 doi:10.20944/preprints201905.0113.v1

Introducing DNA Sequencing to the Next Generation on a Research Vessel Sailing the Bering Sea Through a Storm

Anne-Lise Ducluzeau[1#], Rachel M. Lekanoff[2], Noah S. Khalsa[2], Hillary H. Smith[3], and Devin M. Drown[1,4#]

[1] Institute of Arctic Biology, University of Alaska Fairbanks.
[2] College of Fisheries and Ocean Science, University of Alaska Fairbanks.
[3] College of the Environment, University of Washington.
[4] Department of Biology and Wildlife, University of Alaska Fairbanks.
Address correspondence to Anne-Lise Ducluzeau aducluzeau@alaska.edu and Devin M. Drown, dmdrown@alaska.edu

ABSTRACT

Experiential learning in the field is an opportunity for students to enter the heart of a scientific discipline. Through such experience, they can extract conceptual clues and discover motivational stepping stones that will potentially influence the rest of their education and career choice. Unfortunately, in Biology, the inescapable topic of Next-Generation Sequencing represents a challenge when it comes to create an educational curriculum that aims to provide students with hands-on experience on sequencers. It is an even more difficult task to accomplish if one's purpose was to set such curriculum in a field situation. However, in recent years, educators have seen possibility to bring Next-Generation Sequencing to the reach of students more easily, with the Oxford Nanopore MinION, a low-budget, user-friendly, hand-held sequencer. Academic researchers have illustrated the performances of this device in the field and are inspirational for curricula aiming to take the next generation of scientists in the outdoors. We designed a modular 5-day workshop, with nanopore sequencing to be performed in field conditions. Here we describe the material and methods that lead the students and instructors from sample collection, DNA extraction and preparation for nanopore sequencing with MinION to real-time analysis of the data collected. This curriculum was implemented for the first-time aboard Research Vessel *Sikuliaq* during a transit organized by the STEMSEAS program at Columbia University in collaboration with the University of Alaska BLaST program. The line of investigation formulated for the workshop was an open-ended question that led the students to establish a proof of concept in terms of technology deployment at sea: *what will show metagenomic results from DNA obtained from sea water and sequenced with Oxford Nanopore MinION?* The workshop took place in October 2018 while Research Vessel *Sikuliaq* sailed the Alaskans seas for 7 days. Students successfully used nanopore sequencing for multiple metagenomic seawater samples. Their introductory analysis was consistent with environmental conditions and they were able to present their results by the end of the workshop.

KEYWORDS: Education, genomics, nanopore, oceanography, fieldwork

 doi:10.20944/preprints201905.0113.v1

INTRODUCTION

Field experiences contribute to the success of students enrolled across STEM fields [1]. In their detailed review on the benefits of such field activities, Fleischner et al [2] describes how a non-traditional physical setting helps with the understanding abstract concepts and career orientation. Field-based learning can also enhance peer engagement as well as with faculty and learning content.

Hands-on experiences are also opportunities to incorporate modern technology in the life of future problem-solvers. When it comes to Biology, progress made in DNA sequencing with the development of Next-Generation Sequencing (NGS) devices, has led to the massive integration of genomics across fundamental and applied research. Thereby, the molecular biology processes that lie behind NGS are taught through classes and lab work [3]. Nonetheless, it remains difficult to set a teaching lab experience that would show the process through and through, as NGS sequencers are more than often placed out of students access for their complexity and cost.

In 2014, Oxford Nanopore Technology (ONT) launched the MinION, a DNA sequencer radically different from others commercialized so far. One of its most innovative features is its portability with a weight of less than 90 g. MinION is USB powered which allows it to be controlled by a laptop. The device does not require a traditional laboratory set-up to be implemented and delivers data in real-time. Moreover, MinION's hardware and software designs are user-friendly. Adding to this, MinION is also the most affordable DNA sequencer currently on the market with a cost of $1000. Such attributes designate the MinION as a valuable piece of material for educational purpose and it is of no surprise if educators have opened the door of their classroom to this sequencer shortly after its release. Zaaijer 2016 [4] based the development of a 13-week academic curriculum at Columbia University on the usage of the MinION by the students for the proceeding of two sequencing runs and then, data processing. At the University of Alaska Fairbanks and at the Aberystwyth University, MinION sequencing is explored in shorter format hackathons and Microbiology skills class by Devin Drown and Arwyn Edwards. In the meantime, biologists with a broad range of research goals

have used the MinION to sequence in the field, notably in the outdoors of Antarctica [5], in a rainforest of Ecuador [6], in Western Africa [7] and even, in space [8]. These instances of successful deployments of MinION outside traditional lab set-ups suggest that the device is likely to be fit for the combination of education and research in the field.

With this idea in mind, we set for the development of a modular and reusable workshop integrating Oxford Nanopore MinION for sequencing. The goal of this workshop is to introduce DNA sequencing to undergraduate students in a real-field situation through the format of an open-ended investigation that will take them from sampling through to data analysis.

We field tested this modular workshop aboard a research vessel (R/V) during an event organized by the STEMSEAS program from the University of Pennsylvania and Columbia University. Between two research cruises, a R/V might have to transit from one location to another to meet scientists. The STEMSEAS program takes advantage of these periods to offer undergraduate students the opportunity to experience life and research onboard R/Vs.

