

ARTICLE OPEN



Nanopore sequencing at Mars, Europa, and microgravity conditions

Christopher E. Carr^{1,2,4}✉, Noelle C. Bryan¹, Kendall N. Saboda¹, Srinivasa A. Bhattaru³, Gary Ruvkun² and Maria T. Zuber¹

Nanopore sequencing, as represented by Oxford Nanopore Technologies' MinION, is a promising technology for in situ life detection and for microbial monitoring including in support of human space exploration, due to its small size, low mass (~100 g) and low power (~1 W). Now ubiquitous on Earth and previously demonstrated on the International Space Station (ISS), nanopore sequencing involves translocation of DNA through a biological nanopore on timescales of milliseconds per base. Nanopore sequencing is now being done in both controlled lab settings as well as in diverse environments that include ground, air, and space vehicles. Future space missions may also utilize nanopore sequencing in reduced gravity environments, such as in the search for life on Mars (Earth-relative gravito-inertial acceleration (GIA) $g = 0.378$), or at icy moons such as Europa ($g = 0.134$) or Enceladus ($g = 0.012$). We confirm the ability to sequence at Mars as well as near Europa or Lunar ($g = 0.166$) and lower g levels, demonstrate the functionality of updated chemistry and sequencing protocols under parabolic flight, and reveal consistent performance across g level, during dynamic accelerations, and despite vibrations with significant power at translocation-relevant frequencies. Our work strengthens the use case for nanopore sequencing in dynamic environments on Earth and in space, including as part of the search for nucleic-acid based life beyond Earth.

npj Microgravity (2020)6:24; <https://doi.org/10.1038/s41526-020-00113-9>

INTRODUCTION

Life as we know it uses nucleic acids as the basis for heredity and evolution. Life beyond Earth might utilize identical or similar informational polymers due to the widespread synthesis of common building blocks, common physicochemical scenarios for life's origin(s), or common ancestry via meteoritic exchange, most plausible for Earth and Mars. Beyond the search for life, sequencing is of high relevance for supporting human health on Earth and in space, from detecting infectious diseases, to monitoring of biologically-based life support systems.

Nanopore sequencing¹, as commercialized by Oxford Nanopore Technologies, is a promising approach that is now used ubiquitously in the lab and in the field. McIntyre et al. reported a single mapped read obtained via nanopore sequencing during parabolic flight, obtained across multiple parabolas². Vibration of flow cells revealed that 70% of pores should survive launch, consistent with later successful nanopore sequencing on the International Space Station (ISS)³. However, we are not aware of any nanopore experiments that attempted to quantify the impact of vibration while sequencing.

Here we test the impacts of: (1) altered g level, (2) vibration, and (3) updated chemistry/flow cells.

RESULTS

Parabolic flight

Flight operations were conducted on November 17, 2017 onboard a Boeing 727-200F aircraft (G-Force One®, Zero Gravity Corporation). Four sets of parabolas were performed with 5, 6, 4, and 5 parabolas respectively (Fig. 1a). The first set targeted, in order, Mars g , Mars g , Lunar g , 0 g , and 0 g (Fig. 1b). All other parabolas targeted 0 g . The flight profile was segmented into periods of

"transition", "parabola", "hypergravity", and "other" (typically, gentle climb, descent, straight and level flight, or standard rate turns) on the basis of accelerometer measurements⁴. Sequencing was also performed on the ground prior to the flight as a control.

Sequencing

Sequencing of control lambda deoxyribonucleic acid (DNA) was performed for a total of 38 min on the ground and 103 min during flight, on the same flow cell, resulting in 5293 and 18,233 reads for ground (Supplementary Fig. 1) and flight (Fig. 1c; Supplementary Fig. 1) respectively, of which 5257 and 18,188 were basecalled (Supplementary Tables 1, 2). Of the flight reads, 14,431 fell wholly within a phase of flight, including parabola (404), hypergravity (1996), transition (7), and other (12,024). Sequencing reads were obtained during all parabolas, including under Mars, lunar/Europa, and zero- g conditions (Fig. 2). The g levels achieved during each parabola were previously reported⁴. For the purposes of statistical analysis, mux reads (Fig. 1c, black horizontal lines) were excluded to avoid any sequencer start-up effects.

