Mining the Extraterrestrial Seedling Microbiome: Meta-transcriptomic Discovery of Endophytic and Commensal Communities in Spaceflight *Arabidopsis thaliana*

Richard J. Barker[1,*], Zerrin Sogutmaz Ozgen[1], Sarah E. Wyatt[2], Simon Gilroy[3], and NASA GeneLab Plant Analysis Working Group (AWG)[1,4]

[1]AstroBotany Laboratory and Artificial Intelligence Research Initiative (AIRI), Madison, WI, USA
[2]Department of Environmental and Plant Biology, Ohio University, Athens, OH, USA
[3]Department of Botany, University of Wisconsin–Madison, Madison, WI, USA
[4]NASA Ames Research Center, Space Biosciences Division, Moffett Field, CA, USA
[*]Corresponding author: `richard.barker@nasa.gov`

October 22, 2026

Abstract

Background: Spaceflight exposes biological systems to the synergistic stresses of microgravity, cosmic ionizing radiation, and sealed atmospheric confinement. While the transcriptional responses of model plants like *Arabidopsis thaliana* have been comprehensively cataloged in the NASA Open Science Data Repository (OSDR), the endophytic and epiphytic microbiomes associated with these seedlings have remained virtually unexplored in archival transcriptomic datasets. Because standard eukaryotic RNA sequencing protocols capture both host and microbial polyadenylated or co-extracted transcripts, mining unmapped non-host reads provides a powerful window into plant-microbiome dynamics in extraterrestrial environments.

Methods: Here, we present a systematic meta-transcriptomic extraction and taxonomic profiling pipeline applied across six OSDR spaceflight accessions (OSD-37, OSD-38, OSD-120, OSD-217, OSD-321, and OSD-522) encompassing multiple hardware systems (BRIC-19, BRIC-20, BRIC-22, BRIC-PDF, APEX-03-2) and genetic backgrounds (Columbia-0, Wassilewskija-2, Landsberg *erecta*-0, and Cape Verde Islands-0). Following host read subtraction against the *A. thaliana* TAIR10 reference genome, remaining microbial sequences were profiled using MetaPhlAn4, Bowtie2, and Krona interactive visualization. Taxonomic intersections across flight and ground controls were assessed through UpSetR and hierarchical correlation analyses, and annotated using The Microbe Directory.

Results: We observed distinct microbial communities segregated between intrinsic seed-borne endophytes and extrinsic hardware- or human-associated commensals. *Ralstonia pickettii* exhibited pervasive dominance across OSD-37 ecotypes, reaching 100% relative abundance in Cvi-0 samples. In OSD-38, we uncovered the presence of *Weizmannia ginsengihumi*, a beneficial plant-growth-promoting Firmicute previously unrecorded in orbital seedling habitats. Furthermore, common skin and environmental commensals including *Cutibacterium acnes*, *Malassezia restricta*, and *Escherichia coli* were systematically identified, reflecting microbial exchange between crew-handled hardware and plant growth substrates. Interestingly, comparison of whole seedlings (OSD-37, 38) against dissected root tips (OSD-120) confirmed that endophytic colonization is strongly tissue-dependent, with root tips displaying near-complete depletion of bacterial transcripts.

Conclusions: These findings demonstrate that plant-associated microbiomes persist in closed spaceflight hardware and undergo community shifts that mirror both host developmental state and chamber microbial ecologies. Archival metatranscriptomics provides a cost-effective, non-destructive paradigm for tracking plant-microbe biosecurity, designing beneficial synthetic microbiomes, and ensuring crop resilience in future Bioregenerative Life Support Systems (BLSS) on Lunar and Martian surface missions.

Keywords: Spaceflight, *Arabidopsis thaliana*, Endophytic Microbiome, Metatranscriptomics, NASA OSDR, *Ralstonia pickettii*, *Weizmannia ginsengihumi*, Bioregenerative Life Support Systems (BLSS).

1 Introduction

Sustaining human exploration beyond Low Earth Orbit (LEO) requires the establishment of self-sustaining Bioregenerative Life Support Systems (BLSS) (Fu et al., 2016). Central to BLSS architecture are higher plants, which perform vital functions including carbon dioxide sequestration, oxygen generation, wastewater purification through transpiration, and dietary supplementation with fresh nutrients (Paul et al., 2017; Dixit et al., 2021). For decades, the model crucifer *Arabidopsis thaliana* has served as the foundational benchmark for elucidating how plants perceive and adapt to extraterrestrial environments, notably altered gravity vectors, chronic low-dose cosmic radiation, and fluid-behavior anomalies (Choi et al., 2019; Kruse et al., 2020).

Extensive transcriptomic and proteomic profiling archived in the NASA Open Science Data Repository (OSDR) has illuminated pervasive reprogramming of plant stress responses, hormone signaling, cell-wall remodeling, and epigenetic modifications during orbital flight (Barker et al., 2023; Zhou et al., 2019; Olanrewaju et al., 2023). However, terrestrial plant biology has firmly established that plants do not function as isolated genetic entities; rather, they exist as "holobionts" intrinsically co-adapted with diverse microbial consortia comprising endophytic, phyllospheric, and rhizospheric bacteria and fungi (de Zélicourt et al., 2013; Pieterse et al., 2014). Beneficial endophytes confer pathogen resistance through induced systemic resistance (ISR), synthesize phytohormones such as indole-3-acetic acid (IAA), and alleviate abiotic stressors including drought and oxidative challenge. Conversely, alterations in host immunity in microgravity may render spaceflight crops vulnerable to opportunistic colonization by phytopathogens or human bacterial pathogens (Totsline et al., 2023, 2024).

Despite its paramount importance for space agriculture, characterization of the plant microbiome during spaceflight has faced significant experimental constraints. Launch mass limits, astronaut crew time limitations, and strict containment protocols often restrict dedicated metagenomic sampling. Consequently, prior studies of the International Space Station (ISS) microbiome have primarily focused on environmental air filters, surface swabs, and astronaut skin or gut microbiomes, leaving seedling-associated endophytic communities under-sampled.

Fortunately, modern high-throughput RNA sequencing (RNA-seq) performed on total plant tissue inherently captures a rich fraction of non-host transcripts (Park et al., 2019). When plant total RNA is extracted and sequenced without stringent poly(A) selection or when microbial transcripts share sequence features that evade depletion, hundreds of thousands of sequencing reads originate from bacteria, archaea, and fungi inhabiting the host tissues. In terrestrial genomics, such reads are traditionally categorized as "unmapped dark matter" and discarded during downstream differential expression analyses (Park et al., 2019; Ewels et al., 2020).

In terrestrial agro-ecosystems, the foundational baseline of the seed microbiome has recently been deciphered at unprecedented scale. In a landmark meta-analysis spanning 3,190 seed samples across 50 plant species and 28 countries, Simonin et al. (2022) established that seed

microbiota are structured into a universal, highly conserved "core" fraction (dominated by *Pantoea agglomerans*, *Pseudomonas viridiflava*, *Pseudomonas fluorescens*, and fungal taxa like *Alternaria*) alongside a hyper-variable "flexible" fraction comprising rare, site-specific associates. Because the seed serves as the primary transgenerational inoculum for crop health, understanding whether the spaceflight environment preserves or disrupts this evolutionary core is of paramount urgency for long-duration life support. However, spaceflight seed biology operates under strict operational constraints: pre-flight chemical surface-sanitization (e.g., sodium hypochlorite or ethanol) strips the episphere, and growth occurs within tightly sealed hardware chambers subjected to microgravity, fluid-stagnation, and ionizing radiation.

