Life Sciences in Space Research 18 (2018) 80–86

Contents lists available at ScienceDirect

Life Sciences in Space Research

journal homepage: www.elsevier.com/locate/lssr

Sequencing nothing: Exploring failure modes of nanopore sensing and implications for life detection

Alexandra Pontefract[a,b,*], Julie Hachey[c], Maria T. Zuber[b], Gary Ruvkun[a], Christopher E. Carr[a,b]

[a] Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, United States
[b] Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, United States
[c] ReadCoor, 840 Memorial Dr., Cambridge MA 02139, United States

ARTICLE INFO

Keywords:
Nanopore sequencing
Life detection
Planetary protection
Nucleic acids

ABSTRACT

The detection of extant life is a major focus of many planned future planetary missions, a current challenge of which is the ability to target biomarkers capable of providing unambiguous evidence of life. DNA sequencing is increasingly recognized as a powerful tool for life detection for planetary exploration missions; beyond use of sequence information to determine the origins of the sample (e.g., extant life or forward contamination), recent advances in the field have enabled interrogation of single molecules, with or without amplification. The focus of this work is on failure modes, specifically the issues encountered when there is no-to-low input DNA into a sequencing device, and the potential for the generation of sequencing artifacts that could be interpreted as a false positive. Using Oxford Nanopore Technologies (ONT) MinION, we assess whether single molecule sequencing, involving no amplification, generates noise signals that could be misinterpreted in the context of a planetary exploration mission, and also whether the ability of the instrument to handle these types of situations could make it feasible for clean room monitoring. Utilizing quality score filtering techniques in place at the time of this experiment, runs containing only initial flowcell chemistry and/or library reagents generated 5 passing reads out of a total of 3568 measured reads, and contained estimated sequences with low complexity that did not map to the NCBI database. The noise characteristics in all instances suggest that quality thresholds were appropriately chosen by ONT: new chemistry and basecalling workflows have shown further suppression of noise sources, which completely mitigate the generation of spurious reads.

1. Introduction

One of the current challenges of life detection beyond Earth is our ability to target biomarkers that can provide unambiguous evidence of life. The sequencing of informational polymers (e.g. DNA) has been suggested as a tool for unambiguous life detection (e.g. Carr et al., 2017; Fairén et al., 2017; John et al., 2016) owing to several factors: the high information content of a DNA sequence, as compared with its biochemical detection, the ubiquitous generation of prebiotic molecules such as nucleic acids and sugars within stellar nebula (Ehrenfreund et al., 2002), the probability that life would develop some type of hereditary informational storage system (Benner, 2010), and the potential for the transference of material (e.g. microorganisms) between the inner terrestrial planets early on in the evolution of our solar system – i.e. lithopanspermia (Crick and Orgel, 1973; Horneck, 2006), meaning that putative Mars life could have a shared ancestry with life on Earth. Though instruments common to planetary exploration payloads, such as gas chromatograph-mass spectrometers (GC–MS), are capable of detecting these informational molecules, nucleic acid sequencing is advantageous as it offers a high level of unambiguity, being able to distinguish between abiotic and biotic nucleic acid polymers, moreover allowing for the determination of the origins of the sample – whether it be extant life or the detection of forward contamination – and is thus capable of detecting false positives; sequencing could also address the question of shared ancestry vs. a second origin of life. Life detection instruments targeting DNA and/or related polymers are currently in active development (Carr et al., 2017, 2016), however many challenges remain, i.e., the implementation of biological reagents, robust extraction method(s), the ability to sequence non-standard bases or polymers, as well as data analysis and reduction pipelines that would allow us to cope with data transmission limitations.

Beyond the challenges mentioned above, we must also address the potential issues that might arise from failing to detect life – i.e. sequencing nothing, which for traditional sequencing platforms is not

* Corresponding author.
E-mail address: apontefr@mit.edu (A. Pontefract).

https://doi.org/10.1016/j.lssr.2018.05.004
Received 2 March 2018; Received in revised form 11 May 2018; Accepted 16 May 2018

done. Modern next generation sequencing (NGS) relies upon the amplification of DNA, either during library preparation or during the sequencing process, which, especially in the absence of input DNA (which would simply not be sequenced), produces artifacts (random DNA) and possible sequencer malfunctions. These issues become problematic in instances where either very low amounts of DNA must be quantified, or when the absence of any DNA must be accurately determined, i.e. in clean room surveys and in life detection missions (La Duc et al., 2009, 2014). Current clean room practices for the quantification of the presence of viable bacteria include either swabbing and cultivation (e.g. (Venkateswaran et al., 2016), or the use of propidium monoazide (PMA), followed by PCR amplification and sequencing using NGS technology (Sielaff et al., 2017; Vaishampayan et al., 2013), which require significant sampling to acquire enough material for downstream analyses (La Duc et al., 2014).