Vibration

Zero-phase filtering effectively removed frequencies at or below 10 Hz (Supplementary Figs 2–6). Filtered root-mean-square (RMS) vibration varied throughout the flight and showed clear deviations associated with parabolas (Figs 1c, 2a; Supplementary Fig. 2), indicating a smoother environment during freefall. Remaining aircraft-associated vibrations were largely in the 10 Hz to 1 kHz band with peaks at 116–128, 250–270, 495–496, 580–680, and 876 Hz (Supplementary Fig. 6). During zero- g parabolas, the magnitude of the residual g level and vibrations were comparable (Fig. 2a).

¹Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA. ³Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁴Present address: Georgia Institute of Technology, ESM Building, Room G10, 620 Cherry St NW, Atlanta, GA 30332, USA. ✉email: cecarr@gatech.edu

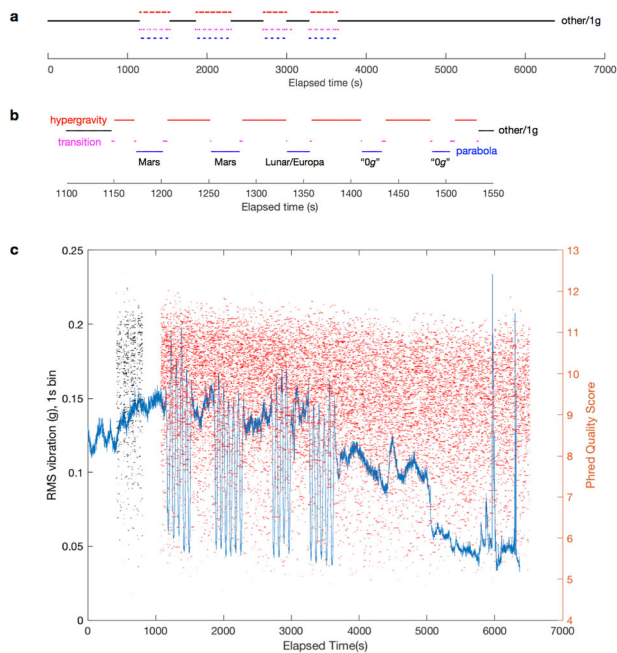


Fig. 1 Single molecule sequencing during parabolic flight. **a** Phases of flight timeline (black: other/1g; red: hypergravity; magenta: transition; blue: parabola). **b** Phases of flight for first set of parabolas. **c** Vibration (blue line, left axis) and sequencing reads measured during flight; each read is represented by a horizontal line (mux = black, run = red) at its representative read quality score, q_p .

Integrated read-level analysis

Stepwise linear regression was used to determine whether time and RMS vibration could predict median sequence quality (Supplementary Fig. 1), the Phred quality score^{5,6} associated with the average per-base error probability of a given read (see Materials and Methods). Unlike ground operations, where time was the only significant predictor of sequence read quality ($p = 0$), time, g level, and their combined effects were predicted to be significant indicators during flight (all $p < 10^{-4}$; Supplementary Tables 3, 4). However, in both cases, the variance explained was small ($adj. R^2 = 0.060$ and 0.275 , respectively, for ground and flight).

In order to elucidate the role of g level on read quality, those reads falling wholly within an individual phase of flight were examined using a one-way ANOVA, with Tukey's Honest Significant Difference (HSD) post hoc analyses (Supplementary Table 5). Sequence quality was significantly different during each phase of flight, with the lowest read quality during parabolas ($q_p = 8.3$) and the highest quality ($q_p = 8.7$) during hypergravity (Supplementary Fig. 7).

