Development of a Nucleic Acid-Based Life Detection Instrument Testbed

Conference Paper

Author(s):
Bhattaru, Srinivasa A.; Tani, Jacopo; Saboda, Kendall; Borowsky, Jonathan; Ruvkun, Gary; Zuber, Maria T.; Carr, Christopher E.

Development of a Nucleic Acid-Based Life Detection Instrument Testbed

Srinivasa Aditya Bhattaru
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge, MA 02139, USA
+1-631-413-7286
sbhattar@mit.edu

Jacopo Tani
ETH Zürich
Sonneggstrasse 3
8092 Zürich, Switzerland
+41-44-632-01-29
tanij@ethz.ch

Kendall Saboda
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge, MA 02139, USA
+1-631-413-7286
kendalls@mit.edu

Jonathan Borowsky
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge, MA 02139, USA
+1-617-253-0786
jonathan3141@gmail.com

Gary Ruvkun
Massachusetts General Hospital
185 Cambridge Street
Boston, MA 02114, USA
+1-617-726-5959
ruvkun@molbio.mgh.harvard.edu

Maria T. Zuber
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge, MA 02139, USA
+1-617-253-3206
zuber@mit.edu

Christopher E. Carr
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge, MA 02139, USA
+1-617-253-0786
chrisc@mit.edu

Abstract— Future space instruments will explore increasingly complex questions about our universe, including the origin of life on Earth and the presence of life elsewhere. These instruments will likely integrate chemical and biological subsystems that will face unique challenges; existing protocols typically require non-stabilized components and manual handling. The Search for Extra-Terrestrial Genomes (SETG) instrument is being developed for *in situ* extraction and sequencing of nucleic acids as a biomarker of life on other planetary bodies. Such sequencing is being implemented using a nanopore-based device, the Oxford Nanopore Technologies (ONT) MinION; as such, it needs to integrate many benchtop-based protocols. Here we describe an automated testbed, designed and built to automate and rapidly prototype extraction, library preparation, and sequencing protocols that could be used in our instrument. The system is designed to be modular with respect to components, facilitating hardware and software modifications with minimal system impact, while also precise across multiple test runs, allowing for accurate evaluation of the impact of varying system inputs as well as exploration of system failure modes and potential solutions. We also present testing results from each of the three primary subsystems (extraction, library preparation, and sequencing) as well as a plan for and initial data on subsystem integration into an end-to-end system. The extraction subsystem is able to match or approach nucleic acid yields attained via manual testing for *B. subtilis* spores in water (15%) and spores in basalt (12%). The library preparation subsystem can successfully prepare a library of *E. coli* DNA that can be identified after sequencing. The loading/sequencing subsystem has successfully automated sequencer loading, resulting in a sequencing run producing 1.4 billion bases after 1 day of sequencing from a pre-prepared sample. These testing results provide valuable data about the challenges of biological protocol automation, while directly informing future design decisions for SETG. In the process, the lessons learned from this milestone are relevant to the technological development of future planetary science instruments that take advantage of molecular biology techniques.

TABLE OF CONTENTS

1. INTRODUCTION

Determining beyond doubt that life is not exclusive to planet Earth would represent a remarkable discovery. In order to accurately identify extant life, it is necessary to search for life via unambiguous biomarkers; many biomarkers can suggest the presence of life but cannot provide umabiguity as to the presence of life or the lack of contamination. One of the most unambiguous biomarkers is the presence of informational polymers, such as DNA [1]. All life as we know it uses informational polymers to store hereditary information, and there are no known abiotic ways to form DNA strands of non-trivial length. Additionally, there is evidence to indicate that life could have existed on other planets, emerging on early Earth due to meteoric exchange or through a common, external origin [2–6]. Mars, in particular, is a compelling location in the search for extant life. Current evidence indicates that Mars was more habitable for life in the past [7]. Additionally, a significant amount of meteoric transfer occurred in the Late Heavy Bombardment phase; evidence suggests that nearly 1 billion tons of rock were exchanged between Mars and Earth in their early history [8]. Thus, a positive detection of informational polymers on Mars would strongly suggest the previous presence of life. Additionally, since all known life shares some common base sequences as well as some distinctive ones [9], obtaining nucleic acid

sequences and comparing them to native Earth data would allow us to explore if detected extant life is Martian or was transferred from Earth as spacecraft forward contamination. Informational polymers are thus an attractive biomarker to pursue for astrobiology.

As such, the Search for Extra-Terrestrial Genomes (SETG) project was founded to develop the technology to detect, identify, and map extant life through informational polymers, with the intent of flying an instrument on a future planetary science mission [10, 11]. This project takes advantage of emerging technologies in life detection. Recent advances have reduced the size, mass, and power requirements of sequencing technology, leading to the development of portable sequencers such as the Oxford Nanopore Technologies (ONT) MinION sequencer. The MinION is a portable, power efficient, sequencing platform that provides data and analysis in real-time [12]. This enables systems to build around that platform to search for nucleic acids *in situ* under the constraints of spacecraft instrumentation.

In order to develop this instrument, individual components need to be designed, tested, and validated under analogue conditions. Sequencing nucleic acids has been done for decades in laboratories, but as far as we are aware it has not been achieved in an automated, integrated instrument that performs all functions required for *in situ* analysis. The SETG project is currently pursuing the assembly and testing of an end-to-end automated system; the envisioned instrument will have to operate autonomously on a planetary surface. By testing this capability, we both validate our current protocols and inform future custom designs necessary for each subsystem.

Here we describe an automated prototyping testbed. This testbed is not the final end-to-end system; it is a platform used to explore the challenges that emerge through protocol automation, as well as a starting point to explore designs for custom-designed, integrated subsystems. We present the system overview in Sec. 2, followed by a description of the extraction, library preparation, and sequencer/loading subsystems. Implementation is discussed in Sec. 3, while testing is discussed in Sec. 4. The results of our initial testing are described in Sec. 5, after which are our discussions and conclusions in Sec. 6 and Sec. 7, respectively.

