

Approaches for Surveying Cosmic Radiation Damage in Large Populations of *Arabidopsis thaliana* Seeds – Antarctic Balloons and Particle Beams

Brandon Califar^{1,3,4}, Rachel Tucker¹, Juliana Cromie¹, Natasha Sng¹, R. Austin Schmitz¹, Jordan A. Callaham¹, Bradley Barbazuk^{2,3,5}, Anna-Lisa Paul^{1,2,3,4}, and Robert J. Ferl^{1,2,3,4,5}

¹Horticultural Sciences, University of Florida, Gainesville, FL; ²Program in Plant Molecular and Cellular Biology, University of Florida, Gainesville, FL; ³The Genetics Institute, University of Florida, Gainesville, FL; ⁴Program in Genetics and Genomics, University of Florida, Gainesville, FL; ⁵Interdisciplinary Center for Biotechnology and Research, University of Florida, Gainesville, FL.

ABSTRACT

The Cosmic Ray Exposure Sequencing Science (CRESS) payload system was a proof of concept experiment to assess the genomic impact of space radiation on seeds. CRESS was designed as a secondary payload for the December 2016 high-altitude, long-duration south polar balloon flight carrying the Boron and Carbon Cosmic Rays in the Upper Stratosphere (BACCUS) experiment. Investigation of the biological effects of Galactic Cosmic Radiation (GCR), particularly those of ions with High-Z and Energy (HZE), was of interest due to the genomic damage this type of radiation inflicts. The biological effects of radiation above Antarctica (ANT) were studied using *Arabidopsis thaliana* seeds and compared to a simulation of GCR at Brookhaven National Laboratory (BNL) and to laboratory control seeds.

The CRESS payload was broadly designed to 1U CubeSat specifications (10 cm x 10 cm x 10 cm, ≤1.33 kg), maintained 1 atm internal pressure, and carried an internal cargo of 580,000 seeds and twelve CR-39 Solid-State Nuclear Track Detectors (SSNTDs). Exposed BNL and ANT M₀ seeds showed significantly reduced germination rates and elevated somatic mutation rates when compared to non-irradiated controls, with the BNL mutation rate also being higher than that of ANT. Genomic DNA from plants presenting distinct aberrant phenotypes was evaluated with whole-genome sequencing using PacBio SMRT technology, which revealed an array of structural genome variants in the M₀ and M₁ plants. This study was the first whole-genome characterization of space-irradiated seeds and demonstrated both the efficiency and efficacy of Antarctic long-duration balloons for the study of space radiation effects on eukaryote genomes.

Key words: Radiation Damage; Cosmic Radiation; Arabidopsis Seeds; High-Altitude Balloon; Antarctica; Mutant Screening; Somatic Mutations; PacBio SMRT sequencing; Genomic Rearrangements; Structural Variants; Advanced Life Support; Spaceflight

Correspondence to: Robert J. Ferl
Interdisciplinary Center for Biotechnology
Research
University of Florida
Gainesville, FL 32601
Telephone: 352-273-0829
E-mail: robferl@ufl.edu

INTRODUCTION

Galactic Cosmic Rays (GCR) and Solar Cosmic Radiation (SCR) constitute two major risks for life forms of all types in the spaceflight environment. The biological effects of GCR and SCR at various dosage levels have been studied extensively to elucidate the mechanisms involved in repairing the DNA Double Strand Breaks (DSBs) typical of these exposures (Arena et al., 2014; Kavanagh et al., 2013). Interactions between these two types of Ionizing Radiation

(IR) and microgravity-induced effects on the DNA damage repair response have been studied in both microgravity simulations and true spaceflight experiments, but no consistent synergistic effects have been identified (Moreno-Villanueva *et al.*, 2017). Further research and application of this knowledge will contribute to efforts to effectively safeguard the genomic integrity of astronauts and other organisms beyond the protection of Earth's magnetic field (Durante, 2014; Kennedy, 2014). Research into the effects of radiation remains a high priority for the further advancement of crewed space exploration, both in terms of reducing risk to the mental and physical health of humans, as well as the viability of plants and other attendant biology vital to the support of long duration missions and exploration (NASA, 2018).

Ions of high-Z (charge) and energy (HZE) constitute 1% of GCR; the remainder is a combination of protons (85-90%), helium (11%), and electrons (1%) (Badhwar and O'Neill, 1992). HZE include the electron-stripped nuclei of Lithium ($Z = 3$) and all heavier elements, and pose significant danger to biological samples due to their capability for high levels of Linear Energy Transfer (LET), which causes damage along their paths. These impacts can be detected either in real-time or post-experimentally, and an array of hardware capable of measuring IR are available (Benton and Benton, 2001). Columbia Resin 39 (CR-39) is one such detector, allowing for post-exposure characterization of IR impacts. CR-39 is a polyallyl diglycol carbonate (PADC) polymer, which passively records HZE passage via bond breakage within its superstructure (Cartwright *et al.*, 1978).

Efforts to replicate the spaceflight radiation environment in an accurate manner for purposes of advancing the field of radiation biology continue to be made in the NASA Space Radiation Laboratory (NSRL) at Brookhaven National Laboratory (BNL) (Miller and Zeitlin, 2016). Bombardment of biological samples with radiation of known ion composition and dosage rate allows for tightly-controlled experimental designs, the results of which can be extrapolated to support downstream studies of DNA damage resulting from cosmic radiation in more complex environments (Kiffer *et al.*, 2017; Suman *et al.*, 2017).

The upper polar stratosphere of Antarctica provides a useful environment for longer-term studies of IR and its effects, without requiring flight on a rocket beyond Low Earth Orbit. Access to this radiation environment is accomplished using various types of scientific balloons (Smith and Sowa, 2017). An average dose rate of $5.47 \pm 0.31 \mu\text{Gy/hr}$ was calculated from data acquired in the 2003 balloon flight of the TRACER (Transition Radiation Array for Cosmic Energetic Radiation) instrument hardware in this environment (Ave *et al.*, 2011; Benton, 2012). These low dosages thus require a longer exposure time to reach the same absorbed dose as an artificial radiation source of increased intensity.

Seeds, including those of the model organism *Arabidopsis thaliana*, have historically served as a space-efficient biological recorder of radiation impact events in spaceflight conditions, with assessments of biological effects occurring post-exposure through germination rates of irradiated seeds and frequencies of mutation (Kranz, 1986; Tepfer and Leach, 2017; Tepfer *et al.*, 2012). The small genome size of *A. thaliana* (~135 Mbp) and the concomitant high gene density of 4.4 kb/gene (Wang *et al.*, 2015) facilitate the identification of HZE-mutagenized individuals through an increased probability for coding sequence loss or rearrangement to result from damage. Additionally, recent studies have compared radiation environments with varying LET and the effects that each had on the stability of the *A. thaliana* genome, and found that HZE-induced damage was more closely clustered in the genome and resulted in a greater loss of heterozygosity than γ -radiation (Huefner *et al.*, 2014). Exposure to various ions with identical LET values has also been shown to produce identical effects on mutation rates and flowering (Kazama *et al.*, 2008). However, there has not yet been a whole-genome interrogation of genomic structural variants (SVs) and mutations using third-generation (single-molecule) sequencing approaches in morphologically distinct *A. thaliana* mutants resulting from exposure to IR.

The Cosmic Ray Exposure Sequencing Science (CRESS) payload system was developed as an end-to-end proof of concept experimental payload aimed at examining large numbers of seeds exposed to space radiation environments. The hardware was developed to fly on high-

altitude balloon flights launched from Antarctica. The design of the hardware was based on components that could also be adapted for other experimental needs. The post-flight procedures were designed to assay the effort and components necessary for the screening and processing of large quantities of seeds. Long-read DNA sequencing technologies were applied to tissues arising from somatic mutants in order to assess the ability of the exposure and screening procedures to provide novel insights into the genomic impacts of space radiation exposure.

MATERIALS AND METHODS

Plant Material

The *Arabidopsis thaliana* ecotype Columbia-0 (Col-0, CS70000) seed stock was purchased from the Arabidopsis Biological Resource Center (ABRC) through The Arabidopsis Information Resource (TAIR) website. Seeds were then produced *en masse* by allowing these Col-0 plants

to self-pollinate in growth chambers at $19 \pm 2^\circ\text{C}$ under 24 h fluorescent light at approximately $80 \mu\text{mol m}^{-2} \text{s}^{-1}$ PAR. Harvested seeds were cleaned, placed in a desiccator for 5-7 days at room temperature, and then stored at 4°C in an O-ring-sealed microcentrifuge tube until use. All experimental and control seed originated from the CS70000 seed stock.