With the advent of the Oxford Nanopore Technologies (ONT) MinION, where each detected single molecule of DNA is fed through a biological nanopore causing a disruption of pore currents that is nucleobase-specific, we have an instrument that is capable of analyzing low-to-no amounts of input DNA (i.e. Mojarro et al., 2018) without the use of amplification, thus supporting experiments in situations where the amount of input DNA is unknown and/or below detection limits, i.e. planetary exploration mission. As the MinION is a new (and continually developing) technology however, it is currently unclear what failure modes might look like in this device, and how that might impact results either for use in life detection missions, or for determinations of the efficacy of clean-room sterilization procedures. Here we assess failure modes of the MinION, specifically addressing issues encountered when there is no known DNA input, and whether this situation can lead to an incidence of false-positives due to the generation of sequencing artifacts.

2. Sequencing platform and methodology

When preparing a sample for sequencing on the MinION, extracted double-stranded DNA is prepared to enable reading of a single strand of each library molecule (1D) or, to enable reading of both strands, through incorporation of a hairpin at one end of each library molecule (2D; no longer available). These methods were conducted using library kits that can perform, respectively, either transposase-based or end-blunted ligation (Fig. 1A, B). Before loading the sequencing library, a flowcell, consisting of ~2048 (can vary from 800 to 2048) pores with an embedded sensor array, is primed through the addition of running buffer. Once primed, the library can then be loaded onto the flowcell and sequencing can proceed. DNA strands are fed through the nanopores (Fig. 1C) and, based on the disruption of the ionic current (Fig. 1D), are basecalled using the ONT software Metrichor (now Albacore), which at the time of this experiment had just been upgraded to a recurrent neural network (RNN) (Fig. 1C). A Phred quality score is applied to the raw data, which is a measure of the accuracy of each called base, and is logarithmically related to the probability of an error in basecalling (Ewing et al., 1998). At the time of sequencing, Metrichor applied passing quality (Q) scores of $Q > 6$ for 1D single strand sequencing, and $Q > 9$ for 2D double strand consensus sequencing, the latter of which corresponded to just under 90% per base accuracy: sequences are then divided between pass and fail accordingly based on the Q score corresponding to the average per base accuracy of each read.

Within the flowcell, an ionic current is being generated across the nanopores at all times regardless of whether or not that pore is active, i.e. has a DNA strand running through it. These CsgG E. coli pores are stabilized by the presence of DNA, but when idle, their inherent looped-structure causes them to destabilize and wobble, which can result in the generation of signals detectable above the background noise that may pass the quality score threshold (Clive Brown – ONT Community Board 2017). Experiments have also shown that a specific type of passing "junk" reads frequently originate from pores that overproduce relative to others in the flowcell, indicating that certain pores can be somewhat defective and potentially need be removed from downstream analyses (Mojarro et al., 2018). In order to examine these occurrences in a methodological manner, we conducted experiments on the ONT MinION Mk 1-B with R9 (FLO-MIN104) flowcells. Four cases were explored (Fig. 1E), and in each case the number of 1D and 2D passing and failing reads and bases were recorded, as well as Q-scores, along with a control run using sheared bacteriophage lambda DNA, to showcase what a DNA-containing run looks like with nanopore. Any passing reads were analyzed using NCBI BLASTn (default parameters) and DUST scores were computed using MATLAB: DUST scores are based on the frequency of tri-nucleotide repeats within a sequence, and are scaled from 0 to 100, with low-complexity sequences scoring highly. Passing and failing reads were run on Kaiju (Menzel et al., 2016) for characterization using inexact k-mers at the protein level against the NCBI BLAST nr + euk: Bacteria, Archaea, Viruses, Fungi and microbial eukaryotes database using default parameters. Finally, an additional run (Run 5), re-examining Run 2 was conducted to explore how the newer pore structure, chemistry (R9.4, FLO-MIN105) and base-calling (Albacore v. 2.1.10) may alleviate any issues with false positives seen in the original four runs.

3. Results and discussion

This experiment, conducted across five independent runs (excluding the control), garnered a total of 3568 measured reads, of which 5 reads passed the quality score filter (Table 1). Runs 1 and 2 of the experiment, conducted without the use of library reagents, established a baseline for the instrument. In both runs, no 2D passing reads were detected, whereas 2 passing 1D reads were generated in each situation. In Run 1 the passing 1D reads were distinctive, consisting of a long series of A's and T's, while the Run 2 passing reads were characterized by an over-representation of thymine calls (Fig. 2). Runs 3 and 4 had fully primed flowcells, along with either 2D or 1D library preparation reagents respectively, and in these instances only the 1D prep run generated a passing read. Run 5 sought to replicate the run with the highest spurious read generation (Run 2), only with new chemistry (described above), and tellingly only generated 1 failed read after 5 hours of sequencing. Alternatively, the control run – essentially Run 3 with DNA – showed a short sequencing run of sheared lambda DNA; garnering almost an equal number of passing and failing reads, ~30,000 for each. Passing reads mapped with 99% accuracy to bacteriophage lambda using the NCBI database, and had a DUST score of 1.79.