2. SYSTEM OVERVIEW

In this section we provide a functional overview of the automated SETG testbed, detailing the role each subsystem plays in the bigger picture. The system is comprised of three main subsystems: extraction, library preparation, and loading/sequencing. These three subsystems work in series to take an environmental sample and produce a mapped DNA sequence (Figure 1).

Extraction Subsystem

The extraction subsystem receives as an input an environmental sample with cells/spores and produces as an output free DNA in solution. Our current methodology for extraction is cell lysis via the use of Claremont BioSolutions OmniLyse® (OL) system. The OL device is a small bead filled chamber with a rotor, powered by a small motor; the motor agitates the beads, which lyse cells in solution in the chamber. Input and output ports allow a user to fill the device, while a two-wire electrical connection allows for powering the motor. The manual protocol is described in [13] and results are expressed in terms of yield percentage, defined as the ratio

of input to output DNA mass. Automatically extracting a similar percentage of DNA to the manual protocol, as well as a sufficient amount of total DNA to enable subsequent subsystems, would constitute an initial success.

Library Preparation Subsystem

Library preparation (library prep) receives as an input free DNA from the extraction subsystem and produces as output prepared, ready-to-sequence DNA. In order for DNA to be sequenced, it needs to be sufficiently prepared into a library. The protocol we implement in the SETG project is determined by the MinION sequencer model and the flow cell chemistry. The current state of the art methods for library prep are (a) via manual handling and (b) by using large automated devices, such as the Andrew Alliance liquid handling robot. Smaller automated devices such as the VolTRAX are in development, but have their own limitations and are not directly transferable to a space grade instrument. The initial goal with this subsystem would be an automated purification and library preparation of free DNA.

Loading/Sequencing Subsystem

The loading/sequencing subsystem receives as an input the prepared DNA from the library prep system and produces a DNA sequence. This subsystem involves priming, loading, and running our MinION nanopore sequencer. The MinION utilizes a Flow Cell unit that contains a polymer membrane, perforated by biological nanopores made of proteins. A constant voltage is applied to the pores, allowing a current to pass through that is disrupted when nucleic acids pass through (Figure 2). The perturbation of that current allows for the sensing and identification of the nucleic acids. The sequencer is primed and loaded via pipette before being started through computer control. The success of this step is dependent on the ability to load a sample without introducing bubbles and on the computer's ability to run the MinKNOW sequencing program. Successfully and cleanly loading the system and running the software would allow us to move forward with the system.

3. IMPLEMENTATION

The full system design is presented below, with relevant components also shown (Figure 3). Each subsystem can be explored independently for performance and failure modes, or they can be run sequentially as one complete system.

General

- LabSmith reservoirs, tubing, fittings, interconnects, breadboards, and 3-way valves were used as the physical foundation for the testbed. Since they are extensible, modular, and designed for sub-mL scale fluid movement, these components were ideal for the initial structure of the testbed.
- A Lee Valve-Driven Pressure Delivery system, custom built by in-house, provides air pressure and vacuum to reservoirs as one of several fluid delivery methods. LabSmith 1 mL reservoirs are designed to be pressurized, and as such can be used to move fluid into a reservoir or out of one.
- The system is controlled via a data acquisition board (DAQ) connected to a testing desktop computer. The board generates analog and digital inputs and outputs, as well as several connections for pulse width modulation (PWM) timing.
- Silicone heaters are used for heating reagents and eluates. These heaters are actuated by custom-designed and built control boards that integrate thermistor based temperature sensing.
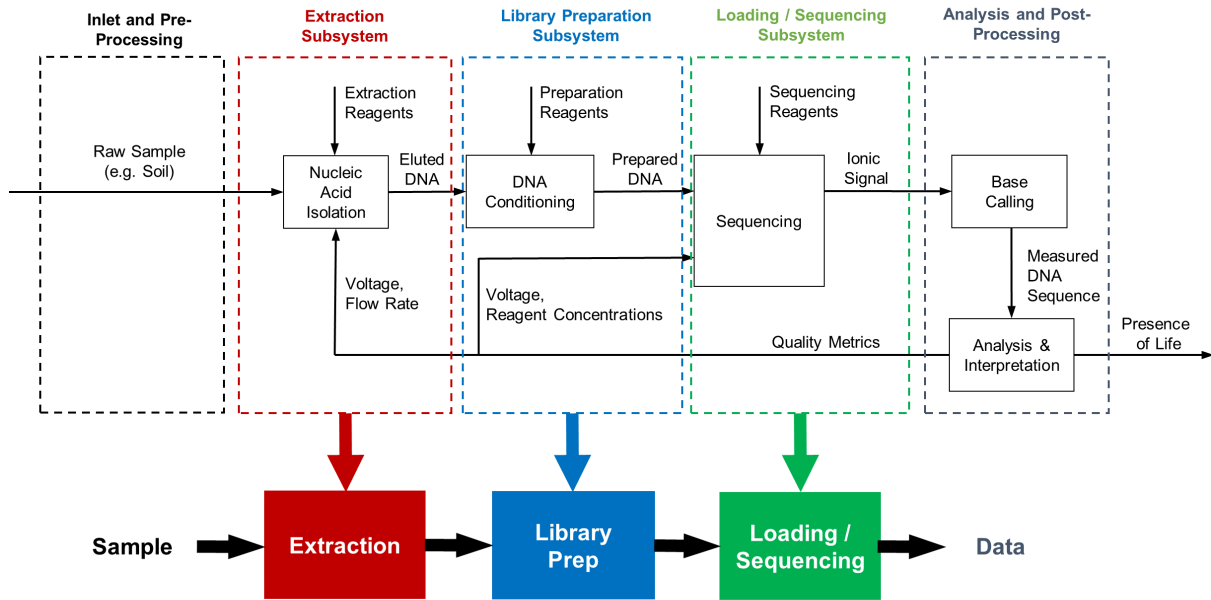
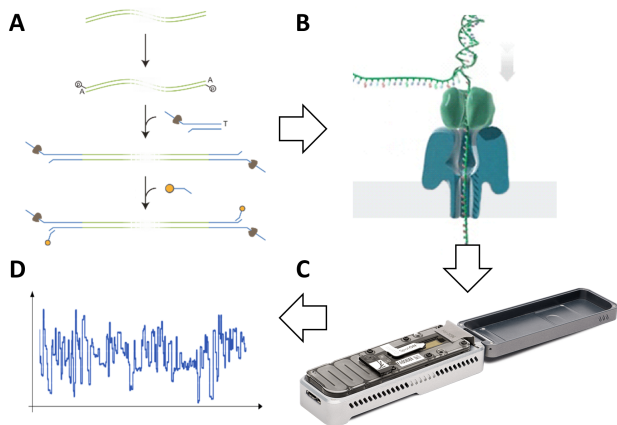**Figure 1**. A system overview of the automated SETG testbed.