Hardware and Payload Configuration for High-Altitude Balloon Flight Over Antarctica

Hardware design for the CRESS payload was conducted with several critical specifications in mind. The containment vessel was designed to maintain an internal pressure of approximately 1 atm, record the impacts of IR, meet the 1U CubeSat standards, and carry a well-organized seed container to facilitate highly controlled post-flight processing and analysis. The payload layout can be seen in Figure 1. The basic unit of the payload is the seed cassette, which is composed of

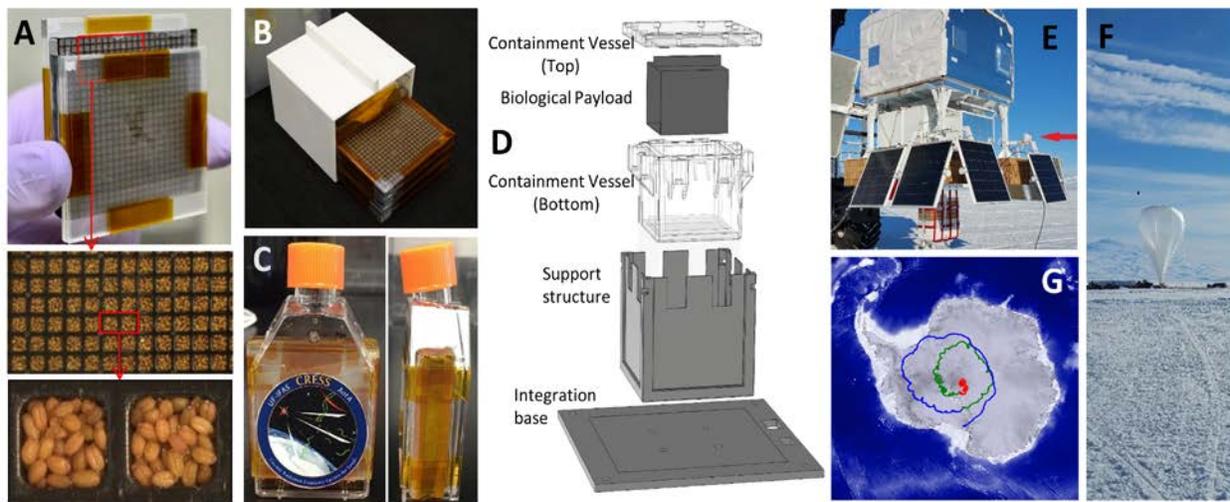


Figure 1. Hardware and payload design for Brookhaven National Laboratory (BNL) exposure and Antarctica (ANT) high-altitude balloon flight. (A) A representative image of a 24 x 24 bin seed tray flanked by CR-39 Solid-State Nuclear Track Detector (SSNTD) sheets. The seed trays are modified 1536-well plates. Magnified views of the seed trays are shown below, including a close view of two individual bins. A single bin contains approximately 250 *Arabidopsis thaliana* seeds. (B) The custom 3D-printed holder used in the Antarctica payload and the biological payload it contained. (C) The modified T-25 flask containing the biological payload for irradiation at the NASA Space Radiation Lab (NSRL), located at the Brookhaven National Laboratory. (D) A diagram illustrating the individual components and configuration of the entire ANT payload for the high-altitude balloon flight in Antarctica. (E) The ANT payload (indicated by a red arrow) integrated onto the Boron And Carbon Cosmic Rays in the Upper Stratosphere (BACCUS) support structure. (F) Inflation of the balloon on the ice. (G) The path of the BACCUS payload over Antarctica, where blue indicates earlier time points and red indicates time points at the end of the flight. [Photos in E and R courtesy of Scott Miller – NASA CSBF]

a seed tray flanked by sheets of CR-39 Solid-State Nuclear Track Detector (SSNTD) material [Track Analysis Systems Ltd] (Figure 1A). The seed trays are modified Nunc™ 1536-bin plates [ThermoFisher Scientific]. Each seed tray was hand-cut from the original 1536-bin plate and contains 576 bins (24 x 24 bin orientation). Each bin holds an average of 250 vertically stacked seeds (Figure 1A bottom). To secure the seeds in the tray, a polyolefin sealant film [USA Scientific] was placed over each loaded seed tray. The ANT biological payload of approximately 500,000 *A. thaliana* ecotype Col-0 seeds was composed of four seed cassettes. To determine the locations of potential IR hits, CR-39 sheets were placed adjacently to the four seed trays – three on each end of the stack, and two between each tray. All four cassettes were inserted within a custom-made 3D-printed holder (Figure 1B). For the NASA Space Radiation Lab (NSRL) experiment, a single seed cassette, such as that shown in Figure 1A, was secured inside a modified T-25 flask (Figure 1C) and presented to the beam.

Figure 1D shows the components of the ANT payload containment system. The biological payload was bolted into a 3D printed containment vessel made from VeroClear™ resin [Stratasys], which was constructed using an Objet Eden 3D printer [Stratasys] utilizing PolyJet™ technology. This design was necessary to ensure that the containment vessel could hold 1 atm of internal pressure. To allow ease of integration onto the BACCUS hardware, the containment vessel was set into a 3D-printed support structure and then bolted onto the integration plate; both structures were produced using a CraftBot Plus 3D printer

[CraftUnique], which utilizes Filament Deposition Modeling (FDM) technology. The assembled payload was bolted onto a corner of the BACCUS support structure (Figure 1E). Figure 1F shows the balloon being filled with helium on the ice, and Figure 1G shows the path of the balloon over Antarctica.

Exposure to Radiation Environment

For this proof of concept experiment and comparison, the full CRESS payload that flew in Antarctica is referred to as ANT. The payload elements that were exposed at the NSRL at the Brookhaven National Laboratory are referred to as BNL. In the figures and tables ANT will refer to data from the Antarctica exposure while BNL will refer to data from the NSRL exposure.

BNL

A single seed cassette, which contained approximately 145,000 seeds, was encased in three CR-39 sheets on each side, and this assembly was placed inside a modified T-25 flask (Figure 1C). Particle accelerator time was reserved at NSRL and a specific ion dosage designed to simulate GCR was requested, as outlined in Table 1. The holder was oriented perpendicularly to the radiation source, and particles were projected in unidirectional paths so that ions passed in a straight line through the assembly. The beam had an area of effect of 20 cm x 20 cm, such that only a portion of the total dosage would impact the biological payload. Seeds were exposed to H, He, O, and Ti ions via the particle accelerator and further details of the entire BNL mixed-beam can be found in Table 1.

Table 1. Profile of the mixed-beam for cosmic radiation simulation used in the Brookhaven National Laboratory (BNL) seed treatment.

Ions	Energy (MeV)	Dose (cGy)	Dose rate
H	250	10.60	0.166
He	250	3.170	0.166
O	350	1.080	0.166
Ti	300	0.735	0.166

ANT

The assembled CRESS payload was en route to Antarctica via the NASA Wallops Flight Facility and New Zealand from November 2, 2016 to November 16, 2016. The payload was removed from cold stowage and integrated onto the BACCUS experimental hardware six hours before launch (Figure 1F). The high-altitude balloon supporting the experiments was then launched from the McMurdo polar station in Antarctica on November 28, 2016 and maintained an altitude between 36 and 40 km above sea level for most of the 30-day float period (Kim *et al.*, 2017) (Figure 1E, F, G). After this exposure to the upper-stratospheric radiation environment, the experiment landed and was recovered. The payload was stored at -20°C until it could be transported. The payload was returned to Gainesville, Florida on March 29, 2017; all seeds remained stored in the hardware at -20°C during shipment and after arrival.

Payload De-Integration

From the ANT experiment, the top seed cassette (#1) was removed from the hardware and stored at 4°C for downstream assessment, while the other three seed cassettes remained in -20°C storage. Seed tray #1 was chosen due to its position at the top of the biological payload, where it was likely to receive a greater amount of radiation than the trays below. The top three CR-39 SSNTDs placed above this tray (#8-10) were also removed and stored at 4°C for further analysis. Eighteen bins were randomly selected from both the BNL and ANT seed trays, with 36 total bins being used for phenotypic screening.

Etching of CR-39

The CR-39 sheets were etched with 6M NaOH for 8 hours at 82°C. Etching in a heated NaOH solution increases the diameter of the aperture created by the penetration, and this remains proportional to the size of the incident particle (Cartwright *et al.*, 1978). Post-etching images of CR-39 track detectors were taken with an Olympus SZX12 stereomicroscope. The etched CR-39 sheets in Figure 2A-B show sections of CR-39 from the BNL and ANT exposures, respectively. The images in Figure 2C were taken

under an Olympus SZBX12 compound microscope and illustrate the post-etching track shape expected from particles impacting the plate in a non-perpendicular manner. In addition, the etched BNL CR-39 sheet showed a uniform angle of particle tracks, and these tracks covered almost all the seed tray bins. This was due to the unidirectional source of radiation from the particle accelerator. However, on the etched ANT CR-39, the angle of tracks was less uniform and X-ray imaging allowed better visualization of radiation impacting the payload in the upper stratosphere.

Seed Bin and CR-39 Alignment

An etched track detector from the ANT payload (#8) was scanned via X-ray Computed Tomography with the Phoenix v|tome|x m scanner [General Electric] and Datos Acquisition Software v. 2.4.0 [General Electric]. A 100 kV X-ray beam was directed at the CR-39 detector at a focus-object-distance (FOD) of 7.32 mm, and a 109x magnified image was transmitted onto a detecting plate opposite the X-ray source at a focus-detector-distance of 799.92 mm. The sheet was rotated approximately 360° and approximately 2500 images were taken with a pixel size of 200 μm^2 . Images were then imported into the Volume Graphics (v. 3.2) software and compiled into a three-dimensional digital reconstruction of the CR-39 sheet, with a pixel-to-voxel ratio of 1:1 (Figure 2D)

Seed Processing

Irradiated seeds were removed in a single mixed batch from a randomly selected bin of the seed tray, hereinafter referred to as the M_0 generation of seed as per convention. This process is illustrated in Figure 3A. M_0 seed batches were then transferred into individual 1.5 mL microcentrifuge tubes for wet sterilization. Briefly, seeds were soaked in 70% ethanol for 5 minutes, followed by 10 minutes in 50% bleach with 2 drops of Tween-20 added per 5 mL of solution, and rinsed with sterile water 6 times to remove residual chemicals (Figure 3B). After sterilization, M_0 seeds were stored in sterile water and vernalized for 4-5 days at 4°C before planting.