Such stage for fieldwork with nanopore sequencing was the opportunity to bring the students of our workshop to set a proof of concept about embarked genomics. Indeed, every year, ocean sciences take researchers at sea onboard R/Vs for cruises that can extend from a week to several months. Because of the prohibitive cost of a research cruise, intense efforts of sampling are proceeded during these periods. Nonetheless, for marine biologists interested in genomics, access to results is postponed until they return on shore for the difficulty of taking DNA sequencers onboard. This implementation of our workshop was aimed at demonstrating that the portability of Oxford Nanopore Technology devices could modify marine biologists' workflow in a drastic way, accelerating investigations.

Hence, we set up an experimental plan that would enable the students to deploy multiple sequencing stations for real-time analysis of seawater metagenomics collected at several water depths. For this field case, 11 undergraduate students embarked R/V *Sikuliaq* for a transit between in Nome and Seward, in October 2018. They were enrolled in diverse programs ranging from Anthropology to Geology passing by Biology. Despite adverse weather, the cohort of students successfully deployed the sequencing stations and analyzed their results. They observed the disruption of the natural stratification of marine microbial communities

consistent with a mixing effect of strong winds on the water column and thus, demonstrated the relevance of using Oxford Nanopore Technology onboard research vessels.

INTENDED AUDIENCE AND PREREQUISITE STUDENT KNOWLEDGE:

We designed this workshop to introduce a user-friendly DNA sequencing technology to a public of undergraduate students with various sets of background courses, skills, knowledge, and previous field experience. Thus, no prerequisite is necessary.

LEARNING TIME:

Below, we report the succession of the workshop steps with their duration.

DAY 1:

- Sample collection (60 minutes) - conductivity, temperature and depth (CTD) instrument cast

DAY 2:

- Short lecture (60 minutes) – background material, optional

DAY 3:

- Workshop introduction (10 minutes)
- Introduction of nanopore sequencing and familiarization with the MinION device (20 minutes)
- Training with pipettes (15 minutes), optional
- Flow cell priming training (15 minutes)
- DNA Library preparation and flow cell loading (30 minutes)
- Familiarization with MinKNOW software and start of the run (10 minutes)
- Data collection in real-time and analysis of a first batch of results (variable)

DAY 4-5:

- Analysis of all the entire sequencing runs and preparation of figures (variable)
- Preparation of a 10 minutes presentation (variable).

LEARNING OBJECTIVES:

Following the completion of this workshop, students will be able to:

- Understand sample collection and processing for genomics,
- Understand the nanopore sequencing workflow including DNA library preparation and software control,
- Interpret metagenomic data in terms of understanding biodiversity,
- Design independent research projects that integrates nanopore sequencing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

METHODS

PROJECT DESCRIPTION:

We provide an overview of the workshop in Figure 1. Briefly, on Day 1, R/V Science engineer and assistant instructor cast a CTD instrument and students collect seawater samples from several depths of the water column. Subsequently, instructors filter the samples onto 0.22uM filters. They perform DNA extractions with Qiagen DNeasy Powerwater kit and clean-up the extracts with magnetic beads. In order to prepare 400 ng aliquots of DNA, instructors measure the concentration of cleaned-up extracts with a Qubit device. Instructors can perform these DNA preparation steps either on the first day or the second.