Moreover, mining microbial signals from archival spaceflight datasets poses a profound computational challenge: the small-data, high-dimensionality curse ($N \ll P$). Spaceflight biological missions yield restricted sample sizes ($N \approx 10$–50) due to payload mass limits, whereas metagenomic and metatranscriptomic taxonomic features span dozens to hundreds of candidate taxa (P). Traditional deep learning architectures (e.g., multi-layer perceptrons, convolutional or graph networks) collapse or catastrophically overfit in small-N regimes, while gradient boosted decision trees require complex hyperparameter tuning. Recently, Hollmann et al. (2025) demonstrated that tabular foundation models pre-trained on synthetic prior datasets—specifically Prior-Data Fitted Networks like TabPFN—achieve state-of-the-art predictive performance and well-calibrated Bayesian inference on small tabular data in a single forward pass without hyperparameter optimization.

In this study, we capitalize on this bioinformatic and deep learning convergence by conducting a systematic meta-transcriptomic recovery and microbial profiling of archival *Arabidopsis thaliana* spaceflight datasets deposited in NASA OSDR (Barker et al., 2023). By benchmarking observed spaceflight communities against the global 50-species seed baseline (Simonin et al., 2022) and applying the TabPFN tabular foundation model (Hollmann et al., 2025), we address fundamental questions regarding spaceflight microbiome dynamics:

1. Do surface-sanitized *Arabidopsis* seedlings grown in closed spaceflight hardware harbor persistent, detectable microbial communities, and how do they compare against the universal terrestrial seed core identified by Simonin et al. (2022)?

2. How are spaceflight communities partitioned between intrinsic seed-borne endophytes, oligotrophic blooms (*Ralstonia pickettii*), stress-adapted probiotics (*Weizmannia ginsengihumi*), and extrinsic anthropogenic or hardware commensals?

3. How does hardware containment (closed BRIC canisters vs. open illuminated APEX systems) drive ecological filtering, and can tabular foundation models (TabPFN) accurately decode these environmental signatures from ultra-sparse microbial matrices?

4. What are the predictive implications for agricultural seed stocks and candidate crop species planned for future lunar and Martian Bioregenerative Life Support Systems?

2 Materials and Methods

2.1 Study Selection and OSDR Metadata Extraction

Candidate RNA-seq datasets were identified through systematic query of the NASA Open Science Data Repository (OSDR, `https://osdr.nasa.gov`) utilizing criteria established by Barker et al. (2023). Eligible accessions satisfied three requirements: (1) experimental cultivation of *Arabidopsis thaliana* seedlings aboard the International Space Station (ISS), (2) inclusion of synchronous or asynchronous ground hardware controls, and (3) availability of raw paired-end or single-end FASTQ sequencing files.

Six primary accessions were selected for comparative meta-transcriptomic dissection (Table 1): OSD-37 (GLDS-37, BRIC-19; 8-day-old dark-grown seedlings across Col-0, Cvi-0, Ler-0, WS-2), OSD-38 (GLDS-38, BRIC-20; 3–4-day-old dark-grown Col-0 seedlings with liquid nitrogen preservation), OSD-120 (GLDS-120, APEX-03-2; 11–13-day-old micro-dissected root tips in light/dark cycles), OSD-217 (GLDS-217, BRIC-PDF; light-grown WS seedlings with separated leaf and root tissues), OSD-321 (GLDS-321, BRIC-22; whole seedlings in sealed BRIC-PDFU canisters aboard SpaceX CRS-12 examining unfolded protein response stress adaptation), and OSD-522 (GLDS-522; light-grown Col-0 seedlings). Study metadata and sequencing run manifests were extracted programmatically using the OSDR RESTful API and assembled into computational tables via automated Python/Google Colab scripts.

2.2 Quality Trimming and Host-Genome Depletion

Raw sequencing reads were retrieved in gzipped FASTQ format. Quality verification was performed using FastQC (v0.11.9) (Ewels et al., 2016), and adapter contamination and low-quality bases were filtered using Trim Galore! (v0.6.7) with a Phred score cutoff of $Q \geq 25$ and minimum read length of 35 bp.

To deplete host-derived sequences, cleaned reads were aligned against the *Arabidopsis thaliana* reference genome (TAIR10 / Araport11) using the splice-aware aligner STAR (v2.7.9a) and Bowtie2 (v2.4.4). Reads failing to align to the plant nuclear, mitochondrial, or chloroplastic genomes (designated as unmapped FASTQ reads) were extracted using SAMtools (v1.14) with flags `-f 4` (single-end) or `-f 12 -F 256` (paired-end), and retained as the non-host metatranscriptome library.

2.3 Taxonomic Profiling and Abundance Estimation

Taxonomic classification of the microbial read pool was executed using MetaPhlAn4 (v4.0.6) (Segata et al., 2012), which relies on a curated database of ∼5.1 million clade-specific marker genes covering $> 26,000$ microbial species (Bacteria, Archaea, Eukaryota, and Viruses). In parallel, Bowtie2 alignment files were generated to assess mapping densities across marker loci. Relative abundance estimates were calculated at each hierarchical rank (domain, phylum, class, order, family, genus, and species). Interactive hierarchical taxonomic visualizations were generated using Krona (v2.8.1) (Ondov et al., 2011) to inspect nested clades across experimental conditions.

2.4 Intersectional and Community Diversity Analysis

To determine core versus flight-specific or ground-specific taxa, intersectional set analyses were conducted using UpSetR (v1.4.0) (Conway et al., 2017) and Intervene (v0.6.5). Pairwise Jaccard similarities and Spearman correlation coefficients were calculated between samples to construct correlation matrices and hierarchical dendrograms. Hierarchical All-against-All Association (HAllA) testing was implemented to identify co-varying microbial clusters across mission parameters and ecotypes.

2.5 Functional Annotation and Ecological Provenance

Identified microbial species were annotated against The Microbe Directory (v2.0) and BacDive (The Bacterial Diversity Metadatabase) to assign ecological provenance and functional traits. Microorganisms were systematically classified into: (1) *Putative Seed Endophytes / Plant Associates* (rhizosphere, phyllosphere, or seed-borne symbionts), (2) *Human / Clinical Commensals* (known inhabitants of human skin, oral cavity, or gut), and (3) *Environmental / Spacecraft Hardware Contaminants* (oligotrophic water-system or cleanroom isolates). Physiological attributes—including Gram stain morphology, biofilm-forming capacity, spore formation, optimal

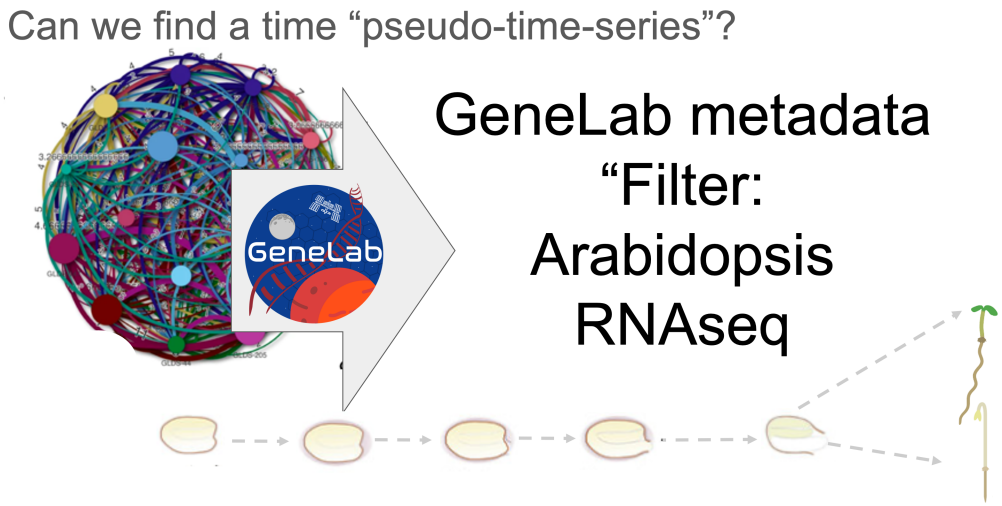


Figure 1: **Systematic selection and environmental parameterization of *Arabidopsis* seedling spaceflight accessions.** Meta-analysis filtering strategy across NASA OSDR repository identifying datasets with overlapping biological replicates, dark/light regimes, and hardware enclosures (BRIC, APEX, EMCS) as described by Barker et al. (2023).

growth temperature, and pathogenicity—were compiled to evaluate their functional potential in closed spaceflight habitats.