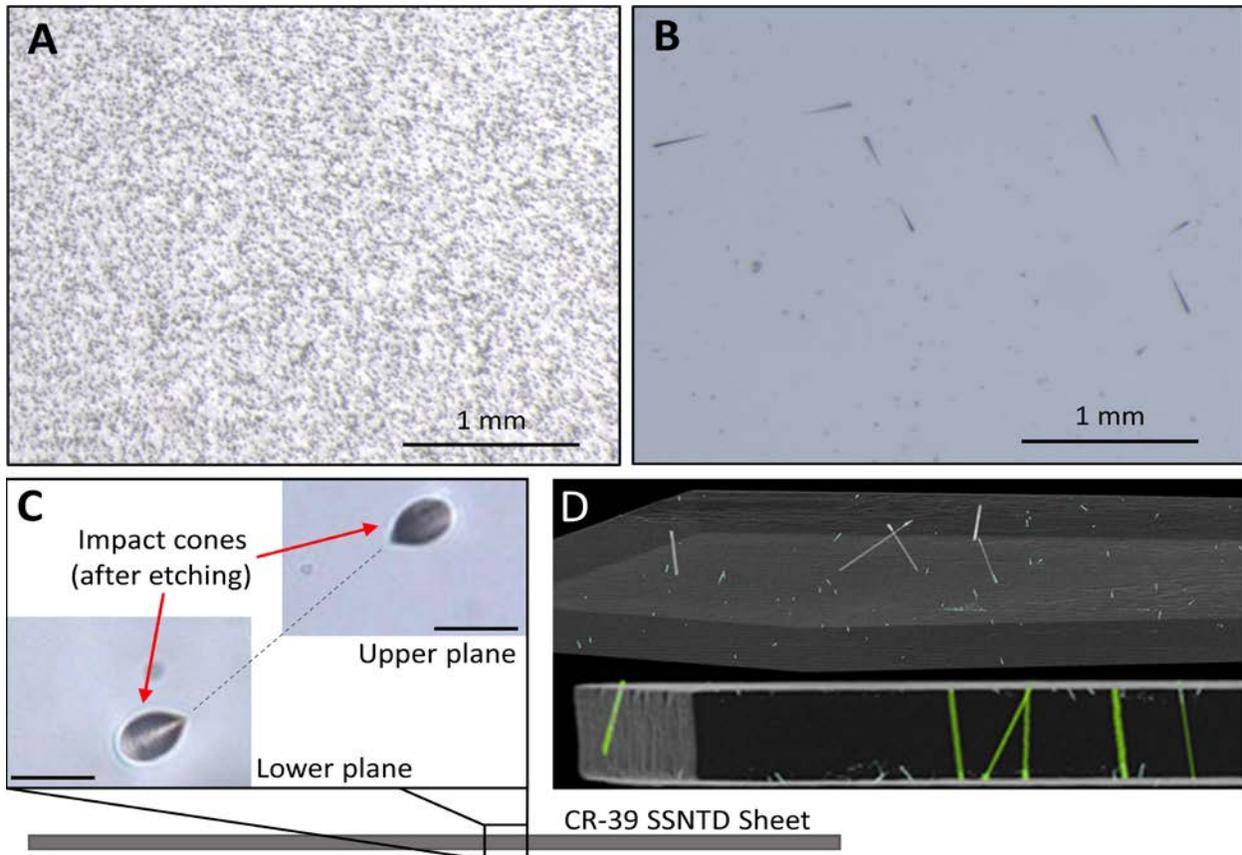


Figure 2. Etching and examination of CR-39 SSNTDs to assess impact locations. (A) A representative CR-39 SSNTD exposed to intense mixed-beam radiation at BNL and etched showed densely distributed and unidirectional particle tracks. Scale bar = 1 mm. (B) A representative ANT CR-39 SSNTD exposed to cosmic radiation in the upper stratosphere showed sparse entry and exit cones at varying angles. Scale bar = 1 mm. (C) Conical holes in both the upper and lower planes of the etched CR-39 SSNTDs, viewed under higher magnification, illustrate the tracks that are indicative of a particle’s passage through the sheet. Scale bars = 25 μm . (D) An image resulting from the X-ray Micro-Computed Tomography (Micro-CT) scan of a CR-39 SSNTD, illustrating how the paths of ions through the sheet can be mapped with high resolution.

Screening Protocols

Plate configurations and planting

A total of 122 M_0 seeds were planted on each Petri plate (150 mm x 15 mm Petri dishes [Fisher Scientific]) under sterile conditions (Figure 3C). A 0.5% PhytigelTM-based nutrient media supplemented with 0.5x Murashige and Skoog (MS) salts [Caisson Laboratories], 0.5% (w/v) sucrose [Sigma], and 1x Gamborg’s Vitamin Mixture [Caisson Laboratories] was used. Each bin of seeds required approximately two plates. A gridded label was adhered to the bottom of each plate, and one seed was planted per grid-square (1 cm^2) to ensure equidistant spacing and consistent seed counts. After planting, plates were sealed

with porous tape [MicroporeTM] to prevent contamination, labelled, and grown in growth chambers at $19 \pm 2^\circ\text{C}$, under 24-h fluorescent light at approximately $80 \mu\text{mol m}^{-2} \text{s}^{-1}$ PAR.

Scoring for germination and somatic mutation

Plates were scored for germination 14 days after planting and storage in the growth chamber (Figure 3D). Both the total number of seeds planted on the plate and the number that germinated were recorded. Plates were also generally screened at the 14-day time point to identify signs of phenotypic differences. Examples of phenotypic aberrations that were scored for include: microgrowth (crown diameter < 8 mm), presence of a single cotyledon only, absence of plant structures, and alterations of

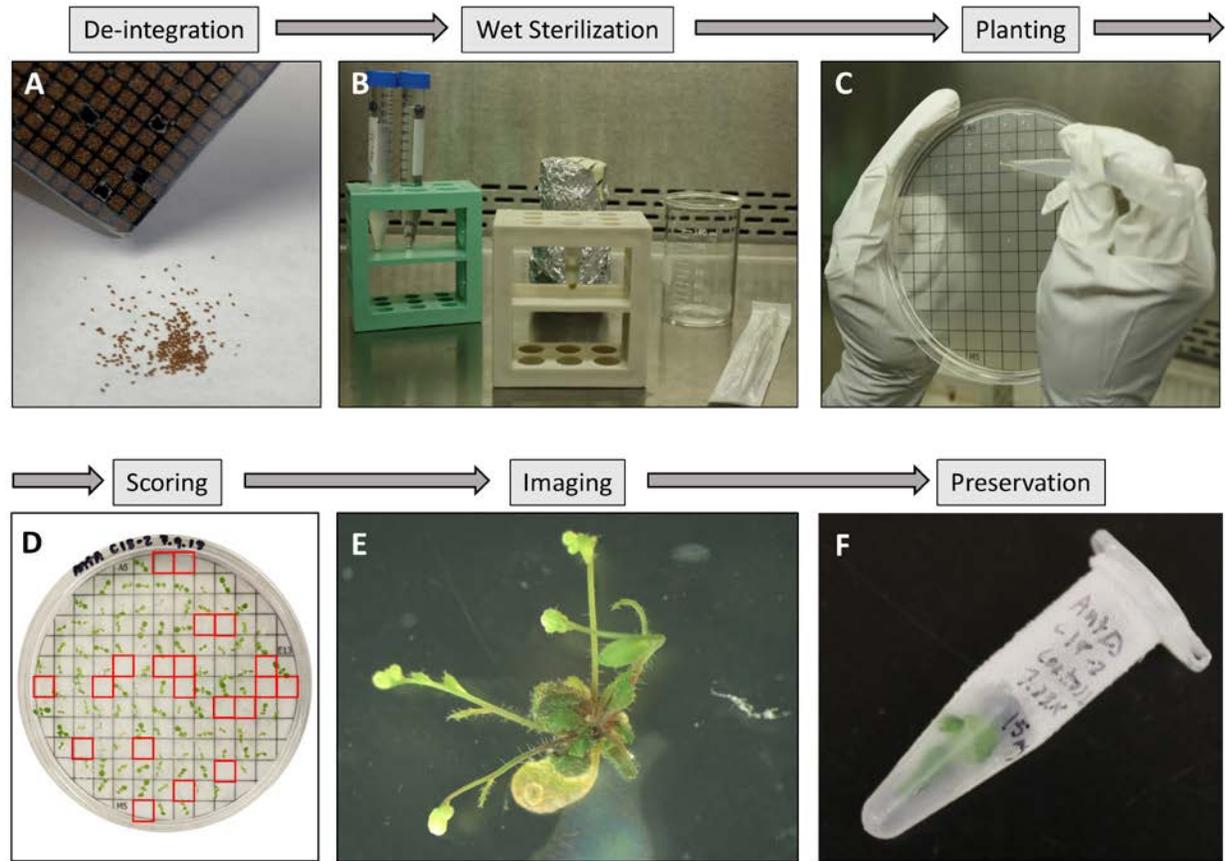


Figure 3. Seed processing and screening procedures. (A) M₀ seeds were de-integrated from specific bins in the seed tray(s) and organized into individual microcentrifuge tubes. **(B)** De-integrated seeds were wet sterilized with 70% ethanol and 50% bleach. **(C)** Once sterilized, seeds were planted on gridded Petri plates containing 0.5% Phytigel™ nutrient media. **(D)** After 14 days of growth, M₀ plants were scored for germination and phenotypic traits; red-outlined grids highlight non-germinated seeds. **(E)** Atypical plants were imaged and tracked for further observation. **(F)** Atypical plants that displayed severely inhibited growth and development were frozen with liquid nitrogen.

pigmentation. Atypical plants demonstrating sufficient health and size after scoring were transferred to soil for the generation of M₁ seeds via self-pollination, while plants with stunted growth and/or absent organs were transferred to individual Phytigel™ plates for further growth and observation. Both of these groups were grown in the same conditions described above. Eventually, plants that displayed severely inhibited growth and development were flash-frozen in liquid nitrogen and stored at -80°C for subsequent DNA extraction. Any plates displaying contamination were discarded and no data were recorded.

To allow study of the effects of radiation damage on subsequent generations, a subset of

bins from the BNL seed tray were dedicated to producing M₁ generation seeds. Seeds harvested from the BNL seed tray were chosen for this analysis as the etched BNL CR-39 detectors showed a dense and uniform distribution of impacts that covered almost every bin in the seed tray. M₀ seeds allotted for M₁ analyses were de-integrated, wet-sterilized, and planted directly onto soil. M₀ seeds remained separated by bin throughout seed processing and growth, and seeds harvested from M₀ plants were stored according to the bin from which their parental plant originated. A total of 13 M₁ seed stocks were collected, and 244 seeds were planted from each stock. M₁ plants were screened using the same parameters as M₀ plants.

Photography

Survey images of all plates were taken at 4-, 8-, and 14-day time points using an Axiocam 506 Color camera (Carl Zeiss Microscopy). Plants that displayed atypical phenotypes at the time of scoring were also imaged at later time points under an Olympus SZX12 stereomicroscope (Figure 3E) and tracked until maturity or until preservation (Figure 3F). Anomalous phenotypes that were discernible prior to the 14-day old time point were tracked and imaged at earlier time points.

