



Article

<https://doi.org/10.1038/s41467-023-42104-w>

Microbial adaptation to spaceflight is correlated with bacteriophage-encoded functions

Received: 25 February 2023

Iris Irby & Jared T. Broddrick

Accepted: 27 September 2023

Published online: 15 May 2024

Evidence from the International Space Station suggests microbial populations are rapidly adapting to the spacecraft environment; however, the mechanism of this adaptation is not understood. Bacteriophages are prolific mediators of bacterial adaptation on Earth. Here we survey 245 genomes sequenced from bacterial strains isolated on the International Space Station for dormant (lysogenic) bacteriophages. Our analysis indicates phage-associated genes are significantly different between spaceflight strains and their terrestrial counterparts. In addition, we identify 283 complete prophages, those that could initiate bacterial lysis and infect additional hosts, of which 21% are novel. These prophage regions encode functions that correlate with increased persistence in extreme environments, such as spaceflight, to include antimicrobial resistance and virulence, DNA damage repair, and dormancy. Our results correlate microbial adaptation in spaceflight to bacteriophage-encoded functions that may impact human health in spaceflight.

Human space exploration is at a turning point. The push for long-duration, crewed missions to the moon and beyond, coupled with the emergence of commercial space tourism is redefining the risks associated with space travel. Adverse health effects due to host-microorganism interactions is acknowledged by the NASA Human Research Program, and currently this risk is accepted for short duration spaceflight. However, as flight durations increase and the population that experiences spaceflight expands beyond professional astronauts, it is of increasing importance to understand the interplay between microbial adaptation to the spaceflight environment (e.g., microgravity, ionizing radiation, altered host-microbe interactions, etc.) and human health.

Microorganisms on the International Space Station (ISS) are undergoing adaptation and host-mediated propagation throughout the environment^{1–3}. Studies to date suggest microbes on the ISS, predominantly derived from the crew microbiome, are enriched in functions related to metal ion tolerance, dormancy, and antibiotic resistance^{3,4}. A primary mechanism by which bacteria adapt to novel environments and stress is through genome plasticity, the ability to uptake exogenous genetic material⁵. The presence of temperate

bacteriophages, or prophages, in bacterial genomes is one of the main factors that enable genome plasticity⁶. The entrance of a phage into a bacterial genome is performed through the process of lysogeny, forming a “dormant” prophage containing the entirety of the bacteriophage DNA that is not actively bactericidal, but can undergo replication into a mature virus. These prophages can remain dormant for extended periods of time, or switch into their active (lytic) state, typically after intense DNA damage. Lytic phages can become bactericidal and infect other bacterial cells, taking either their lytic or lysogenic state in their new host⁷.

Prophages dramatically alter their host's physiology. The presence of prophages can disrupt and silence bacterial genes, and they can cause chromosomal rearrangements, leading to changes in gene expression in their bacterial host. Prophages can confer host immunity to other bacteriophages, preventing lytic attack⁸. During the transition to the lytic state, prophages can copy adjacent bacterial genes, integrating them into the viral genome. This genetic material is transferred to new bacterial hosts⁹. Examples of phage transduced functions include virulence and antimicrobial resistance genes, toxins, effector proteins, and regulatory proteins⁸.

¹Space Biosciences Research Branch, NASA Ames Research Center, Moffett Field, CA, USA. ²School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA. e-mail: jared.t.broddrick@nasa.gov

Prophages also affect microbial populations at the community level. The introduction of bacteriophages into soil samples significantly changed the microbial composition and diversity of the soil microbial community¹⁰. In aquatic environments, bacteriophage lyse 20–40% of bacterial biomass per day¹¹. In mouse model studies, both temperate and lytic phages influence the host gut microbiome, with lytic phages leading to a more beneficial community, and temperate phages leading to dysbiosis¹².

The role of bacteriophage in the adaptation of bacteria to the spaceflight environment is unknown. Previous studies have highlighted the importance of microbial communities in spaceflight, with changes noted in the composition of both the gut and skin microbiomes of astronauts¹³. While other studies have noted the presence of prophages in a single bacteria species isolated from the ISS¹⁴, in this work we show a comprehensive analysis of the prevalence and characterization of prophage that inhabit the spacecraft-built environment.

Results

Genome summary and sampling cadence and locations

The International Space Station (ISS) is a unique built environment consisting of 388 cubic meters of habitable space. Crew size has ranged from 3 to 10 over the 22 years of continuous habitation. Surfaces within the occupied modules have been continuously sampled and analyzed for microbial populations to understand the microbiome of the built spacecraft environment. We set out to analyze the genomes that resulted from these experiments to assess the role of bacteriophage in the spacecraft-built environment.

We identified 10 bacterial strains of interest (see Methods): *Acinetobacter pittii*, *Bacillus amyloliquifaciens*, *Bacillus thuringiensis*, *Enterobacter bugandensis*, *Klebsiella quasipneumoniae*, *Micrococcus luteus*, *Paenibacillus polymyxa*, *Pseudomonas fulva*, *Staphylococcus epidermidis*, and *Staphylococcus saprophyticus*. These 10 species represent isolates from sampling campaigns spanning from 2006 to 2018 from eight locations within the ISS (Fig. 1) and a Crew Resupply Vehicle while it was docked with the ISS. Overall, 245 spaceflight genomes and metagenome-assembled genomes¹⁵ were included in our analysis. A full table with links to genomes and data sources can be found in Supplementary data 1.