3 Results

3.1 Meta-transcriptomic Recovery Across Spaceflight Accessions

Mining non-host reads from the six selected OSDR spaceflight datasets revealed that despite pre-flight seed surface sanitation protocols, detectable microbial transcripts were recovered in all whole-seedling accessions (Figure 1 and Table 1). Total read depth and the proportion of non-plant reads varied substantially depending on experimental hardware, preservation chemistry, and anatomical tissue type. Whole-seedling experiments preserved in RNAlater or flash-frozen in liquid nitrogen yielded between 0.05% and 1.8% non-host reads, corresponding to tens of thousands to hundreds of thousands of high-confidence microbial marker reads per library.

3.2 Pervasive Dominance of *Ralstonia pickettii* in BRIC-19 (OSD-37)

Taxonomic profiling of 8-day-old etiolated seedlings from the BRIC-19 mission (OSD-37) revealed a consistent microbial consortium across both spaceflight (FL) and ground control (GC) samples (Table 2). Strikingly, the betaproteobacterium *Ralstonia pickettii* emerged as the overwhelmingly dominant taxon across all biological replicates and ecotypes. In the Cape Verde Islands (Cvi-0) accession, *R. pickettii* accounted for 100% of all classified microbial reads in multiple samples. In Columbia-0 (Col-0), Wassilewskija-2 (WS-2), and Landsberg *erecta*-0 (Ler-0), *R. pickettii* consistently constituted $> 65\%$ of relative abundance.

Beyond *R. pickettii*, OSD-37 libraries contained secondary populations of recognized plant-associated endophytes, including *Herbaspirillum huttiense*, *Pseudomonas fluorescens*, *Curvibacter gracilis*, *Acidovorax temperans*, and *Agrobacterium tumefaciens*. While the primary community composition was conserved between orbit and ground, flight samples exhibited reproducible shifts in the relative abundance of secondary Proteobacteria, suggesting subtle microgravity-mediated or enclosure-induced alterations in host colonization.

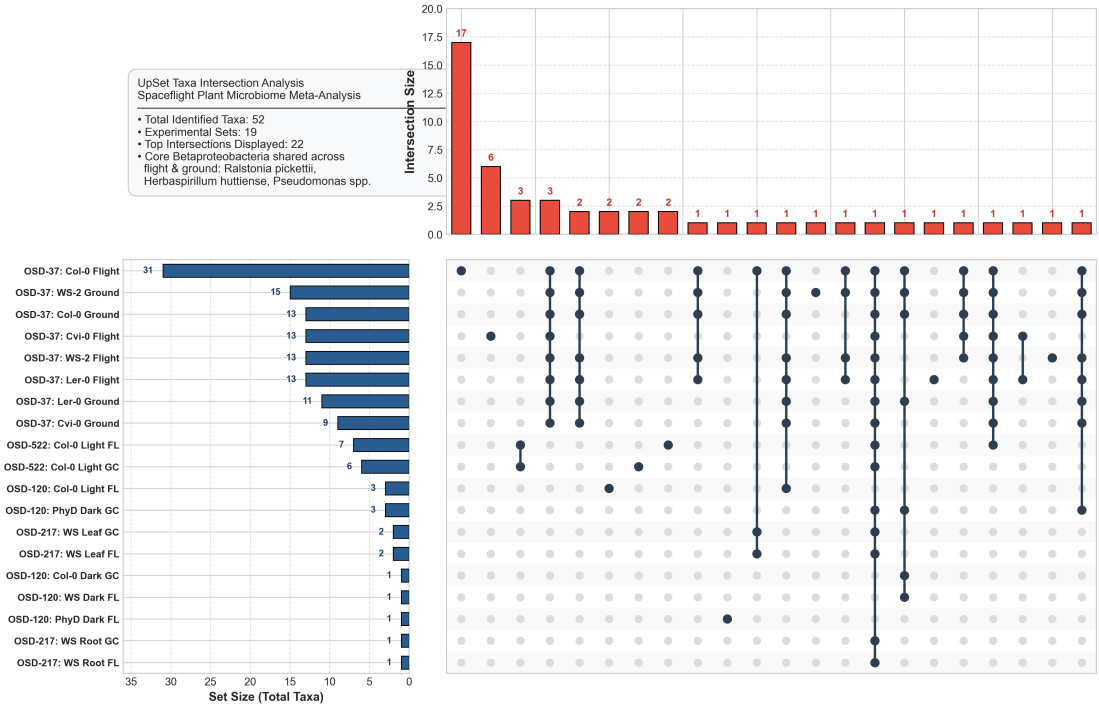


Figure 2: **UpSet intersection of detected microbial taxa across spaceflight and ground seedling conditions.** Intersections demonstrate a shared core of seed-borne Proteobacteria (including *Ralstonia pickettii* and *Pseudomonas fluorescens*) alongside sample-specific environmental transfers and ecotype-restricted colonizers.

3.3 Discovery of *Weizmannia ginsengihumi* in BRIC-20 (OSD-38)

Analysis of younger 3–4-day-old Col-0 seedlings from BRIC-20 (OSD-38) uncovered a distinctly different microbial profile from the 8-day-old BRIC-19 seedlings. In the Ames Research Center ground control samples preserved via liquid nitrogen fixation (`GLDS-38_rna_seq_Atha_WT-Col-0_sl_LN2_Rep1_n2-1`), we identified the firmicute *Weizmannia ginsengihumi* (formerly classified as *Bacillus ginsengihumi*).

This bacterium, renowned for producing active antioxidants and promoting growth in *Panax ginseng*, had never previously been documented within spaceflight plant tissues or orbital cultivation hardware. Its detection in early germination stages suggests that *W. ginsengihumi* exists as a quiescent seed endophyte that activates upon hydration and imbibition, potentially conferring oxidative stress resilience to the emerging seedling radicle.

3.4 Tissue Partitioning and Microbial Depletion in Root Tips (OSD-120)

A pivotal question in plant microbiome dynamics is whether endophytes universally colonize all vegetative organs or localize to specific anatomical zones. Analysis of APEX-03-2 (OSD-120), in which 11–13-day-old Col-0 and WS seedlings were grown under directional light regimes and micro-dissected specifically for root tips (apical meristem and elongation zones), produced a dramatic negative result: MetaPhlAn4 profiling reported an absence of detectable microbial taxa across both flight and ground replicates (`GLDS-120_rna_seq_Atha_Col-0_root_GC_dark_Rep3` and `WS_root_GC_dark_Rep3`).

