NUCLEIC ACID EXTRACTION AND SEQUENCING FROM LOW-BIOMASS SYNTHETIC MARS ANALOG SOILS. A. Mojarro[1], J. Hachey[2], R. Bailey[3], M. Brown[3], R. Doebler[3], G. Ruvkun[2], M. T. Zuber[1], and C. E. Carr[1,2,*] [1]Department of Earth, Atmospheric and Planetary Science, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA, 02139; [2]Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA, 02114; [3]Claremont Biosolutions, 1182 Monte Vista Ave, Upland, CA, 91786
[*]E-mail: chrisc@mit.edu

Introduction: Life on Mars, if it exists, may share a common ancestry with life on Earth due to the estimated billion tons of rock transferred between Mars and the Earth during the late heavy bombardment period [1]. Therefore, biological informational polymers (e.g., deoxyribonucleic acid - DNA, and ribonucleic acid - RNA) have the potential to provide unambiguous evidence of any ancestrally-related life on Mars and discern any forward contamination [2].

We are developing an integrated nucleic acid extraction and sequencing instrument, the Search for Extra-Terrestrial Genomes (SETG), for life detection on Mars and Ocean Worlds [3-7]. Our goal for technology readiness level (TRL) 6 is to achieve a sensitivity of one part per billion. This corresponds to 10^4 *Bacillus subtilis* spores (~4 megabase genome, selected to represent a "worst case" extraction scenario) in our nominal 50 mg soil sample size (2×10^5 spores/g). We have defined detection as sequencing of 10^6 bases. Given the current state of the art in single molecule sequencing, including post-extraction losses, this goal requires at least a 5% extraction yield (~2 pg DNA) [3].

On Earth, metagenomic analysis of DNA extracted from environmental samples can be difficult [8,9]. Complex matrices (e.g., oxidized soils, sea water, etc.), low cell concentrations [10], and tough-to-lyse organisms pose numerous challenges for acquiring DNA. Challenges mostly derive from a combination of competitive adsorption of DNA to soils [11] , DNA destruction by soils [9], and/or hardy cells resistant to lysis.

Here we present DNA extraction results from Mars-relevant cell concentrations, e.g. similar to those observed in the Atacama Desert [12], in synthetic Mars analog soil [13] using a custom extraction cartridge (Figure 1a) developed by Claremont BioSolutions, LLC (CBIO). Furthermore, we characterize the sequencing efficiency of the current generation of Oxford Nanopore Technologies (ONT) MinION Mk-1b sequencers and R9.4 flowcells for low-input sequencing. We employ a "pore maintainer" (*Enterobacteria phage*, Lambda DNA) to preserve nanopore viability and sequence 0.2 ng of *B. subtilis* DNA (equivalent to 5% yield from 10^6 *B. subtilis* spores) which would otherwise produce insufficient high-quality detections.

Methods: *Extraction:* The extraction cartridges (Figure 1a) were based on CBIO's OmniLyse® mechanical cell disruption system. Coupled with binding and elution buffers, we conducted rapid (5 minutes) solid-phase nucleic acid extraction from tough-to-lyse organisms [14]. Work by Mojarro *et al.* on synthetic Mars analog soils and the commercial OmniLyse® systems has led to the development of the modified extraction protocols employed here (Figure 1b) [15,16].

