Single-molecule long-read methylation profiling reveals regional DNA methylation regulated by Elongator Complex Subunit 2 in *Arabidopsis* roots experiencing spaceflight

Mingqi Zhou[1], Alberto Riva[2], Marie-Pierre L. Gauthier[3], Michael P. Kladde[3], Robert J. Ferl[1,4,], Anna-Lisa Paul[1,2,*]*

[1]Department of Horticultural Sciences, University of Florida, 2550 Hull Road, Fifield Hall, Gainesville FL, 32611, USA

[2]Interdisciplinary Center for Biotechnology Research, University of Florida, 2033 Mowry Road, Gainesville FL, 32610, USA

[3]Department of Biochemistry and Molecular Biology, University of Florida, 2033 Mowry Rd, Gainesville, FL 32608, USA

[4]Office of Research, University of Florida, 1523 Union Rd, Grinter Hall, Gainesville, FL, 32611, USA

***Correspondance: Robert J. Ferl** (robferl@ufl.edu), **Anna-Lisa Paul** (alp@ufl.edu)

Abstract

Flap-Enabled Next-Generation Capture (FENGC), a novel targeted DNA capture technique, has been applied to Arabidopsis root samples from the Advanced Plant Experiment-04 - Epigenetic Expression (APEX-04) experiment onboard the International Space Station (ISS). Single-molecule long-read profiling reveals regional DNA methylation alteration that was not previously uncovered by population-based methods during spaceflight.

DNA methylation epigenetically modulates gene expression and impacts chromatin structure to function in physiological adaptation to environmental changes in eukaryotes

(Zhang et al., 2018). Plants will play an important role in the human life support systems envisioned for spaceflight and planetary exploration habitats (Wheeler, 2010; Zabel et al., 2016). Spaceflight presents a unique environmental challenge for plants, an environment that is outside the evolutionary experience of any terrestrial organism. Epigenetic change plays an important role in environmental adaptation, and it is crucial to understand the role of epigenetics in plant adaptation to this novel environment (Paul et al., 2021; Zhou et al., 2019). The Advanced Plant Experiment-04 - Epigenetic Expression (APEX-04) experiment onboard the International Space Station (ISS) examined the spaceflight-altered cytosine methylation in *Arabidopsis thaliana* using *elp2-5*, a mutant of an epigenetic regulator, Elongator Complex Subunit 2 (ELP2), along with Col-0, a wild-type control (Paul et al., 2021). Genome-wide analysis revealed remodeled transcriptomes and DNA methylomes during spaceflight, remodeling that was regulated by ELP2. Phenotypically, root growth in *elp2-5* did not set up a typical spaceflight directionality, and a majority of differentially expressed genes in *elp2-5* roots were also differentially methylated between spaceflight samples and ground controls (Paul et al., 2021).

Flap-Enabled Next-Generation Capture (FENGC) is a novel, strand-specific DNA capture method that allows DNA methylation analysis with long reads of up to ~1 kb in user-defined target regions (Zhou et al. 2022). FENGC allows sensitive characterization of contiguous single molecules over long ranges, thereby providing a deeper insight into the methylation state afforded by averaged state of methylation obtained by multiple short reads. In addition, compared with whole genome bisulfite sequencing (WGBS), FENGC increases coverage depth in regions of interest, which improves data quality and avoids cost of reads in unwanted areas. The accuracy and efficiency of enzymatic methyl-seq, which is part of the FENGC procedure, has been verified in plants (Feng et al., 2020); however, FENGC has yet to be applied to methylation analysis in plants.

Therefore, we set out to demonstrate the effective application of FENGC in plant methylation contexts. The observed phenotypical and molecular alteration in roots of Col-0 and *elp2-5* grown on ISS in APEX-04 present a well-characterized, short-read

methylation dataset for comparison of methylation profiles assayed by FENGC single-molecule long reads.