Integrated base-level analysis

Tombo⁷ was used to associate raw ionic current signals with specific genomic bases, and the number of reads aligning was similar to the number of reads with Phred quality scores^{5,6} > 6.5 . The percentage of bases that aligned to the lambda genome via tombo⁷ was 87.8% and 89.7% for ground and flight, respectively (Supplementary Table 1). Average coverage for tombo-aligned bases was adequate to sequence the lambda genome many times over during each parabola (Fig. 2d) and the coverage was largely explained by parabola duration ($adj. R^2 = 0.807$; Supplementary Table 6). Ionic current levels associated with unique subsequences (k-mers) were similar between ground and flight conditions (Supplementary Fig. 8).

By aligning ionic current signals to bases, tombo allowed us to measure the translocation time associated with each base (Supplementary Fig. 9), the time required for the motor protein, acting as a ratchet, to move the DNA strand one base into the nanopore. Translocation here refers to motion of the motor protein relative to the DNA strand, and not the total time to get through the nanopore, which requires many translocation steps. The inverse of translocation time is a direct measurement of sequencing rate for a given nanopore (bases/s).

Translocation times were similar, but statistically longer, during flight as compared to ground. Despite the nearly sixfold (5.89) average higher RMS vibration during flight compared to ground (Fig. 3a), the probability densities for translocation time are strikingly similar (Fig. 3b; Supplementary Fig. 9). However, base translocation times were significantly different (Kolmogorov–Smirnov, two-tailed, $p = 0$, test statistic 0.0306), with a slight shift toward longer translocation times during flight. Notably, the median base translocation times were identical (seven samples or 1.8 ms) and the means only differed by 0.125 ms (2.2786 ms, ground; 2.4035 ms, flight). Thus, translocation times were robust to large variations in vibration.

Ionic current noise is the variation in the flow of ions passing through the nanopore, measured here at the per-base level as a normalized signal standard deviation determined by tombo through optimal alignment of measured ionic current to a genomic sequence⁷. A stepwise linear regression was performed to determine if time, RMS vibration, or their combined effects were significant predictors of ionic current noise during ground (Supplementary Fig. 10; Supplementary Table 7) and flight (Supplementary Fig. 10; Supplementary Table 8) operations. Flight analysis included the additional variable g level.

For ground operations, the impact of time alone was not significant. However, both vibration ($p = 0.0018$) and the interaction effect of time and vibration ($p = 0.041$) were significant predictors of ionic current noise (Supplementary Table 7). However, the explanatory power of the regression was low ($adj. R^2 = 0.009$). Conversely, time was the only significant predictor of the effect on ionic current noise during flight. Neither RMS vibration, g level, nor any of their respective combined effects had significant impacts on ionic current noise (Fig. 4; Supplementary Table 8).

Because time was a significant indicator of ionic current noise during flight, it was necessary to assess whether the effect could be attributed to a specific phase of flight (Supplementary Table 9). Tukey's HSD post hoc test demonstrated that out of all six possible pairwise comparisons, only one, parabola vs. transition, was not significant (two-sided $p = 0.345$; other $p < 10^{-3}$). Ionic current was significantly lower in hypergravity, parabola, and transition phases as compared to other. Ionic currents during hypergravity phases were, on average, lower than all other phases (Supplementary Table 9; Supplementary Fig. 11). Thus, while the impact of phase of flight on read quality showed a trend toward higher read quality with higher g level (Supplementary Fig. 7), no such pattern was observed with ionic current (Supplementary Fig. 11).

DISCUSSION

The Mars 2020 rover, currently enroute to Mars, is expected to touch down in Jezero Crater in February, 2021. While this mission will not attempt to detect extant life, it represents a new era in the search for life beyond Earth. Ambiguous or positive results in the search for ancient life could usher in a new era of life detection efforts, including instrumentation aimed at measuring the presence of nucleic acids, one of the “smoking gun” pieces of evidence for life beyond Earth⁸. In preparation for future life detection missions targeting DNA, we explored the capabilities of nanopore sequencing, and present results demonstrating its successful performance while experiencing aircraft vibrations