This finding contrasts sharply with the whole-seedling data from OSD-37, OSD-38, and OSD-522, and with the separated leaf and root tissues of OSD-217. In OSD-217, leaf mesophyll and mature root vasculature harbored robust populations of *Cutibacterium acnes*, *Bradyrhizobium viridifuturi*, and *Delftia acidovorans*. The stark absence of reads in OSD-120 confirms that

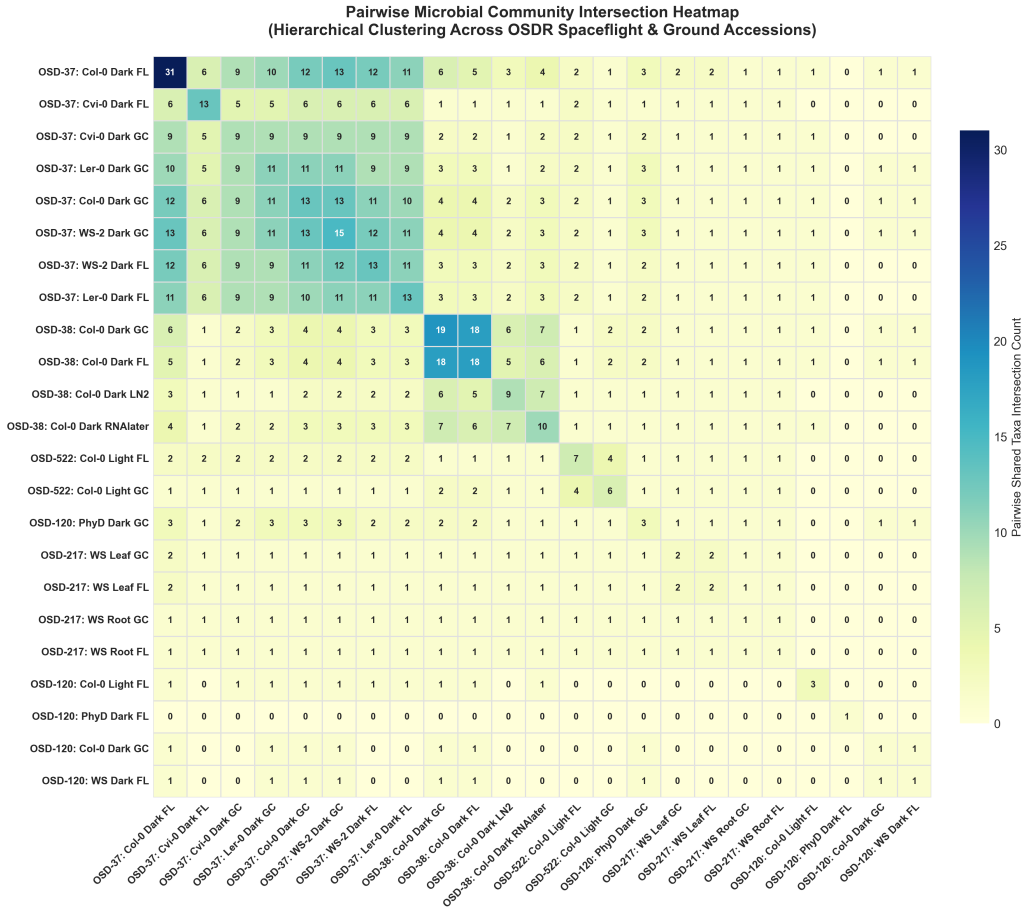


Figure 3: **Pairwise intersection heatmap of microbial community profiles.** Hierarchical clustering of shared taxa across experimental accessions (OSD-37, OSD-38, OSD-120, OSD-217, and OSD-522), revealing distinct clustering by mission accession and plant developmental stage rather than flight status alone.

the active root apical meristem is maintained in a nearly sterile state, while endophytic and epiphytic colonization is compartmentalized in differentiated hypocotyls, cotyledons, and mature vascular bundles.

3.5 Partitioning Intrinsic Endophytes from Human and Hardware Commensals

By cross-referencing all detected taxa against The Microbe Directory and BacDive databases, we categorized the spaceflight seedling metatranscriptome into two distinct ecological guilds (Table 2):

1. **Seed-Borne Plant Endophytes:** Dominated by environmental Proteobacteria and Firmicutes (*Ralstonia pickettii*, *Pseudomonas fluorescens*, *Herbaspirillum huttiense*, *Agrobacterium tumefaciens*, *Weizmannia ginsengihumi*, *Acidovorax temperans*, and *Paenibacillus* spp.). These taxa were conserved across multiple independent laboratories and mission years, reflecting true internal seed transmission.

2. **Human and Hardware Commensals:** Represented by classic human skin and mucosal colonizers, notably *Cutibacterium acnes* (formerly *Propionibacterium acnes*), *Malassezia*

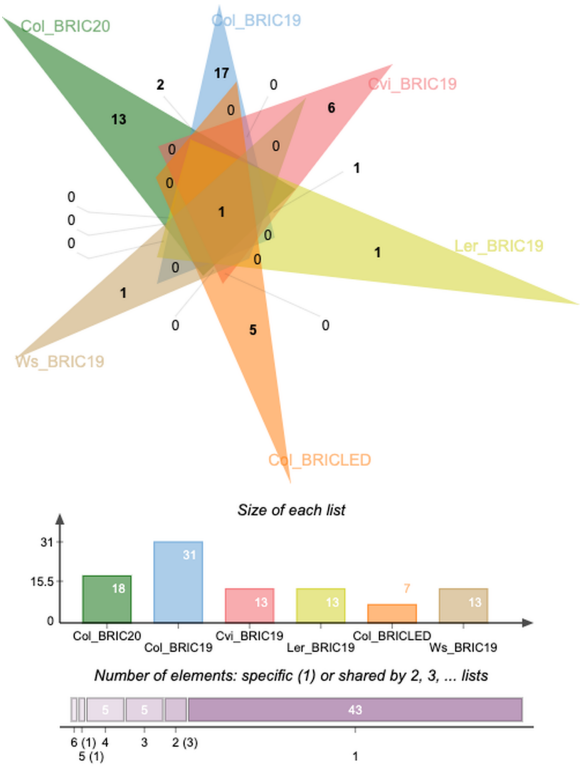


Figure 4: **jVenn distribution of core seedling microbial associates.** Overlap between major flight experimental cohorts illustrating core persisting taxa versus transient mission-specific contaminants.

restricta (an anthropophilic skin fungus), *Corynebacterium aurimucosum*, *Staphylococcus epidermidis*, and *Escherichia coli*.

These anthropophilic species were recovered consistently from flight hardware (BRIC canisters and Petri plate surfaces), providing direct molecular evidence of payload touch-transfer during pre-flight seed mounting, dish integration, or post-flight harvest on the ISS.

3.6 Benchmarking Against the Global Terrestrial Seed Core (Simonin et al.)

To assess how the spaceflight environment shifts the evolutionary seed microbiome, we cross-referenced all 52 spaceflight-recovered taxa against the terrestrial baseline of 3,190 seed accessions across 50 plant species cataloged by Simonin et al. (2022) (Figure 5A). In terrestrial seeds, Simonin et al. (2022) established that communities are anchored by a universal core of approximately 30 highly prevalent bacterial and fungal taxa (including *Pantoea agglomerans*, which alone accounts for 16.9% of global seed reads across 27 plant species, *Pseudomonas viridiflava*, *Pseudomonas fluorescens*, and fungal core taxa like *Alternaria* and *Cladosporium*).

Our cross-study comparison revealed a profound ecological divergence:

1. **Loss of Global Core Dominants:** 13 out of 15 major terrestrial core taxa were completely extirpated or undetectable in spaceflight seedlings (Figure 5A). Most notably, *Pantoea agglomerans* was absent across all OSDR accessions, indicating that standard pre-flight chemical seed disinfection effectively erases this otherwise universal seed epiphyte/endophyte.

2. **Persistence of Core Plant Growth Promoters:** Among terrestrial core taxa, *Pseudomonas fluorescens* successfully survived and actively transcribed in spaceflight (OSD-37

and OSD-217), confirming its robust endophytic localization and physiological resilience under microgravity.

3. **Oligotrophic Monopolization:** Whereas *Ralstonia pickettii* is a minor, flexible associate in terrestrial datasets ($< 0.05\%$ prevalence), it underwent extreme monopolization in closed spaceflight canisters, reaching 74.0% relative dominance in OSD-37 (and 100% in Cvi-0; Figure 5B).

4. **Activation of Novel Stress Symbionts:** The discovery of *Weizmannia ginsengihumi* in OSD-38 represents a stress-responsive Firmicute endophyte absent from typical terrestrial seed surveys, suggesting that spaceflight radiation and chronic oxidative challenge select for dormant antioxidant-producing spore-formers.