The target panel for capture and assay contained 108 genic regions ranging from 509 to 704 nt (Supplemental Table 1), with mean of 580 nt and median of 572 nt, within the promoter or gene body regions of epigenetically regulated genes derived from our APEX-04 spaceflight and related experiments (Paul et al., 2017a; Paul et al., 2021; Paul et al., 2013; Zhou et al., 2019). The capture was performed as a single-tube multiplex reaction followed by enzymatic methyl-PCR (Oligos are listed in Supplemental Table 2). The FENGC products were subjected to PacBio long-read HiFi sequencing. The samples consistently yielded ~90% on-:off-target ratio and, on average, 82% targets were detected by at least 10 reads after filtering and combining the four biological replicates from each sample (Figure 1A, 1B, and Supplemental Table 3). The GC content of the 108 targets ranged from 19% to 52%, which positively correlated with the aligned reads number (Figure 1C, Supplemental Figure 1), which was opposite to FENGC results in DNA materials with GC content above 50% in most of targets (Zhou et al. 2022). This indicates that both high and low GC content of targets influenced the reads yield of FENGC products.

One major advantage of single-molecule profiling of DNA methylation is that the contiguous long reads allow assessment of every cytosine within same DNA molecule, thereby presenting a coherent regional methylation status of that entire gene element within a cell. In addition, the average methylation levels in cytosines across all molecules can be derived from the data for comparison to WGBS data. For the present study, we applied long-read, single-molecule sequencing to gene targets showing both significant differential expression and changes in average cytosine-5 methylation levels between spaceflight and ground control. These single-molecule investigations further revealed space-altered distributions of the proportion of methylated sites per molecule.

We present four examples associated with defense responses to stimuli associated with spaceflight (Supplemental Figure 2), the methylation levels of which exhibited correlation coefficients above 0.96 between any two biological replicates and above 0.95 between FENGC and WGBS results (Figure 1D). The first three examples examine

FENGC data for CG methylation sites as no significant differences were observed for CHG and CHH (data not shown). *CML46* (AT5G39670), a gene encoding a calmodulin-like protein showing significant transcriptional down-regulation during spaceflight in Col-0 but not *elp2-5* plants, was assayed in a promoter region containing a MYB-like DNA binding element (Supplemental Figure 3A). More than 99% of the mapped reads are full-length, containing all CG sites in this region of *CML46* (Supplemental Figure 3B). Flight Col-0 Roots (FCR) and Ground control Col-0 Roots (GCR), but not roots from *elp2-5* (FER and GER), showed a significant difference in the CG methylation level in an area flanking the MYB-like element, which contained five CG sites (Supplemental Figure 3C, in dashed frame). Consistent with these data, CG methylation proportion per molecule in this target region was significantly higher in FCR than GCR (Supplemental Figure 3D). A significant difference between GCR and GER was also observed. FENGC showed more molecules containing consecutively methylated CG sites in FCR (12%) compared to GCR (6.9%) in the area flanking this *cis*-regulatory element (Supplemental Figure 3E). Consistency between transcriptional gene expression, population-based methylation level derived from WGBS, and FENGC-derived methylation proportion per molecule support the biological conclusion that ELP2 played a negative role in expression and a positive role in promoter methylation of *CML46* in spaceflight.

ABCB17 (AT3G28380), an ATP-binding cassette transporter gene that also showed a significantly lower transcriptional expression level in FCR than GCR (Supplemental Figure 2), was assayed in a gene body region (Supplemental Figure 4A). More than 98% of the mapped reads contained all CG sites in this region (Supplemental Figure 4B). The averaged regional CG methylation levels were in a same trend, which did not show significant changes between any pair of samples (Supplemental Figure 4C); however, the CG methylation proportion per molecule in this target region was significantly higher in FCR than in GCR (Supplemental Figure 4D). The FENGC single-molecule data showed that 3.8% of the reads in FCR possessed at least 6 consecutive methylated CG sites while in GCR the percentage was 0.8% (Supplemental Figure 4E). Biologically, *ABCB17* was repressed by spaceflight and ELP2 contributed to its transcription and gene body methylation.