Prophage region identification and comparison

We extracted prophage regions from flight genomes for each species using PHASTER^{16,17} and summed the prophage regions identified (categorized by PHASTER as complete, questionable, or incomplete), as these region numbers are proportional to the number of prophages that, at some point, lysogenized and entered the bacterial genome. For comparison, we obtained all terrestrial, complete genomes for these species from NCBI¹⁸ and determined the number of prophage regions for these genomes (NCBI accessions can be found in Supplementary data 1). Average number of prophage regions in the flight genomes ranged from 0.75 per genome for *M. luteus* to 7 for *P. fulva* (Fig. 2). For *M. luteus*, 3 of the 8 flight isolates had two possible incomplete prophage regions per genome, but the PHASTER results suggested they were low confidence annotations and prophage analysis by PhiSpy¹⁹, which employs more stringent criteria, did not identify any prophage regions. Thus, *M. luteus* was removed from further analysis. Of the nine remaining species, all had multiple incomplete prophage regions and at least one complete phage, which, in theory, could convert to a lytic phenotype. In addition, five species presented with poly-lysogeny defined as containing multiple complete prophages. There were no universal trends in the number of prophage regions in the flight genomes versus the terrestrial isolates (Fig. 2), with four species having significantly more prophage in flight genomes, four with more in terrestrial, and one that was insignificant (Fig. 2, Mann–Whitney, two-sided, $p < 0.05$).

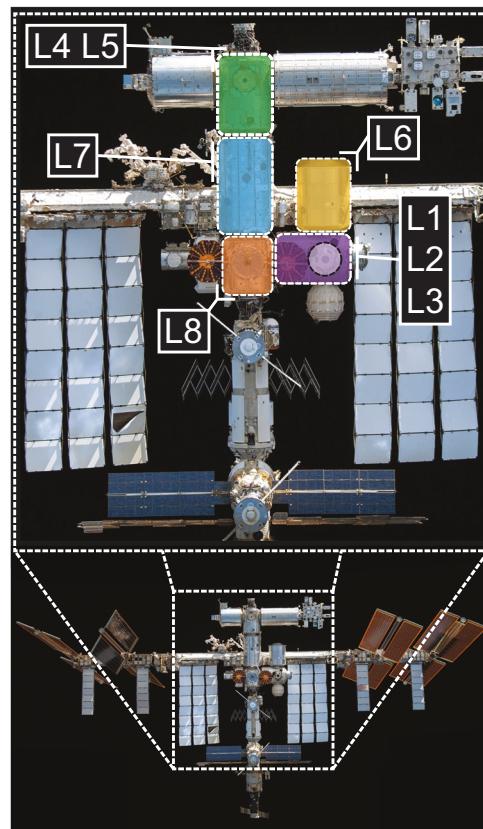


Fig. 1 | Sampling locations on the International Space Station. Purple shaded area: Node 3; sampling locations: Cupola (L1), Waste and Hygiene Compartment (L2), Advanced Resistive Exercise Device (ARED) (L3), Green shaded region: Node 1; sampling locations: dining table (L4), overhead 4 (L5), Yellow shaded region: Permanent Multipurpose Module; sampling locations: (L6), Blue shaded region: US Lab sampling locations: Lab Overhead 3 (L7), Orange shaded region: Node 2; sampling locations: crew quarters (L8). Images courtesy of NASA.

Due to the consistency in flight isolate prophage numbers for most of the species investigated (Fig. 2), coupled with the sampling cadence on the ISS (Supplementary data 1), we explored the genomic diversity between the flight isolates and the terrestrial strains. We used a pan-genome approach to compare the similarity within the flight strains for each species, compared to the pan-genome for all strains of the same species (all genomes in NCBI with an Assembly level at or above scaffold). We defined the core pan-genome as genes present in at least 90% of the genomes, with the remaining genes defined as the accessory pan-genome. We looked at two metrics for our analysis. First, we looked at the ratio in the size of the flight-strain pan-genome versus the all-strain pan-genome, working from the assumption that the more similar the flight strains, the larger the flight-strain core genome. Second, we took the fraction of the flight-strain core pan-genome versus the total flight-strain pan-genome (core plus accessory), as the smaller the accessory genome the less genetic diversity between the ISS strains. For high-confidence prophage regions (prophage calls consistent between PHASTER and PhiSpy algorithms), we removed the prophage regions from the genomes (replaced with N) to capture phage-independent genome variability.

Flight strains for four of nine species (*B. amyloliquifaciens*, *K. quasipneumoniae*, *S. epidermidis*, and *S. saprophyticus*) had a core pan-genome fraction of 0.90 or less (Supplementary data 1). The accessory pan-genome of the remaining five species was 3% or less of the total pan-genome, suggesting a high level of genetic similarity between the flight strains for these species. The ratio of the flight strains' core pan-genome versus the total species core pan-genome was