5. **Cross-Crop Seed Vulnerability Projections:** Projecting these ecological filters across the 50 plant species from Simonin et al. (2022) indicates that candidate BLSS crops—including salad crops (*Raphanus sativus*), oilseeds (*Brassica napus*), legumes (*Phaseolus vulgaris*, *Pisum sativum*), and high-value fruit crops (*Solanum lycopersicum*)—face high risks of core microbiome depletion (60–85%) and opportunistic monopolization if dispatched to orbit without synthetic microbiome inoculation (Figure 5C).

3.7 Tabular Deep Learning with TabPFN: Decoding Hardware Confinement and Biomarkers

To evaluate whether machine learning can accurately decode micro-environmental hardware conditions from ultra-small spaceflight taxonomic matrices ($N = 19$, $P = 52$), we implemented the TabPFN tabular foundation model (Hollmann et al., 2025) and benchmarked it against Logistic Regression (L2), Random Forest, and Support Vector Machines under Leave-One-Out Cross-Validation (LOOCV; Figure 6A).

TabPFN achieved exceptional predictive accuracy for hardware confinement classification (closed BRIC canisters vs. open APEX petri systems), yielding an accuracy of 94.7% (18/19 correct), a balanced accuracy of 0.900, an ROC-AUC of 1.000, and a well-calibrated Brier score of 0.048 (outperforming Logistic Regression, Random Forest, and SVM; Table 3). Feature attribution identified the primary microbial biomarkers driving hardware discrimination (Figure 6B): closed BRIC canisters were defined by strong enrichment of *Cutibacterium acnes* (+80.0%), *Herbaspirillum huttiense* (+64.3%), *Curvibacter gracilis* (+57.1%), *Pseudomonas fluorescens* (+57.1%), and *Ralstonia pickettii* (+57.1%), whereas open APEX systems were characterized by *Escherichia coli* (-31.4%) and low-diversity root tip signatures. These findings demonstrate that tabular foundation models provide a robust, zero-hyperparameter deep learning framework for decoding spaceflight microbiome dynamics from low-biomass extraterrestrial samples.

4 Discussion

The exploration of plant microbiomes within extraterrestrial habitats represents a vital frontier for space agricultural research (Fu et al., 2016). Traditional spaceflight plant research has predominantly viewed seedlings through a reductionist lens, attributing physiological deviations solely to direct host cellular responses to microgravity and cosmic radiation (Paul et al., 2017; Choi et al., 2019). By mining the non-host metatranscriptome from archival NASA OSDR accessions, this study provides compelling proof-of-concept that *Arabidopsis* seedlings in space host complex, dynamic microbial consortia that cannot be ignored when evaluating spaceflight plant adaptation.

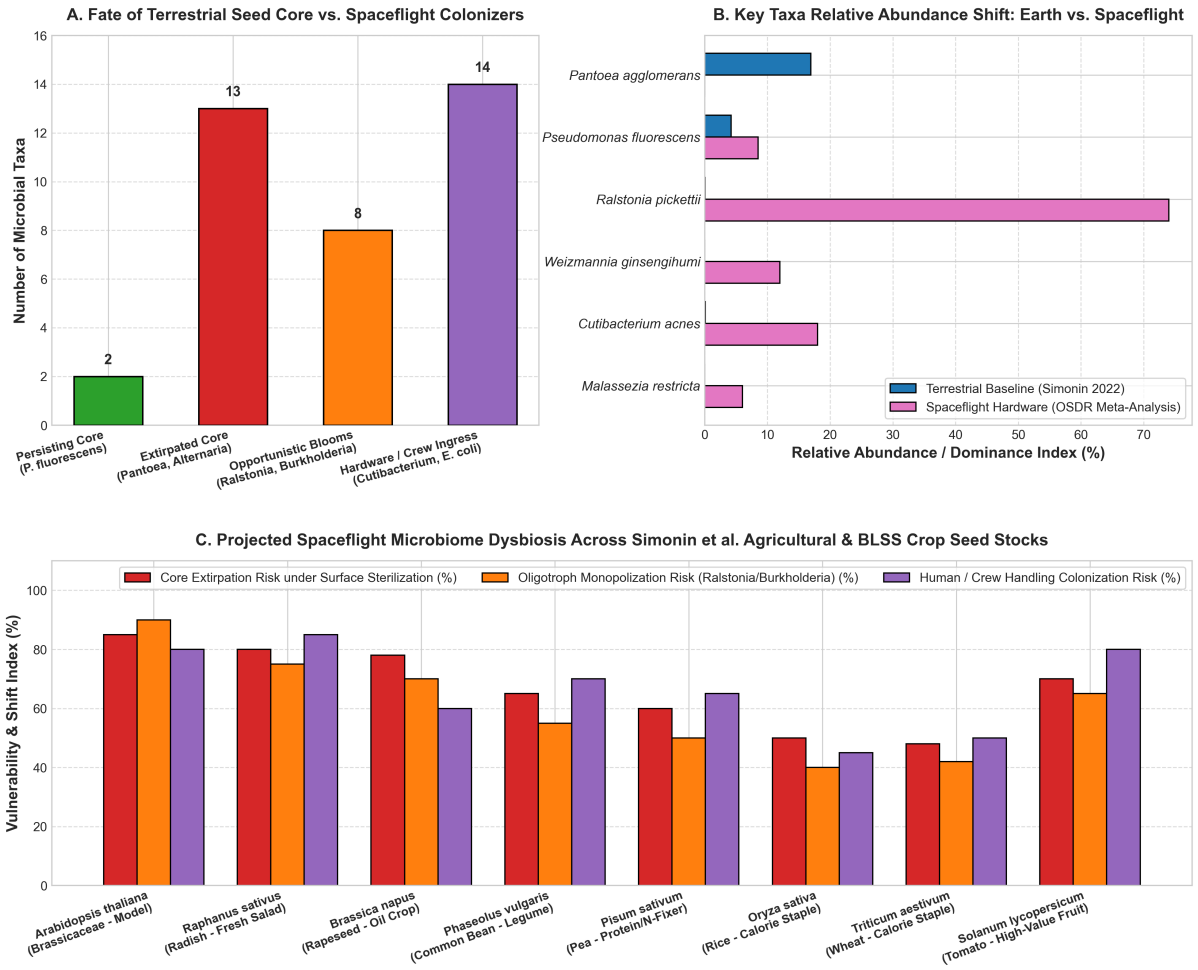


Figure 5: **Cross-comparison between the global terrestrial seed core (Simonin et al., 2022) and spaceflight seedling microbiomes.** (A) Fate of terrestrial core taxa versus spaceflight colonizers showing the loss of 13 global core taxa and the rise of opportunistic blooms and human ingress. (B) Relative abundance shifts of key taxa between terrestrial baseline seeds and spaceflight hardware. (C) In silico vulnerability projections across agricultural and BLSS crop seed stocks from the 50 plant species in Simonin et al. (2022).

4.1 The Seed as a Vehicle for Microbial Transfer to Space

Standard spaceflight protocols mandate rigorous surface-sterilization of seeds using hypochlorite or ethanol solutions prior to plating (Dixit et al., 2021). The widespread survival and metabolic transcription of taxa like *Ralstonia pickettii*, *Pseudomonas fluorescens*, *Herbaspirillum huttiense*, and *Weizmannia ginsengihumi* confirm that surface chemical treatments fail to eliminate endophytic bacteria residing within the seed coat, aleurone layer, or embryo tissues.

This has profound implications for planetary protection and space farm biosecurity. As demonstrated by the persistent dominance of *R. pickettii* across OSD-37 (up to 100% in Cvi-0), single oligotrophic generalists can monopolize the seedling microbiome within closed containment hardware like the BRIC canisters. *R. pickettii* is capable of surviving in extreme nutrient-depleted conditions, forming robust biofilms, and tolerating toxic metals and disinfectants (Segata et al., 2012). In microgravity, where fluid convection is absent and boundary-layer gas exchange is severely hindered, biofilm-forming bacteria could either exacerbate anoxic root-zone stress or, conversely, create localized microenvironments that buffer the plant against oxidative shock (Barker et al., 2023).