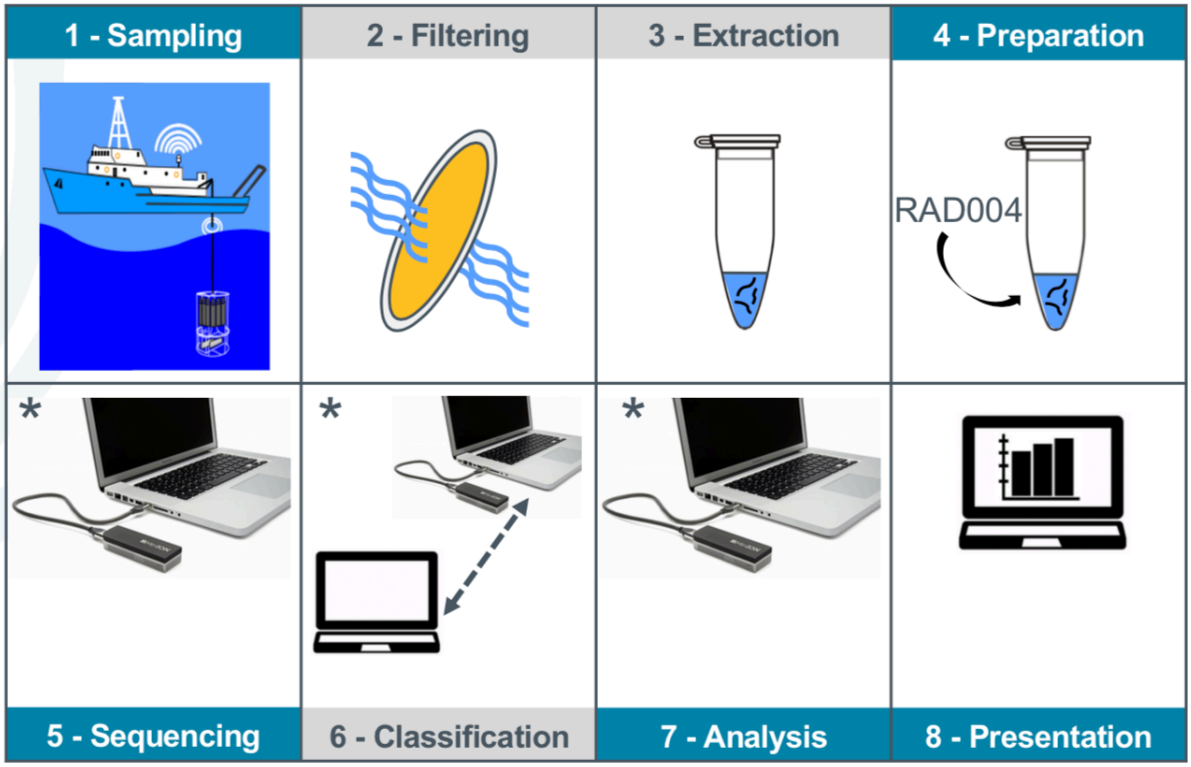

Figure 1. Workshop workflow: analysis of sea water samples processed for the collection in real-time of metagenomics data with Oxford Nanopore MinION. Steps designed to be performed by students are indicated in blue and by instructors, in grey. Asterisks indicate steps of real-time access to data and results.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

On Day 2, instructors can choose to give a lecture about DNA sequencing and environmental DNA, in order to provide students with a background about the DNA sequencing experiment to come.

At the beginning of Day 3, instructors introduce nanopore sequencing technology to the students with the support of a slide presentation (Appendix 1) and hands-on training (e.g. pipetting, flow cell priming). Then, each group receives 400 ng of aliquoted DNA and completes the preparation of a DNA library with the Oxford Nanopore Rapid Sequencing Kit (SQK-RAD004) following the manufacturer recommendations, (included in the slide presentation, Appendix 1). students load their library and start the sequencing run with the "live basecalling" option activated which means that data become available for analysis immediately. MinKNOW GUI interface allows students to follow the progression of the run in real-time.

At this point, instructors describe the bioinformatic pipeline that will be performed part by themselves and part by the students. Within the first hour of sequencing, during a break, instructors collect data from each station (DNA sequence reads as fastq files) and proceed to taxonomic classification with Kraken [9] using two databases: minikraken_4Gb for bacteria, archaea and viruses and a custom-made database encompassing NCBI Refseq mitochondrial sequences for eukaryotes. The instructor redistributes the created classification reports to each original sequencing station.

When the workshop resumes, students visualize the classification of their reads through the R package Pavian [10]. Students report their results on a "sequencing board" through the guidance of a set of questions (Appendix 2). This bioinformatic workflow can be iteratively conducted during the sequencing run to follow the progression of results. Students and instructors choose to stop the runs at any moment upon review of runs progress.

Students can decide to continue the activity over the next couple days where they can further analyze the totality sequencing results and create figures based on the exportation of CSV tables from Pavian. Finally, the students prepare a presentation for entire group on the last day of the transit. This presentation must include a description of the entire workflow that led to the results and the results *per se*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

MATERIAL:

We provide a list of material in Appendix 3. Adjust quantities of samples and material in consideration of the number of student groups expected for the workshop. This field case consisted in the deployment of six regular sequencing stations (12 students) and of one extra station including the Oxford Nanopore MinIT (instructors). For the travel, the overall material was arranged in two plastic cargo boxes (e.g. Actions Packers, 25 gallons each), a foam-padded Pelican case for laptops (iM2975) and an additional small cargo box (4 gallons) for power adapters.

Based upon the design of this workshop, the budget for the sequencing of one sample with one sequencing station is ~$1033. We provide a detailed breakdown of this cost estimate in Table 1. Note that the final cost includes the consumables necessary for the sample preparation, the sequencing consumable but excludes the cost of reusable equipment such as computers, peristaltic pump, Oxford Nanopore MinIONs, ThermoFisher Qubit and miscellaneous laboratory equipment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

Table 1 – Estimated budget for the deployment of one Oxford Nanopore MinION station for the metagenomic sequencing of one sample.

Company	Product	Cat. No	Price per unit (USD)	Unit quantity	Amount needed for ONT protocol	Cost
Amazon	5L foldable drinking water container bags		$8	1	1	$8
Millipore	Mixed Cellulose GSWP 0.22 µm filters	GSWP04700	$147	100	1	$1.47
Fisher scientific	Lysozyme	BP535-1	$111	1 g	1 mg	$0.11
Sigma-Aldrich	Mutanolysin	M9901-5KU	$115	5 KU	0.3 KU	$7
Sigma-Aldrich	Lysostaphin	L7386-1MG	$84	500 U	24 U	$4
Qiagen	RNAse A	1007885	$222	2.5 ml	4 µl	$0.35
Qiagen	DNeasy PowerWater Kit	14900-50-NF	$497	50	1	$10
MagBio Genomics	Magbio High Prep PCR magnetic beads	DT-70005	$315	5 ml	30 µl	$2
ThermoFisher	Qubit HS reagent kit	Q32854	$281	500	1	$0.6
Oxford Nanopore	Flow cells		$900	1	1	$900
Oxford Nanopore	ONT RAD004 kit	SQK-RAD004	$599	6	1	$100
Project cost per group per run:						**$1033.5**

180

 doi:10.20944/preprints201905.0113.v1

STUDENT INSTRUCTIONS:

Instructors provide specific instructions regarding experimental and bioinformatic tasks through a slide presentation (Appendix 1). For general instructions: at the beginning of the workshop, instructors advise students to work in pairs as we found that this configuration is favorable for peer-learning and for involving each student in a significant amount of hands-on tasks. For training at the flow cell priming, instructors make students aware that they have been provided with old consumables and thus, can feel comfortable while getting familiar with the material.