The generated passing reads from all experimental runs were first analyzed for pore-location using the CarrierSeq algorithm (Mojarro et al., 2018) to determine if they originated from a so-called "bad" pore, and thus could be discounted using these means, but did not reveal any such correlation (Fig. 3). The reads were then run through the NCBI BLASTn database and on Kaiju and in both cases failed to map (i.e. reads were low complexity and thus were filtered out as non-informative), indicating that even though they passed the quality score filters, it was still possible to discern the reads as false positives within the context of this experiment, resulting from wobbling of the CsgG pore (see above), or in the case of Runs 3 and 4, potentially also resulting from backwards missteps of the helicase, referred to here as "chattering" (Caldwell and Spies, 2017). Moreover, the structure of some of the reads, i.e. long A-T runs, or high T composition is consistent with spurious detections (i.e., not actual DNA). The above results would easily be accepted as the detection of a false positive within a

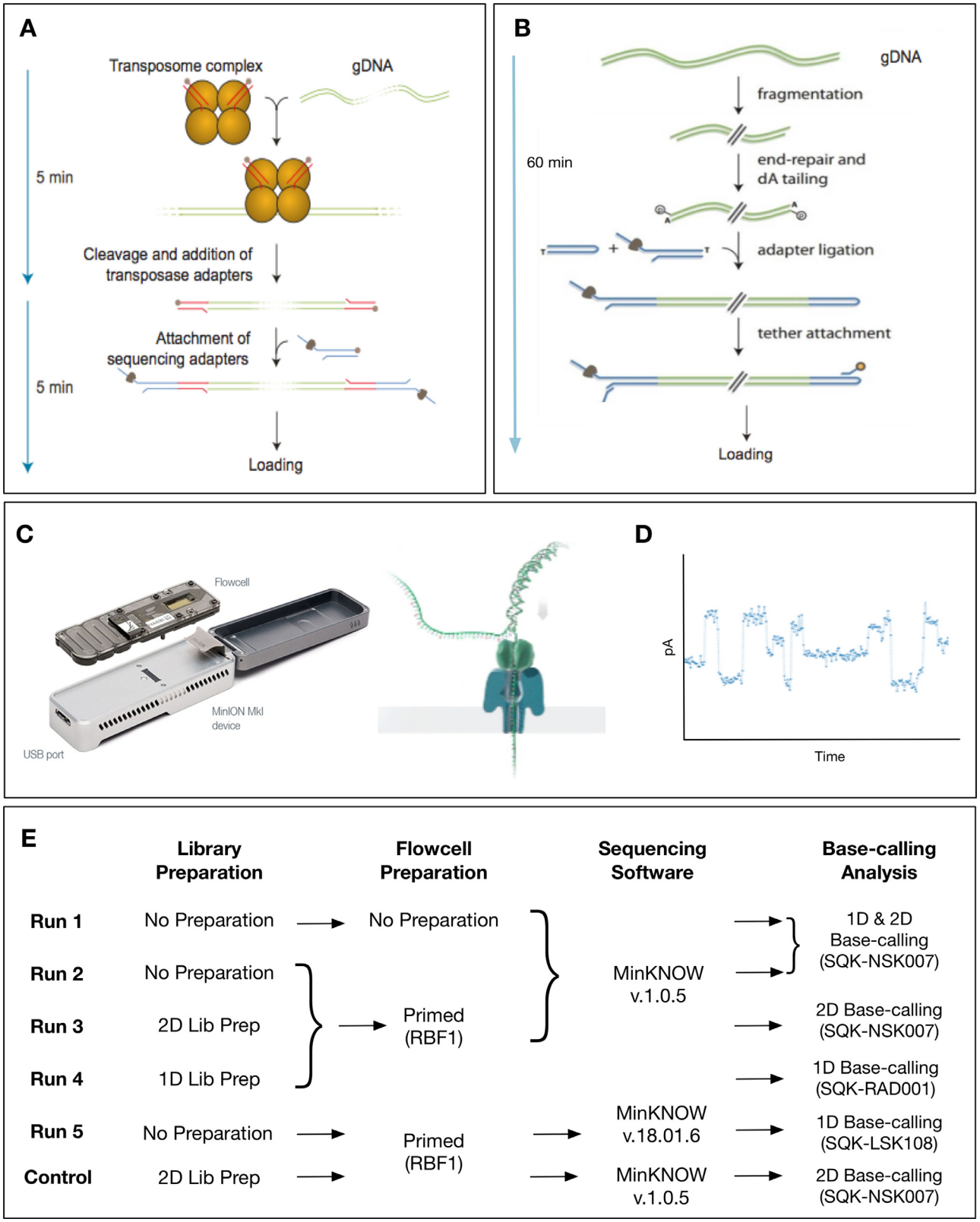

Fig. 1. (A) Schematic showing 1D library prep through the use of a transposome complex. (B) 2D Library preparation methodology in use at the time of the experiment. (C) Prepared DNA library is then loaded into a primed flowcell (left), DNA is fed through a biological nanopore (middle), where the transit of different bases within the pore produces fluctuations in ionic current (D), the readout of which is then base-called (right). (E) Workflow for experiment detailing the library and flowcell preparation utilized. No preparation indicates a state where no library was used, i.e. the case of Run 1 indicates a sequencing run where there was no input into the system and the flowcell was not primed. 2D and 1D library preparation reagents were applied to molecular grade water. Runs 1–4 and the control utilized Metrichor v. 1.107 for basecalling, and Run 5 used Albacore v. 2.1.10. Image credit (A–C): Oxford Nanopore Technologies.

Table 1
Sequencing statistics for each type of sequencing run with no DNA, including statistics for the control run with sheared Bacteriophage lambda DNA.

	Base-Calling Analysis	Measured Reads	Measured Bases	Pass Reads	Pass Bases	Passing Q-Score (avg)	Fail Reads	Fail Bases	Failing Q-Score (avg)
Run 1	1D	812	883,558	2	487	6.7	809	883,071	3.9
	2D	812	596,128	0	0	–	812	596,558	4
Run 2	1D	1608	1,084,494	2	9105	7.3	1606	1,075,389	4.8
	2D	1608	1,246,044	0	0	–	1608	1,246,044	3.9
Run 3	2D	663	2,278,106	0	0	–	663	2,278,106	2.9
Run 4	1D	484	1,295,373	1	192	8.25	483	1,295,181	3.9
Run 5	1D	1	260	0	0	–	1	260	4.9
Control	2D	63,290	266,429,803	32,619	162,325,515	8.3	30,671	104,104,288	5.7