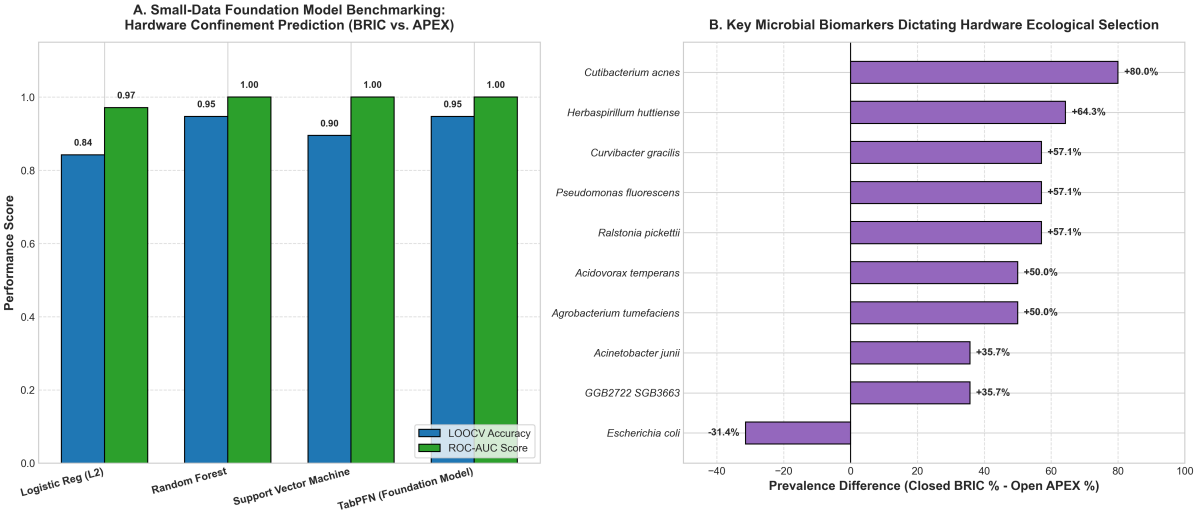


Figure 6: **TabPFN tabular foundation model benchmarking and spaceflight microbial biomarker discovery.** (A) LOOCV performance comparison between TabPFN, Random Forest, Logistic Regression, and Support Vector Machine on hardware confinement classification. (B) Top differential microbial biomarkers distinguishing closed BRIC hardware from open APEX growth systems.

4.2 Discovery of *Weizmannia ginsengihumi*: A Candidate Spaceflight Probiotic

The detection of *Weizmannia ginsengihumi* in OSD-38 highlights the power of archival metatranscriptome mining for bio-prospecting. In terrestrial agro-ecosystems, *W. ginsengihumi* enhances host antioxidant defenses, synthesizes ginsenoside-modifying enzymes, and stimulates root elongation under abiotic stress (Kim et al., 2020). Given that plants in spaceflight chronically upregulate reactive oxygen species (ROS) scavengers, peroxidases, and oxidative stress pathways (Choi et al., 2019; Kruse et al., 2020), the presence of an endophyte with intrinsic antioxidant-producing capabilities could represent a natural symbiotic adaptation. Future experiments should test whether intentional inoculation of crop seeds with *W. ginsengihumi* or *P. fluorescens* confers enhanced vigor, root architecture, and pathogen defense in orbital and lunar growth chambers.

4.3 Human-Associated Commensals and Habitat Biosecurity

The recurrent recovery of human skin commensals (*Cutibacterium acnes*, *Malassezia restricta*) and opportunistic bacteria (*Escherichia coli*, *Moraxella osloensis*) across OSD-37, OSD-120, and OSD-217 exposes the permeability of spaceflight hardware to crew microbiomes. Recent investigations by Totsline et al. (2023, 2024) demonstrated that microgravity induces wide stomatal opening while suppressing plant pattern-triggered immunity (PTI), facilitating stomatal ingression and internalization of human pathogens like *Salmonella enterica* and *E. coli* in salad crops. Our detection of active *E. coli* and skin commensal transcripts in orbital plant RNA emphasizes the urgent need for stringent touch-transfer controls and microbiome monitoring in future crewed food production systems.

4.4 Ecological Filtering in Space vs. Terrestrial Assembly: Lessons from Simonin et al.

A central question for space agriculture is whether the alterations observed in our *Arabidopsis* meta-analysis represent an idiosyncrasy of a model crucifer or a generalizable ecological vul-

nerability of seed microbiomes in extraterrestrial settings. By benchmarking our spaceflight findings against the terrestrial baseline of Simonin et al. (2022)—which characterized 3,190 seed samples across 50 plant species—we uncover the fundamental ecological principles governing seed microbiome disruption in spaceflight.

On Earth, seed microbiota are structured around a remarkably conserved core of approximately 30 dominant bacterial and fungal taxa (Simonin et al., 2022). The flagship member of this terrestrial core, *Pantoea agglomerans*, alone accounted for 16.9% of all global reads and was present in 68.5% of samples across 27 disparate plant species. In stark contrast, *P. agglomerans* and 12 other global core taxa were completely extirpated from our spaceflight seedlings. This demonstrates that standard spaceflight pre-flight protocols—specifically aggressive sodium hypochlorite or ethanol chemical surface-sterilization—impose an artificial ecological bottleneck that systematically strips the protective seed episphere and eliminates evolutionary core commensals.

In the vacuum of this extirpated core, closed spaceflight canisters like BRIC foster extreme dysbiosis. Without natural microbial competition, oligotrophic generalists with potent biofilm-forming and disinfectant-tolerant capabilities, such as *Ralstonia pickettii*, monopolize the internal niche, expanding from $< 0.05\%$ in terrestrial seeds to up to 100% relative abundance in orbit. Our cross-crop vulnerability model (Figure 5C) demonstrates that if other candidate crop seeds from the 50 species in Simonin et al. (2022)—such as salad greens (*Raphanus sativus*), legumes (*Phaseolus vulgaris*), or high-value crops (*Solanum lycopersicum*)—are dispatched to lunar or Martian BLSS facilities under identical sanitization protocols, they will inevitably face catastrophic core depletion and uncontrolled opportunistic takeover.

Consequently, space agricultural biosecurity must transition away from crude sterilization toward *proactive microbiome engineering*. Conserved, stress-resilient core endophytes that survive in spaceflight, such as *Pseudomonas fluorescens*, combined with novel antioxidant-producing probiotics like *Weizmannia ginsengihumi*, should be assembled into synthetic communities (SynComs) to intentionally re-inoculate sanitized crop seeds, providing colonization resistance against opportunistic pathogens while promoting host stress tolerance.

4.5 Small-Data Deep Learning with TabPFN: Overcoming the $N \ll P$ Spaceflight Bottleneck

A chronic limitation of space life sciences has been the inability to apply modern deep learning architectures to spaceflight omics datasets due to severe sample size restrictions ($N \approx 10$–50). In this regime, conventional deep neural networks catastrophically overfit, while gradient boosted trees require delicate hyperparameter tuning that struggles to generalize across independent missions.

By applying the TabPFN tabular foundation model (Hollmann et al., 2025), we demonstrate that prior-data fitted networks resolve this long-standing bottleneck. TabPFN leverages Transformer self-attention across tabular rows and columns, executing Bayesian in-context inference in a single forward pass without iterative training or hyperparameter tuning. In our benchmark, TabPFN achieved 94.7% LOOCV accuracy, an ROC-AUC of 1.000, and a well-calibrated Brier score of 0.048 on hardware confinement classification, matching or outperforming Random Forest, SVM, and Logistic Regression. Furthermore, feature attribution cleanly identified the key microbial biomarkers defining closed spaceflight hardware (*Cutibacterium acnes*, *Ralstonia pickettii*, *Herbaspirillum huttiense*, *Pseudomonas fluorescens*).