FACULTY INSTRUCTIONS:

Instructions provided in Appendix 4 cover all the tasks that instructors will have to execute: preparation of laptops, protocols (sampling, filtering, DNA extraction and clean-up, quality control and storage), lab preparation and bioinformatics. Additionally, we suggest the following directions for instructors to apply during the sequencing experience. If a pipetting training is included, instructors spend time with each group of students to give some feedback on students technique. Instructors proceed the same way during the flow cell priming training period which necessitates the application of an alternative way to use pipettes, the “Quick method”. This technique consists in the distribution of volumes through the rotation of the plunger instead of the usual pressing movement. This technique is recommended for the flow cell priming step that requires a gentle release of the priming buffer. At the end of the flow cell priming training, instructors gather the used flow cells and distribute the new flow cells along with the library loading mix. During the DNA libraries preparation, instructors circulate from a group to another with the FRA tube so each group of students can withdraw the appropriate volume of reagent. Instructors ask to the students to keep the FRA reaction in their tube rack until all student pairs are ready to proceed together to the incubations at 30°C and 80°C. Instructors also provide the RAP reagent to one group at a time. When starting the runs, instructors ask the students to leave “real time basecalling” option ON.

SAFETY ISSUES:

Students work with DNA extracts and Oxford Nanopore Technology reagents that are not hazardous. While students did not carry the DNA extractions, clean-up and quantifications

 doi:10.20944/preprints201905.0113.v1

in this field case, exposure to molecular biology reagents has to be considered for the instructors: Qiagen DNeasy PowerWater and Qubit assay kits contain some flammable and irritant for the skin and eyes reagents. MagBio magnetic beads are not hazardous. Instructors must dispose wastes in consideration of regional regulations. For this field case, instructors and students respected at all time the safety instructions communicated by the R/V personal. It included multiple safety and ship evacuation trainings.

RESULTS

The activity took place during a one week long transit in the Bering Sea and Alaskan Peninsula onboard R/V *Sikuliaq*, starting from Nome, AK and ending in Seward, AK in October 2018. We prepared this workshop for the participation of 12 undergraduate students. Students came from a diversity of educational disciplines including science communication, anthropology and political sciences, environmental geosciences, geophysics, biology, fisheries, zoology and environmental sciences. Pipetting skills ranged from none to good. Notably, 2 students had previously been involved in DNA sequencing projects for the characterization of microbial communities from the environment. One of them had previously used the Oxford Nanopore MinION.

A succession of two storms marked the first 3 days of the transit and as a consequence, magnified the level of seasickness experienced by the STEMSEAS students and instructors. In consequence, the original layout of the workshop was modified and implemented as it follows. During the first day (Oct. 3rd 2018), instructors proceeded to samples collection themselves after the students who were fit had prepared the CTD instrument. A cast performed at Station 002 (63° 0.042N/167° 0.029W) where total depth is 28m, allowed the collection of about 3L of seawater from 7 depths (5.0m, 8.6m, 10.3m, 14.2m, 17.2m, 20.3m and 22.6m) of the water column. At the same time, the instrument measured physical oceanographic properties, including water temperature and salinity. The assisting instructor filtered 1.1L to 1.5L of each sample and stored the filters overnight at 4°C.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

On the second day, lead instructor prepared the samples: the DNA was extracted, cleaned up and measured for concentrations that ranged between 54 and 110 ng/µl. DNA extracts were then stored at 4°C overnight.

In the morning of the third day, still sailing in unfavorable conditions, the lead instructor opted for proceeding to the library preparation herself in consideration of the discomfort created by pipetting in such conditions and which could have impacted the course of the workshop. The assisting instructor prepared the lab and material necessary for the workshop. At the beginning of the workshop, the instructors observed that students properly set-up the sequencing stations on their own. This unexpected display of early autonomy from students was for us one confirmation of the intuitive and appealing aspects of MinION platforms for educational purpose.

Lead instructor started the workshop with a short background lecture and then, proceeded to the hands-on activities. Instructors skipped the pipetting training originally designed as a drop forming exercise, and instead slightly increased the duration of the flow cell priming training. It is at this time that students non-familiar with pipettes were introduced to technique basics along with the “Quick method”. Lead instructor used the seventh platform to show the sample loading process and then guided the students verbally while circulating with the assisting instructor from a group to another for the start of the sequencing run. Instructors observed a real ease of the students to use MinKNOW software for the launching of their sequencing run.