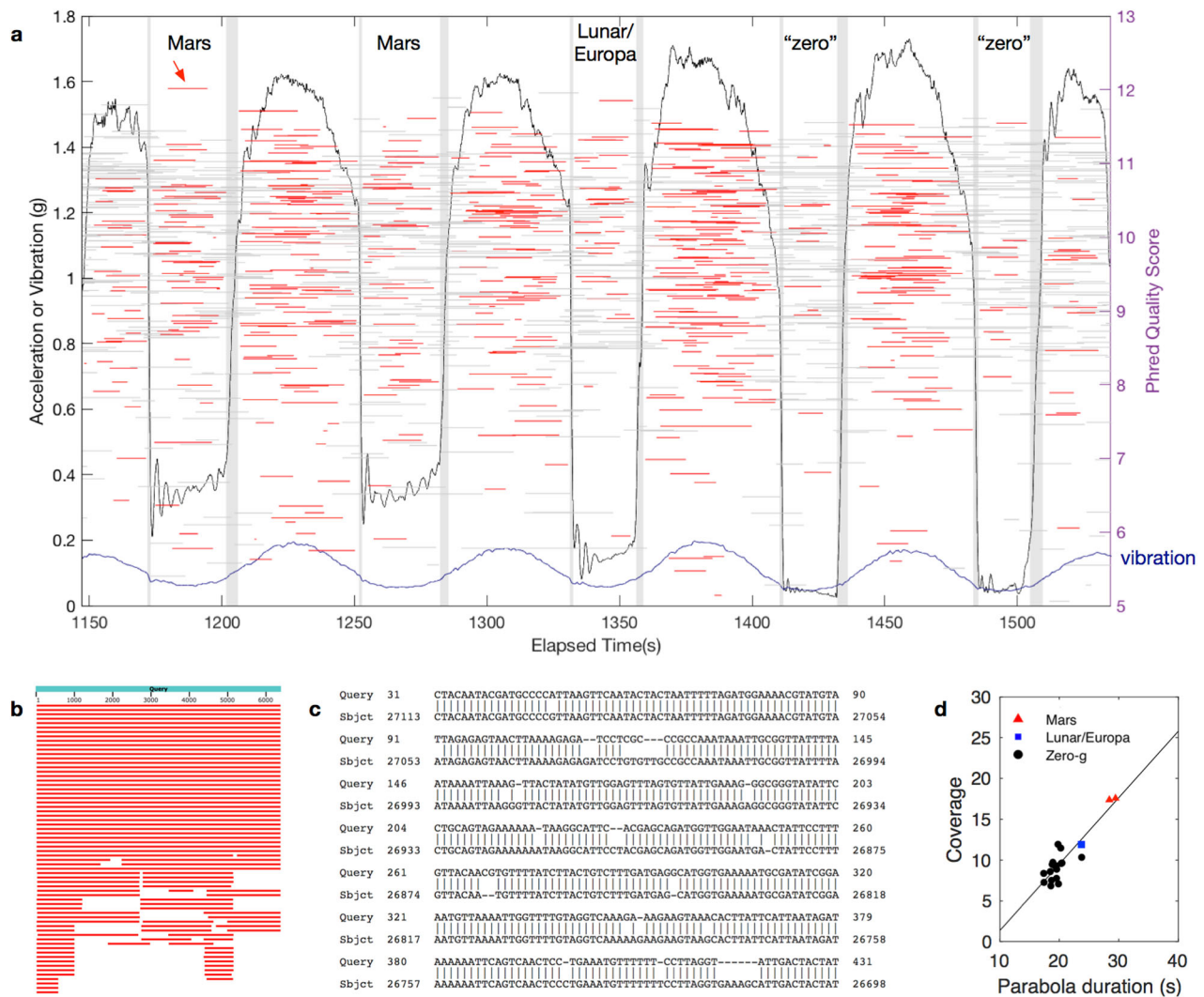


Fig. 2 Sequencing in reduced gravity. **a** g level achieved (black line) and RMS vibration (1 s bins, blue line) and associated sequencing reads acquired during first set of parabolas. Each read is represented by a horizontal line (gray: partially or completely in transition period; red: completely in non-transition period) at its representative read quality score, q_p (right axis). Vertical gray bands demarcate transitions between phases of flight. **b** Top scoring BLAST results for highest quality "Mars" read, indicated via arrow in **a**, length 6402. **c** Start of best match sequence alignment, to J02459.1 *Enterobacteria phage lambda*, complete genome, length 48502 (range 20562–27113, score 8907 bits(9877), expect 0.0, identities 6108/6651 (92%), gaps 395/6651 (5%), strand Plus/Minus). **d** Average genomic coverage of lambda for all parabolas based on tomo-aligned bases.