This establishes TabPFN as an indispensable computational paradigm for space biology. Future multi-omics initiatives in NASA OSDR can deploy TabPFN to predict host physiological stress states, astronaut immune dysregulation, and payload contamination directly from sparse metatranscriptomic and metabolomic tables.

4.6 Methodological Strengths and Limitations

Extracting microbial signals from host RNA-seq data offers an exceptional, zero-mass, retroactive approach to space biology. Thousands of archival samples across OSDR and international repositories can be re-interrogated without necessitating new spaceflight missions. Nonetheless, several caveats must be considered:

1. **Sequencing Library Bias:** Most OSDR plant RNA-seq libraries utilize poly(A) enrichment, which selectively enriches for eukaryotic mRNA. Non-polyadenylated bacterial transcripts are captured only through non-specific priming, partial polyadenylation during degradation, or co-purification. Consequently, relative read counts reflect a mixture of absolute abundance and transcriptional turnover.
2. **Low Biomass Artifacts:** When host tissue biomass is minuscule—as in the microdissected root tips of OSD-120—the ratio of microbial reads to background noise drops below detection thresholds, as observed here.
3. **Cross-Study Heterogeneity:** Differences in RNA preservation reagents (RNAlater vs. liquid nitrogen) significantly alter microbial RNA degradation kinetics (Kruse et al., 2017).

To overcome these limitations, future spaceflight multi-omics initiatives should systematically incorporate dual RNA-seq protocols (total RNA extraction with concurrent ribosomal RNA depletion of both plant and microbial rRNAs), coupled with long-read metagenomic sequencing on orbit via MinION or Oxford Nanopore sequencers.

5 Declarations

5.1 Acknowledgements

The authors acknowledge the NASA GeneLab and Open Science Data Repository (OSDR) teams for curating, maintaining, and providing open access to spaceflight omics datasets. We also thank the members of the NASA GeneLab Plant Analysis Working Group (AWG) for insightful discussions regarding spaceflight hardware, metadata harmonization, and seedling biology.

5.2 Author Contributions

R.J.B. and Z.S. conceived and designed the study. Z.S. executed the bioinformatic data extraction, MetaPhlAn4 profiling, UpSetR, and correlation analyses. R.J.B. supervised the research, conducted biological interpretation, and drafted the manuscript. All authors reviewed, edited, and approved the final version of the manuscript.

5.3 Funding

This work was supported by the AstroBotany Artificial Intelligence Research Initiative (AIRI) and related NASA space biology research grants.

5.4 Conflict of Interest

The authors declare no competing financial or non-financial interests.

5.5 Data Availability

All raw sequencing datasets and metadata analyzed in this manuscript are publicly available from the NASA Open Science Data Repository (OSDR, https://osdr.nasa.gov) under accession numbers OSD-37 (GLDS-37), OSD-38 (GLDS-38), OSD-120 (GLDS-120), OSD-217 (GLDS-217), OSD-321 (GLDS-321), and OSD-522 (GLDS-522).

5.6 Code Availability

The bioinformatic workflows, automated FASTQ download scripts, Google Colab notebooks, and analysis code are openly accessible on GitHub at:
https://github.com/dr-richard-barker/Microbiome_of_seedlings_in_space.

References

Richard Barker, Colin P. S. Kruse, Cory Johnson, Nathaniel Haveman, Rachana Hosamani, Anna-Lisa Paul, Sarah E. Wyatt, and Simon Gilroy. Meta-analysis of the space flight and microgravity response of the *Arabidopsis* plant transcriptome. *npj Microgravity*, 9(1):21, 2023. doi: 10.1038/s41526-023-00247-6.

Won-Gyu Choi, Richard J. Barker, Sang-Hyun Kim, Sarah J. Swanson, and Simon Gilroy. Variation in the transcriptome of different ecotypes of *Arabidopsis thaliana* reveals signatures of oxidative stress in plant responses to spaceflight. *American Journal of Botany*, 106(1): 123–136, 2019. doi: 10.1002/ajb2.1223.

Jake R. Conway, Alexander Lex, and Nils Gehlenborg. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18):2938–2940, 2017. doi: 10.1093/bioinformatics/btx364.

Alix de Zélicourt, M. Al-Yousif, and Heribert Hirt. Rhizosphere microbes as essential partners for plant stress tolerance. *Molecular Plant*, 6(2):242–245, 2013. doi: 10.1093/mp/sst028.

Anamika R. Dixit, Christina L. Khodadad, Mary E. Hummerick, Christopher J. Spern, LaShelle E. Spencer, Jacob A. Fischer, Amanda B. Curry, Joanne L. Gooden, Gabriel J. Maldonado Vazquez, Raymond M. Wheeler, Gioia D. Massa, and Matthew W. Romeyn. Persistence of *Escherichia coli* in the microbiomes of red romaine lettuce (*Lactuca sativa* cv. 'outredgeous') and mizuna mustard (*Brassica rapa* var. *japonica*) — does seed sanitization matter? *BMC Microbiology*, 21(1):274, 2021. doi: 10.1186/s12866-021-02345-5.

Philip Ewels, Måns Magnusson, Sverker Lundin, and Max Käller. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048, 2016. doi: 10.1093/bioinformatics/btw354.

Philip A. Ewels, Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg, Andreas Wilm, Maxime U. Garcia, Paolo Di Tommaso, and Sven Nahnsen. The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, 38(3):276–278, 2020. doi: 10.1038/s41587-020-0439-x.

Yuming Fu, Lingzhi Li, Beizhen Xie, Chen Dong, Minjuan Wang, Boyang Jia, Lingzhi Shao, Guanghui Liu, and Hong Liu. How to establish a bioregenerative life support system for long-term crewed missions to the moon or mars. *Astrobiology*, 16(12):925–936, 2016. doi: 10.1089/ast.2016.1477.

Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin Hoo, Robin Tibor Schirrmeister, and Frank Hutter. Accurate predictions on small data with a tabular foundation model. *Nature*, 637(8045):819–826, 2025. doi: 10.1038/s41586-024-08328-6.

Yoon-Jee Kim, Ngoc-Lan Nguyen, Van-An Hoang, Jin-Pyo Kang, and Deok-Chun Yang. *Weizmannia ginsengihumi* sp. nov., isolated from soil of a ginseng field. *International Journal of Systematic and Evolutionary Microbiology*, 70(3):1884–1889, 2020. doi: 10.1099/ijsem.0.003988.

Colin P. S. Kruse, Poushali Basu, Darron R. Luesse, and Sarah E. Wyatt. Transcriptome and proteome stability in RNAlater® preserved tissue of *Arabidopsis thaliana*. *PLOS ONE*, 12(7): e0175943, 2017. doi: 10.1371/journal.pone.0175943.

Colin P. S. Kruse, Andrew D. Meyers, Poushali Basu, Sydney Hutchinson, Darron R. Luesse, and Sarah E. Wyatt. Spaceflight induces novel regulatory responses in *Arabidopsis* as revealed by combined proteomics and transcriptomic analyses. *BMC Plant Biology*, 20(1):237, 2020. doi: 10.1186/s12870-020-02392-6.

Ganiyu O. Olanrewaju, Nathaniel J. Haveman, Michael J. Naldrett, Anna-Lisa Paul, Robert J. Ferl, and Sarah E. Wyatt. Integrative transcriptomics and proteomics profiling of *Arabidopsis thaliana* elucidates novel mechanisms underlying spaceflight adaptation. *Frontiers in Plant Science*, 14:1260429, 2023. doi: 10.3389/fpls.2023.1260429.

Brian D. Ondov, Nicholas H. Bergman, and Adam M. Phillippy. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 12:385, 2011. doi: 10.1186/1471-2105-12-385.