Figure 2. Overview of the nanopore sequencing process, adapted from ONT images; DNA is prepared (A) and fed to a nanopore (B), which is embedded in the MinION Flow Cell membrane (C). The MinION sequences and produces ionic signals (D) that can be analyzed to estimate DNA sequences.

- Almost all components are controlled, via the DAQ or through serial communication, by a computer running MATLAB Simulink. Simulink is a real-time, block-based graphical programming environment, designed for modeling of dynamic systems. It was chosen due to its modularity, its suite of tools for live data plotting and analysis, and ability to facilitate rapid prototyping.
- Generic power supplies provide several relevant voltages for the system, including 6.6 V and 12 V.

Extraction

Extraction comprises the first two regions in the testbed overview (Figure 3). The extraction-specific hardware components are as follows:

- The primary form of fluid delivery in the extraction subsystem is a Harvard Apparatus Syringe Pump. The pump drives a BD 3 mL syringe connected to the system via 1/16" tubing.
- The OL device is the central component of extraction. The device is driven by a PWM board, allowing for dynamic, automated motor speed control.
- A BioChem 8-way valve is used for fluid path selection.

The system is centered around the syringe pump mixing reagents and delivering samples to the OL device. This allows us to explore the effects of long tubing as well as automatic OL device automation.

Library Preparation

Library Preparation comprises the next four regions in the testbed overview, including Phases II, IV, V, and VI (Figure 3). Library prep specific hardware is as follows:

- Two magnetic assemblies were constructed to assist in actuating the paramagnetic Ampure XP suspended bead solution. These assemblies are comprised of motors with small magnets attached via a prong, driven by off the shelf motor driver boards.
- BioChem one-way valves force unidirectional flow for several reagent reservoirs, preventing unintended backflow.

This subsystem is driven primarily by the pressure delivery system moving microliter scale fluids. Since this subsystem with the largest number of components and the smallest reagent volumes, we can explore the effects of automation on precise fluid delivery and sample loss.
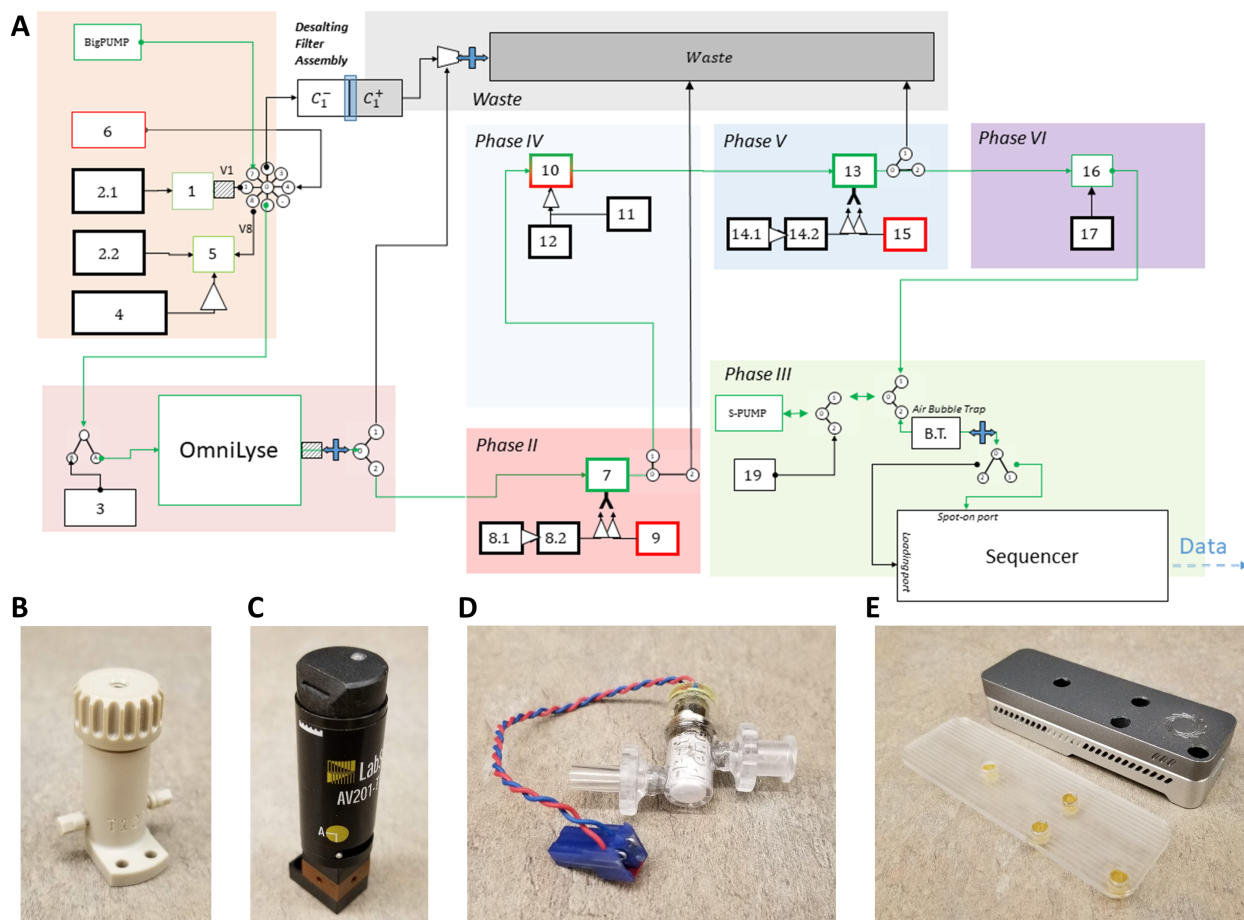
Figure 3. **A representative overview of the automated testbed design (A). Each numbered box represents a reservoir (B), each set of three connected circles represents a three-way valve (C), and major components like the OL device (D) and sequencer (E) are labeled.**