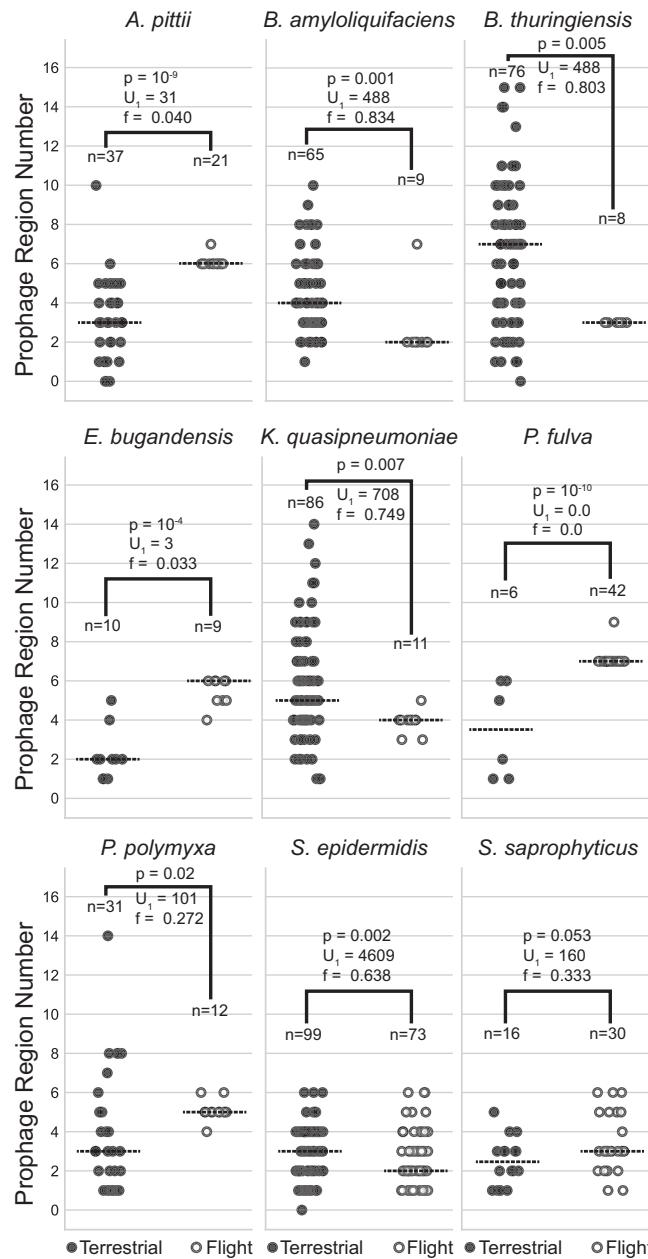


Fig. 2 | ISS (flight) versus terrestrial prophage counts. The median is indicated by a dashed horizontal line for each condition (flight or terrestrial) for each species. The number of genomes in the flight and terrestrial categories are indicated adjacent to each group of data. Two-tailed Mann–Whitney p value, for H_0 , $n_{\text{flight}} = n_{\text{terrestrial}}$, and the U -statistic and Common Language Effect Size (f) for the terrestrial condition are shown. Source data are provided as a Source Data file.

larger for all species, as expected for strains continuously sampled from the same environment. However, the magnitude of the ratio mirrored the change in the core pangenome fraction with the *K. quasipneumoniae*, *S. epidermidis*, and *S. saprophyticus* flight strains having similarly sized core pangenome as the all-strain pangenome (1.06, 1.01, and 1.03, respectively). *S. epidermidis* had the most longitudinal sampling cadence with 30 separate samples taken over 12 years (Supplementary data 1), which correlated with the highest level of genetic diversity between flight genomes (flight versus all core ratio of 1.01 and a flight ISS core fraction of 0.54). However, there were no clear trends in the pangenome comparison versus sampling cadence for the remaining eight species. For example, the 21 flight strains for *A. pittii* were sampled across three campaigns, two of which were 3 years apart.

Still, these genomes showed the highest pangenome core ratio (1.21) and a flight core versus total pangenome fraction of 0.97, with only 89 genes in the accessory pangenome. *S. saprophyticus*' 29 flight genomes came from a similar sampling cadence while having the second most genetic diversity among flight strains. The full genome diversity comparison can be found in Supplementary data 1.

We then applied the pangenome approach to determine if prophages contribute to variation between flight and terrestrial strains of the same species. Again, using all genomes in NCBI with an Assembly level at or above scaffold, but not removing the high confidence prophage regions, we identified Clusters of Orthologous Groups (COG) functions that were significantly different (q value < 0.05) between the flight and terrestrial genomes. For both flight and terrestrial genomes, there were significant differences in gene clusters associated with the phage-associated COG Mobilome (Fig. 3). Terrestrial genomes were enriched for transposases across multiple species, including *A. pittii*, *B. thuringiensis*, *K. quasipneumoniae*, and *S. epidermidis*, compared to spaceflight genomes. Spaceflight genomes were enriched for Mu-like phage structural proteins in *A. pittii*, *E. bugandensis*, and *P. fulva* and enriched for general phage structural proteins in *A. pittii*, *E. bugandensis*, *K. quasipneumoniae*, *P. fulva*, and *S. epidermidis*. Only *S. saprophyticus* contained prophages that were not significantly different between flight and terrestrial genomes. In addition, *A. pittii*, *B. amyloliquifaciens*, *B. thuringiensis*, *E. bugandensis*, and *P. fulva* indicated unique phage-associated gene clusters in both flight and terrestrial genomes suggesting a possible divergence in prophage regions resulting from spaceflight.

Complete prophage classification

To investigate the novelty of the flight prophages, we analyzed the complete prophage regions in comparison to known phage. In total, 283 complete prophage regions were identified in the 9 bacterial species' flight genomes and all 283 were consistent between PHASTER and PhiSpy, suggesting high confidence prophage annotations. We clustered these complete prophages into 71 bins of >90% similarity, and determined the hypothetical phage family of each cluster by NCBI nucleotide BLAST search [Megablast²⁰] of viruses in the RefSeq Genome Database²¹ and the IMG Viral (VR) Database²². The IMG Viral database was the more comprehensive of the two with an average query identity score (nucleotide identity multiplied by alignment percentage) of 55%, compared to 28% for the RefSeq viral genomes. Of the 71 clusters, 11 were unclassified (15%) based on having a total query identity of less than 40% with any phage in the RefSeq or IMG VR Databases. Of the remaining 60 clusters, all complete prophages were found to be tailed phages of class Caudoviricetes (Supplementary data 1). Of these 71 intact prophage clusters, 21 could be classified as novel species and 26 as novel genera based on current suggested cutoffs of 95% query identity for species and 70% for genera²³ (Fig. 4a).

Next, we determined the homology of the complete prophage regions to bacterial genomes in the RefSeq database. As lysogenic prophage regions are incorporated into the bacterial genome, homologous regions may be present in terrestrial genomes in the bacterial database. The average query identity increased from 55% in the viral database to 80% with the bacterial database. Using the same cutoffs as above, 21 of the 71 clusters would be considered novel at the 95% query coverage cutoff and 16 would be considered novel at the 70% cutoff (Fig. 4b). Of the nine species, only *B. thuringiensis* lacked a novel complete prophage at the 70% cutoff (hereafter: non-redundant prophage region), suggesting unique prophage-associated genomic content was broadly observed in the flight genomes (Table 1).