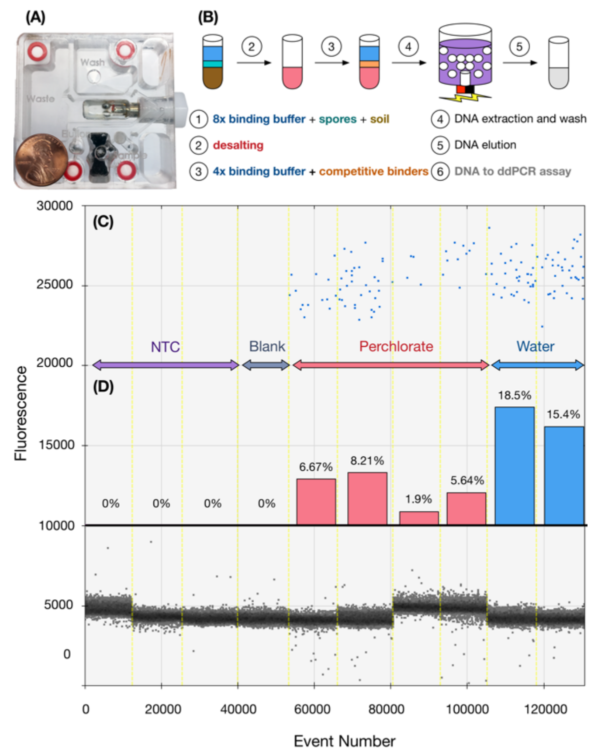

Figure 1: (a) CBIO extraction cartridge (b) Extraction protocol (c) Raw ddPCR data, events with fluorescence greater than 20,000 represent one spore detection (d) DNA yield results, 1.56×10^4 *B. subtilis* spores in 50 mg of soil and water. NTC and Blank measure reagent and hardware contamination.

Approximately 1.56×10^4 *B. subtilis* ATCC 6633 spores (SBS-08, NAMSA) and 50 mg of a "perchlorate" soil, an analog for the Phoenix lander site [13] was homogenized in 800 µL of 8x CBIO binding buffer (n=4) followed by desalting in an Amicon® Ultra-0.5 column (Z740183). The soil/spore mixture was then re-suspended in 800 µL of 4x CBIO binding buffer and 4 µg of random hexamer primers (C1181, Promega), homogenized, and processed in our extraction cartridge resulting in 200 µL of CBIO elution buffer. Moreover, a blank extraction (n=1) and ddPCR water control (n=3) were used to identify any sources of contamination.

Extracted *B. subtilis* DNA was quantified using droplet digital polymerase chain reaction (ddPCR) with single copy primers targeting *B. subtilis* (SpaC gene, forward: TGA GGA AGG ATGG GAC GAC A, reverse: AAC AGA TTG CTG CCA GTC CA). ddPCR partitions a sample into ~20,000 water-oil emulsion droplets to create a massively parallel number of PCR reactions. Droplets containing an amplified product can be measured as highly fluorescent due to presence of a DNA-specific fluorescent dye. When a solution is adequately dilute, such as in this case, each droplet will contain either zero or one copy of a gene sequence (according to the Poisson distribution). Since spores of *B. subtilis* contain only one copy of their genome, one fluorescent droplet represents one spore (Figure 1c). Extraction yield was calculated as extracted DNA / input DNA.

Sequencing: The MinION sequencer and flowcell performs strand sequencing by monitoring changes in an ionic current caused by the translocation of a polymer through a nanopore. Before DNA can be sequenced, it must undergo a "library preparation" step where it is modified into a readable format for sequencing. We prepared 0.2 ng of *B. subtilis* ATCC 6633 spore DNA within 250 ng of Lambda DNA using a transposase-based kit (SQK-RAD001), loaded half of the library (~125 ng), and sequenced on an Intel® Compute Stick (STK2MV64CC) for 7.1 hours. The resulting raw detections were basecalled (convert raw current measurements into base pairs) on ONT's Metrichor cloud service. Additional analysis was conducted using PoreTools, Burrows-Wheeler Aligner (BWA), MASH, and the Basic Local Alignment Search Tool (BLAST).

Results and Discussion: Extraction yields of *B. subtilis* spore DNA in water (Figure 1d) were consistent with prior work on higher concentrations (~15% yield at 10^8 spores) [15]. Low extraction yields from spores are likely due to small acid-soluble proteins that bind DNA [15]. In contrast, DNA yields from vegetative *Escherichia coli* in water are around 80-90% [14]. All extractions from the perchlorate soil satisfied our 5% requirement except one sample (1.9%) for which a pump malfunctioned at the elution step.