Across the platforms loaded with samples, immediately, one showed a screen feedback indicating that a poor run was to be expected most likely because of a deficient library preparation. Nonetheless, the student and lead instructor decided to let the run progress and discussed about the specific factors that were indicating difficulties. Subsequently, this station, loaded with the 10.3 m depth sample provided no data *i.e*, less than 4000 passed reads that could not be retrieved as fastq file.

At this point, instructors and students decided to take time off that lead instructor used to collect the data produced during the first 30 minutes of the runs. The fastq files of each platform were combined and processed through Kraken to generate two classification reports:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

one for bacteria, archaea and viruses with minikraken_4Gb database and another for eukaryotes with the RefSeq custom made mitochondrial database. This analysis took about 20 minutes for the 6 stations. As a result, Kraken classified 0.6% to 1.3% of the reads for the for microbial communities and 8.4% to 11.6% of the reads for the eukaryotic communities. When the workshop resumed, this first batch of result was available for the students to review with the R package Pavian (10). Taxonomic names of bacteria such as *Planktomarina, Pelagibacter, Croceibacter atlanticus, Octadecabacter arcticus* were in support of the oceanic origin of the samples (Figure 2).

Unfortunately, the eukaryotic communities classified with the custom made RefSeq mitochondrial database, remained more enigmatic for the difficulty of linking the taxonomic names to actual animals, plants and algae despite resources about Alaskan species of metazoa brought onboard.

Students communicated their sequencing yields and major results to the instructors whom filled up the sequencing board with temporary results. After 7 hours of sequencing, instructors decided to end the runs for (i) the risks of letting the stations unmonitored

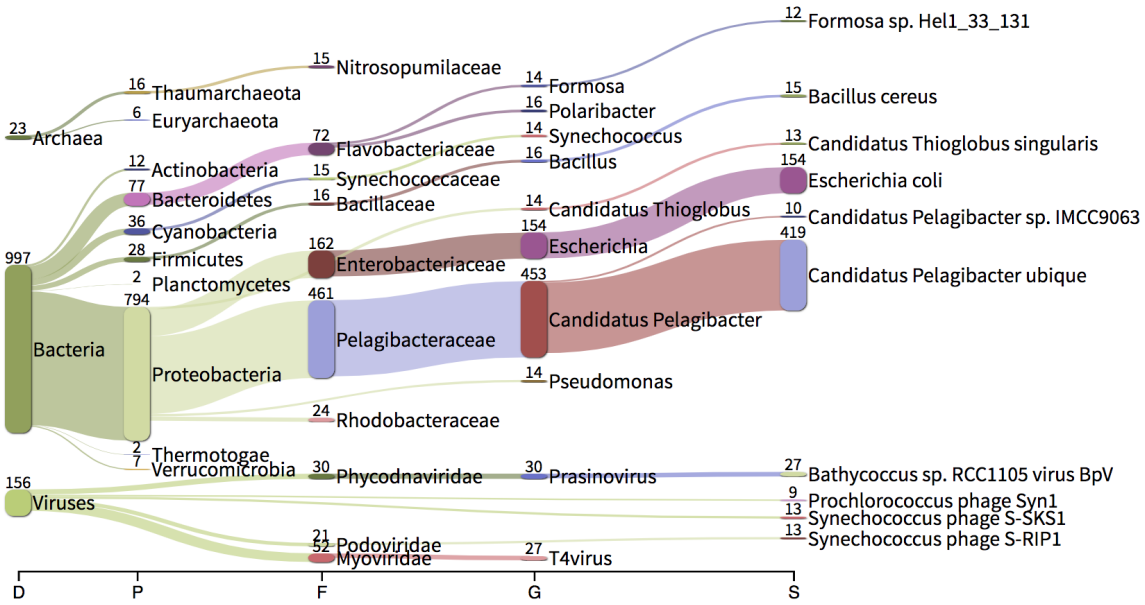

Figure 2. Visualization of results processed during the workshop: Sankey diagram (R package Pavian) of the microbial community at 22.6m depth. Numbers indicate number of classified reads, letters stand for: D, domain; P, phylum; F, family; G, Genus; S, species.

 doi:10.20944/preprints201905.0113.v1

overnight while the storm was still ongoing (ii) the significantly reduced sequencing speed observed within the past hour and (iii) the amount of accumulated data that was satisfactory for the next step of the activity (32 000 to 132 000 reads).

For the next couple of days, students had to choose between different afternoon options including Geology, Oceanography, Biology and Science Communication. Two students decided to proceed to the analysis of the data collected on the 6 stations. Instructors provided them with the corresponding Kraken reports. One student reviewed the resources brought onboard about marine viruses, microbial communities. The other student and assistant instructor took over the abundance analysis through 'R studio' (Figure 3).

Results showed very similar communities through the 6 different depths of the water column. The two students formed the hypothesis that such pattern could be consistent with an extensive mixing of the water column due to the strong winds experiences at the time of the sampling. The natural stratification of microbial communities could have been disrupted by such weather conditions and led to appear as a homogeneous community. Instructors prompted the 2 students to look for support of this hypothesis through additional data available onboard. Through discussions with oceanographers and the assisting instructor, the two students retrieved (i) salinity and temperature data measurements collected during the sampling CTD cast that showed uniformities for these physical properties as well and (ii) an additional dataset related to samples collected in the same area, a couple of months prior to the workshop showing a stratification of communities over 3 depths of the water column.