Loading/Sequencing

Loading/Sequencing is the final subsystem and is represented as Phase III in the testbed overview (Figure 3). Major components are presented below:

- Fluid movement is controlled through a LabSmith 80 uL syringe pump, designed to move small volumes of fluid precisely.
- An air bubble trap removes bubbles from sample before delivery, preventing loading inefficiencies due to bubbles in the Flow Cell membrane.
- A custom-designed and built interface is used to connect the base LabSmith system with the ports on a MinION Flow Cell.
- The MinION sequencer is used as the last major component of the automated testbed. The sequencer is powered and run by a sequencing computer running the MinKNOW software program, which records sequence data and monitors the flow cell.
- The UDOO X86 single board computer was chosen as the sequencing computer for the automated testbed, primarily due to its small footprint (120mm x 85mm).

The loading/sequencing system is driven by the LabSmith syringe pump, which both primes the flow cell with priming fluid before loading the completed library from library prep.

The subsystem's relative simplicity with respect to library prep or extraction means that it is able to explore other points of failure such as air bubble introduction and sample mis-delivery.

4. TESTING

Design, development, and testing of the automated testbed was done in parts, focusing on one subsystem at a time, to isolate potential issues. Each subsystem was broadly tested first for successful fluidic movement, then for low nucleic acid loss through movement, then finally for subsystem functionality. This enabled a more structured debugging process and a more smooth integration process.

Extraction

Extraction was the first subsystem that was explored. The process had previously been largely automated, but at the time the testbed was developed, additional aspects such as pre-lysis desalting were still being automated. Thus the testbed was intended to focus on the automation of library preparation and loading, as well as how those systems integrate with extraction. The system uses a syringe pump to move reagents from LabSmith reservoirs through the OL device (Figure 3D). The protocol is as follows, adapted from

CBIO:

1) Before testing, the extraction subsystem is flushed with sodium pyrophosphate (NaPP) as a pre-wash.

2) The input sample (regolith or water) is suspended in 1x CBIO binding buffer, along with competitive binders (810 uL).

3) The slurry is pushed back and forth through an OL device for 2 minutes as it runs, creating shear forces that act to lyse cells, before being pushed to waste.

4) 1x CBIO binding buffer is pushed through the OL device and to waste to wash the system (3 mL).

5) Elution buffer is delivered to the OL device and heated (200 uL).

6) The elution buffer is pushed through the device to the next subsystem.

Due to ongoing work done by our laboratory and our partners Claremont BioSolutions, we have an extensive data set for extractions in a bench-top environment [13]. Thus, we can compare the yield efficiency of the automated testbed to our previous manual testing results.

Initial extraction tests used long fragments of *E. coli* DNA suspended in pure water. To test nucleic acid sample loss, free DNA retention was tested before lysed spore DNA retention. The previously described protocol was designed and modified through this process to account for sample loss and yield efficiency, which included timing changes as well as a sodium pyrophosphate pre-wash (NaPP). NaPP is known to inhibit binding of nucleic acids to surfaces by occupying the same charge sites as DNA phosphate groups. Flushing the system with NaPP prior to extraction allows more DNA to pass through the system without loss.

Library Preparation

Library prep was developed as the final subsystem. Currently, there is no routine method for assessing the efficiency of ligating sequencing adapters to DNA without sequencing. Thus, determining whether or not library preparation was successful would require sequencing the resulting library. To explore the complicated failure modes of this system, library preparation was split into three sub-subsystems. An abridged version of the protocol is as follows, adapted from ONT and the Loman Lab in the University of Birmingham [14]:

1) *Post-Extraction DNA Concentration*

   – The input eluate is introduced to a reservoir of Ampure XP solution (70 uL).

   – A magnet is moved such that it holds the Ampure beads in place on the inside of the reservoir, creating a bead pellet on the reservoir wall.

   – The supernatant is removed and the chamber is washed with ethanol twice (2x 200 uL).

   – The magnet is moved away. Water is added to the bead solution and heated (50 uL).

   – The magnet is moved back. The beads are pulled back to the wall.

   – The water is pushed along as eluate.

2) *DNA Preparation into Library*

   – End preparation buffer (7 uL) and enzymes (3 uL) are added to the eluate, to prepare the ends of the nucleic acid strands.

   – Ligation master mix (80 uL), ligation enhancer (2 uL), and nucleic acid adapters (40 uL) are added to the solution and allowed to incubate.

3) *Library Purification*

   – The solution is introduced to another reservoir of Ampure XP solution (91 uL).

   – A magnet is moved such that it holds the Ampure beads in place on the inside of the reservoir, creating a bead pellet on the reservoir wall.

   – The supernatant is removed and the chamber is washed with adaptase binding buffer twice (2x 150 uL), re-suspending and re-pelleting the Ampure beads each time.

   – The magnet is moved away. Elution buffer (15 uL) and water (25 uL) are added to the bead solution.

   – The magnet is moved back, re-pelleting the Ampure beads.

   – The eluate is pushed to the next subsystem.

The system required moving many reagents, without the capability for syringe contamination. As such, the system was designed to use pressure delivery, which was the simplest method to move fluid without reusing syringes. One way valves are used throughout the system to prevent backflow, two heaters are used to heat reactions, and two magnets are implemented to actuate the Ampure XP bead solution.

Each of the individual sub-subsystems was first tested for fluid movement and nucleic acid retention individually before being tested as a complete unit. Design iterations were implemented after initial issues of fluid retention and splattering emerged, a major concern in a subsystem working with small volumes of reagents. After exploring these failure modes and potential solutions via new reservoir designs, the final system was tested by taking a sample of free DNA sheared to 6 kilobases, preparing it for nanopore sequencing, loading the library manually into a MinION, and running the sequencer. The results are described below.