Finally, we attempted to elicit a lytic phenotype from the flight strains. Seven isolates from four different BSL1 species [*B. amyloliquifaciens* (IIF7SC-B1 and IIF7SW-P4), *P. fulva* (F8_7S_9B), *P. polymyxa* (IIF5SW-B4 and IIF8SW-P3), and *S. saprophyticus* (F6_7S_P5 and F5_7S_P13)] were provided by the NASA Jet Propulsion Laboratory. As

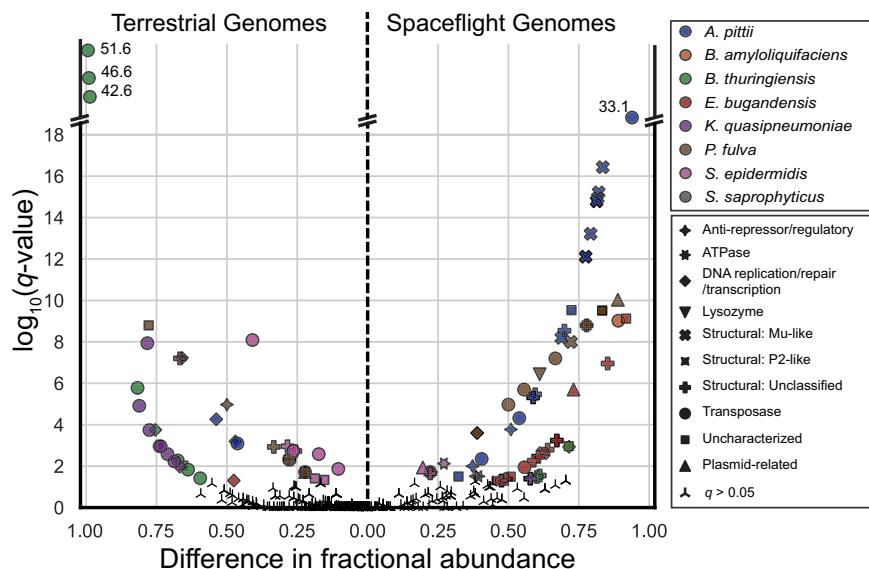


Fig. 3 | Mobilome-associated Clusters of Orthologous Groups (COG) functions significantly different in either flight or terrestrial genomes [q value < 0.05 , two-tailed Rao score test p value corrected for multiple testing (see Methods)].

General categories of phage genes are given different shapes, while the different

species are show in different colors. $\log_{10}(q\text{-values}) > 18$ have their values shown adjacent to their respective data point. Source data are provided as a Source Data file.

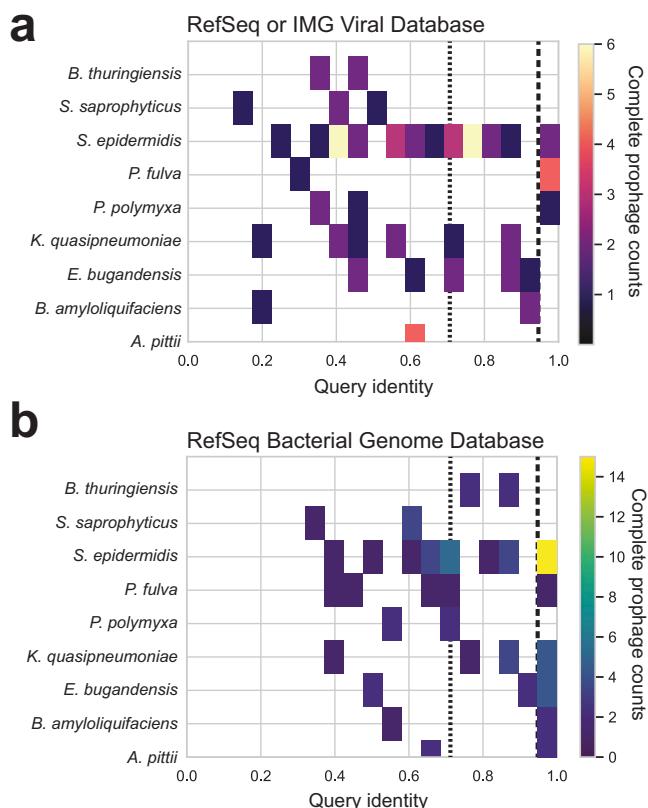


Fig. 4 | Homology of complete prophage regions versus viral and genome databases. **a** Query identity (nucleotide query coverage fraction \times by identity fraction) for complete prophage regions in each species queried against the RefSeq and IMG viral databases. The higher value between the two databases was plotted. **b** Query identity (nucleotide query coverage fraction \times by identity fraction) for complete prophage regions in each species queried against the RefSeq bacteria database. Vertical dashed lines indicate the 0.95 and 0.70 query identity cutoffs. Source data are provided as a Source Data file.

the phage induction mechanisms have not yet been explored in these species, we subjected them to two common DNA damaging stimuli, mitomycin C and UV exposure. For all strains, log-phase cultures treated with mitomycin C (500, 100, 20, 4, and 0.8 ng mL⁻¹) showed a dose-response decrease in culture viability, as measured by OD600, ~8 h post-mitomycin C addition (Supplementary Fig. 1). In contrast, cell viability was more variable following UV-C radiation (254 nm; 5, 25, 50, 100, 200 J m⁻²). Both *S. saprophyticus* species showed dose-dependent effect on the growth curve (b–c). This was similar to both *B. amyloliquifaciens* isolates for the initial 12 h of growth, after which the growth curves were indistinguishable between treatments, which coincided with the formation of a biofilm (Supplementary Fig. 2d, e). The data for *P. polymyxa* was inconclusive as this species' biofilm growth phenotype made it difficult to standardize culture starting densities and resulted in non-uniform light attenuation during OD600 determination (Supplementary Fig. 2f, g). *P. fulva* growth curves showed very little change in response to UV expose, which may suggest enhanced UV resistance in this ISS isolate (Supplementary Fig. 2a).

For all species with more than one isolate, the complete prophages were identical in both strains, preventing us from testing for bacteriophage infection via the soft-agar overlay method. Instead, we attempted to purify phage particles post-mitomycin C addition via filtration (0.22 μ m) and precipitation (8% PEG8000, 1 M NaCl). While we were unable to recover phage particles, our attempt was cursory, and a more comprehensive exploration of phage induction in these isolates is needed.