cleanroom scenario, but would potentially be ambiguous if the result was generated as part of a life-detection experiment, i.e. extraction from soil with full library preparation. To be accepted within such a scenario these sequences would likely need to be subjected to a higher quality score filter (e.g. Q > 9) and would need to be highly prevalent in the data, where the increased incidence of low-complexity sequences could then be reminiscent of a eukaryotic-like organism (Toll-Riera et al., 2011). Moreover, this type of situation could be further elucidated by conducting a subsequent sequencing run without ligation, which would then indicate if these low-complexity sequences were due to issues with nanopore chemistry, or actually the result of *in situ* DNA. Rapid changes in the ONT platform – i.e. updates to pore chemistry and software used for basecalling – meant that results generated from Runs 1–4, and the ambiguity resulting from them, might have been largely addressed. Indeed, when Run 2 (which generated the largest number of spurious reads: 1608) was repeated with the new flowcell (R9.4) and basecalling software (Albacore), the run (Run 5) generated one solitary failed read. This read consisted of 260 bases containing an over-representation of thymine, and a DUST score of 10.26, somewhat more complex than a tri-nucleotide repeat, yet still low complexity.

Just as false positives might be generated, we also wanted to explore the potential for the generation of false negatives – sequences that should have passed the quality score filter, but did not. In our specific set of experiments, false negatives could only exist through the presence of contaminating DNA in either Runs 3 or 4, as no input DNA was used. Based on the passing reads, this did not occur, however, understanding what constitutes a failing read would allow us to distinguish between the potential occurrence of false negatives in a planetary context (e.g. highly damaged DNA resulting in a low quality score) and true negatives (spurious reads that were rejected by the system). Failing reads that were below the quality score threshold were run on Kaiju at the protein level, with a minimum match length of 11 and a minimum match score of 75, with 5 allowed mismatches. In general, only a small percentage (< 5%) of the reads mapped. To visualize these positive hits, a Krona plot was created from the 26 mapped failed reads from Run 4 (Fig. 4), generating 6 "identified" species all present at 17% abundance, mapping most closely to species within the Eukaryota and the Actinobacteria – the latter comprised of *Mycobacterium* and *Streptomyces*. This even species distribution reinforced the conclusion that these were "nonsense" reads that belonged in the Fail folder, and did not represent false negatives. The failed read generated in Run 5 was common enough to map to the entirety of the Bacterial kingdom, but did not yield any specific hits either within Kaiju or within the NCBI database, the latter of which failed to generate any analyses.