Loading/Sequencing

The loading/sequencing subsystem had to be designed had to be designed around a complex system: the MinION sequencer. The loading/sequencing subsystem was designed and built to interface with the MinION while modifying it as little as possible (Figure 3E), centered around a custom fluidic interface. The interface has four ports that correspond to four ports on the MinION Flow Cell: the Spot On Port, the Loading Port, and two Waste Ports. The MinION has

four holes drilled through the top that allow the interface to connect with the Flow Cell. The system is actuated by the previously described 80 ul LabSmith syringe, and is supplemented by reservoirs, valves, and an air bubble trap. The protocol is as follows:

1) The priming buffer mix is loaded into the MinION.

2) The elution is mixed with running buffer to become the library and loaded into the MinION sequencer.

3) The MinION is activated and sequencing begins.

The pre-built nature of the MinION sequencer leads to some advantages, namely ease of validation; a previously prepared library could be automatically loaded into a Flow Cell and accurately compared to a manual load procedure. There are no new chemical elements or processes in the loading/sequencing subsystem, and as such most failure modes could be identified as primarily mechanical in nature. Fluid movement was tested first before full sequencing. Initial testing led to interface redesigns, improved flow speed control, and better interface clamping methods. The results of the first successful automatic load are described below.

5. RESULTS

Extraction

Several sets of results are available for the extraction subsystem. These results show both a progression and expansion of extraction capability. All tests were conducted with $\sim 10^8$ *B. subtilis* spores as an input, with results expressed as [DNA Mass Out]/[DNA Mass In] (Figure 4).
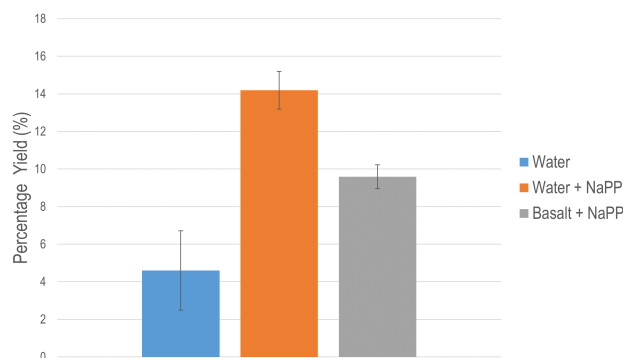
Figure 4. **Extraction yields with water, with and without using NaPP as a pre-wash, as well as basalt. Extractions with NaPP match manual extraction yield numbers.**

The results show that water extraction yields are close to 14% while basalt yields are at a lower 9%, which match initial manual yield data [13]. Yields prior to using NaPP as a pre-wash to prevent DNA binding are shown for comparison.

Library Preparation

Initial tests were conducted for each sub-subsystem. The primary focus of initial testing was on the automatic Ampure XP bead cleanup. This protocol comprised the first and third sub-subsystem, so maximizing yield was imperative. Manual yield values, measured as DNA Mass In/DNA Mass Out, were estimated at 70%, though with experience we have

been able to achieve yields up to 90%; this was the baseline value the automated bead cleanup was trying to achieve. Barring failures, initial automated cleanup yields were 41% (Figure 5). Upon identifying issues with fluid delivery, the system was further refined through reduced pressure and the design of new, low-fluid-retention reservoirs. Implementing these changes brought automated bead cleanup yields to 74%, albeit with higher variance.
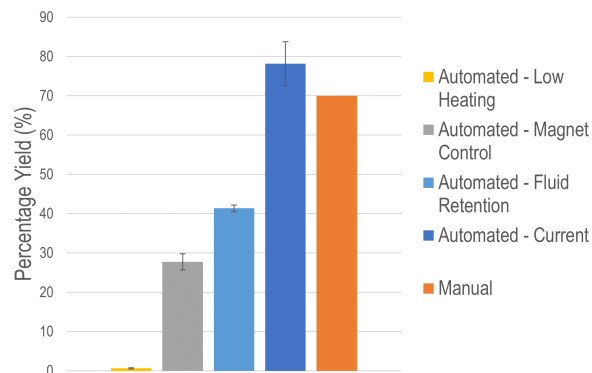
Figure 5. **Automated bead cleanup yields, compared to the lower bound of manual tests as well as an exploration of potential failure modes.**

The full system was then implemented and tested. The first successful library preparation was conducted with 1124 ng of *E. coli* DNA sheared to 6 kb as an input. Due to the integrated nature of library preparation, individual performances of the sub-subsystems is not available. The 73 uL library, with an estimated 230 ng of DNA, was loaded manually into a MinION sequencer and run for 2.5 hours, purely to confirm if library preparation had successfully occurred; the sequencing data is presented below.

The large amount of DNA in the library indicates that the subsequently low amount of data was due to low adaptation efficiency of library preparation (Figure 6A). In addition, the low length of DNA indicates that heavy shearing occurred during library prep or storage (Figure 6B); this is less relevant for environmental samples, since we expect those samples to be heavily sheared as is [10, 11]. 4.9 megabases were sequenced over 4300 reads, however, and mapping the reads afterwards indicated that they originated from *E. coli*. As such, the data indicates that the system can successful prepare DNA for nanopore sequencing.