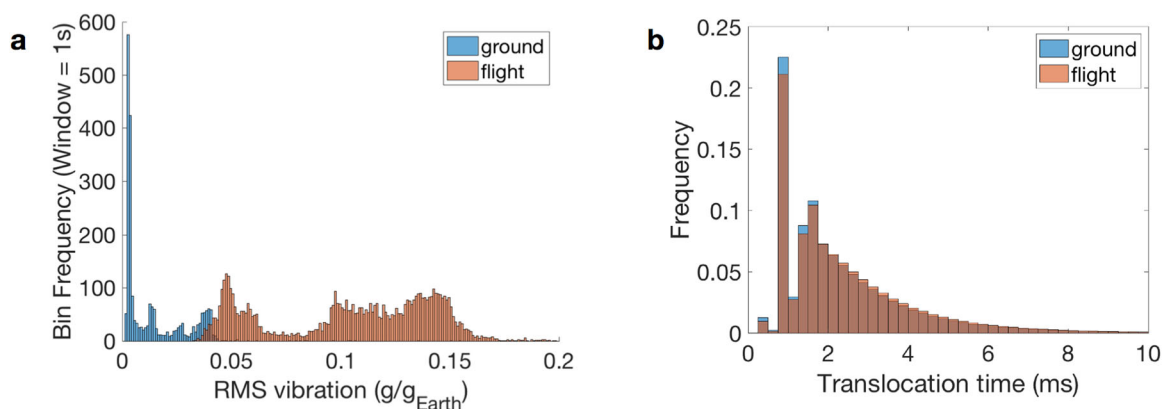


Fig. 3 Translocation time is weakly or not affected by vibration. **a** RMS vibration distributions for ground and flight. **b** Nanopore translocation time as measured by alignment of ionic current to the genomic reference: distribution for <10 ms. Ground (blue), flight (light brown), both (dark brown).

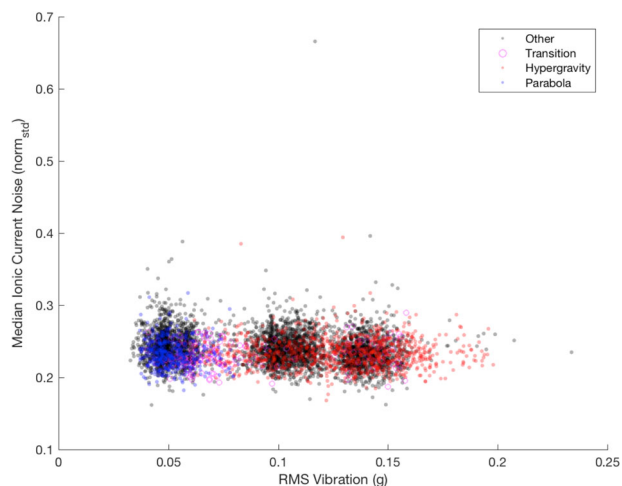


Fig. 4 RMS vibration and median ionic current noise during flight. The single 1 s period with median ionic current noise >0.5 has a median absolute deviation (MAD) of >15 and is therefore an outlier (typically defined as $\text{MAD} > 3$).

and under altered g levels, including those that would be encountered on the surface of Mars, the Moon, and/or Jupiter's moon Europa. Due to the limitations of parabolic flight, our zero- g conditions involved mean acceleration around $4\times$ higher ($0.041 \pm 0.005 \text{ g}$)⁴ than at the surface of Saturn's moon Enceladus (0.011 g).