On the last afternoon of the workshop, the 2 students prepared a 12 slides presentation and on the last day of the transit, they presented the overall results to the rest of the students and R/V crew member.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

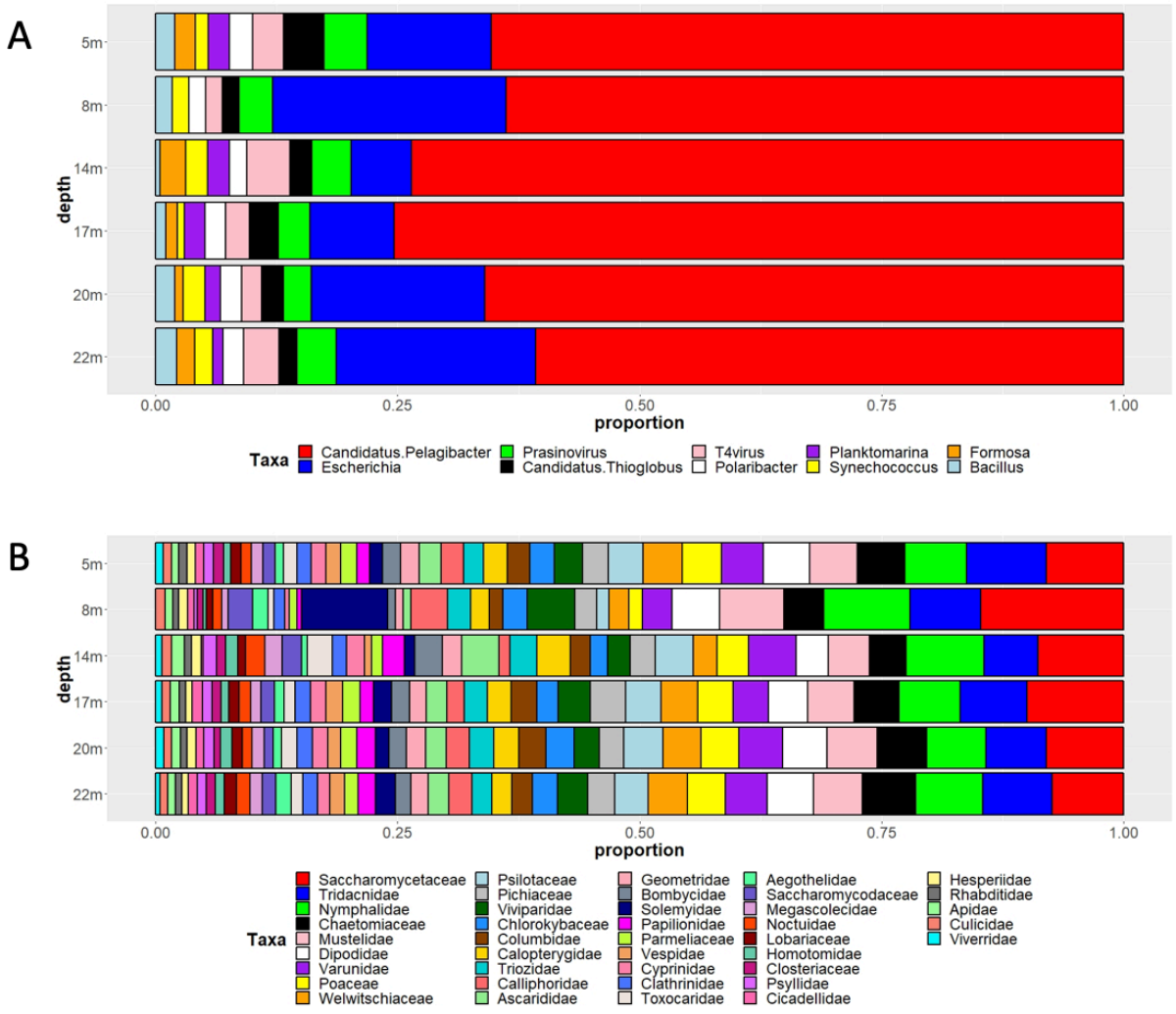

Figure 3. Representation of the results based on the analysis proceeded during the optional workshop times. Communities composition of samples collected at 6 different depths of the water column and which DNA libraries successfully provided reads during the workshop. Classified reads ranged between 0.6% and 1.3% of the provided datasets for the microbial and viral community (minikraken_4Gb database) and between 8.4% and 11.6% of the provided datasets for the eukaryotic community (RefSeq custom made mitochondrial database). Visualization created with R Studio and based on originals produced onboard: only the colors were changed. (A) Genera of the for microbial community, (B) families of the eukaryotic community.

303

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

DISCUSSION:

POSSIBLE MODIFICATIONS:

This workshop is modular. For instance, sample filtering, DNA extraction and/or clean-up steps could be performed by students depending on time, goals and environments. Secondly, depending on the level of Biology knowledge and lab experience of the expected students, instructors can choose to implement lectures and basic lab skills training prior to the workshop *i.e*, on day 1 or 2.