Loading/Sequencing

The loading test data presented below is from a manually prepared library. We prepared a 72 ul library with 772.5 ng of sheared Lambda phage DNA, which was loaded into the input reservoir of the loading subsystem, a stand-in for the output reservoir of the library preparation subsystem. Upon starting and running the loading test, the subsystem successfully primed and loaded the flow cell without any large-scale problems; a small bubble formed in the flow cell membrane, obscuring 12.5 percent of pores, a frequent concern even in manual handling. Despite the bubble, the test produced a wealth of data, producing over 900*M* events in twenty hours. The test was only run for 24 hours rather than a full 48 hours since, judging by previous sequencing run experiences, little data would be produced the next day. Live base calling was not implemented, since the UDOO board computer was likely not powerful enough to sequence and
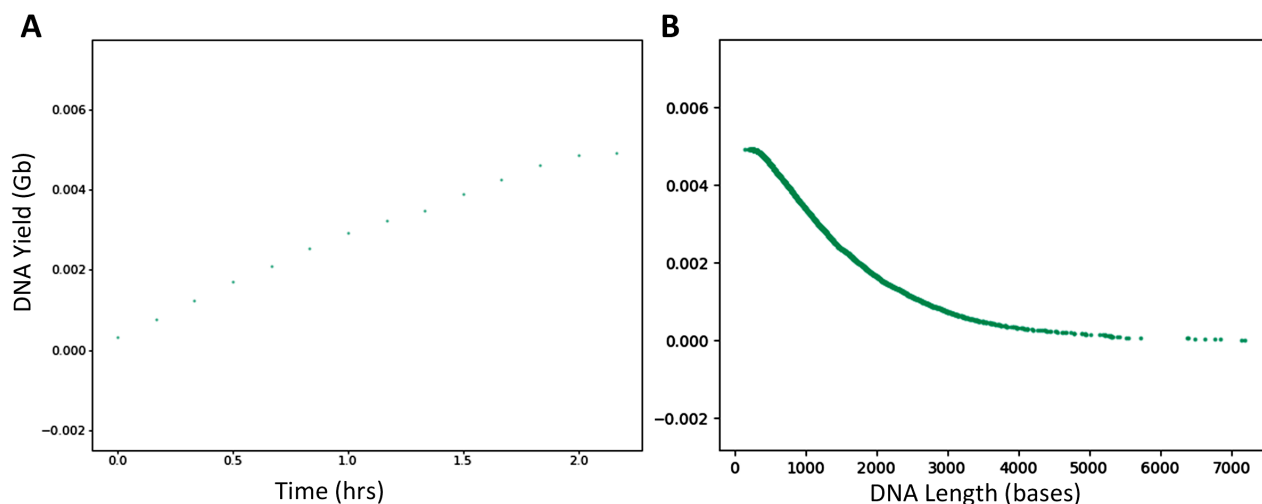
Figure 6. **First Library Preparation Cumulative Yield by Time and Read Length**

base call simultaneously. Raw reads were basecalled offline using ONT Albacore V2.01.

Base calling the data showed that 1.39 gigabases were read and identified (Figure 7A). These bases were spread over 212,000 reads, with lengths as high as 57,000 bases (Figure 7B). While better manual loading and sequencing results have been acquired, this data set is acceptable considering that the protocol has yet to be optimized for the automated subsystem.

6. DISCUSSION

These results show successful tests with all three automated subsystems, which significantly advances the individual processes of the future SETG instrument. Importantly, however, lessons learned and failure modes explored via the testbed will inform future design iterations moving forward. Throughout development, concerns and challenges that were addressed provided insight into potential hurdles in system advancement.

The extraction subsystem faced several challenges with clogging and valve wear by moving Martian simulant through the fluid pathway. This is one of several motivating factors in designing new extraction protocols that limit soil movement. DNA loss in fluid lines also led to an exploration of how to reduce nucleic acid binding to tubing, as well as a requirement to minimize fluid line length. New extraction protocols are being designed by the SETG team with these requirements in mind, which will address problems in future designs.

Loading/sequencing experienced several failures in fluid movement due to insufficient sealing between the interface and Flow Cell, leading to an exploration of interface port geometry and clamping mechanisms. The generation of air bubbles during fluid movement confirmed the requirement for bubble management, currently implemented as a microfluidic bubble trap.

Splattering and liquid retention in library preparation reduced nucleic acid yield due to small reagent volumes, leading to the development of low retention delivery reservoirs and the development of low-splatter general reservoirs. Exploring fluid delivery failure modes also indicated the need for precise

pressure control, leading to the development of electrolytic pump sets derived from a CBIO design. Being small, light, and atmosphere-independent, they offer ways to analyze how pressure affects fluid delivery at microliter scale.

These explorations highlight some of the advantages of the testbed: modularity and independence. There are, however, limitations to the system. The testbed layout is optimized for rapid modification and access to fluid lines and other elements, not for space. Thus, the testbed currently occupies more space than a final instrument would. Cleaning runs are required after a test due to the non-disposable nature of its components, functionally extending the system downtime between runs. Fluid movement in two of the three subsystems is operated via syringes, which is not immediately translatable to a space-based system. The testbed also has no real mechanisms for reagent storage, isolation, or stabilization. As such, this system is not representative of a complete SETG prototype; the testbed is a system to analyze potential failures and to test potential solutions for both SETG and other instruments.

Next Steps

While the current state of the automated testbed represents significant technical advances, there still exist open challenges to address. Desalting has not been implemented into the extraction system, which is important for processing environmental samples and removing elements that inhibit extraction, and heating the OL device has not been integrated, which improves elution of DNA off of the OL beads. The library preparation subsystem's pressure delivery unit has difficulty in managing small volumes of fluid, which motivates the ongoing move to electrolytic pumps for fluid movement. Finally, loading/sequencing is highly dependent on its air bubble trap for loading, and would benefit from a smaller, more compact bubble management solution.

In addition to individual subsystem challenges, integration will be required in order to achieve an end-to-end automated system. While each individual subsystem is functional, testing is required to ensure that the interfaces available connect smoothly. Integration testing is currently ongoing for library preparation into loading, and will continue afterwards as extraction is integrated as well.
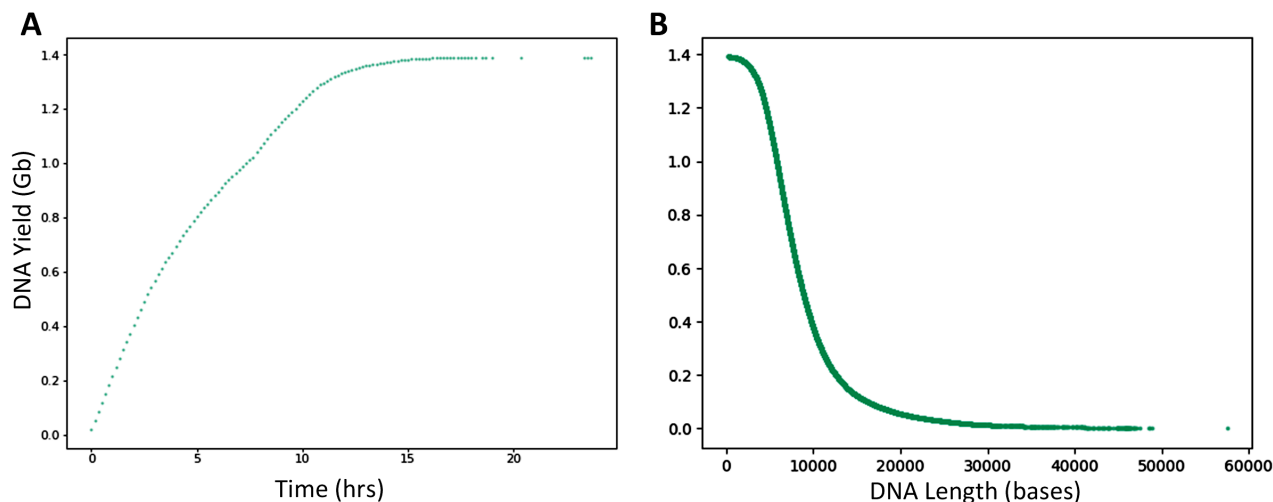
Figure 7. **First Loading Test Cumulative Yield by Time and Read Length**