Several factors may have influenced the overall quality of our nanopore sequencing data. The DNA sequencing library used in this experiment was stored for 72 h prior to loading onto the flow cell. As such, the sample was potentially subjected to degradation, which could impact read quality, and may have resulted in the loss of ligated adaptors. Such conditions would negatively impact the proper loading of DNA into the individual nanopore. In addition, there is an expected degradation of the flow cell over time during sequencing, which could explain some of the time-related trends, independent of any effects of vibration or acceleration. Despite sequencing for a limited time at any given g level during parabolic flight, the operation of the MinION for sustained periods on the ISS³ gives us confidence that extended periods of reduced g level does not negatively impact nanopore sequencing. In addition, it provides confidence in nanopore sequencing as a viable life detection technology in very low but nonzero g environments, such as Enceladus.

Because zero-phase filtering of vibration data effectively removed frequencies at or below 10 Hz (Supplementary Figs 5, 6), filtered vibration measurements did not reflect frequencies where sensor data would be inaccurate due to the non-unity frequency response of the sensor near DC (0 Hz). In addition, this filtering ensured that we could assess the independent effects of g level and vibration.

The peaks in the vibration spectrum occurs at frequencies relevant to nanopore sequencing. Despite this, vibration did have any significant impacts on sequence quality nor on ionic current noise, except during ground-based sequencing, where the explanatory power of vibration was negligible ($<1\%$; Supplementary Table 7).

Random vibration at translocation-relative frequencies could exert a minor interfering effect on translocation, although any impact in changes in vibration during flight did not translate into any consistent or large changes in ionic current noise due to the small (0.125 ms) mean difference in translocation times despite nearly a sixfold change in RMS vibration.

Higher g levels tended to be associated with higher read quality, although the effect size is small ($\Delta q_p = 0.4$, hypergravity – parabola; Supplementary Table 5, an upper bound of $\sim 0.2/g$). The

smallest mean values of ionic current noise were also observed during hypergravity (0.282; Supplementary Fig. 11). Although statistically larger, the difference between largest mean value for ionic current noise (phase other) was miniscule (0.003).

Nanopore sequencing is compatible with many life detection missions from the perspective of mass ($\sim 100 \text{ g}$), size, and power ($<2 \text{ W}$). Recent work also suggests that MinION electronics and flow cell components would survive radiation doses consistent with life detection missions to Mars, Venus, and Enceladus, although not Europa, without additional shielding⁹. Our work shows that sequencing on all these worlds, including Europa, could be feasible from a g level perspective. In addition, the robustness to vibration suggests that operation concurrent with other mission activities, such as drilling or operation of other instrument payloads, could occur without any substantial negative impacts. In addition, our work highlights the potential for nanopore sequencing on Earth and beyond in mobile and dynamic environments such as on passenger aircraft, drones, wheeled vehicles, ships, buoys, underwater vehicles, or other platforms.

METHODS

Acceleration measurement and flight profile segmentation

The flight profile was segmented as described in Carr et al.⁴ from acceleration data collected using a metal-body Slam Stick X™ (Mide Technology Corp.). The accelerometer was mounted next to the MinION to a common baseplate, using double sided sticky tape (3M 950) to provide a near-unity vibration frequency response. Vibrations were measured with the internal triaxial piezoelectric accelerometer (TE Connectivity Ltd., 832M1) at a frequency of 5 kHz.

Sequencing

Sequencing libraries were prepared using DNA derived from *Enterobacteriaceae* *phage lambda* (NEB N3011S), fragmented using a g-TUBE™ (Covaris® 520079) with the 6 kb protocol. Next, the libraries were prepared using the 1D ligation method (SQK-LSK108) using a “one-pot” barcoding protocol¹⁰ and stored at 4°C for $\sim 72 \text{ h}$ prior to the flight. At the time of storage, the total library DNA was estimated to be 440 ng at $31.4 \text{ ng}/\mu\text{l}$ as assessed by fluorometry (ThermoFisher Qubit® 3.0 Fluorometer with Qubit™ dsDNA HS Assay Kit, Q32854).