Similarly, in the promoter region of *DEFL* (AT4G22217), a defensin-like gene (Supplemental Figure 5A and 5B), the patterns of averaged CG methylation levels were highly similar and no significant alterations were detected between any two samples (Supplemental Figure 5C); however, CG methylation proportion per molecule was significantly higher in FER than GER (Supplemental Figure 5D and 5E). This higher methylation proportion correlates with the lower transcription level of this gene in FER compared to GER (Supplemental Figure 2). ELP2 enhanced expression and repressed promoter methylation of *DEFL* in spaceflight. These examples suggest that FENGC single-molecule level methylation profiling can be more revealing of consequential methylation changes compared with population-based methylation methods such as WGBS, especially when alterations potentially occur in a subpopulation of cells within the sample.

The fourth example examines another advantage of long-read, single-molecule FENGC assay of DNA methylation in plants, which is that all three methylation sequence contexts (CG/CHG/CHH) can be identified and plotted for the same molecule, thereby aggregating the entire methylation status of a gene region. For this example, we chose a promoter fragment containing a MeJA-responsive element in *PDF1.1* (AT1G75830), a gene coding a pathogenesis-related protein belonging to plant defensin family (Figure 1E). This target was captured with more than 92.6% full-length, aligned reads containing all cytosines in three sequence contexts within this region (Figure 1F). The average methylation level per site over all sequenced molecules showed significant differences at CHG and CHH sites between GER and GCR, as well as in CHH sites between FCR and GCR (Figure 1G), correlating well with the previous WGBS data (Figure 1D) (Paul et al., 2021). For methylation proportion per molecule, significant differences were detected in all pairwise comparisons between flight conditions and genotypes at cytosines in all three methylation contexts (Figure 1H). The patterns of methylation proportion in DNA molecules demonstrated that three cytosine contexts showed a trend of co-methylation in this region among cells in the population (Supplementary Figure 6). Within each molecule, methylation proportion of CG is higher than CHG in 97.9% reads, and that of CHG is higher than CHH in 87.5% reads (Supplemental Figure 7). The CG methylation proportions were above 0.57 in all reads

of GCR and above 0.28 in GER, while FCR and FER had reads with CG methylation proportion as low as 0. These data indicate that more molecules were hypomethylated in this region in *elp2-5* compared with Col-0 and in spaceflight compared with ground control.

To examine the spatial distribution of these methylation differences, three separate but correlated panels were plotted for CG, CHG and CHH methylation in the same DNA molecule per row (Figure 1I). In these plots, black areas are unmethylated regions, red areas are methylated regions, and gray areas are transition zones between methylated and unmethylated sites. It is visually apparent that spaceflight results in more unmethylated regions in Col-0, with the deepest loss of methylation concentrated near the MeJA-responsive element. It is also visually apparent that *elp2-5* shows a deeper loss of methylation in that region. For the CG methylation context, a footprint showing consecutive non-methylation of the first 2 CG sites in flanking area of the MeJA-responsive element was identified in 8.4%, 18.0%, 39.4% and 44.0% reads in GCR, FCR, GER and FER, respectively, quantifying the loss of CG methylation in this area. Interestingly, 60%, 86.5%, 97.5% and 97.6% reads with consecutive non-methylation in the second and the third CG sites in GCR, FCR, GER and FER, respectively, were also included in this footprint and showed consecutive non-methylation of the first 3 CG sites. In addition, CHG and CHH methylation proportions were significantly lower in the footprint area in the reads containing this footprint (Supplementary Figure 8). These were consistent with previously observed high frequency of co-methylation/demethylation in nearby cytosines in *Arabidopsis* (Tran et al., 2018), and demonstrated that FENGC provides further opportunities for studying all three types of co-methylation in same molecules.

