envision a programming model to simplify implementation of such strategies. The model could provide primitives to encapsulate alternative realizations of the same DNA processing task as small migratable entities, allowing them to move across mobile and cloud domains or run in parallel if necessary. It could be further extended to account for the fact that certain DNA processing problems can be answered with different quality, trading specificity or sensitivity for computational, memory or energy performance. For instance, one of the

most general questions a user may wish to ask is *"what's in my pot"* [9], which is to report all known organisms whose DNAs have been found in a sample. The question can be answered by classifying detected organisms at the species level (i.e. fine-grained assignment) or just at the family level (i.e. coarse-grained assignment). The fine-grained assignment will typically require large reference databases and compute intensive sequence comparison. However, the process can be accelerated by using alternative strategies with lower precision (e.g., with data sketching and clever indexing). By implementing such multiple strategies, we could open more flexible execution paths. For example, in cases where resources are scarce, "approximate" tasks could provide less precise but potentially useful information, while waiting until resources are sufficient for a detailed answer.

One additional advantage of having task-based analytics layer is the ability to easily deploy workflows with stream processing and speculative execution capabilities (the techniques known to improve resource utilization). We note that while many of the existing algorithms, especially involving querying of reference databases, can be cast into this model with little or no effort, some other (e.g. construction of DNA assembly graphs, clustering, etc.) would require additional research and reformulation.

Workflow Manager: The goal of the workflow manager is to orchestrate all aspects of the DNA sequencing and analysis workflow, taking into account multiple static and dynamic factors that a user might encounter during a sequencing experiment. Many of the limitations that we discussed earlier fall into this category. Energy consumption, network connectivity, bandwidth variation, flow cell degradation, etc. all contribute to dynamically changing conditions of an experiment. Hence, the workflow manager should carefully monitor these variables and continuously make decisions on what the best course of actions is. This leads to several design questions. For example, how to monitor an experiment including not only the basic properties of a mobile device, e.g. how much energy is left or what the current network condition is, but also status of a flow cell and the quality of reads it is producing. Currently, very limited resources are available regarding flow cell monitoring, which we believe is an interesting research opportunity.

Once we have monitoring capabilities, the second question is how best to utilize available resources to get a desired outcome. This is especially important since a user conducting a DNA analysis in the field often has to make critical decisions based on limited information. For example, suppose a user is conducting DNA analysis in a remote area where there is no network connectivity. The user might wonder if she has enough energy to finish pending DNA analysis on her laptop and use the resulting data to adjust her experiment (e.g. collect more samples, etc.), or if she should move to an area where there is network connectivity to offload the analysis to a cloud and then continue the experiment. The workflow manager should either assist users to make well-informed decisions, or be intelligent enough to make decisions on its own, without requiring any user intervention. In cases where DNA analysis algorithms are amenable for partitioning or migrating across different domains, the workflow manager could make fine-grained decisions of moving various computational tasks across domains directly.

5 FINAL REMARKS

The proposed architecture is our attempt to introduce a more systematic and scalable approach to mobile DNA analytics, by tapping into concepts known from mobile computing. One potential caveat is that the proposed architecture would require reimplementation of some of the existing bioinformatics tools to fully leverage facilities in the DNA analytics layer. However, we believe this is feasible considering that portable sequencers are relatively new technology, with almost no algorithms designed for mobile systems. Currently, our team uses the latest release of MinION to investigate the questions we pose in this paper in the context of environmental DNA analysis.

REFERENCES

[1] M.D. Cao, D. Ganesamoorthy, A.G. Elliott, H. Zhang, M.A. Cooper, and L.J. Coin. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION sequencing. *GigaScience*, 5(1), 2016.

[2] S.L. Castro-Wallace, C.Y. Chiu, K.K. John, et al. Nanopore DNA sequencing and genome assembly on the International Space Station. *bioRxiv*, 2016. doi:10.1101/077651.

[3] A. Edwards, A.R. Debbonaire, B. Sattler, L.AJ Mur, and A.J. Hodson. Extreme metagenomics using nanopore DNA sequencing: A field report from Svalbard, 78 N. *bioRxiv*, 2017. doi:10.1101/073965.

[4] A. Edwards, A. Soares, S. Rassner, , P. Green, J. Felix, and A. Mitchell. Deep sequencing: Intra-terrestrial metagenomics illustrates the potential of off-grid nanopore DNA sequencing. *bioRxiv*, 2017. doi:10.1101/133413.

[5] N.R. Faria, E.C. Sabino, M.R.T. Nunes, L.C.J. Alcantara, N.J. Loman, and O.G. Pybus. Mobile real-time surveillance of Zika virus in Brazil. *Genome Medicine*, 8(1), 2016.

[6] J.L. Gardy and N.J. Loman. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics*, page nrg.2017.88, 2017.

[7] F.C. Hewitt, S.L. Guertin, K.L. Ternus, K. Schulte, and D. R. Kadavy. Toward rapid sequenced-based detection and characterization of causative agents of Bacteremia. *bioRxiv*, 2017. doi:10.1101/162735.

[8] S.S. Johnson, E. Zaikova, D.S. Goerlitz, Y. Bai, and S. W. Tighe. Real-time DNA sequencing in the Antarctic Dry Valleys using the Oxford Nanopore sequencer. *Journal of Biomolecular Techniques*, 28(1), 2017.

[9] S. Juul, F. Izquierdo, A. Hurst, X. Dai, A. Wright, E. Kulesha, R. Pettett, and Turner. D. J. What's in my pot? Real-time species identification on the MinION. *bioRxiv*, 2015. doi:10.1101/030742.

[10] Metrichor Ltd. Metrichor, an Oxford Nanopore Company. https://metrichor.com/, 2017.

[11] H. Lu, F. Giordano, and Z. Ning. Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics & Bioinformatics*, 14(5), 2016.

[12] A. Pomerantz, N. Penafiel, A. Arteaga, L. Bustamante, F. Pichardo, L.A. Coloma, C.L. Barrio-Amoros, D. Salazar-Valenzuela, and S. Prost. Real-time DNA barcoding in a remote rainforest using nanopore sequencing. *bioRxiv*, 2017. doi:10.1101/189159.

[13] J. Quick, N.J. Loman, S. Duraffour, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), 2016.

[14] C. Quince, A.W. Walker, J.T. Simpson, N. J. Loman, and N. Segata. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 2017.

[15] M.R. Stratton, P.J. Campbell, and P.A. Futreal. The cancer genome. *Nature*, 458(7239), 2009.

[16] Oxford Nanopore Technologies. Oxford Nanopore. https://nanoporetech.com, 2017.

[17] Oxford Nanopore Technologies. Voltrax. https://nanoporetech.com/products/voltrax, 2017.

[18] M.C. Walter, K. Zwirglmaier, P. Vette, S. A. Holowachuk, K. Stoecker, G. H. Genzel, and M. H. Antwerpen. MinION as part of a biomedical rapidly deployable laboratory. *Journal of Biotechnology*, 250, 2017.

[19] E. Waltz. Portable DNA sequencer MinION helps build the Internet of Living Things. *IEEE Spectrum*, 2017.

[20] R.R. Wick, L.M. Judd, and K.E. Holt. Comparison of Oxford Nanopore basecalling tools. https://github.com/rrwick/Basecalling-comparison, 2017.