A flow cell (FLO-MIN106 R9) was loaded on the ground and sequencing performed using an offline version of MinKNOW 1.7.14 in the flight hardware configuration while the aircraft was on the ground. After 38 min, sequencing was stopped. In flight, sequencing was reinitiated around 12 min prior to parabolic flight maneuvers, and continued for a total of 103 min before termination. After the flight, basecalling was performed with ONT Albacore version 2.3.1 with quality filtering disabled.

Sequence data processing

To quantify adaptor sequences, fastq output was trimmed using Porechop (<https://github.com/rwick/Porechop> docker container quay.io/biocontainers/porechop:0.2.3_seqan2.1.1-py35_2). Original untrimmed fast5 reads were aligned to the reference genome (NEB Lambda, equivalent to NCBI NC_001416.1 with mutations 37589 C→T, 45352 G→A, 37742 C→T, and 43082 G→A) using tomo (docker container quay.io/biocontainers/ont-tombo:1.5-py27r351h24bf2e0_0)⁷ with the `—include-event-stdev` option.

Sequencing and acceleration data integration

A custom script was used to parse tomo-processed fast5 files to characterize each read and each tomo-aligned genomic base within each read (Supplementary Data). A representative read quality score was calculated as $q_p = -10\log_{10}(\bar{p})$, where \bar{p} is the mean of the per-base error probability $p = 10^{-q/10}$, where q is the per-base Phred quality score^{5,6} estimated via basecalling. Read timings were adjusted by offsets to align genomic and accelerometer data (Supplementary Table 2). Each read and base was assigned one of the following states (parabola, transition, hypergravity, other) on the basis of the periods.txt file produced by prior analysis⁴ and available online at <https://osf.io/nk2w4/>.

Vibration data processing

A vibration equivalent to g level (Earth-relative gravity) was computed as $g = \sqrt{g_x^2 + g_y^2 + g_z^2}$ to provide a measure of vibration that is independent of the Slam Stick X™ orientation.

The vibration power spectral density (PSD) for g was computed using Welch's method (MATLAB *pwelch()* function) with default parameters (Supplementary Fig. 4). Filtering was then performed for two reasons: (1) to eliminate vibration data where the frequency response of the piezoelectric accelerometer is not unity, and (2) to analyze vibration at frequencies related to timescales at which base translocation occurs during nanopore sequencing, which are overwhelmingly <10 ms (Fig. 3b, Supplementary Fig. 9). The g level equivalent vibration g was filtered with a high pass infinite impulse response filter (Supplementary Fig. 5) that was generated with MATLAB's *designfilt()* function (stopband 5 Hz @ 60 dB attenuation, passband 10 Hz with unity ripple, sample rate 5 kHz). Filtering was performed using the MATLAB *filtfilt()* function, which uses forward and reverse filtering to achieve zero-phase delay. The PSD was computed as before for the resulting filtered g level equivalent vibration g_f (Supplementary Fig. 6). RMS vibration was computed in 1 s bins from g_f using the MATLAB *rms()* function. An overview of vibration is shown in Supplementary Fig. 2 for flight and Supplementary Fig. 3 for ground.

Sequencing read quality regression analysis

Sequencing read times were adjusted by an offset to place sequencing reads into the accelerometer elapsed time (Supplementary Table 2). A time series of median read quality was estimated in 1 s bins by computing the median of q_p for all reads covering the bin. Stepwise linear regression, via the MATLAB *stepwiselm()* function, was used to evaluate the impact of time, RMS vibration, and g level (flight only) on sequence quality, as measured by median q_p (Supplementary Tables 3, 4). For flight, the regression time was restricted to a maximum elapsed time of 4000 s to eliminate potential confounding effects of the aircraft descent and landing.

Sequencing read quality phase of flight analysis (flight only)

To assess differences in read quality as a function of phase of flight, we performed a one-way analysis of variance (ANOVA) via the MATLAB *anova1()* function on the non-mux reads (Supplementary Table 5), excluding reads in transition periods due to their low number (7) and short length. To compare group means we then used Tukey's Honestly Significant Difference test (MATLAB *multcompare()* function), which accounts for multiple testing and is conservative for one-way ANOVA with different sample sizes.