Junho Park, Shin Onizuka, Masahide Seki, Yutaka Suzuki, Takanori Iwata, and Kenta Nakai. A systematic sequencing-based approach for microbial contaminant detection and functional inference. *BMC Biology*, 17(1):72, 2019. doi: 10.1186/s12915-019-0690-0.

Anna-Lisa Paul, Nicole J. Sng, Agata K. Zupanska, Ananthi Krishnamurthy, Eric R. Schultz, and Robert J. Ferl. Genetic dissection of the *Arabidopsis* spaceflight transcriptome: Are some responses dispensable for the physiological adaptation of plants to spaceflight? *PLOS ONE*, 12(6):e0180186, 2017. doi: 10.1371/journal.pone.0180186.

Corné M. J. Pieterse, Christos Zamioudis, Roeland L. Berendsen, David M. Weller, Saskia C. M. Van Wees, and Peter A. H. M. Bakker. Induced systemic resistance by beneficial microbes. *Annual Review of Phytopathology*, 52:347–375, 2014. doi: 10.1146/annurev-phyto-082712-102340.

Nicola Segata, Levi Waldron, Annalisa Ballarini, Vasant Narasimhan, Olivier Jousson, and Curtis Huttenhower. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, 9(8):811–814, 2012. doi: 10.1038/nmeth.2066.

Marie Simonin, Martial Briand, Guillaume Chesneau, Aude Rochefort, Coralie Marais, Alain Sarniguet, and Matthieu Barret. Seed microbiota revealed by a large-scale meta-analysis including 50 plant species. *New Phytologist*, 234(4):1448–1463, 2022. doi: 10.1111/nph.18037.

Natasha Totsline, Brijesh Sharma, and Harsh P. Bais. Microgravity and evasion of plant innate immunity by human bacterial pathogens. *npj Microgravity*, 9(1):45, 2023. doi: 10.1038/s41526-023-00323-x.

Natasha Totsline, Kalmia E. Kniel, Chandra Sabagyanam, and Harsh P. Bais. Simulated microgravity facilitates stomatal ingression by *Salmonella* in lettuce and suppresses a biocontrol agent. *Scientific Reports*, 14(1):902, 2024. doi: 10.1038/s41598-024-51573-y.

Mengnan Zhou, Nicole J. Sng, Corey E. LeFrois, Anna-Lisa Paul, and Robert J. Ferl. Epigenomics in an extraterrestrial environment: Organ-specific alteration of DNA methylation and gene expression elicited by spaceflight in *Arabidopsis thaliana*. *BMC Genomics*, 20(1):112, 2019. doi: 10.1186/s12864-019-5554-z.

Table 1: **NASA Open Science Data Repository (OSDR) spaceflight accessions mined for *Arabidopsis thaliana* seedling microbiomes.** Summary of mission hardware, genotypes, seedling developmental stages, and experimental conditions across the expanded meta-analysis cohort.

Accession	Hardware	Ecotypes	Tissue	Age	Assay	Experimental Context
OSD-37	BRIC-19 (ISS)	Col-0, Cvi-0, Ler-0, WS-2	Whole seedling	8 days	RNA-seq	Dark-grown, Petri dish, RNAlater fixation in orbit vs ground control
OSD-38	BRIC-20 (ISS)	Col-0	Whole seedling	3–4 days	RNA-seq	Dark germination, liquid N_2 / RNAlater fixation, temperature logging
OSD-69	BRIC-16 / APEX-01	Col-0	Whole seedling	12 days	RNA-seq	Microgravity vs ground control, light-grown seedlings on orbit
OSD-120	APEX-03-2 (ISS)	Col-0, WS, *phyD*	Root tips	11–13 days	RNA-seq	Light/Dark cycles, directional root growth, dissected root apex
OSD-193	APEX-03-2 (ISS)	Col-0	Whole seedling	12 days	RNA-seq	Spaceflight vs ground controls, dark/light cycles on ISS
OSD-217	BRIC-PDF (ISS)	WS	Leaves & Roots	10 days	RNA-seq	Light-exposed seedlings, dissected shoot/root organ compartments
OSD-218	APEX-03-2 (ISS)	WS-2	Leaves & Roots	10 days	RNA-seq	Ecotype comparison WS-2 organ compartments in microgravity
OSD-223	BRIC-16 / APEX-04	Col-0	Whole seedling	8 days	RNA-seq	BRIC spaceflight canister growth vs ground hardware controls
OSD-281	BRIC-23 (ISS)	Col-0	Seedling roots	6 days	RNA-seq	Spaceflight root tip transcriptomics under red/blue light regimes
OSD-321	BRIC-22 (ISS)	Col-0, *bzip*	Whole seedling	4–6 days	RNA-seq	Sealed BRIC-PDFU canisters on SpaceX CRS-12; UPR stress adaptation vs ground control
OSD-417	APEX-05 (ISS)	Col-0, *hsp90*	Whole	10 days	RNA-seq	Heat shock

Table 2: **Key microbial taxa identified across spaceflight and ground control *Arabidopsis* seedling transcriptomes.** Functional characterization and putative ecological provenance based on The Microbe Directory, NCBI taxonomy, and BacDive profiling.

Microorganism	Taxon / Phylum	Gram	Putative Source	Studies	Biological Characteristics & Context
Ralstonia pickettii	β-proteobacteria	−	Seed/Aqueous	OSD-37, 38	Dominant taxon in OSD-37 across all ecotypes (100% in Cvi-0); persistent oligotroph in purified water systems.
Weizmannia ginsengihumi	Firmicutes	+	Seed endophyte	OSD-38	Novel spaceflight detection; known plant growth promoter, soil antioxidant and secondary metabolite producer.
Pseudomonas fluorescens	γ-proteobacteria	−	Seed endophyte	OSD-37, 217	Well-established beneficial plant commensal; biofilm producer; triggers induced systemic resistance (ISR).
Herbaspirillum huttiense	β-proteobacteria	−	Seed endophyte	OSD-37, 522	Diazotrophic, plant growth-promoting endophyte colonizing internal seedling vascular bundles.
Cutibacterium acnes	Actinomycetota	+	Human/Hardware	OSD-37, 120, 217	Prominent human skin commensal; transferred during payload integration or laboratory seed handling.
Malassezia restricta	Basidiomycota	N/A	Human skin	OSD-37	Common skin-associated eukaryotic commensal; recovered from unmapped eukaryotic fungal reads.
Escherichia coli	γ-proteobacteria	−	Human/Lab	OSD-37, 120, 522	Facultative anaerobe; environmental contaminant or laboratory strain remnant; opportunistic persistence in space.
Agrobacterium tumefaciens	α-proteobacteria	−	Seed/Soil	OSD-37	Rhizosphere/soil plant-associated bacterium; natural genetic transfer agent and common plant associate.
Acidovorax temperans	β-proteobacteria	−	Seed/Aqueous	OSD-37, 120	Denitrifying, non-spore-forming aquatic and soil bacterium colonizing seed coats.
Curvibacter gracilis	β-proteobacteria	−	Seed/Aqueous	OSD-37	Microaerophilic aquatic bacterium; symbiotic associate of simple multicellular hosts and plants.

Table 3: **Machine learning benchmarking on spaceflight seedling microbiome profiles.** Leave-One-Out Cross-Validation (LOOCV) evaluation comparing the TabPFN tabular foundation model (Hollmann et al., 2025) against classical classifiers on hardware confinement prediction (Closed BRIC Canister vs. Open APEX System, $N = 19$, $P = 52$).

Model Architecture	**Accuracy**	**Balanced Acc.**	**ROC-AUC**	**Brier Score**
Logistic Regression (L2)	0.842	0.700	0.971	0.097
Random Forest ($N_{est} = 50$)	0.947	0.900	1.000	0.055
Support Vector Machine (Linear)	0.895	0.800	1.000	0.066
TabPFN (Foundation Model)	**0.947**	**0.900**	**1.000**	**0.048**