These changes in average cytosine methylation level or methylation proportion per molecule could be due to modulation by ELP2 in a diverse manner during spaceflight. Four targets shown here are associated with plant defense signaling, which is largely utilized in spaceflight adaptation of plants and may be involved in regulation of root growth phenotypes during spaceflight (Califar et al., 2020; Paul et al., 2017b). Targets possessing changes that cannot be observed in a population-based average level but

are revealed by consecutive methylation in each same molecule would be novel candidates for further investigation of downstream pathways elicited by ELP2. Whereas short-read WGBS only detects average methylation level alteration in cytosines, targeted long-read, single-molecule profiling further reveals the detailed epigenetic modifications occurring in cell subpopulations and identifies epialleles with low frequency, which provides additional power over population-based data for methylation analysis and identification of epigenetically modified substrates.

Supplemental Information

Supplemental Materials and Methods

Supplemental Figure 1. Correlation between numbers of filtered HiFi reads aligned to targets and the GC content of target regions in each sample.

Supplemental Figure 2. Expression level of AT5G39670, AT3G28380, AT4G22217, and AT1G75830 in roots of Col-0 and elp2-5 in APEX04 spaceflight experiment.

Supplemental Figure 3. FENGC results for AT5G39670.

Supplemental Figure 4. FENGC results for AT3G28380.

Supplemental Figure 5. FENGC results for AT4G22217.

Supplemental Figure 6. Correlation between proportions of different types of methylated cytosines within the same molecule in the captured region of AT1G75830.

Supplemental Figure 7. Proportion of methylated CG, CHG and CHH sites within the same molecule in the captured region of AT1G75830.

Supplemental Figure 8. Comparisons of methylation proportions per molecule in reads with a footprint and in all reads.

Supplemental Table 1. Information of 108 capture targets.

Supplemental Table 2. Oligos used in this study.

Supplemental Table 3. HiFi reads on target and targets detected after removing reads with more than three consecutive methylated CHH sites.

Supplemental Table 4. Number of filtered HiFi reads aligned to 108 targets.

Funding

This work was supported by NASA grants NNX14AT24G and 80NSSC19K0130 to A.-L.P. and R.J.F. Development of FENGC was supported by Defense Threat Reduction Agency grant HDTRA1-16-1-0048 awarded to M.P.K. as a co-PI and NIH grant R01 CA155390 awarded to M.P.K.

Author contributions

M.Z., R.J.F. and A.-L.P. designed this study. M.Z. conducted the experiments and analyzed the data. M.Z. and M.P.K. developed the FENGC methodology. A.R., M.-P.L.G. and M.P.K. contributed to the FENGC oligo design program. M.Z., M.P.K., R.J.F. and A.-L.P. wrote the manuscript.

Accession numbers

Demultiplexed CCS reads data files of PacBio Sequel IIe SMRT sequencing have been deposited to BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA857037 in National Center for Biotechnology Information (NCBI).

Acknowledgments

We thank members of UF Space Plants Lab for helpful discussion and members of UF ICBR, NextGen DNA sequencing (RRID:SCR_019152), Bioinformatics (RRID:SCR_019120) and Gene Expression and Genotyping (RRID:SCR_019145)

cores, for their support and services. We also thank Jason O. Brant for his assistance in developing the FOLD program. A patent has been submitted as an international application (No. PCT/US2022/020624, filed March 16, 2022). M.Z. and M.P.K. are two of the co-inventors of this application.

Legends

Figure 1. Application of flap-enabled next-generation capture (FENGC) in root samples of *Arabidopsis* experiencing spaceflight.

(A) Percentage of filtered HiFi reads on target for all 108 targets. Reads with more than three consecutive methylated CHH sites were removed. GCR, Ground Col Roots; FCR, Flight Col Roots; GER, Ground control *elp2* Roots; FER, Flight *elp2* Roots.

(B) Percentage of targets detected by at least 1 filtered HiFi read or at least 10 filtered HiFi reads.