4. Conclusions

Though this experiment was conducted with in the last two years, rapid developments in the sequencing technology of the MinION have resulted in evolved flowcell chemistry, the removal of 2D sequencing, and an increase in basecalling accuracy from 85% in the R9 chemistry for 1D reads (Jain et al., 2017) to 97% accuracy with the new $1D^2$ analysis using R9.5 chemistry (ONT Community Board communication). Moreover, the basecalling software has been reworked and is now conducted using a program called Albacore (v.2.1.3) which utilizes an updated version of the Recurrent Neural Net; the quality score threshold used for 1D reads is now Q > 7. Oxford Nanopore Technologies has also moved towards a more rigid-pore design, aiding in the mitigation of false signal generation resulting from pore wobbling. The benefits that these adjustments to the pore design and software have afforded are clearly visible in our results from Run 5, where spurious read generation was almost completely absent – a reduction of two orders of magnitude from the original experiments – which speaks to the robustness of this type of technology for the detection of low amounts of DNA.

With proper Q-score filtering, it is clear that single molecule sequencing produces little noise, and even small numbers of single molecule reads mapping to an organism can provide unambiguous detection of that organism (e.g. Mojarro et al., 2018). The noise characteristics in all instances explored in this work suggest that the quality thresholds (Q > 6 for 1D, Q > 9 for 2D) were appropriately chosen by ONT at the time: their improvements to basecalling and the increase of the quality threshold from 6 to 7 has only served to increase the robustness of this technology. This technology should prove attractive to those involved in the perfection of clean room practices, where small amounts of eDNA can be spiked into a known substrate of DNA such as bacteriophage lambda, and thus be sequenced without the need for amplification. This would enable high resolution, quick turnover, monitoring of microbial populations in clean room facilities as well as the ability to monitor changes in microbial populations over time, such as in situations associated with human space exploration (e.g. Castro-Wallace et al., 2017). Importantly, this sequencing technology could represent a viable option for unambiguous detection of nucleic acid-based life, and could also be a valuable tool for planetary protection; allowing for clear interpretations of false-positives resulting from both spurious read generation (the focus of this paper) and potential forward contamination, determined through post-sequencing analyses (i.e. Carr et al., 2016). Currently, this technology does present some limitations due to a reliance on biologically-based reagents and flowcells which would need to be protected during processes such as dry heat sterilization of the instrumentation, as well as exposure to intense radiation, such as would be found in the Jovian system (Carr et al., 2016). However, movement towards lyophilized reagents has provided increased stability over a wider temperature range, and work towards the generation of a solid-state single-molecule sequencer is underway (e.g. Dekker, 2007; Di Ventra and Taniguchi, 2016; Goto et al., 2016; Kawai et al., 2016).

Run 1: Stock Flowcell

>Sequence_1: 472 nucleobases; DUST Score: 32.21
TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATATGTGTAATAATAATAATAATAATAATAATAAT
AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA
TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT
AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA
TAATAATAATAATAATAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAC

>Sequence_2: 10 nucleobases; DUST Score: 7.14
GGGCCAGGCC

Run 2: Primed Flowcell

>Sequence_1: 4554 nucleobases; DUST Score: 33.15
TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATTGTTGTTATTGTTGTTGATGATAT
TGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATATGATG
ATGATGATATGATATTGTGTGTGTGATATTATTGTGTGTGTGTGTGTGTGTGTGTGTGTTGTGTTATTATTATTA
TTGTTATTGTTATTATTATTGTGTGTGTGTGTGTGTGTGTGTTATTATTGTGTTGTGTGTGTGTGTGTGTGT
GTGTGTGGT...GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGC

>Sequence_2: 4551 nucleobases; DUST Score: 25.79
TTTATTGTTATTATTGTTATTGTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT
GTTATTGTTCGTATTATTATTATTATTGTTATTGTTATTATTATTATTATTATTGTTATTATTATTCATTGATGATAT
TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT
ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTGTTA
TTATTATTATTATTATTATTATTATTATTATTATTATTATTGATTATTATTATTATTATTATTATTATTGTTATTATTATT
ATTATTATTATTATTAT...TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTGTTATTA
TTGTTATTATTGTTATTATTATTATTATTATTATTATTGTTATTATTGTTATTATTATTATTATTATTATTATTATT
ATTATTATTATTATTATTGTTATTATTGTTATTATTATTATTATTATTGTTATTATTATTGTTATTATTATTATTAT
TATTGTTATTATTATTATTGTTATTGTTATTATTATTATTATTGTTATTATTATTGTTATTATTATTATTATTGTTGTT
ATTATTGTTATTATTGTTATTGTTATTATTATTATTATTATTATTATTGTTATTG

Run 3: 2D Ligation

>Sequence_1: 192 nucleobases; DUST Score: 31.26
GCATCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA
TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
ATAATAATAATAATAATAATAATAATA