Our sequencing metrics indicate that current nanopore sequencing is inefficient. From the theoretically available library (Table 1), only 8.7 x 10^{-5} % was sequenced. The causes of this detection remain to be explored; we speculate library may be lost due to adsorption to flow cell channels or other surfaces, the translocation-regulating motor protein may me knocked off of library molecules, or some library molecules floating in the headspace may simply never reach the nanopores. Although the theoretical sensitivity of single-molecule nanopore sequencing is astounding, it currently requires appreciable amounts of input DNA to compensate for these post-extraction losses.

Library to Sequening	**Mass (ng)**	**Estimated Bases**
Enterobacteria phage	125	2.47E+14
Bacillus subtilis ATCC 6633	0.10	1.97E+11
Sequencing Data	**Measured Reads**	**Measured Bases**
Total Raw Data	49,076	2.32E+08
High Quality Data	31,654	2.02E+08
Identified Species (mismatch)	**Mapped Reads**	**Bases**
Enterobacteria phage (12.7%)	31,624	2.02E+08
*Bacillus subtilis subsp. Spizizenii str. W23** (15.1%)	30	1.07E+05

Table 1: Sequencing metrics. *Closest genome reference for *B. subtilis* ATCC 6633, 99.35586 symmetric identity.

We mapped all high-quality reads (Table 1) (Metrichor pass data) to either a *B. subtilis* and Lambda reference genomes using the BWA analysis package. Based on library stoichiometry, we expected 1 *B. subtilis* detection per 1251 Lambda detections. BWA mapping revealed 1:1886 bases and 1:1055 reads (better performance likely due to shearing). We then confirmed positive identification of *B. subtilis* using MASH.

Conclusions & Summary: In this study, we have validated our extraction protocols to Mars-relevant cell concentrations and characterized sequencing efficiency. Our results suggest employing a pore-maintainer can improve low-input sequencing, however, future studies will focus on improving sequencing efficiency, required to achieve our future sensitivity goals. In addition, our sequencing results simulate the potential for a genomics-based instrument to detect low-levels of contamination, relevant to planetary protection.

Acknowledgements: Extractions were by AM, library prep and sequencing by JH. This work was supported by NASA MatISSE award NNX15AF85G.

References: [1] Gladman B.J. et al. (1996) *Science*, *271*, 1387. [2] Isenbarger T.A. et al. (2008) *Orig. Life Evol. Biosph.*, *38*, 517–533. [3] Carr C.E. et al. (2016) *2016 IEEE Aerospace Conf.*, 1–15. [4] Carr C.E. et al. (2013) *2013 IEEE Aerospace Conf.*, 1–12. [5] Carr C.E. et al. (2013) *Astrobiology*, *13*, 68–78. [6] Carr C.E. et al. (2013) *Astrobiology*, *13*, 560–569. [7] Lui C. et al. (2011) *2011 IEEE Aerospace Conf.*, 1–12. [8] Lever M.A. et al. (2015) *Front. Microbio.l*, *6*, 1–25. [9] Herrera A. and Cockell C.S. (2007) *J. of Microbiol. Methods*, *70*, 1–12. [10] Barton H.A. et al. (2006) *J. of Microbiol. Methods*, *66*, 21–31. [11] Takada-Hoshino Y. and Matsumoto N. (2004) *Microb. and Environ.*, *19*, 13–19. [12] Azua-Bustos A. et al. (2012) *FEBS Lett.*, *586*, 2939–2945. [13] Schuerger A.C. et al. (2012) *Planet. and Space Sci.*, *72*, 91–101. [14] ClaremonBio (2011) 1–4. [15] Mojarro A. et al. (2016) *LPS XLVII*, Abstract # 1643. [16] Mojarro A. et al. (2016) *3rd IPM* Abstract # 4095.