In regards of the experimental aspect of the activity, samples other than seawater can be considered provided that they will contain sufficient amounts of DNA for subsequent library preparations. Also, the DNA clean-up step is optional but strongly recommended as sample purity is critical for the sequencing yield with Oxford Nanopore technology. Finally, the Oxford Nanopore kit used for this field case, SQK-RAD004, is the fastest preparation commercialized by the company (expect 30 min). Considering another type of chemistry is of course possible but will increase significantly the duration of the hands-on part of the workshop, except in the case of SQK-RBK004 (expect 1 hour).

Concerning Oxford Nanopore equipment, an alternative to the “classic” flow cell is the now released Flongle. This new consumable features fewer nanopore channels compared to the classic flow cell (128 versus 512) but its cost is also substantially less ($90/flow cell) and will lead to a significant decrease of a station’s budget. These factors make it exceptional for educational and outreach applications.

Regarding the bioinformatics, various pipelines could be adopted. In order to avoid classification of some reads as *Escherishia coli* with Minikraken database (Figure 3), we suggest adding an adapter trimming step with PoreChop (https://github.com/rrwick/Porechop) to the pipeline. Kraken v1 used in this field testing has a confidence scoring option embedded in the kraken-filter script. This option increases the precision of results at the disadvantage of sensitivity. We did not apply this option here as our datasets were small. However, in future implementations of the workshop, if dataset sizes allow, we will use the confidence scoring option (--confidence option) with the new version of the classifier, Kraken2.

LESSONS LEARNED:

Better yields for the sequencing runs. In this field case, the sequencing runs performed lower yields than what was expected by instructors based on preparatory experiments. We see two possible reasons for this. First, it is likely that the cold chain of the RAD004 kits has been interrupted during the travel from Fairbanks to Nome (6 hours) because of an insufficient amount of blue ice packs (total of 3) in a too small styrofoam box. In the future, we intend to reuse, Oxford Nanopore shipping material when travelling or to have the chemistry directly delivered on location if possible. Second, it is possible that impurities within the samples remained after the DNA clean-up step. However, this seems less likely to us as preparatory experiments conducted with seawater samples have shown us that a single clean-up step was sufficient to reach the required level of purity. Nonetheless, from now on, we will include an Agilent Nanodrop device to our portable laboratory set-up in order to control this aspect of sample preparation.

Flexibility of activity's plan. Fieldwork situations are subject to inherent factors that cannot be controlled and can impact operations and people. Fortunately, in our case, the unfavorable weather did not affect the sampling in a timely manner and our initial 5 day schedule was respected. It seems worthy to us to mention that we designed the duration of the activity to be shorter than the overall time spent at sea. Our goal was to create room in our schedule to prevent stress if for instance the sampling was delayed. Nonetheless, task adjustment and rescheduling had to be adjusted considering the STEMSEAS team's altered fitness. Additionally, we would also emphasize the importance of preparing back-up solutions for every step, in order to insure instructor ability to adjust to the widest spectrum possible of field conditions. This includes multiple sampling plans, samples collected prior to the activity, DNA extracts and datasets.

Of the benefits of additional data for a broader perspective on results. It is likely that weather conditions at the time of sampling led the sequencing results to depict the disruption of the natural stratification of microbial communities. In order to clarify this observation, students engaged in cross-field conversations with onboard oceanographers and collected a set of complementary data (CTD measurements of water temperature and salinity). They also

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0113.v1

addressed assisting instructor who had a previous experience in our sampling area and doing so, they obtained an additional sequencing dataset. As a consequence, students enabled themselves to consider their results in a more comprehensive context and to consolidate their analysis. We consider the gathering and discussion of these additional data as a turning point in the engagement of students. While at first considered as adverse, the challenging conditions of the activity created a sense of the unexpected for us all. We feel that for the students, it particularly deepened their investigative experience and promoted a sense of ownership towards the results of the overall activity.

The cost of taxonomic classification without an internet connection. While at sea, the R/V *Sikuliaq* offers a limited access to internet for onboard personal and researchers. Nonetheless, more often than not, access to internet is just impossible at sea. To adapt to such remote conditions, we relied on the offline version of MinKNOW to execute the sequencing runs and we came to use a local solution *i.e*, a server encompassing two small size databases for the classification of metagenomic reads.

Through preparatory experiments, instructors had observed that results provided with the minikraken_4Gb database for bacteria, archaea and viruses were consistent with expectations but those collected through the custom made RefSeq mitochondrial database were at best approximative to plainly aberrant. For the workshop, instructors decided to primarily focus on the for microbial community and to use the eukaryotic classification as an opportunity for a discussion about the influence of database over results.

"*What will show metagenomic results from DNA obtained from sea water and sequenced with Oxford Nanopore MinION?*" was the question that students addressed during this field-case. Answering to this involved for them to depict a marine landscape through organisms identified with Kraken and two databases. Technically, the activity demonstrated that it was possible: the workflow was entirely executed to its completion and results were consistent with a marine environment.