Coverage of genomic-aligned bases

Base times were adjusted by an offset to place each tomo-aligned base into the accelerometer elapsed time (Supplementary Table 2). Coverage was estimated as the sum of tomo-aligned bases within a given phase of flight divided by the lambda genome size (48,502 bases). Stepwise linear regression, via the MATLAB *stepwiselm()* function, was used to evaluate the relationship between coverage and parabola period (Supplementary Table 6).

Base ionic current noise regression analysis

A time series of ionic current noise was estimated in 1 s bins by computing the median of ionic current (tomo norm_std output) for all bases within a bin. Stepwise linear regression, via the MATLAB *stepwiselm()* function, was used to evaluate the impact of time, RMS vibration, and g level (flight only) on median ionic current noise (Supplementary Tables 7, 8). For flight, the regression time was restricted as stated above.

Base ionic current noise phase of flight analysis (flight only)

To assess differences in ionic current as a function of phase of flight, we performed a one-way ANOVA via the MATLAB *anova1()* function on the non-mux tomo-aligned bases (Supplementary Table 9). To compare group means we then used Tukey's Honestly Significant Difference test as above.

Does flight vs. ground impact translocation time?

A Kolmogorov–Smirnov test was performed with the MATLAB *kstest2()* function on the base translocation times for ground vs. flight.

Preprint

A previous version of this manuscript was published as a preprint¹¹.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Raw and calibrated data are available via the Open Science Framework at: <https://osf.io/n6krq/> and <https://osf.io/nk2w4/>.

CODE AVAILABILITY

The MATLAB scripts implementing our analysis are available at: <https://github.com/CarrCE/zerogseq>.

Received: 10 January 2020; Accepted: 11 August 2020;

Published online: 07 September 2020

REFERENCES

- Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
- McIntyre, A. B. R. et al. Nanopore sequencing in microgravity. *npj Microgravity* **2**, 16035 (2016).
- Castro-Wallace, S. L. et al. Nanopore DNA sequencing and genome assembly on the International Space Station. *Sci. Rep.* **7**, 18022 (2017).
- Carr, C. E. et al. Acceleration profiles and processing methods for parabolic flight. *npj Microgravity* **4**, 14 (2018).
- Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
- Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
- Stoiber, M. H. et al. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. Preprint at <http://biorxiv.org/lookup/doi/10.1101/094672> (2016).
- Neveu, M., Hays, L. E., Voytek, M. A., New, M. H. & Schulte, M. D. The ladder of life detection. *Astrobiology* **18**, 1375–1402 (2018).
- Sutton, M. A. et al. Radiation tolerance of Nanopore sequencing technology for life detection on Mars and Europa. *Sci. Rep.* **9**, 5370 (2019).
- Quick, J. One-pot ligation protocol for Oxford Nanopore libraries. Preprint at <https://doi.org/10.17504/protocols.io.k9acz2e> (2018).
- Carr, C.E. et al. Nanopore sequencing at Mars, Europa, and microgravity conditions. Preprint at <https://www.biorxiv.org/content/10.1101/2020.01.09.899716v1>.

ACKNOWLEDGEMENTS

We thank the MIT Media Lab Space Exploration Initiative for providing the parabolic flight. This work was supported by NASA award NNX15AF85G. N.C.B. was supported by NASA Postdoctoral Fellowship award 80NSSC17K0688.

AUTHOR CONTRIBUTIONS

C.E.C. designed the experiment, C.E.C., K.S., S.A.B., and N.C.B. built and tested the hardware, N.C.B. and M.T.Z. collected the data, C.E.C. and N.C.B. processed the data. G.R. advised on the experiment design. C.E.C. and N.C.B. wrote, and all authors edited and approved, the paper. C.E.C. is the guarantor.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41526-020-00113-9>.

Correspondence and requests for materials should be addressed to C.E.C.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020