Nomenclature for files

Immediately following seed planting, plates were labelled by generation (e.g., M₀, M₁), test group, (e.g., ANT, BNL), bin (i.e., the location of the bin on the seed tray), and the numerical ID of the Petri plate on which the plant was grown. Individual plants were also denoted by their location within the Petri plate grid, which was labelled using the Cartesian coordinate system with alphabetical labels along the y-axis and numerical labels along the x-axis. This system of nomenclature was developed to organize image files over the course of the screening process and to serve as an effective tracking method from seed de-integration to DNA sequencing. All image filenames included the generation and origin of the seedling, as well as the date of imaging and the age of the seedling at the time of imaging. A sample file name of 180812_M₀_ANT_F3-2_L12_8D would indicate that the image was taken August 12, 2018 of plant L12 at 8 days of age that originated from the F3 bin of the ANT seed tray.

DNA Extraction

Genomic DNA was extracted from selected mutants, which are displayed in Figure 4, for sequencing. Using the aforementioned system, these mutants were named: Mutant1-M₀ANT W3-2 D7 (Figure 4A), Mutant2-M₀ANT H16-2 K9 (Figure 4B), Mutant4-M₁BNL H13-1 H10 (Figure 4C), Mutant5-M₁BNL P18-3 F3 (Figure 4D), and 6 pooled mutants collectively referred to as Mutant7-M₀ANT-Pooled-Mutants (Figure 4E-J). An ethanol precipitation-based method adapted from a previous publication was used (LeFrois *et al.*, 2016). Briefly, either fresh or previously

frozen mutant plants were ground directly in liquid nitrogen without rinsing in wash buffer. The aforementioned protocol, beginning with the addition of the lysis buffer, was followed accordingly until the phenol/chloroform/isoamyl alcohol extraction step, which was eliminated. After the pellets were suspended in Tris/EDTA/RNase A (TER) buffer and incubated at room temperature for 20 minutes, 7.5M ammonium acetate (AmOAc) and 95% ethyl alcohol (EtOH) were added to the tubes, inverted to mix, and placed at -20°C overnight. As described in the protocol, the tubes were then spun at 4°C for 10 minutes, the supernatant decanted, and the pellets thrice rinsed with 70% EtOH. The samples were dried on the benchtop for 15 minutes and suspended in 0.25x TE buffer. The concentrations of the DNA samples were quantified using a Qubit fluorometer [Invitrogen] and these values are displayed in Table 2.

DNA Sequencing

Barcoded, large-insert (10 kb) libraries were constructed using 600 ng of pure, high MW DNA (>40 kb) from the five different *Arabidopsis* mutants (Mutant1-M₀ANT W3-2 D7, Mutant2-M₀ANT H16-2 K9, Mutant4-M₁BNL H13-1 H10, Mutant5-M₁BNL P18-3 F3, and Mutant7-M₀ANT-Pooled-Mutants). DNA preparations typically had an ABS 260/280 ratio of 1.8-2.0, and an ABS 260/230 ratio of 2-3. The integrity of the samples was evaluated on the Agilent TapeStation using a genomic tape. A slightly modified protocol from the one recommended by PacBio for multiplex SMRT sequencing of bacterial genomes (manual PN 101-069-200-02) was used, in conjunction with barcodes from the Barcoded Adaptor kit 8B (PN101-475-100). Briefly, DNA (150 µl) was sheared down to 10 kb using Covaris g-TUBES (Covaris Inc. Cat# 520079), using two passes at 7,000 rpm. The resulting size of the fragments was verified on the Agilent TapeStation (genomic tape). The library construction reactions for individual samples consisted of the following sequential steps: ExoVII treatment, DNA damage repair, end repair, and blunt-end ligation of barcoded SMRTbell adaptors. After ligation, all samples were pooled, AMPure-cleaned, and submitted to an ExoIII/ExoVII treatment to eliminate excess adaptors and any damaged DNA. This procedure

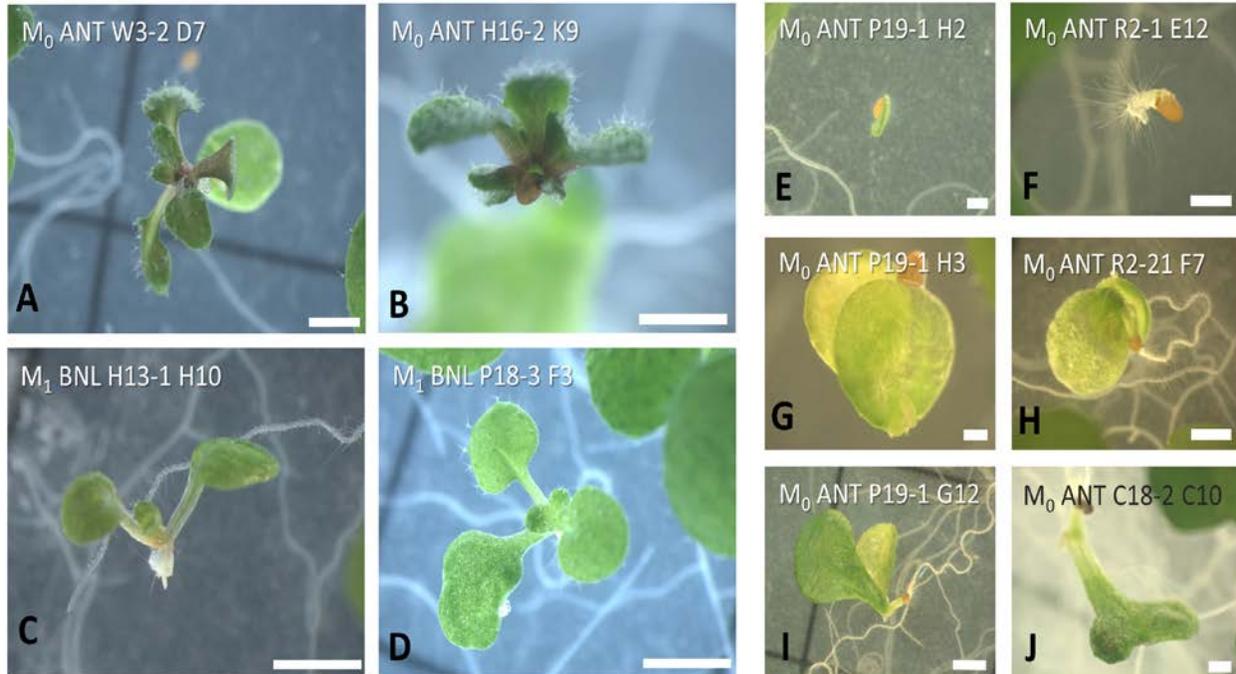


Figure 4. Plants exhibiting mutant phenotypes selected for PacBio sequencing. Plants with a sufficient amount of fresh tissue weight to provide 1 µg of DNA for library generation (A-D) were sequenced individually. (A) Plant M₀ ANT W3-2 D7 and (B) M₀ ANT H16-2 K9 both exhibited development without growth of a primary root. (C) M₁ BNL H13-1 H10 had delayed root growth by 14 days. (D) M₁ BNL P18-3 F3 had only a single cotyledon, which then showed malformation. Plants from (E-J) were pooled due to the small amount of DNA extracted from each sample. (E) M₀ ANT P19-1 H2 developed a single cotyledon with no other shoot or root structures, whereas (F) M₀ ANT R2-1 E12 grew only a primary root. (G) M₀ ANT P19-1 H3 cotyledons emerged without root growth. (H) M₀ ANT R2-2 F7 and (I) M₀ ANT P19-1 G12 both had cotyledons and a primary root only. (J) M₀ ANT C18-2 C10 exhibited a fused cotyledon morphological mutation. All images are of plants aged 14 days old, with the exception of (B) for which only a 16-day-old image was available. Scale bars = 2 mm.

Table 2. Quantification of gDNA extractions via Qubit.

Sample	Fresh weight (mg)	Qubit Conc. (ng/µl)	Volume (µl)	DIN	Sample mass (ng)
Mutant1-M ₀ ANT W3-2-D7	10	24.2	28	6.6	678
Mutant2-M ₀ ANT H16-2-K9	10	42.0	28	6.4	1176
Mutant4-M ₁ BNL H13-1-H10	10	25.6	30	6.6	768
Mutant5-M ₁ BNL P18-3-F3	20	49.4	30	6.8	1482
Mutant7-M ₀ ANT (Pooled)	15	15.7	53	5.9	833

DIN: DNA Integrity Number

resulted in 350 ng of adaptor-ligated SMRTbell library. The final library was further size-selected in the SageELF™ instrument (Cat# ELD 7510), using 0.75% agarose gel cassettes and the 1-18 kb v2 cassette definition program. Fractions in wells 5-9 contained library fragments in the 5-20 kb range with ~10 kb average (TapeStation). The desired SageELF fractions were cleaned using AMPure magnetic beads (0.6:1.0 beads to sample ratio) and eluted in 15 µl of 10 nM Tris HCl, pH 8.0. Sequencing was done using the Chemistry Bundle 5.0.1, which includes Sequencing and Binding kits 2.1 in combination with the SMRTLink 5.1 software. Library at 8 pM was applied to the PacBio SEQUEL sample plate by diffusion loading for sequencing on an LR-SMRT cell with 20-hr data collection. All other steps for sequencing were performed according to the recommended protocol by the PacBio sequencing calculator. The sequencing output was 8.04 Gb with an average polymerase read length of 14.7 kb. Approximately 6.7 Gb of the data were contained in the demultiplexed reads.