(C) Correlation between numbers of total filtered HiFi reads of all samples aligned to targets and the GC content of target regions. Log transformation of reads number was done by Log_2(read number +1). R-squared value and *p* value of logarithmic regression analysis are shown.

(D) Correlation coefficients between biological replicates of FENGC and between FENGC and whole-genome bisulfite sequencing (WGBS) for average methylation levels in cytosines.

(E) Genic region of AT1G75830 (PDF1.1). The target region is indicated by brown rectangle (target ID: AT1G75830-1 in Supplemental Table 1). A MeJA-responsive element is indicated by green brick. The transcription start site is indicated by the arrow.

(F) Percentage of full-length reads containing all CG, CHG or CHH sites in the target region, and truncated reads with 1 or more cytosines missing. For CHH sites, 1 site missing is allowed in the full-length reads.

(G) Average CG, CHG or CHH methylation levels over all sequenced molecules of the target region. The relative position of MeJA-responsive element shown in **(E)** is indicated by green brick. The significant differences of methylation levels between samples for CHG and CHH sites in the target region are shown. One-way ANOVA with post-hoc Tukey HSD test; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Black arrows indicate the differentially methylated cytosines (DmCs) detected in WGBS using the threshold of

methylation difference > 0.2 and $p < 0.01$. All six sites were significantly different between GER and GCR.

(H) Distribution of proportion of methylated CG, CHG or CHH sites per molecule in filtered full-length HiFi reads aligned to the target region. The box indicates the upper and lower quartiles of the data set. The whiskers extend to the largest or the smallest data point that is within 1.5 times the interquartile range. The data points of methylation proportion per molecule are represented by dots. The line and "x" in the box indicate median and mean, respectively. The significant differences of methylation proportion are shown (Kruskal-Wallis test; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

(I) Heatmaps showing the cytosine methylation in single molecules. Reads from the four biological replicates (Supplemental Table 4) were combined to generate heatmaps for each sample. If the number of combined reads exceeded 1,000, 1,000 reads were randomly selected and plotted. The three heatmap panels show the patterns of methylation at CG, CHG, and CHH sites (vertical white lines) as indicated at bottom along the same DNA molecule in each row. Red indicates two consecutively methylated cytosines, black indicates two consecutively unmethylated cytosines, and gray indicates borders between methylated and unmethylated cytosines. Reads were hierarchically clustered using Kendall's Tau based on CG methylation. The relative positions of MeJA-responsive element shown in **(E)** and differentially methylated cytosines detected by WGBS shown in **(G)** in the target region are indicated by green brick and black arrows, respectively. A footprint (blue ovals) with consecutive protection against methylation of the first 2 CG sites in the area flanking the MeJA-responsive element is present in 8.4%, 18.0%, 39.4% and 44.0% molecules in GCR, FCR, GER and FER, respectively. One-way ANOVA with post-hoc Tukey HSD test was performed for pairwise comparisons of percentage of reads with this footprint. FCR vs GCR, $p < 0.05$; GER vs GCR, $p < 0.001$; FER vs FCR, $p < 0.001$.

References

Califar, B., Sng, N.J., Zupanska, A., Paul, A.-L., and Ferl, R.J. (2020). Root skewing-associated genes impact the spaceflight response of *Arabidopsis thaliana*. Front. Plant Sci. **11**:239. 10.3389/fpls.2020.00239.

Feng, S., Zhong, Z., Wang, M., and Jacobsen, S.E. (2020). Efficient and accurate determination of genome-wide DNA methylation patterns in *Arabidopsis thaliana* with enzymatic methyl sequencing. Epigenetics Chromatin **13**:42. 10.1186/s13072-020-00361-9.

Paul, A.-L., Zhou, M., Callaham, J., Reyes, M., Stasiak, M., Riva, A., Zupanska, A., Dixon, M., and Ferl, R. (2017a). Patterns of *Arabidopsis* gene expression in the face of hypobaric stress. Aob Plants 910.1093/aobpla/plx030.