Long-read assemblies provide insight into mobile genetic elements

We explored the contribution of genome assembly level on the prophage predictions by both the PHASTER and PhiSpy algorithms. From the seven isolates we sequenced two strains, *P. fulva* F8_7S_9B and *S. saprophyticus* F5_7S_P13, via Nanopore MinION. *P. fulva* was chosen as all the spaceflight strains are poly-lysogenic with a non-redundant prophage region in each strain (Cluster 52, Table 1). *S. saprophyticus* F5_7S_P13 is a member of Cluster 41 containing a non-redundant prophage region (Table 1).

Assemblies for both species resulted in a single contig for the bacterial chromosome. *S. saprophyticus* F5_7S_P13 contained three

Table 1 | Non-redundant prophage regions in spaceflight isolates

Species	Cluster	Representative Phage	Genomes in Cluster	Size (kb)	Family	Query Identity ^a		
						NCBI Virus	IMG Virus	NCBI Bacteria
<i>A. pittii</i>	2	IIF1SW-P2 sequence:120779-155538	13	34.8	Caudoviricetes	0%	65%	59%
	3	IIF1SW-P4 sequence:16202-50961	8	34.8	Caudoviricetes	0%	65%	59%
<i>B. amyloliquefaciens</i>	6	IIF7SW-B4 sequence:76-30790	2	30.7	Caudoviricetes	0%	55%	21%
<i>E. bugandensis</i>	12	IFACSW-P1 sequence:830462-877738	7	47.3	Caudoviricetes	11%	52%	47%
	17	IF2SW-F2 sequence:448858-496167	2	47.3	Caudoviricetes	11%	51%	47%
<i>K. quasipneumoniae</i>	24	IIIF3SW-P1 sequence:1181253-1232366	2	51.1	Unclassified	4%	37%	20%
<i>P. fulva</i>	52	F8_8S_9P sequence:173154-179665	1	6.5	Caudoviricetes	0%	56%	33%
<i>P. polymyxa</i>	26	IIF5SW-B4 sequence:979867-1033318	5	53.5	Caudoviricetes	0%	56%	33%
	28	IIF2*SW-P4 sequence:600724-654175	2	53.5	Caudoviricetes	0%	66%	29%
<i>S. epidermidis</i>	34	IF4SW-P4 sequence:652269-693911	2	41.6	Caudoviricetes	20%	58%	55%
	38	s34 sequence:116777-162391	2	45.6	Caudoviricetes	16%	68%	41%
	62	s51 sequence:616-49418	1	48.8	Caudoviricetes	11%	65%	54%
	69	s17 sequence:35-44746	1	44.7	Unclassified	29%	40%	39%
	74	F6_3S_P_7 sequence:72993-118974	1	46	Caudoviricetes	14%	48%	45%
<i>S. saprophyticus</i>	41	F6_7S_P_5 sequence:592884-635056	2	42.2	Caudoviricetes	43%	63%	25%
	42	F5_7S_P5C sequence:174-42331	5	42.2	Caudoviricetes	43%	63%	25%
	70	s43 sequence:757637-811609	1	54	Unclassified	5%	32%	12%
	71	F5_7S_P11B sequence:38923-82523	1	43.6	Caudoviricetes	50%	61%	24%

^aNCBI Megablast query coverage × alignment.

plasmids of 30, 33 and 36 kb in length. For this species, all five of the short-read assembly prophage regions mapped to the long-read assembly. However, one short-read incomplete prophage region was missing in the long-read assembly PHASTER analysis, despite the entire prophage region mapping to the long-read assembly (Supplementary data 1). This prophage region was present in the PhiSpy prophage calls for the long-read assembly. In addition, the long-read prophage analysis suggested the complete prophage from this species was truncated in the short-read assembly (42.1 kb versus 56.2 kb). Interestingly, the incomplete prophage #5 from the short-read prophage analysis (Accession: NZ_JAFDPF010000007.1, sequence: 10127 to 19977) was located on the 36 kb plasmid.

The prophage analysis for *P. fulva* F8_7S_9B suggested the short-read assembly led to fragmented prophage analysis. The long-read assembly identified 9 prophage regions, compared to 7 in the short-read assembly, with one short-read incomplete prophage region missing in the long-read assembly analysis (Supplementary data 1). Similar to *S. saprophyticus*, the entire prophage region mapped to the long-read assembly with no clear reason for its exclusion from the PHASTER prophage calls. This prophage region was not identified by PhiSpy in the long-read analysis. Long-read prophages #1, 5 and 6 were absent in the short-read analysis due to the prophage regions being split across separate contigs in the short-read assembly. Of the two complete prophages in *P. fulva* F8_7S_9B, a 62.9 kb long-read prophage region was split into two short-read prophage regions due to lying on two adjacent contigs. The other complete prophage region was annotated as a single 47.6 kb region in the long-read assembly but as two prophage regions in the short-read assembly (intact: 35.6 kb and incomplete: 12.1 kb region). The long-read assemblies can be found on NASA's Open Science Data Repository (OSD-582, DOI: 10.26030/v6s6-w170) and NCBI (see Data Availability Statement).