Sequencing Data Analysis

The demultiplexed subreads from this PacBio SEQUEL run were aligned to the TAIR10 release of the *A. thaliana* ecotype Col-0 reference genome (Berardini *et al.*, 2015) with the coNvex Gap-cost alignMents for Long Reads (NGMLR) mapping software (Ver. 0.2.7) using default parameters (Sedlazeck *et al.*, 2018). The output was then submitted to NGMLR's companion SV detection software Sniffles (Ver. 1.0.8), which was run with a threshold of 5 supporting split-reads required to call an SV (default = 10). Two simulated control datasets were also derived from publicly available data (PRJNA314706: SRR3405265, SRR3405288, and SRR3405289; PRJNA418774: SRR6298199), both being *A. thaliana* ecotype Col-0 resequencing experiments, with each of these subsets having approximately 1.5 Gb total, similar to that of the average sample in our dataset (Chin *et al.*, 2016; Liang *et al.*, 2018). The resultant datasets were hand-curated to remove artefactual SVs and those unrelated to IR damage: translocations between the nuclear and organellar genomes, SVs shared among all of the samples, and spurious SV calls that resulted from regions of excessive (e.g., >500x) coverage. The Integrative Genomics Viewer (Ver. 2.4) was used

in conjunction with the filtered datasets to examine SVs less than 1Mb in size (Robinson *et al.*, 2017), and Circos (Ver. 0.69-2) was used for the visualization of translocations (Krzyszowski *et al.*, 2009). The SAMtools utility set (Ver. 1.9) was used for all file format conversions, some of which were required for compatibility (Li *et al.*, 2009).

RESULTS

Germination Rates

For each experimental group, 18 bins were de-integrated from the seed trays and used for determining rates of germination. Each bin contained an average of 250 seeds. As can be seen in Figure 5A, M₀ BNL and ANT seeds displayed mean germination rates of 76.4% and 82.5%, respectively. In the non-irradiated controls, the mean germination rate was 98%, which is consistent with reports from other studies (Kazama *et al.*, 2008). In both M₀ BNL and ANT, the germination rate was significantly lower than the non-irradiated controls (Student's *t*-test: $p < 0.01$). However, no significant differences in germination rate between the M₀ BNL and ANT samples were identified (Student's *t*-test: $p > 0.05$).

To determine the effects of radiation on the next generation of seeds, the germination rates for M₁ BNL seeds were also measured. The box-and-whisker plot in Figure 5B shows that germination rates for the M₁ generation of sampled BNL bins were similar to the control group (at 98%) and significantly higher than the M₀ BNL seeds (Student's *t*-test: $p < 0.01$).

Frequency of Somatic Mutation

Mutation rate was calculated based on the number of phenotypic aberrations observed per 100 germinated seeds to avoid data skewing due to plants that displayed multiple aberrations. As was mentioned above, the mutation rate was calculated from screening 18 randomly selected bins, each containing roughly 250 seeds. Phenotypic aberrations were defined as observable traits that deviated in a significant manner from that of normally developing seedlings. These included color differences, developmental anomalies, and substantial size differences. Absence of true leaves, an ablated primary root or a substantial delay in root growth, and lack of cotyledons were

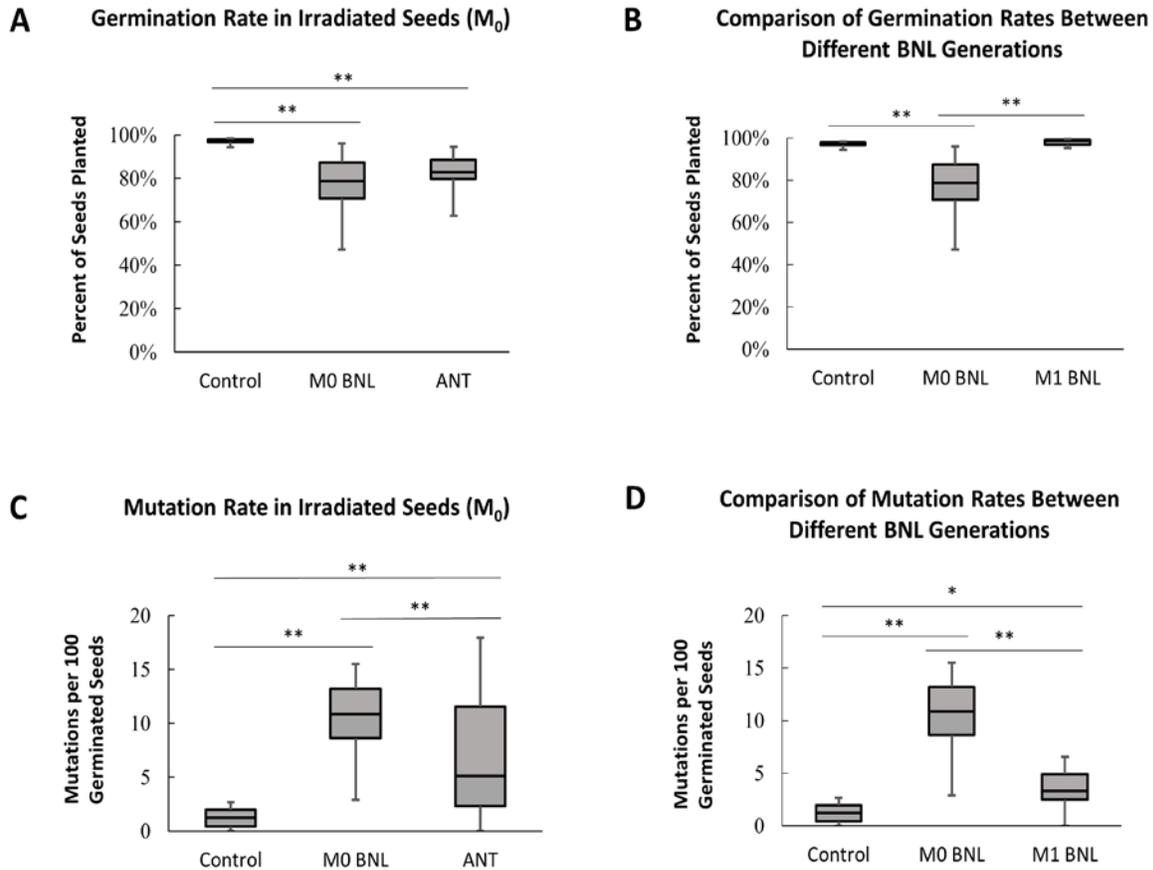


Figure 5. Germination and mutation rates in irradiated seeds and plants. *A. thaliana* seeds irradiated at BNL and ANT exhibited significantly different germination and mutation rates from a non-irradiated control group at 14 days post-planting. (A) M_0 Seeds exposed to heavy ions from a laboratory particle accelerator at BNL had a mean germination rate of 76%, whereas naturally occurring stratospheric radiation (ANT) resulted in a mean germination rate of 82%. The non-irradiated control had a mean germination rate of 98%. These results were statistically significant when compared to a control group. (B) Data for germination rates were also collected for the next generation (M_1) of irradiated BNL seeds. Germination rates for the M_1 BNL seeds were similar to the control group (at 98%) and significantly higher than the M_0 BNL seeds. (C) Mutation rate was calculated based on the number of phenotypic aberrations observed per 100 germinated seeds. The mean mutation rates for the M_0 of BNL and ANT were 12 and 4 mutations per 100 germinated seeds, respectively. In the control, the mean mutation rate was significantly lower at 1 per 100 seeds. (D) The BNL M_1 showed a mean rate of 6 mutations per 100 seeds, which was significantly higher than the control, but significantly lower than the BNL M_0 plants. Significance was determined via a Student's *t*-test and is denoted by asterisks (* = $p < 0.05$ and ** = $p < 0.01$).

the developmental aberrations most prominently observed in the plants grown from both ANT and BNL M_0 seeds.

The box-and-whisker plot in Figure 5C shows the mean numbers of mutations observed in the M_0 plants per 100 germinated seeds, as well as the

overall distribution of mutation rates among bins. The non-irradiated controls showed a mean mutation rate of 1 event per 100 germinated seeds, whereas in the irradiated BNL and ANT M_0 samples, the mean mutation rates were significantly higher than the control group at 12

and 4 aberrations per 100 seeds, respectively (Student's *t*-test: $p < 0.01$). Significant differences in mutation rate between M_0 BNL and ANT were also observed (Student's *t*-test: $p < 0.01$).

To determine radiation-induced effects on the next generation of seeds, mutation rates were also determined for the M_1 generation of BNL plants. The box-and-whisker plot in Figure 5D shows that BNL M_1 sample had a significantly higher mutation rate than the non-irradiated control plants (Student's *t*-test: $p < 0.05$) at a mean of 6 aberrations per 100 germinated seeds, which was also significantly lower than that observed in the M_0 BNL seeds (Student's *t*-test: $p < 0.01$).

Genomic Damage

As a proof of concept for the application of long read DNA sequencing technologies for detecting genomic structural variations in irradiated tissues, genomic DNA from the following plant samples was isolated and sequenced (see Figure 4):

- Mutant1- M_0 ANT W3-2 D7 (Figure 4A)
- Mutant2- M_0 ANT H16-2 K9 (Figure 4B)
- Mutant4- M_1 BNL H13-1 H10 (Figure 4C)
- Mutant5- M_1 BNL P18-3 F3 (Figure 4D)
- Pool of 6 M_0 ANT mutants (Figures 4E, 4F, 4G, 4H, 4I, 4J)

Structural variants were detected in the genomic DNA from all five samples, with sizes that ranged from less than 100 bp to greater than 1 Mbp in length. Due to DNA repair of DSBs via non-homologous end joining (NHEJ) having potentially resulted in translocations and intrachromosomal SVs, these genomic rearrangements were of primary interest in identifying the genomic effects of IR in the sequenced samples. In Figure 6, Circos plots graphically represent the translocations detected in the genome (Krzywinski *et al.*, 2009). Breakpoints of translocations that were bridged during sequencing are indicated by lines joining one chromosome to another. In the individual ANT mutants (Mutant1- M_0 ANT W3-2 D7 and Mutant2- M_0 ANT H16-2 K9) that were sequenced, 15 translocations (TLs) indicated by blue lines and 11 TLs indicated by red lines were detected in each mutant, respectively (Figure 6A). The six ANT M_0 mutants (M_0 ANT P19-1 H2,

M_0 ANT R2-1 E12, M_0 ANT P19-1 H3, M_0 ANT R2-1 F7, M_0 ANT P19-1 G12, and M_0 ANTC18-2 C10) from which DNA was pooled and sequenced (M_0 ANT Pooled Mutants), exhibited 12 TLs whose breakpoints are indicated by green lines (Figure 6B). In the individually sequenced M_1 BNL mutants (Mutant4- M_1 BNL H13 1H10 and Mutant5- M_1 BNL P18-3 F3), 15 TLs indicated by orange lines and 9 TLs indicated by black lines were detected in the respective mutants (Figure 6C). As expected, the SV detection rate in the simulated Col-0 control datasets was much lower, with only 1 detected SV remaining after the aforementioned filtering steps (Figure 6D).