Paul, A.-L., Zupanska, A.K., Schultz, E.R., and Ferl, R.J. (2013). Organ-specific remodeling of the *Arabidopsis* transcriptome in response to spaceflight. BMC Plant Biol. **13**:112. 10.1186/1471-2229-13-112.

Paul, A.-L., Haveman, N., Califar, B., and Ferl, R.J. (2021). Epigenomic regulators *Elongator Complex Subunit 2* and *Methyltransferase 1* differentially condition the spaceflight response in *Arabidopsis*. Front. Plant Sci. **12**:691790. 10.3389/fpls.2021.691790.

Paul, A.-L., Sng, N.J., Zupanska, A.K., Krishnamurthy, A., Schultz, E.R., and Ferl, R.J. (2017b). Genetic dissection of the *Arabidopsis* spaceflight transcriptome: Are some responses dispensable for the physiological adaptation of plants to spaceflight? PLoS One **12**:e0180186. 10.1371/journal.pone.0180186.

Tran, H., Zhu, H., Wu, X., Kim, G., Clarke, C.R., Larose, H., Haak, D.C., Askew, S.D., Barney, J.N., Westwood, J.H., et al. (2018). Identification of differentially methylated sites with weak methylation effects. Genes (Basel) 910.3390/genes9020075.

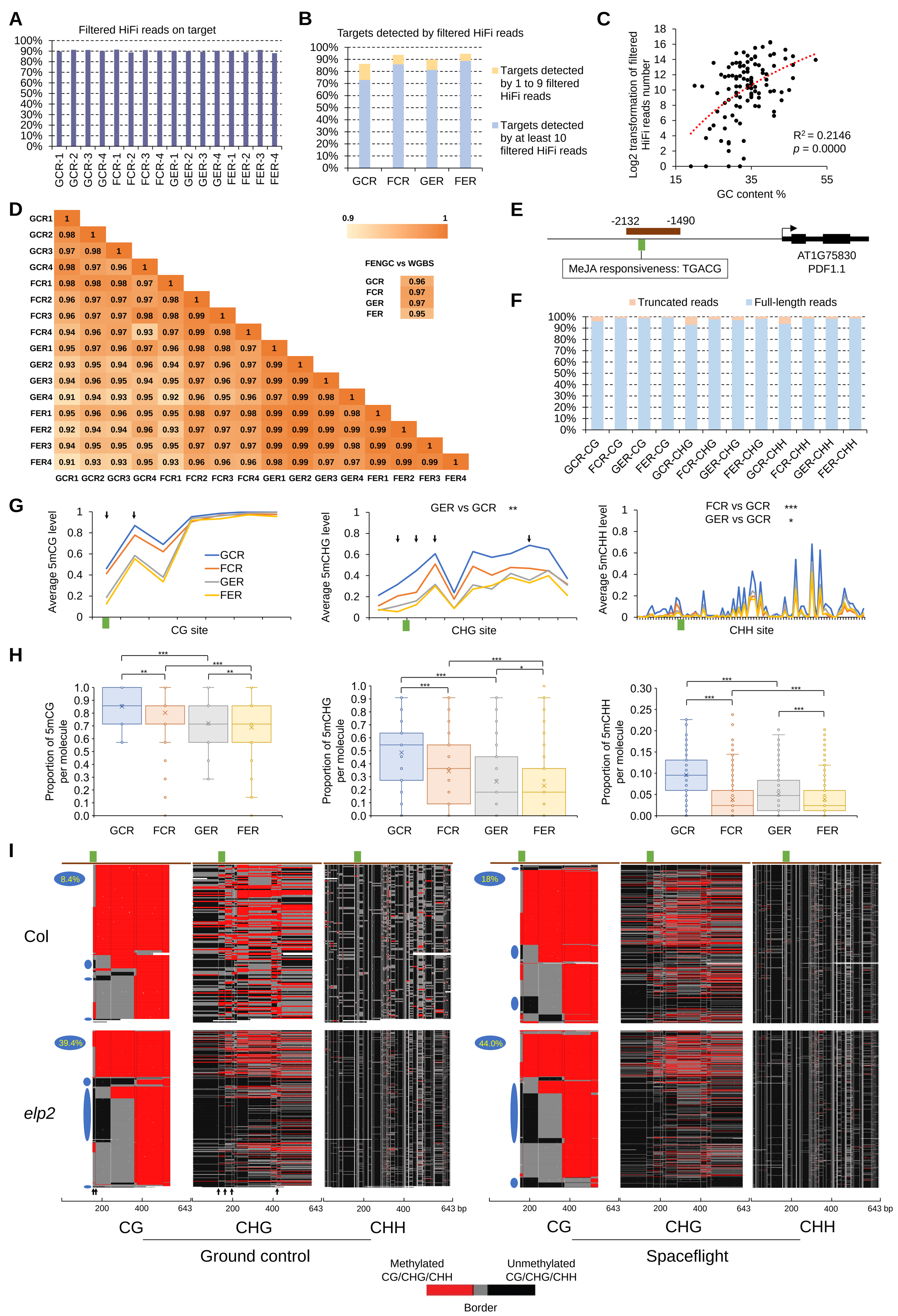
Wheeler, R. (2010). Plants for human life support in space: From Myers to Mars. Gravit. Space Biol. **23**:25-35.