Functional analysis of prophage region-encoded proteins

We selected a representative strain from each species and curated the prophage-encoded bacterial gene annotations to assess the functional contribution of phages to their hosts. Of the functionally annotated content, 46% was uncharacterized or hypothetical proteins. Of the remaining content, 20% was annotated as related to phage processes

such as structural proteins (e.g., capsid, tail, etc.), lysogeny, and phage replication (e.g., integrases, terminases). The remaining content was annotated as microbial proteins, conferring a variety of capabilities to the host species (Fig. 5). The total number of prophage-encoded bacterial proteins that could be ascribed a function ranged from ~60 for both *Staphylococcus* species to a high of 102 for *E. bugandensis* (707 total for all species). The most prominent functional category contained proteins related to Signal Transduction and Transcriptional Regulation (104 proteins across all species), followed by DNA Repair and Replication (91 proteins), and Cell Wall Biogenesis and Degradation (65 proteins). To investigate the subset of prophage-encoded proteins that provide unique capabilities to their host, we extracted prophage-encoded annotations that lacked functional homology to host genome-encoded proteins (Fig. 5). We computed a conservation score for each category defined as the fraction of prophage-encoded proteins that were functionally unique. While only consisting of five proteins across three species, Secretion Systems had the highest conservation score of 1.0 (100% of the prophage-encoded proteins conferred a unique capability to their host). Translation, Ribosomal Structure and Biogenesis, as well as Lipid and Fatty Acid Metabolism also had high conservation scores (0.9 for both categories). The prominent prophage-encoded functional categories of Signal Transduction and Transcriptional Regulation, DNA Repair and Replication, and Cell Wall Biogenesis and Degradation had conservation scores of 0.6, 0.51, and 0.48, respectively.

Many of the functions conferred by prophage-encoded proteins are correlated to increased fitness in competitive environments. All nine species had putative virulence related genes encoded on prophage regions, with 45% of these conferring unique functions to their hosts. Based on the assessment of antibiotic resistance genes using ABRicate, *P. polymyxa* prophage contained an aminoglycoside resistance gene (*aadK*), and *S. epidermidis* prophage contained two confirmed antibiotic resistance genes, a macrolide resistance gene [*mph(C)*] and an efflux pump [*msr(A)*]. Approximately 33% of the putative AMR proteins were possible resistance pumps for macromolecules. *A. pittii*, *B. amyloliquefaciens*, and *E. bugandensis* all contained prophage-encoded pyocin-related defense proteins. Both *Staphylococcus* species contained prophage-encoded pathogenicity

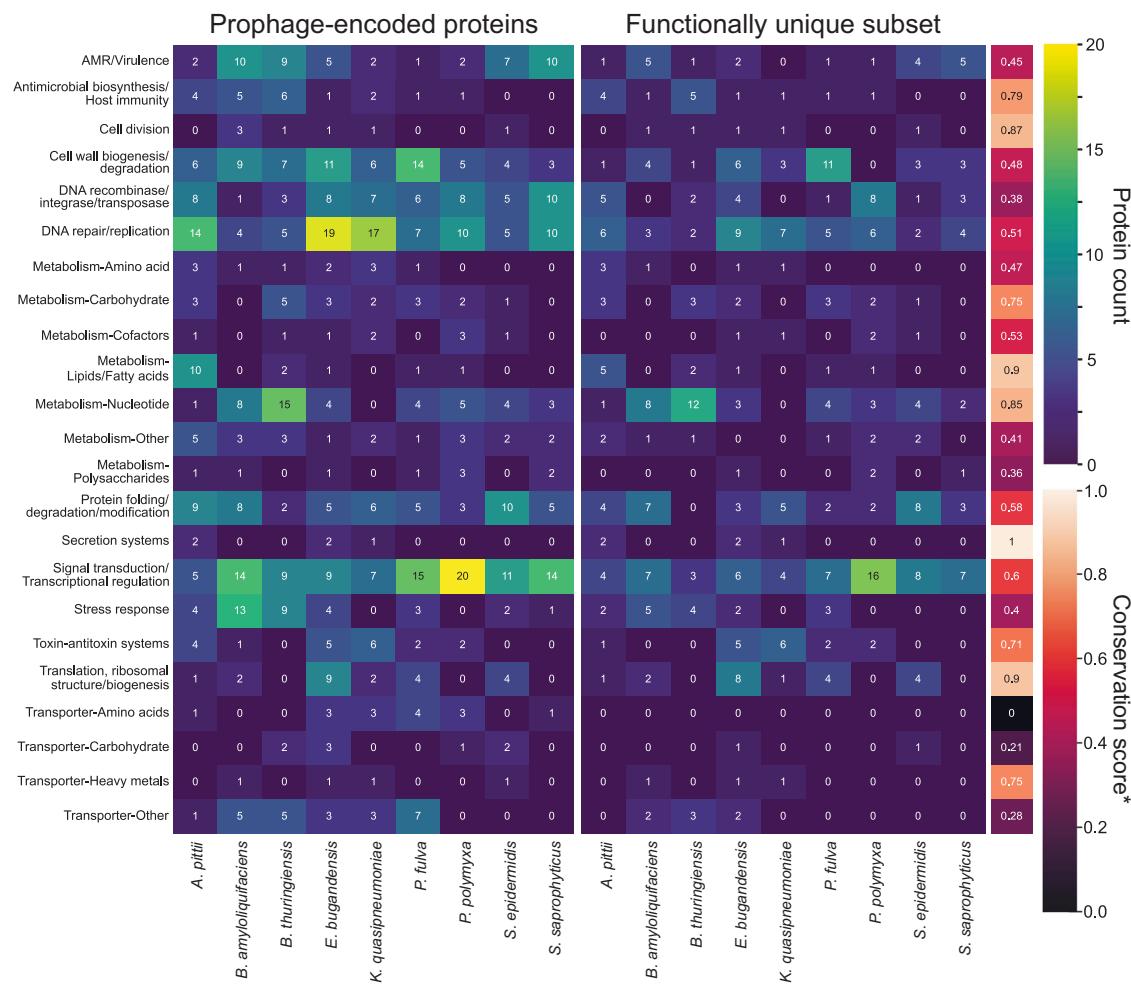


Fig. 5 | Species-specific prophage-encoded protein functions. Left panel: all prophage encoded proteins in each species. Right panel: the subset of prophage-encoded proteins that confer a unique function to their host. *Conservation score:

mean of the functionally unique protein count divided by the total prophage protein count for each functional category. Phage-related proteins and uncharacterized/hypothetical proteins were excluded from the figure.

island genes and enzymes associated with putative virulence or antibiotic resistance functions (Supplementary data 1). *A. pittii*, *B. amyloliquifaciens*, and *B. thuringiensis* contained prophage-encoded enzymes for deoxynucleotide sugar biosynthesis found in both antimicrobial biosynthetic pathways, such as streptomycin, as well as polyketide biosynthesis related to pathogenicity and antimicrobial activity²⁴. These enzymes were unique to the prophage regions, with no homologs found in the non-prophage host genomic content.