The total numbers of detected translocations, as well as their distribution, were largely similar between the ANT and BNL mutants. In some genomic regions of both the BNL and ANT mutants, the density of detected translocation breakpoints appeared to be elevated when compared to the remainder of the genome. These regions include the beginning of chromosome 5 (chr5:1-3 Mb) and two additional regions in chromosome 1 (chr1:9-15 Mb, 23-25 Mb). However, while the set of partner breakpoints associated with these potential regions of DSB enrichment demonstrated no fully conserved pattern in their distribution, there did seem to be a tendency for certain regions to be preferred as matches for repair. For example, the breakpoints in the (chr1:23-25 Mb) region connected primarily to corresponding breakpoints in the (chr3:5-17 Mb) region in both individually barcoded ANT mutants, while another connection is shared from this region to the (chr2:3-5 Mb) region in both an ANT mutant (Mutant1- M_0 ANT W3-2 D7) and a BNL mutant (Mutant5- M_1 BNL P18-3 F3). Chromosome 2 showed a larger number of identified breakpoints in the BNL mutants than the individual ANT mutants, but not the pooled ANT mutants.

Structural variants other than translocations were also detected. In Table 3, it is apparent that for most SV types the raw number of SVs that were detected was similar between mutants. A clear enrichment of inverted duplications (INVDUPS) was seen in the ANT Pooled Mutants sample, 96 INVDUPS having been detected in this sample whereas no more than 4 were detected in any other samples. M_0 ANT H16-2 K9 is also distinguished by its overall reduction in the

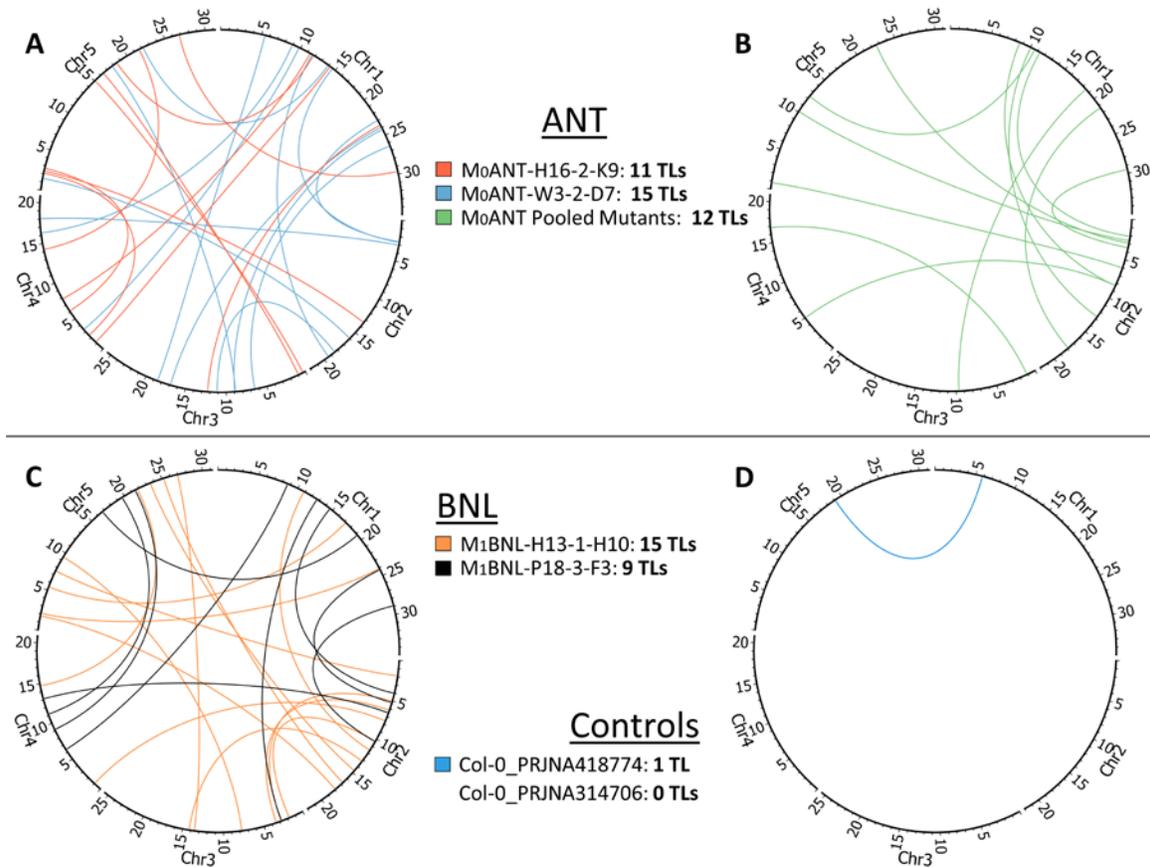


Figure 6. Genomic rearrangements detected in sequenced ANT and BNL mutants. Translocations detected in the sequenced mutants from the (A) ANT M₀ individual mutants, (B) ANT M₀ Pooled Mutants, (C) BNL M₁ individual mutants, and (D) Col-0 control datasets. Plots were generated in Circos from the filtered SV dataset for each sample. Translocation breakpoints detected by Sniffles are connected by links, which are color-coded according to the mutant in which the translocation was detected. Numeric labels on the periphery of the plots correspond to the distance (in Mbp) to that locus from the beginning of that chromosome’s reference sequence. The number of translocations detected in each sample (post-filtering) is indicated next to its name.

Table 3. Structural variants detected in the nuclear genomes of sequenced mutants.

Sample	DEL	DUP	INS	INV	INVDUP
Mutant1-M ₀ ANT W3-2-D7	6	5	5	3	3
Mutant2-M ₀ ANT H16-2-K9	1	1	2	2	2
Mutant4-M ₁ BNL H13-1-H10	5	4	12	7	4
Mutant5-M ₁ BNL P18-3-F3	8	3	7	2	2
Mutant7-M ₀ ANT (Pooled)	3	2	5	5	96

DEL: Deletion, DUP: Duplication, INS: Insertion, INV: Inversion, INVDUP: Inverted Duplication

number of detected SVs compared to the other samples. M₁BNL H13-1 H10 in contrast demonstrates a slight elevation in the number of insertions detected.

DISCUSSION

The methods described here provide a framework in which the effects of radiation can be examined at both the phenotypic and genomic levels in large numbers of seeds. The advantages of technologies such as PacBio sequencing, X-ray computed tomography, and access to deep space radiation analog environments facilitated the process of data collection, from seed irradiation through to post-sequencing bioinformatics.

Overview of Effects of Simulated Cosmic Radiation of BNL and the High-Altitude Radiation Environment of Antarctica

As BNL was designed to serve in part as a positive control, it was expected that M₀ BNL could have a high mutation rate relative to M₀ ANT, and that the quantity of particle radiation events/mm² of CR-39 sheets would also be elevated in the BNL experiment compared to that of the ANT exposure. A high dosage (Table 1) was also used in the BNL exposure due to the total area of the beam (20 cm x 20 cm) being larger than that of the seed tray used (5.4 cm x 5.4 cm). ANT seed tray #1 was selected for these analyses as it was expected to receive the most radiation of the payload, minimizing the dosage gap between the two seed populations being studied. The BNL exposure produced a visibly higher number of events/mm² as visualized through CR-39 SSNTD, as well as more mutations in the BNL M₀ than were seen in the ANT seedlings.

These results indicated that there may be significant variation in the observed biological effects resulting from differences in both the composition and dosage of the radiation presented to each group of seeds. A more thorough understanding of the radiation dosage for each individual bin could allow better explanation of the difference in mutation rates between the two samples, and further research into mapping the CR-39 SSNTDs with X-ray Micro-CT or other scans is necessary to accomplish a complete mapping of radiation events to seed bins.

However, a lack of tracking of seed position and orientation within bins during the de-integration process prohibits conclusions being drawn on any significant protective effects dependent on sub-bin positioning.

Joint United States Air Force/National Oceanic and Atmospheric Administration (USAF/NOAA) Report and Forecast of Solar and Geophysical Activity (RSGA) documents from the time period of the ANT exposure (November 28, 2016 – December 28, 2016), indicate either low or very-low levels of solar activity for all but 2 days with moderate activity levels (November 29-30) in this period (NOAA, 2016). This indicates that the GCR was likely the primary source of IR to which the ANT payload was exposed. Previous studies have also shown that a large range of heavy ions ($Z = 6-28$) contribute to high-altitude HZE (NCRP, 1995). This array of HZE ions contrasts with the radiation environment experienced by the BNL seed cassette, which was composed of H, He, O, and Ti ions ($Z = 1, 2, 8,$ and 22 , respectively). The LET expected of the average HZE event in Antarctica would therefore be higher than that of the average HZE event in the BNL simulation. However, due to the fact that the BNL M₀ exhibited a higher rate of mutations, it appears that the high dosage rate of the exposure relative to that of the ANT exposure contributed more to mutagenesis than differences in HZE ion composition. Deeper and detailed analyses of SSNTDs collected from the ANT payload will also provide information on which ions impacted individual bins.