Finally, it is important to consider the next stage of development after the testbed. An automated benchtop testbed is different from a prototype that can address the challenges of an analog environment. Some components, such as the OL device and the MinION sequencer, have been tested in microgravity and by our own team at Mars atmospheric pressure, but other components have not been tested under operational conditions [15, 16]. The next stage of development will be to use the SETG's project's capabilities for low pressure and low temperature testing to validate each subsystem separately and together. This will require design iterations beyond what is described in this paper and will necessitate a move away from off-the-shelf parts towards custom-designed and built hardware.

7. Conclusions

This paper describes the development of the SETG automated testbed as well as its usage as a vehicle to explore potential future failure modes. The subsystems that comprise the testbed have not yet been fully optimized, but are informing current design iterations. A plan for integration is also laid out and in progress. These steps are essential for SETG instrument development and the lessons learned will assist in creating a more accurate and robust instrument.

Through automating, prototyping, and testing, however, the testbed both furthers the development of the SETG instrument and provides a potential avenue for other instruments to explore development challenges. Outside of the SETG project, other instruments can benefit from a modular benchtop testbed. Components can be rapidly implemented, protocols can be modified immediately, and trade spaces can be fully explored without requiring complete prototype redesigns. Automated testbeds also provide benefits for instruments using biological and chemical components, such as protocol consistency. Continuing optimization of the testbed will also inform instrument requirements for reagent storage and fluid movement.

Future steps beyond integration will include development of an integrated singular instrument that can operate at Mars temperature and pressure. The final instrument will be a powerful tool in the search for extant life on other worlds, and will be more robust due to extensive system development via the SETG testbed. By exploring failure modes using automated benchtop systems, researchers can continue to ask larger, more complex questions about the universe while validating the functionality of their instruments earlier and more efficiently.

References

[1] M. Neveu, L. E. Hays, M. A. Voytek, M. H. New, and M. D. Schulte, "The ladder of life detection," *Astrobiology*, vol. 0, no. 0, p. null, 0, pMID: 29862836. [Online]. Available: https://doi.org/10.1089/ast.2017.1773

[2] M. P. Callahan, K. E. Smith, H. J. n. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House, and J. P. Dworkin, "Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 34, pp. 13 995–13 998, Aug. 2011.

[3] Z. Martins, O. Botta, M. L. Fogel, M. A. Sephton, D. P. Glavin, J. S. Watson, J. P. Dworkin, A. W. Schwartz, and P. Ehrenfreund, "Extraterrestrial nucleobases in the Murchison meteorite," *Earth and Planetary Science Letters*, vol. 270, no. 1, pp. 130–136, 2008.

[4] M. H. Engel and S. A. Macko, "Isotopic evidence for extraterrestrial non- racemic amino acids in the Murchison meteorite," *Nature*, vol. 389, p. 265, Sep. 1997. [Online]. Available: http://dx.doi.org/10.1038/38460 http://10.0.4.14/38460

[5] G. Cooper, C. Reed, D. Nguyen, M. Carter, and Y. Wang, "Detection and formation scenario of citric acid, pyruvic acid, and other possible metabolism precursors in carbonaceous meteorites." *Proceedings of the*

National Academy of Sciences of the United States of America, vol. 108, no. 34, pp. 14 015–14 020, Aug. 2011.

[6] P. Schmitt-Kopplin, Z. Gabelica, R. D. Gougeon, A. Fekete, B. Kanawati, M. Harir, I. Gebefuegi, G. Eckel, and N. Hertkorn, "High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 7, pp. 2763–2768, feb 2010.

[7] J. P. Grotzinger, D. Y. Sumner, L. C. Kah, K. Stack, S. Gupta, L. Edgar, D. Rubin, K. Lewis, J. Schieber, N. Mangold, R. Milliken, P. G. Conrad, D. Des-Marais, J. Farmer, K. Siebach, F. Calef, J. Hurowitz, S. M. McLennan, D. Ming, D. Vaniman, J. Crisp, A. Vasavada, K. S. Edgett, M. Malin, D. Blake, R. Gellert, P. Mahaffy, R. C. Wiens, S. Maurice, J. A. Grant, S. Wilson, R. C. Anderson, L. Beegle, R. Arvidson, B. Hallet, R. S. Sletten, M. Rice, J. Bell, J. Griffes, B. Ehlmann, R. B. Anderson, T. F. Bristow, W. E. Dietrich, G. Dromart, J. Eigenbrode, A. Fraeman, C. Hardgrove, K. Herkenhoff, L. Jandura, G. Kocurek, S. Lee, L. A. Leshin, R. Leveille, D. Limonadi, J. Maki, S. McCloskey, M. Meyer, M. Minitti, H. Newsom, D. Oehler, A. Okon, M. Palucis, T. Parker, S. Rowland, M. Schmidt, S. Squyres, A. Steele, E. Stolper, R. Summons, A. Treiman, R. Williams, A. Yingst, and M. S. Team, "A habitable fluvio-lacustrine environment at yellowknife bay, gale crater, mars," *Science*, vol. 343, no. 6169, 2014.