Control: 2D Ligation

>Sequence_1: 4155 nucleobases; DUST Score: 1.79
GAACTTCGTCGGGTTACGTATTGCTAAACGTCAGAAACGAATGCTGCAGCGTCACAACAATCCAACCGC
CACGTCTGCCTCCACCGCGGCCACGAAAGCGTCAGGCCGCCACTTCAACACAGATGCGGTGGCCTCA
AAGAGGCAGCAAAATCAAAACGAACATCATCAGGAAGTGCCGGTCGTGCAGCTTCCTCGCAGCGGCAG
AAAATTCTGCAGGGCGGCAAAACGTCCGAAGACGAATGCCAGGTCATCTGAAACAACAGGAACGGAGC
GCCTCTGCCGCAGACGGGCAAAAACAACGGCGGCGGGGAGTGCGTCAACGGCATCACGAAGGCGAC
AGAGGCTGCGGGAAGTGCGGTATCAGCATCGCAGAGCAAAGTGCGGCACATGCGTGCAAAATTCGGCA
AACGTGCAGAAGATATAGCTTCAGCTGTCGCGCTTGAGGATGCGGACACAACGAAAGGGGATAGTGCA
GCTCAGCAGTGCAACAGCACGTCTGAAACGCTTGCTGCAACGCCAAAAGGCGGTTAAGGTGGTAATGG
ATGAAACGAACAGAAAGCCCACTGGACGGTCGGCACTGACCGGAACGCCAACCTTTCCAACCGCGC...
CTAAAATTGAATTTAGCAATACCAGATTTTGTAAAACAGCCCTCCTCATATAAAAGTATTCGTTCACTTGATA
AGCGTCGTAATTTCTATCTTTCATCATATTCCTCTGAAAAATCTTCCGAGTTTACTAGGCACTGATACATAA
CTCTTTTCCAATAATTGGGAAGTCATTCAAATCTATAATAGGTTTCAGATTTGCTTCAATAAATTCTGACCAG
CTGCTGAAACGTTGCGGTTGAACTATATTTCCTTATAACTTTTAACGTTCTGTTTATGTT

Fig. 2. Passing reads from four different cases, revealing long runs of low-complexity. Sequences in Run 2 have been abbreviated for length. The low DUST score for Run 1, Sequence 2 is the result of sequence length. A subset of Sequence 1 from the Control run was provided for contrast.

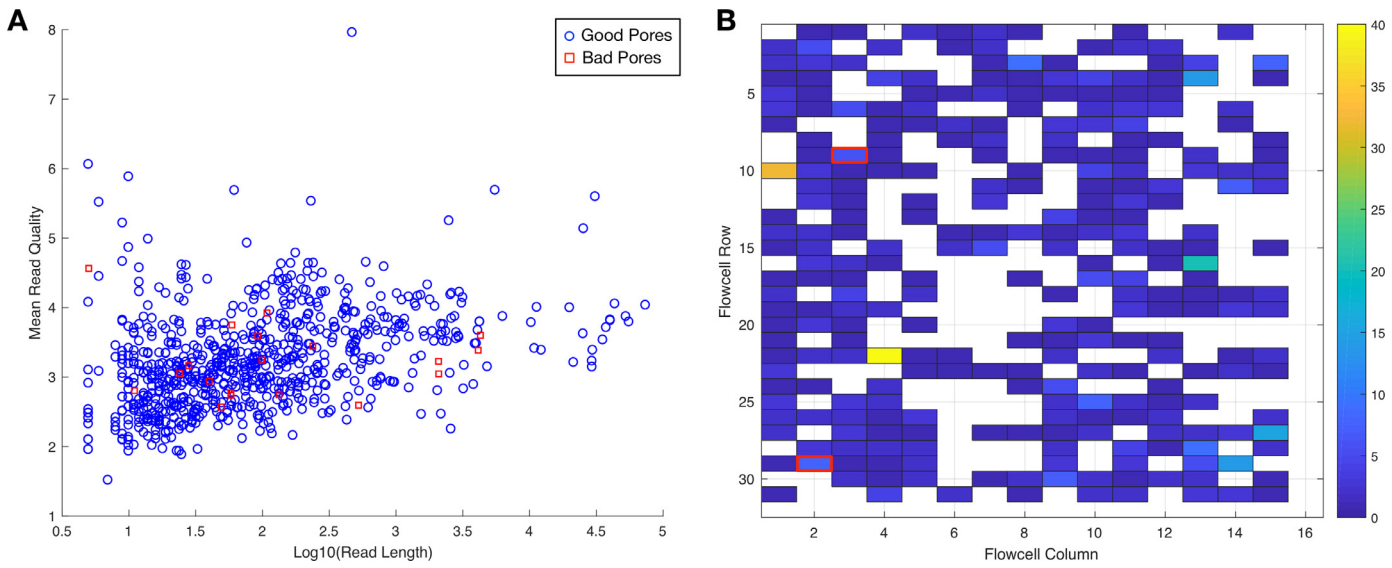

Fig. 3. Pore quality data from Run 1: (A) Plot of read length vs. mean read quality, where circles indicate pores that are producing reads at or near the expected rate, and bad pores (squares) indicate reads from pores producing reads above the expected rate. Note the two data points above Q = 6, which are the two passing reads for this run (B) Histogram of the number of pores per nanopore channel. Colors exceeding 8 correspond to "bad" or overproducing pores. Highlighted channels in red indicate the origin of the two passing reads for this case, identified in (A).