Zabel, P., Bamsey, M., Schubert, D., and Tajmar, M. (2016). Review and analysis of over 40 years of space plant growth systems. Life Sci. Space Res. (Amst) **10**:1-16. 10.1016/j.lssr.2016.06.004.

Zhang, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. **19**:489-506. 10.1038/s41580-018-0016-z.

Zhou, M., Nabilsi, N.H., Wang, A., Gauthier, M.-P.L., Murray, K.O., Azari, H., Owens, W.S., Newman, J.R.B., Pardo-Palacios, F.J., Conesa, A., Riva, A., Clanton, T.L., Reynolds, B.A., Concannon P., Brant, J.O., Bacher, R., Kladde, M.P. (2022). Flap-enabled next-generation capture (FENGC): precision targeted single-molecule profiling of epigenetic heterogeneity, chromatin dynamics, and genetic variation. bioRxiv. doi: https://doi.org/10.1101/2022.11.08.515732.

Zhou, M., Sng, N., LeFrois, C., Paul, A.-L., and Ferl, R. (2019). Epigenomics in an extraterrestrial environment: organ-specific alteration of DNA methylation and gene expression elicited by spaceflight in *Arabidopsis thaliana*. BMC Genomics 2010.1186/s12864-019-5554-z.


A
Filtered HiFi reads on target
B
Targets detected by filtered HiFi reads
Targets detected by 1 to 9 filtered HiFi reads
Targets detected by at least 10 filtered HiFi reads
GCR FCR GER FER
C
Log2 transformation of filtered HiFi reads number
GC content %
R² = 0.2146
p = 0.0000
D
FENGC vs WGBS
GCR 0.96
FCR 0.97
GER 0.97
FER 0.95
E
-2132 -1490
MeJA responsiveness: TGACG
AT1G75830
PDF1.1
F
Truncated reads
Full-length reads
G
Average 5mCG level
CG site
GCR
FCR
GER
FER
GER vs GCR **
Average 5mCHG level
CHG site
FCR vs GCR ***
GER vs GCR *
Average 5mCHH level
CHH site
H
Proportion of 5mCG per molecule
Proportion of 5mCHG per molecule
Proportion of 5mCHH per molecule
I
Col
elp2
8.4%
39.4%
18%
44.0%
CG
CHG
CHH
Ground control
Spaceflight
Methylated CG/CHG/CHH
Unmethylated CG/CHG/CHH
Border