[8] B. J. Gladman, J. A. Burns, M. Duncan, P. Lee, and H. F. Levison, "The Exchange of Impact Ejecta Between Terrestrial Planets," *Science*, vol. 271, no. 5254, pp. 1387 LP – 1392, Mar. 1996. [Online]. Available: http://science.sciencemag.org/content/271/5254/1387

[9] C. R. Woese, O. Kandler, and M. L. Wheelis, "Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya." *Proc Natl Acad Sci U S A*, vol. 87, no. 12, pp. 4576–4579, Jun 1990, 2112744[pmid]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC54159/

[10] C. Carr, A. Mojarro, J. Tani, S. Bhattaru, M. Zuber, R. Doebler, M. Brown, K. Herrington, R. Talbot, C. Fuller, M. Finney, G. Church, and G. Ruvkun, "Advancing the search for extra-terrestrial genomes," in *IEEE Aerospace Conference Proceedings*, vol. 2016-June, 2016.

[11] C. Carr, A. Mojarro, J. Hachey, K. Saboda, J. Tani, S. Bhattaru, A. Smith, A. Pontefract, M. Zuber, R. Doebler, M. Brown, K. Herrington, R. Talbot, V. Nguyen, R. Bailey, T. Ferguson, M. Finney, G. Church, and G. Ruvkun, "Towards in situ sequencing for life detection," in *IEEE Aerospace Conference Proceedings*, 2017.

[12] M. Jain, I. T. Fiddes, K. H. Miga, H. E. Olsen, B. Paten, and M. Akeson, "Improved data analysis for the minion nanopore sequencer," *Nature methods*, vol. 12, no. 4, p. 351, 2015.

[13] A. Mojarro, G. Ruvkun, M. T. Zuber, and C. E. Carr, "Nucleic Acid Extraction from Synthetic Mars Analog Soils for in situ Life Detection," *Astrobiology*, vol. 17, no. 8, pp. 747–760, jul 2017. [Online]. Available: https://doi.org/10.1089/ast.2016.1535

[14] J. Quick, "One-Pot Ligation Protocol for Oxford Nanopore Libraries," 2018.

[15] S. L. Castro-Wallace, C. Y. Chiu, K. K. John, S. E. Stahl, K. H. Rubins, A. B. R. McIntyre, J. P. Dworkin, M. L. Lupisella, D. J. Smith, D. J. Botkin, T. A. Stephenson, S. Juul, D. J. Turner, F. Izquierdo, S. Federman, D. Stryke, S. Somasekar, N. Alexander, G. Yu, C. E. Mason, and A. S. Burton, "Nanopore DNA Sequencing and Genome Assembly on the International Space Station," *Scientific Reports*, vol. 7, no. 1, p. 18022, 2017. [Online]. Available: https://doi.org/10.1038/s41598-017-18364-0

[16] M. Parra, J. Jung, T. D. Boone, L. Tran, E. A. Blaber, M. Brown, M. Chin, T. Chinn, J. Cohen, R. Doebler *et al.*, "Microgravity validation of a novel system for rna isolation and multiplex quantitative real time pcr analysis of gene expression on the international space station," *PloS one*, vol. 12, no. 9, p. e0183480, 2017.

BIOGRAPHY

Srinivasa Aditya Bhattaru received his B.S. in Mechanical Engineering from the California Institute of Technology in 2015 and his Masters in Aero/Astro from MIT in 2018. He is a Graduate Research Assistant and Doctoral Candidate in the Aero/Astro department, and is interested in evolving technologies for planetary science on Mars.

Jacopo Tani received the Laurea Specialistica degree (with honors) in Space Engineering from La Sapienza University in 2009, and the Ph.D. degree in Aeronautical Engineering from Rensselaer Polytechnic Institute in 2015. He is currently an Oberassistant at ETH Zrich and a Registered Professional Engineer in Rome, Italy.

Kendall Saboda received a B.S. (with honors) in chemical engineering from Montana State University in 2016. She is currently a research assistant at MIT and previously worked in a similar role at the Center for Bio-Inspired Nanomaterials as well as in the Plant Pathology department at Montana State University.
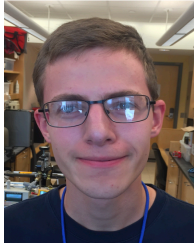
Jonathan Borowsky is a B.A. candidate majoring in biology and physics at Washington University in Saint Louis. He has worked on micro-fluidic design for the SETG instrument and library preparation protocols to expand the applications of nanopore sequencing. He is broadly interested in the intersection of biology and physics.

Maria T. Zuber received the B.A. in Astrophysics from the University of Pennsylvania and Sc.M. and Ph.D. in Geophysics from Brown University. She is the E. A. Griswold Professor of Geophysics in the MIT Department of Earth, Atmospheric and Planetary Sciences, and was previously on the faculty at Johns Hopkins University and served as a Research Scientist at the NASA Goddard Space Flight Center in Maryland. Dr. Zuber has participated in multiple NASA planetary missions aimed at mapping the Moon, Mars, Mercury, and several asteroids, and was Principal Investigator for the Gravity Recovery and Interior Laboratory (GRAIL) mission.

Gary Ruvkun received the A.B. in biophysics from University of California at Berkeley in 1973, and the Ph.D. in biophysics from Harvard University in 1982. He is Professor of Genetics in the Massachusetts General Hospital and at Harvard Medical School, and an associate member of the Broad Institute. Dr. Ruvkun is a member of the National Academy of Sciences and the National Research Council Committee on Planetary Science and Astrobiology. Dr. Ruvkun originated the SETG project and has long been involved in studies of RNA biology, microbial evolution and diversity.

Christopher E. Carr received his B.S. degree in Aero/Astro and Electrical Engineering in 1999, his Masters degree in Aero/Astro in 2001, and his Sc.D. degree in Medical Physics in 2005, all from MIT. He is a Research Scientist at MIT in the Department of Earth, Atmospheric and Planetary Sciences and a Research Fellow at the Massachusetts General Hospital in the Department of Molecular Biology. He is broadly interested in searching for and expanding the presence of life beyond Earth.