Impacts on Germination Rates

Genomic damage, regardless of its source, has been shown to cause inhibition of the germination process; if the genome sustains an irreparable degree of damage the germination process can fail altogether (Waterworth *et al.*, 2015; Waterworth *et al.*, 2016). Germination may not fail if the genomic damage results in a mutator phenotype with increased tolerance of radiation-induced damage (Walker *et al.*, 1997; Zhang *et al.*, 2008). Study of these mechanisms is thus vital to guaranteeing significant rates of seed viability for long-term space exploration missions. The concern of how best to store seeds during these missions arises, as seeds irradiated in monolayers have been shown to exhibit lower germination

rates (Kazama *et al.*, 2008). However, little effort has been directed toward better understanding the extent of shielding provided by packaging many seeds into the same container. Study of other seed trays from the ANT payload could reveal some of these effects in future work, where seed trays further isolated within the payload may demonstrate different germination or mutation rates from peripheral trays. In the present study, the effects of upper stratosphere exposure and simulated GCR exposure on seed germination were assessed using only one tray from each condition.

Analysis of irradiated M_0 ANT and BNL seeds compared to non-irradiated seed samples showed that radiation damage significantly reduced the germination rates of the experimental groups, which is in line with previous studies despite the array of radiation conditions under which they were conducted (De Micco *et al.*, 2011; Kranz, 1986; Tepfer and Leach, 2017; Tepfer *et al.*, 2012). This also served to validate the ability of the BNL mixed-beam to faithfully reproduce this basic GCR effect, to the level that there was no significant difference found between the BNL and ANT M_0 germination rates. Interestingly, germination for the BNL M_1 seed recovered to nearly that of the control group (Figure 5B). It is logical that the recovered germination rate in the M_1 generation would be a result of the decreased viability of M_0 seed due to lethal mutations, when the connections between germination regulation, growth regulation, and cell cycle checkpoints for genomic damage are taken into account (Waterworth *et al.*, 2016; Zhang *et al.*, 2008). Further study using lines deficient in these genes could enable the discovery of new radiation-induced phenotypes.

Frequencies and Types of Observable Mutations in BNL and ANT

Germinated seeds with genomic damage of a high enough level to produce distinct mutant phenotypes were classified and counted to quantify the rates of mutation in each exposure sample. Compared to the non-irradiated control, both the BNL and ANT M_0 plants exhibited significantly higher mutation rates. In contrast to the similarities seen in the germination rate data, observations of the mutation rates in M_0 BNL and M_0 ANT showed that significant differences

existed between them, with the BNL rate being higher than was observed in ANT. This serves as an indicator that differences in downstream biological effects do exist between these radiation sources, and that there is still room for improvement of GCR simulations. Of particular interest is the significant difference in mutation rate between the BNL M_1 samples and the control samples, which clearly demonstrates the presence of heritable mutations that were the result of radiation damage and repair in the genomes of the M_0 BNL seeds.

In the M_0 BNL and ANT test groups, the type of mutations and their frequencies differed. The discolorations observed in the BNL M_0 were most prevalent near the shoot apex, and the phenotype was often accompanied by microgrowth. Deficiencies of cotyledon development were also seen to be enriched in the BNL M_0 seedlings. However, the ANT M_0 also contained a subset of abnormally-colored and stunted mutants, classified based on their cotyledons, which exhibited albinism. Dwarf architecture has been prevalent in past studies, and was statistically significant in its enrichment within the M_0 BNL test group (De Micco *et al.*, 2011; Lake *et al.*, 2009). These findings suggest that there may be links between the levels of growth and discolorations observed. Radiation exposure has previously been connected with abnormal coloration, but this was also shown to be an LET-dependent effect (Kazama *et al.*, 2008). While this could be a product of overall genomic damage, as stress-induced pigmentation changes are well-known to occur (Gangappa and Botto, 2016), it would be expected that the discoloration events would be more similar between samples if this were the case, unless the two types of radiation are significantly dissimilar in terms of LET values. Further study into the compositions of GCR and SCR as well as sampling of IR alongside biological experiments may allow determination of the true nature of these phenotypes.

Impacts of Radiation Environments on Genomic Damage – Preliminary Data and Lessons Learned

Analysis of sequencing data for the purpose of detecting structural variants revealed that irradiation of *A. thaliana* seeds resulted in

complex genomic rearrangements that could be identified at high resolution through long-read sequencing of regions containing SV breakpoints – even in the mutant plants directly arising from the ANT M_0 seeds. Interestingly, regardless of the radiation source to which the seeds were exposed in these experiments, the quantity and distribution of the observed translocation breakpoints are similar (Figure 6). One notable example is a region of chromosome 1 (chr1:9-15 Mb) that contains translocation breakpoints in all sequenced samples, but which also appears to show breakpoint enrichment in the ANT samples compared to BNL samples.

The similarities in observed translocations also indicated the potential for non-random factors having influenced the probability of certain chromosomal rearrangements occurring, such as spatial proximity and chromatin structure of genomic loci when IR-induced DSBs occur. IR can, in addition to impacting the genome directly, lead to the generation of reactive oxygen species that have the potential to create DSBs and other genomic lesions. Seeds have some protections against these threats, such as endogenous antioxidants (Roldán-Arjona and Ariza, 2009). While our approach is unable to distinguish between these direct and indirect types of damage, both are biologically relevant effects of IR and are of interest as a result. Prior investigations into the off-target repair frequencies of DSBs resulting from radiation have found trends similar to those observed in this study (Hada *et al.*, 2011; Lavelle and Foray, 2014). The (chr1:23-25 Mb) region of chromosome 1 serves as an example of this concept, where it shows preferential off-target repair to chromosome 3 in only the ANT mutants. Previous studies have also used high-throughput sequencing methods to map genomic domains with high frequencies of interaction in *A. thaliana* during growth (Feng *et al.*, 2014; Wang *et al.*, 2015); synthesis of these methods could provide further insights into chromatin packaging in *A. thaliana* embryonic nuclei and inform further genomic studies.

Mutant phenotypes in the ANT and BNL M_0 were likely to have occurred due to somatic mutations in the seed embryo, which resulted in the death or delayed growth of certain plant organs, such as in the case of the ANT “rootless” mutants (Figure 4A and B). The resultant

mosaicism offers the unique opportunity to sequence wild-type and mutant genomes within the same plant, but would create issues with tissue amplification measures such as tissue culture, where wild-type cells would be expected to outcompete mutant cells exhibiting deleterious phenotypes. However, the presence of heritable mutations in the BNL M_1 samples, as evidenced by their observed mutation rate (Figure 5D), and the detection of structural variants in the sequenced BNL M_1 individuals (Figure 6C), support the notion that intergenerational screening would be sufficient to affix these IR-induced SVs and allow better study of their biological effects.

Pooling multiple mutants (Figure 4E-J) into a single DNA sample before library preparation showed some potential as a strategy for the detection of SVs in samples that cannot provide enough DNA for individual sequencing libraries. As the mutants that were of the highest interest due to their severe phenotypes fell into this category, it was important to explore methods that could enable their study. It is noted, however, that in all respects other than the heavy enrichment of INVDUPs (Table 3), the SVs detected in the pooled ANT sample (Figure 6B) are similar to those in the individually barcoded samples (Figure 6A, C). Further study can shed light on both the possible causes and biological effects of these INVDUPs, and whether pooling of mutants is a viable option for SV analysis on a larger scale.

The issue of tissue availability was the primary driving force for the development of a yield-optimized DNA extraction protocol. DNA of sufficient quality and quantity for library preparations was successfully isolated from low-biomass samples despite the removal of cleaning steps that would have resulted in significant losses of DNA. In the event of low yields, DNA from multiple mutants could be pooled before barcoding and SVs were detected (Figure 6B), but at the cost of an inability to determine the sample of origin for each SV. The presence of organelle genomes in extracted DNA led to false SV calls due to high coverage of the organelle genomes leading to spurious alignments, but this was considered acceptable given that isolation of nuclei would lead to further losses of material. In future studies, organelle genomic sequences would be excluded from read alignment steps unless they are of particular interest.

Of the detected SVs, there was potential for the blunt-end ligation of SMRTbell adaptors to generate fusion artifacts between DNA fragments during library preparation (Griffith *et al.*, 2018), but no DNA was able to be retained for validation of SVs due to the small amount available. However, it was notable that in the simulated datasets only one SV was detected (Figure 6D). This demonstrated that fusion artifacts were rarely called as SVs in the sequencing output of these other studies that also utilized blunt-end ligation during library preparation (Chin *et al.*, 2016; Liang *et al.*, 2018). In future work, the A/T overhang adaptor ligation method will be used to ameliorate this effect, and it is likely that a small but significant number of SVs detected in this study were artefactual.

The methods described here were developed for the purpose of defining an experimental pipeline that would allow workable experimentation on large numbers of seeds exposed to space radiation, with the goal of using seeds to characterize genomic damage by space radiation. The extended goal will be to create large datasets of space-irradiated genomes in order to deeply characterize genomic damage tendencies that may accumulate in long duration exposures to deep space environments. We found that these test exposures, screening procedures, and sequencing methods provided a tractable experimental path that resulted in clearly different phenotypes during the early development of *A. thaliana*, and the ability to identify genomic structural variants present in those mutant plants. The results illustrate the efficacy of these procedures and offer promise for future study and development of yet more specialized methods for investigating radiation-induced mutations and identification of the causal variants.

These data show that long duration Antarctic balloon flights provide exposure to enough IR to realistically emulate, not just simulate, deep space exposure to IR for the purposes of gathering data on genomic impacts. The balloon flight platform and operations provided enough exposure to make searching for mutations both tractable and productive. The outcomes include the genomic characterization of actual space IR mixes, rather than shorter, high fluence exposures meant to simulate deep space IR. These data clearly demonstrate the potential for long duration and

low dosage rate experiments to provide unique insights on the biological effects of IR on eukaryotes relevant to space exploration. The further development of biological payloads and hardware to facilitate balloon-based sampling of naturally occurring IR can thus allow advances in scientific knowledge as well as protective technologies – without requiring travel beyond the Earth's magnetic field.