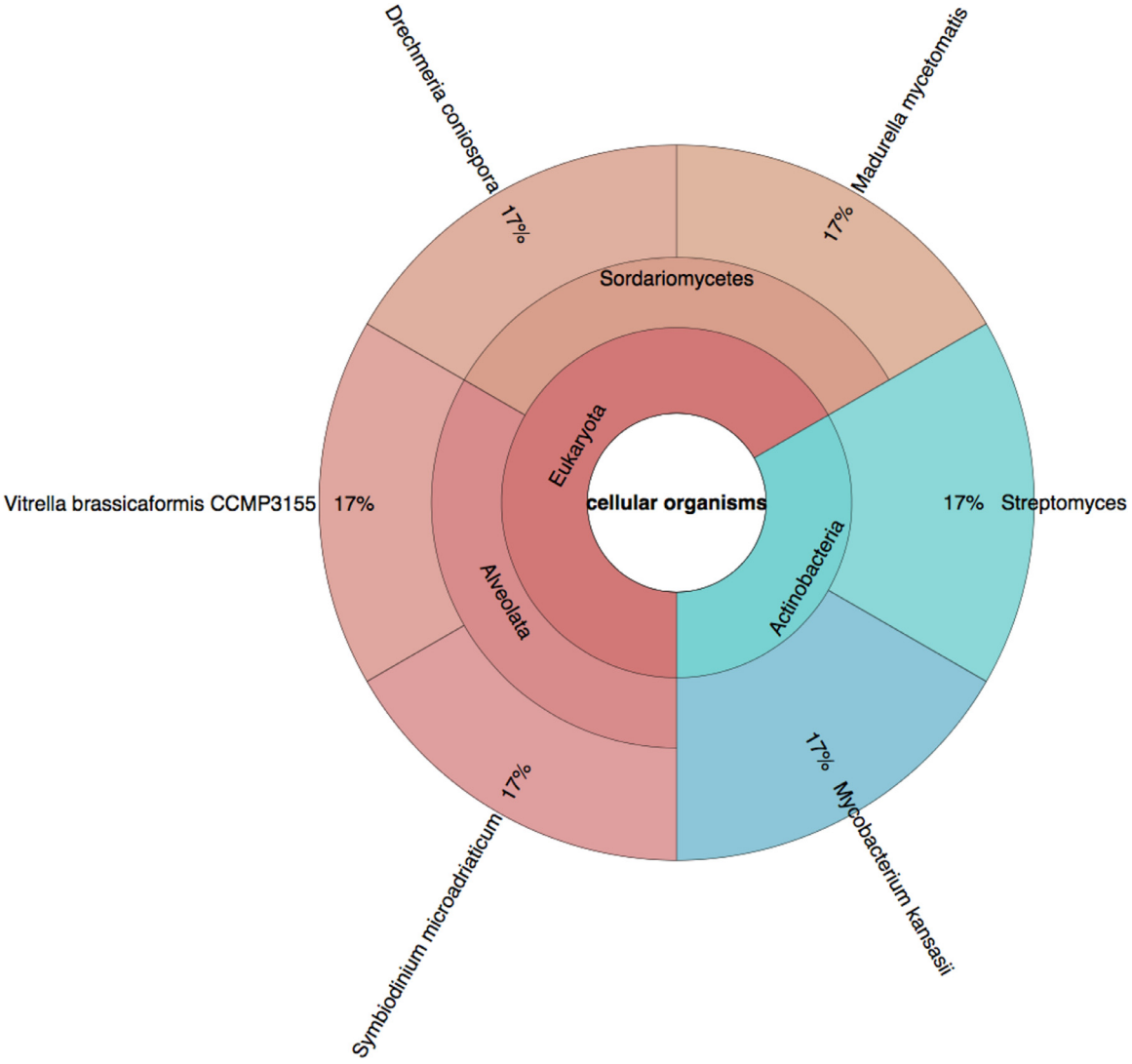

Fig. 4. Krona plot created from Kaiju of the 6 failed reads that mapped from Run 4.

Acknowledgments

This work was supported by NASA MatISSE (NNX15AF85G). Metagenomic data have been added to the NCBI Sequence Read Archive (SRA) under project PRJNA434429 (SRP133105).

Declarations of interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.lssr.2018.05.004.