Additional prophage-encoded gene products may confer fitness in the built spacecraft environment. Stress response proteins included those associated with sporulation and hibernation and metal resistance genes and transporters. Both categories have been shown to be enriched in spacecraft-associated microorganisms³. Four species had prophage-encoded LexA-related transcriptional regulators or repressors involved in the regulation of the SOS response in bacteria, which have recently been shown to be uniquely present in ISS strains of *A. pittii*²⁵. This study also showed the presence of the RelBE toxin-antitoxin system in ISS *A. pittii* strains, which is also prophage-encoded.

Discussion

Prophages are cited to be the largest factor associated with strain variation and diversity within terrestrial bacterial species^{5,7}, which holds in the spaceflight environment as phage-associated genes were significantly different in 8 of the 9 phage-containing species analyzed. Several spaceflight species contained genes related to the production

of Mu-like phage proteins compared to their terrestrial counterparts. Mu-like phages are unique in that they randomly integrate into the host genome, increasing mutation rates and host gene disruption²⁶. Increased genetic diversity via Mu-like phage induced mutation may assist in microbial adaptation to the spaceflight environment and would explain the retention of these phage in flight strains. In comparison, terrestrial genomes contained transposases absent in the flight strains. Transposases are common in dense populations where there is rampant DNA sharing across a population of individuals²⁷. The absence of these transposase regions may be related to the nutrient poor environment of the ISS. The presence of transposases in the terrestrial organisms may also be indicative of active mobile genetic elements, which we were unable to confirm for the spaceflight isolates.

Our analysis suggested a high number of novel prophages in the flight genomes. In total, we identified 283 complete prophages making up 71 unique prophage clusters, 18 of which were hypothesized as novel, and host-related genes making up over a third of the genes present in these prophage regions. While the discovery of novel prophages is not uncommon, with viral dark matter reported to be 70–95% of human associated viromes^{28,29}, a previous viral metagenomic study was able to classify the majority of the viruses in extreme environments³⁰.

Prophages are a primary mechanism of bacterial host gene transfer, to include antibiotic resistance and virulence genes³¹. We identified at least one putative antibiotic resistance or virulence host-related gene within the prophage regions of all species studied. In

comparison, <12% of highly antibiotic resistant ESKAPE pathogens carry a prophage-encoded antibiotic resistance gene³². In addition, lysogens carrying antibiotic resistance genes have been shown to provide a fitness benefit for their host^{31,33}. There are several reports of spaceflight-induced increases in antibiotic tolerance or increased antibiotic resistance gene clusters^{14,25,34–36}. With our identification of confirmed antibiotic genes within prophage regions of *P. polymyxa* and *S. epidermidis*, as well as the conservation of prophage regions within members of the same species, we describe one possible mechanism for this increased antibiotic tolerance.

We also identified categories of genes related to host immunity, stress response, and DNA repair. The presence of prophages within a bacterial host can provide immunity to attack from other bacteriophages or phage-like pyocins by inhibiting lytic phage replication, or repressing other phages^{37,38}. We identified several host immunity genes in the flight prophages that could mediate phage and pyocin resistance. In addition, both *Bacillus* species contained prophage-encoded sporulation genes, related to persistence in harsh terrestrial environments^{39,40} and spaceflight³. Prophage-encoded DNA repair genes were present across several species, notably those activated in response to radiation. UV radiation induced activation of lytic phage is well known; however, the ISS structure shields the interior from UV radiation. The presence of DNA-repair genes may play a role in the protection of the host genome from galactic cosmic rays or general DNA damage from desiccation in the built environment⁴¹.

The association between genomic features enriched in ISS isolates and prophage regions suggest phage are important for adaptation to the spaceflight environment. However, these are correlations, and experimentation is required to directly connect prophage-mediated genetic plasticity with persistence in the built spacecraft environment. In addition, the opportunity for phage infection of microbial strains after they leave the crew microbiome is likely limited. Thus, an important line of inquiry is to assess the lytic-lysogenic traits of microorganisms while they are still host-associated.

Methods

Flight and terrestrial genomes

Candidate flight species to use in our analysis were determined from literature^{42–44}. Isolates from the ISS were identified in the NCBI databases¹⁸ by searching the BioSample database for “geo_loc_name=USA: International Space Station”[attr] AND “SPECIES NAME”. These hits were mapped back to the NCBI Assembly database and the flight genomes were validated and downloaded. The terrestrial genomes were downloaded from the Assembly database with the following conditions: Latest RefSeq only, Exclude anomalous and Taxonomy Check Status OK. The Assembly Level differed based on analysis with Scaffold and above used for pangenome analysis and Complete genome was used for prophage identification. The Assembly Level for all ISS flight genomes were either Scaffold or Contig. Metagenomes were downloaded from the NASA Open Science Data Repository OSD-69 (DOI: 10.26030/cary-q426)⁴⁵. A full list of genomes, reference links and sampling locations can be found in Supplementary data 1.