ACKNOWLEDGEMENTS

This work was conducted with funding derived from University of Florida resources and the donated effort of many people and organizations. The High Altitude Student Platform (HASP) program out of Louisiana State University offered the opportunity to develop and fly an initial balloon payload as an undergraduate project preliminary to the work described here, and was funded in part by the NASA Florida Space Grant Consortium (NNX15_011). The NASA Balloon Payload Office (BPO) out of Goddard Space Flight Center created the opportunity to fly CRESS as a piggyback payload in Antarctica. The entire BPO administration, and especially flight development and operations crews, were exceedingly accommodating and helpful throughout the payload design and implementation process. The BACCUS team created the entire Antarctic balloon mission and graciously accommodated our piggyback payload. Dr. Howard Levine of Kennedy Space Center was instrumental in enabling and supporting the connections between Space Biology and the use of NASA balloons, through a non-reimbursable NASA Space Act Agreement KCA-4572-1. Dr. Ye Zhang of Kennedy Space Center, together with Dr. Levine and Dr. Megumi Hada (PI of the primary beam time) of Prairie View A&M University, created the opportunity to expose CRESS seeds at the NASA Space Radiation Laboratory (NSRL) as a piggyback experiment on an NSRL grant awarded to Dr. Zhang, again using Space Act Agreement KCA-4572-1. The National Science Foundation Office of Polar Programs, particularly Dr. Nature McGinn, guided the CRESS payload through the processes of permitting biological samples in Antarctica. The Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida

conducted the DNA sequencing. Dr. David Moraga and Dr. Yanping Zhang from ICBR guided the application and design of the PacBio sequencing approach used in these preliminary analyses. The sequencing, payload design, fabrication, and development work at the University of Florida was primarily funded by the internal University of Florida accounts of RJF and A-LP.

Accession Codes

Sequence and structural variant data have been submitted to the Gene Expression Omnibus (Series ID: GSE121111) and NASA GeneLab (accession number GLDS-210).

REFERENCES

- Arena C, De Micco V, Macaeva E, Quintens R (2014) Space radiation effects on plant and mammalian cells. *Acta Astronautica* **104**: 419-431
- Ave M, Boyle P, Brannon E, Gahbauer F, Hermann G, Höppner C, Hörandel J, Ichimura M, Müller D, Obermeier A (2011) The TRACER instrument: A balloon-borne cosmic-ray detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654**: 140-156
- Badhwar GD, O'Neill PM (1992) An improved model of galactic cosmic-radiation for space exploration missions. *Nuclear Tracks and Radiation Measurements* **20**: 403-410
- Benton E (2012) Space radiation passive dosimetry. *The Health Risks of Extraterrestrial Environments* <https://three.jsc.nasa.gov/articles/BentonPassive-Dosimetry.pdf>
- Benton ER, Benton EV (2001) Space radiation dosimetry in low-Earth orbit and beyond. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **184**: 255-294
- Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E (2015) The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* **53**: 474-485
- Cartwright BG, Shirk EK, Price PB (1978) A nuclear-track-recording polymer of unique sensitivity and resolution. *Nuclear Instruments and Methods* **153**: 457-460
- Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* **13**: 1050
- De Micco V, Arena C, Pignalosa D, Durante M (2011) Effects of sparsely and densely ionizing radiation on plants. *Radiation and Environmental Biophysics* **50**: 1-19
- Durante M (2014) Space radiation protection: Destination Mars. *Life Sciences in Space Research* **1**: 2-9
- Feng S, Cokus SJ, Schubert V, Zhai J, Pellegrini M, Jacobsen SE (2014) Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in *Arabidopsis*. *Molecular Cell* **55**: 694-707
- Gangappa SN, Botto JF (2016) The multifaceted roles of HY5 in plant growth and development. *Molecular Plant* **9**: 1353-1365
- Griffith PM, Raley C, Sun D, Zhao Y, Sun Z, Mehta M, Tran B, Wu X (2018) PacBio library preparation using blunt-end adapter ligation produces significant artefactual fusion DNA sequences. *bioRxiv*
- Hada M, Zhang Y, Feiveson A, Cucinotta FA, Wu H (2011) Association of inter- and intrachromosomal exchanges with the distribution of low- and high-LET radiation-induced breaks in chromosomes. *Radiation Research* **176**: 25-37
- Huefner ND, Yoshiyama K, Friesner JD, Conklin PA, Britt AB (2014) Genomic stability in response to high versus low linear energy transfer radiation in *Arabidopsis thaliana*. *Frontiers in Plant Science* **5**: 206
- Kavanagh JN, Redmond KM, Schettino G, Prise KM (2013) DNA double strand break repair: a radiation perspective. *Antioxidants & Redox Signaling* **18**: 2458-2472
- Kazama Y, Saito H, Yamamoto YY, Hayashi Y, Ichida H, Ryuto H, Fukunishi N, Abe T (2008) LET-dependent effects of heavy-ion beam irradiation in *Arabidopsis thaliana*. *Plant Biotechnology* **25**: 113-117

- Kennedy AR (2014) Biological effects of space radiation and development of effective countermeasures. *Life Science in Space Research* **1**: 10-43
- Kiffer F, Carr H, Groves T, Anderson JE, Alexander T, Wang J, Seawright JW, Sridharan V, Carter G, Boerma M, Allen AR (2017) Effects of 1H + 16O charged particle irradiation on short-term memory and hippocampal physiology in a murine model. *Radiation Research* **189**: 53-63
- Kim K, Amare Y, Anderson T, Angelaszek D AN, Cheriyian K, Choi GH, Copley M, Coutu S, Derome L, Eraud L, Hagenau L, Han JH, Huh HG, Im S, Jeon JA, Jeong S, Kim MH, Lee HY, Lee J, Lee MH, Liang J, Link JT, Lu L, Lutz L, Mernik T, Mitchell JW, Mognet SI, Morton S, Nester M, Nutter S, Ofoha O, Park IH, Picot-Clémente N, Quinn R, Seo ES, Smith JR, Walpole P, Weinmann RP, Wu J, Yoon YS (2017) Boron And Carbon Cosmic Rays in the Upper Stratosphere (BACCUS). In *Proceedings of Science: 35th International Cosmic Ray Conference*.
- Kranz AR (1986) Genetic and physiological damage induced by cosmic radiation on dry plant seeds during space flight. *Advances in Space Research* **6**: 135-138
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Research* **19**: 1639-1645
- Lake JA, Field KJ, Davey MP, Beerling DJ, Lomax BH (2009) Metabolomic and physiological responses reveal multi-phasic acclimation of *Arabidopsis thaliana* to chronic UV radiation. *Plant, Cell and Environment* **32**: 1377-1389
- Lavelle C, Foray N (2014) Chromatin structure and radiation-induced DNA damage: From structural biology to radiobiology. *The International Journal of Biochemistry & Cell Biology* **49**: 84-97
- LeFrois CE, Zhou M, Amador DM, Sng N, Paul A-L, Ferl RJ (2016) Enabling the spaceflight methylome: DNA isolated from plant tissues preserved in RNAlater™ is suitable for bisulfite PCR assay of genome methylation. *Gravitational and Space Research* **4**: 28-37
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079
- Liang Z, Shen L, Cui X, Bao S, Geng Y, Yu G, Liang F, Xie S, Lu T, Gu X, Yu H (2018) DNA N6-adenine methylation in *Arabidopsis thaliana*. *Developmental Cell* **45**: 406-416
- Miller J, Zeitlin C (2016) Twenty years of space radiation physics at the BNL AGS and NASA Space Radiation Laboratory. *Life Science in Space Research* **9**: 12-18
- Moreno-Villanueva M, Wong M, Lu T, Zhang Y, Wu H (2017) Interplay of space radiation and microgravity in DNA damage and DNA damage response. *npj Microgravity* **3**: 14
- NASA (2018) *Human Research Program Integrated Research Plan*. NASA/HRP-47065-RevJ <https://humanresearchroadmap.nasa.gov/>
- NCRP (1995) *Radiation Exposure and High Altitude Flight*, Commentary #12. Bethesda, MD: US National Council for Radiation Protection and Measurements
- NOAA (2016) Report and Forecast of Solar and Geophysical Activity, National Oceanic & Atmospheric Administration
- Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP (2017) Variant review with the integrative genomics viewer. *Cancer Research* **77**: e31-e34
- Roldán-Arjona T, Ariza RR (2009) Repair and tolerance of oxidative DNA damage in plants. *Mutation Research/Reviews in Mutation Research* **681**: 169-179
- Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods* **15**: 461-468
- Smith DJ, Sowa MB (2017) Ballooning for biologists: mission essentials for flying life science experiments to near space on NASA large scientific balloons. *Gravitational and Space Research* **5**: 52-73
- Suman S, Kumar S, Moon B-H, Fornace AJ, Datta K (2017) Low and high dose rate heavy ion radiation-induced intestinal and colonic

- tumorigenesis in APC1638N/+ mice. *Life Sciences in Space Research* **13**: 45-50
- Tepfer D, Leach S (2017) Survival and DNA damage in plant seeds exposed for 558 and 682 days outside the International Space Station. *Astrobiology* **17**: 205-215
- Tepfer D, Zalar A, Leach S (2012) Survival of plant seeds, their UV screens, and nptII DNA for 18 months outside the International Space Station. *Astrobiology* **12**: 517-528
- Walker KA, Jennings CD, Pulliam J, Ogburn C, Martin GM, Urano M, Turker MS (1997) A cell line selected for resistance to ionizing radiation exhibits cross resistance to other genotoxic agents and a mutator phenotype for loss of heterozygosity events. *Somatic Cell and Molecular Genetics* **23**: 111-121
- Wang C, Liu C, Roqueiro D, Grimm D, Schwab R, Becker C, Lanz C, Weigel D (2015) Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Research* **25**: 246-256
- Waterworth WM, Bray CM, West CE (2015) The importance of safeguarding genome integrity in germination and seed longevity. *Journal of Experimental Botany* **66**: 3549-3558
- Waterworth WM, Footitt S, Bray CM, Finch-Savage WE, West CE (2016) DNA damage checkpoint kinase ATM regulates germination and maintains genome stability in seeds. *Proceedings of the National Academy of Sciences USA* **113**: 9647-9652
- Zhang Y, Rohde LH, Emami K, Hammond D, Casey R, Mehta SK, Jeevarajan AS, Pierson DL, Wu H (2008) Suppressed expression of non-DSB repair genes inhibits gamma-radiation-induced cytogenetic repair and cell cycle arrest. *DNA Repair* **7**: 1835-1845