References

Benner, S.A., 2010. Defining life. Astrobiology 10, 1021–1030.
Caldwell, C.C., Spies, M., 2017. Helicase SPRNTing through the nanopore. In: Proceedings of the National Academy of Sciences. 114. pp. 11809–11811.
Carr, C.E., Mojarro, A., Hachey, J., Saboda, K., Tani, J., Bhattaru, S.A., Smith, A., Pontefract, A., Zuber, M.T., Doebler, R., 2017. Towards in situ sequencing for life detection. In: Aerospace Conference, 2017 IEEE. IEEE.
Carr, C.E., Mojarro, A., Tani, J., Bhattaru, S.A., Zuber, M.T., Doebler, R., Brown, M., Herrington, K., Talbot, R., Fuller, C.W., 2016. Advancing the search for extra-terrestrial genomes. In: Aerospace Conference, 2016 IEEE. IEEE.
Castro-Wallace, S.L., Chiu, C.Y., John, K.K., Stahl, S.E., Rubins, K.H., McIntyre, A.B.R., Dworkin, J.P., Lupisella, M.L., Smith, D.J., Botkin, D.J., 2017. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. Scientific Reports 7, 18022.
Crick, F.H., Orgel, L.E., 1973. Directed panspermia. Icarus 19, 341–346.
Dekker, C., 2007. Solid-state nanopores. Nat. Nanotechnol. 2, 209–215.
Di Ventra, M., Taniguchi, M., 2016. Decoding DNA, RNA and peptides with quantum tunnelling. Nat. Nanotechnol. 11, 117–126.
Ehrenfreund, P., Irvine, W., Becker, L., Blank, J., Brucato, J., Colangeli, L., Derenne, S., Despois, D., Dutrey, A., Fraaije, H., 2002. Astrophysical and astrochemical insights into the origin of life. Rep. Prog. Phys. 65, 1427.
Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. 8, 175–185.
Fairén, A.G., Parro, V., Schulze-Makuch, D., Whyte, L., 2017. Searching for life on Mars before it is too late. Astrobiology 17, 962–970.
Goto, Y., Yanagi, I., Matsui, K., Yokoi, T., Takeda, K.-i., 2016. Integrated solid-state nanopore platform for nanopore fabrication via dielectric breakdown, DNA-speed deceleration and noise reduction. Sci. Rep. 6, 31324.
Horneck, G., 2006. Bacterial spores survive simulated meteorite impact. Biological Processes Associated with Impact Events. Springer, pp. 41–53.
Jain, M., Tyson, J.R., Loose, M., Ip, C.L., Eccles, D.A., O'Grady, J., Malla, S., Leggett, R.M., Wallerman, O., Jansen, H.J., 2017. MinION analysis and reference consortium: phase 2 data release and analysis of R9. 0 chemistry. F1000 Res. 6.
John, K.K., Botkin, D.J., Burton, A.S., Castro-Wallace, S.L., Chaput, J.D., Dworkin, J.P., Lehman, N., Lupisella, M.L., Mason, C.E., Smith, D.J, et al., 2016. The biomolecule sequencer project: nanopore sequencing as a dual-use tool for crew health and astrobiology investigations. In: 47th Lunar and Planetary Science Conference, Lunar and Planetary Institute, Houston, Abstract #2982.
Kawai, T., Taniguchi, M., Ohshiro, T., Oldham, M.F., Nordman, E.S., 2016. Devices, systems and methods for sequencing biomolecules. Google Patents.
La Duc, M.T., Osman, S., Vaishampayan, P., Piceno, Y., Andersen, G., Spry, J., Venkateswaran, K., 2009. Comprehensive census of bacteria in clean rooms by using DNA microarray and cloning methods. Appl. Environ. Microbiol. 75, 6559–6567.
La Duc M.T., Venkateswaran K., and Conley C.A. (2014) A genetic inventory of spacecraft and associated surfaces. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.
Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nature Commun. 7.
Mojarro, A., Hachey, J., Ruvkun, G., Zuber, M.T., Carr, C.E., 2018. CarrierSeq: a sequence analysis workflow for low-input nanopore sequencing. BMC Bioinformatics 19, 108.
Sielaff, A.C., Avila-Herrera, A., Jaing, C., Allen, J.E., Venkateswaran, K., Nicholas, A.B., Singh, N., 2017. Whole metagenome profiles of particulates collected from the International Space Station. Microbiome 5, 81.
Toll-Riera, M., Radó-Trilla, N., Martys, F., Alba, M.M., 2011. Role of low-complexity sequences in the formation of novel protein coding sequences. Mol. Biol. Evol. 29, 883–886.
Vaishampayan, P., Probst, A.J., La Duc, M.T., Bargoma, E., Benardini, J.N., Andersen, G.L., Venkateswaran, K., 2013. New perspectives on viable microbial communities in low-biomass cleanroom environments. The ISME J. 7, 312–324.
Venkateswaran, K., La Duc, M.T., Vaishampayan, P., Spry, J.A., 2016. Microbial life in extreme low-biomass environments: a molecular approach. Manual of Environmental Microbiology, fourth ed. American Society of Microbiology 4.3. 3-1-4.3. 3-11.