Now, from an educational perspective, we feel that having relied on the etymology of taxonomic names and hard-copies of marine related publications fell short at conveying the relevance of the results to the group that was mostly composed of students non-familiar with

 doi:10.20944/preprints201905.0113.v1

marine biology and microbiology. For future workshops, we anticipate to create lines of investigation pertaining to a specific organism or group of organisms. Such focus will minimize the effect of a lack of familiarity with taxonomy in offline internet conditions and will enable every student to mine their dataset with an enhance sense of a task successfully accomplished.

 doi:10.20944/preprints201905.0113.v1

ACKNOWLEDGEMENTS

The authors would like to thank the October 2018 STEMSEAS transit students: Katie-Jo Campbell (Indiana University of Pennsylvania), Demi Carballosa (Miami Dade College), Jennie Humphrey (University of Alaska Fairbanks), Kevin Huo (University of Alaska Fairbanks), James Kowalski (University of Utah), Samuel Radcliffe (Mt. San Antonio College), Alexis Ruiz (Mt. San Antonio College), Myesa Swangel (University of Washington), Jesse Trott (Portland State University). We thank the STEMSEAS program for offering these students the amazing opportunity to participate to the October 2018 transit. We are grateful to Oxford Nanopore Technologies, Ltd for sharing of the offline version of the MinKNOW software and providing a MinIT device and flow cells. We wish to thank the UAF-BLaST program and the Institute of Arctic Biology for their generous support. We would also like to acknowledge the College of Fisheries and Ocean Science at the University of Alaska Fairbanks for their logistical support. Thanks to Elisabeth Rull and Joseph Quigley who braved the waves with their cameras, to the incredible Wes Schult for creating videos and to Eric Collins for providing us with indispensable travel-size microcentrifuge. Finally, but not the least, we would like to profusely thank the entire crew of the R/V *Sikuliaq* for their professionalism and welcoming: it was a lifetime experience to sail this storm with you all. As John Hamill, R/V 2nd mate, says *"Every day is a good day to be at sea, until it's a very bad day"*: you made it all good days. The research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers UL1GM118991, TL4GM118992, or RL5GM118990. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. STEMSEAS is supported by the NSF under Award Numbers 1701174 for the University of Pennsylvania and 1701168 for Columbia University.

REFERENCES

1. Easton E, Gilburn A. 2012. The field course effect: gains in cognitive learning in undergraduate biology students following a field course. *J Biol Educ* 46:29–35.
2. Fleischner TL, Espinoza RE, Gerrish GA, Greene HW, Kimmerer RW, Lacey EA, Pace S, Parrish JK, Swain HM, Trombulak SC, Weisberg S, Winkler DW, Zander L. 2017. Teaching Biology in the Field: Importance, Challenges, and Solutions. *Bioscience* 67:558–567.
3. Hotaling S, Slabach BL, Weisrock DW (2018) Next-generation teaching: a template for bringing genomic and bioinformatic tools into the classroom. *J Biol Educ* 52:301–313
4. Zaaijer S, Columbia University Ubiquitous Genomics 2015 class Y, Erlich Y. 2016. Using mobile sequencers in an academic classroom. *Elife*. doi: 10.7554/eLife.14258.
5. Johnson SS, Zaikova E, Goerlitz DS, Bai Y, Tighe SW. 2017. Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. *J Biomol Tech* 28:2–7.
6. Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, Barrio-Amorós CL, Salazar-Valenzuela D, Prost S. 2018. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *Gigascience*. doi: 10.1093/gigascience/giy033.
7. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouédraogo N, Afrough B, Bah A, Baum JH, Becker-Ziaja B, Boettcher J-P, Cabeza-Cabrerizo M, Camino-Sanchez A, Carter LL, Doerrbecker J, Enkirch T, Dorival IGG, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D, Nitzsche K, Pallash E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan K, Vitoriano I, Yemanaberhan RL, Zekeng EG, Trina R, Bello A, Sall AA, Faye O, Faye O, Magassouba N, Williams C V, Amburgey V, Winona L, Davis E, Gerlach J, Washington F, Monteil V, Jourdain M, Bererd M, Camara A, Somlare H, Camara A, Gerard M, Bado G, Baillet B, Delaune D, Nebie KY, Diarra A, Savane Y, Pallawo RB, Gutierrez GJ, Milhano N, Roger I, Williams CJ, Yattara F, Lewandowski K, Taylor J, Rachwal P, Turner D, Pollakis G, Hiscox JA, Matthews DA, O'Shea MK, Johnston AM, Wilson D, Hutley E, Smit E, Di Caro A,

Woelfel R, Stoecker K, Fleischmann E, Gabriel M, Weller SA, Koivogui L, Diallo B, Keita S, Rambaut A, Formenty P, Gunther S, Carroll MW. 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530:228–232.

8. Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins KH, McIntyre ABR, Dworkin JP, Lupisella ML, Smith DJ, Botkin DJ, Stephenson TA, Juul S, Turner DJ, Izquierdo F, Federman S, Stryke D, Somasekar S, Alexander N, Yu G, Mason CE, Burton AS. 2017. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Sci Rep* 7:18022.

9. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46.

10. Breitwieser FP, Salzberg SL. 2016. Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. *bioRxiv* 084715.

APPENDICES

Appendix 1: Workshop slide presentation

Appendix 2: Analysis questions for students

Appendix 3: List of materials

Appendix 4: Faculty instructions