Prophage identification and classification

Prophage regions for both ISS and terrestrial isolates were identified with PHASTER¹⁷. Results were separated into complete, questionable, and incomplete prophage regions and the prophage regions in fasta format were downloaded from the PHASTER web-based server. In addition, flight and terrestrial genomes, downloaded from NCBI using E-utilities (v 1.77) in GenBank format, were also searched for prophage regions using PhiSpy [v4.2.20,¹⁹] using default parameters. PhiSpy prophage regions consistent with PHASTER prophage regions were analyzed using ABRicate (v1.0.0, Seemann T, Abricate, Github <https://github.com/tseemann/abricate>) with all associated databases^{46–53} to identify antibiotic resistance genes in prophage regions. Fasta files of

the bacterial genomes without prophages were exported and used for the flight isolate genome diversity analysis. Prophage regions from ISS strains were queried against the Virulence Factor Database (accessed April 2023)⁵³ to identify putative virulence factors using NCBI Blast²⁰. Statistical differences in prophage number were determined using a two-tailed Mann–Whitney U rank test, correcting for ties, as implemented in SciPy [v1.7.3,⁵⁴]. The effect size was computed as the common language effect size f value where $f = \frac{U}{n_1 n_2}$. The U-statistic and f value were reported for the terrestrial condition. Prophage regions identified as complete were clustered using MeShClust [v3.0,⁵⁵] with a cut off of 90% nucleotide identity. The nucleotide sequence labeled as the center of each cluster was assessed for homology with known viral and bacterial sequences in the NCBI RefSeq genome database²¹ using BLAST [Megablast²⁰, accessed November 2022] and high confidence genomes in the IMG/VR database [v4.1, accessed July 2023²²]. Phage classification was determined based on the taxonomic similarity to the best hit in the viral database (query coverage percentage multiplied by percent identity > 40%).

Pangenome analysis

Pangenome analysis of the flight and terrestrial genomes from NCBI and the NASA Open Science Data Repository⁵⁶ was completed using the pangenomics workflow^{57–59} in anvi'o [v7.1⁶⁰]. For *Bacillus thuringiensis*, *Klebsiella quasipneumoniae*, and *Staphylococcus epidermidis*, the terrestrial genomes were subsampled into sets of 300 genomes and the pangenome workflow run separately on each subset (2 sets for *B. thuringiensis* and *K. quasipneumoniae* each, and 3 sets for *S. epidermidis*). For the ISS flight genome diversity analysis, the non-prophage containing bacterial genomes output from PhiSpy was used for analysis. Outputs from the anvio-summarize function were parsed for the number of genomes in which each gene was present. The core pangenome was defined as genes present in at least 90% of the genomes (or for ISS species with less than 10 genomes, the cutoff was n-1 genomes, where n is the number of available genomes). The accessory pangenome was defined as the difference between the core pangenome and the total number of genes in the pangenome. The core ratio was the defined as the number of genes in the ISS flight core pangenome versus to all-strains (ISS and terrestrial) core pangenome. The core pangenome fraction was defined as the number of genes in the core pangenome divided by the total pangenome size (core plus accessory).

For the functional enrichment analysis, the full bacterial genome with prophage regions was used for analysis. Functional enrichments between flight and terrestrial strains were determined using the built in anvi-compute-functional-enrichments-in-pan workflow⁵⁹ with COG20_FUNCTION as the annotation source. This workflow used the two-tailed Rao test statistic to determine the significance of a categorical variable (in this case flight versus terrestrial) to explain the functional occurrence. The resulting *p* value was corrected for multiple tests and reported as a *q* value, with a set significance threshold of *q* < 0.05. For species with subsampled genomes, a COG was considered enriched if it was present at a *q* value (*p* value adjusted for multiple tests) of 0.05 or lower for at least two subsets. Enrichment summaries and outputs from the anvio-summarize script were used for downstream analysis.

Mitomycin C and UV treatment of ISS bacterial isolates

Flight isolates of *Paenibacillus polymyxa* (strains IIFSW-B4 and IIF8SW-P3), *Bacillus amyloliquefaciens* (strains IIF7SW-P4 and IIF7SC-B1), *Pseudomonas fulva* strain F8_7S_9B, and *Staphylococcus saprophyticus* (strains F5_7S_P13 and F6_7S_P5) were provided by the Biotechnology and Planetary Protection Group at the NASA Jet Propulsion Laboratory and grown overnight in tryptic soy broth (TSB) at 30 °C. Cultures were diluted to an OD600 of 0.3–0.4 and 4 mL was placed in a 30 mm diameter plastic petri dish. With the lid removed, the cultures were

subjected to aggregate UV doses of 5, 25, 50, 100, and 200 J m⁻² of UV-C radiation from a G3OT8 UV lamp (75 μW cm⁻² at the plate surface). Doses were determined using a 2π UV-B/C light probe connected to a hand-held light meter (ILT2400, International Light Technologies, Peabody, MA, USA) in integrate mode. After each dose was delivered, duplicate 200 μL aliquots were removed from the petri dish and placed into separate wells of a 96-well microtiter plate. Cell growth was determined by microplate reader optical measurements at 600 nm once per hour (SpectraMax 340, Molecular Devices, San Jose, CA, USA). Negative control sample OD600 was subtracted from the time-course culture data prior to analysis.

For the mitomycin C treatment, flight isolates grown overnight in TSB at 30 °C were diluted to 0.3–0.4 OD600 and 200 μL was aliquoted into a 96-well plate. Mitomycin C stock solution in DMSO (10 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) was diluted to 110 μg/mL in TSB before being serially diluted five times at 5× dilution in TSB. 5 μL of these dilutions, along with a TSB control, were added to separate wells of the culture-containing 96-well plate (2 technical replicates per isolate) resulting in a final concentration set of 500, 100, 20, 4, 0.8, and 0.0 ng mL⁻¹. Cell growth was determined by microplate reader optical measurements at 600 nm once per hour (SpectraMax 340, Molecular Devices, San Jose, CA, USA). Negative control sample OD600 was subtracted from the time-course culture data prior to analysis.

For phage isolation attempts, 5 mL of cultures were grown to mid-log phase before being treated with mitomycin C (0.5 μg/mL final concentration). Cultures were returned to the 30 °C incubator overnight. Cells and debris were removed by centrifugation and the supernatant was filtered through a low-protein binding 0.22 μm PES syringe filter. To the clarified supernatant, a PEG8000, NaCl stock solution was added to yield a final concentration of 8% PEG8000, 1 M NaCl. Solutions were incubated overnight at 4 °C, transferred to triplicate 1.5 mL microcentrifuge tubes and centrifuged at 15 000 × g for 30 min at 4 °C. Tubes were visually examined for the presence of phage pellets.

Long-read genome sequencing of bacterial isolates

Flight isolates of *Paenibacillus polymyxa* (strains IIFSW-B4 and IIF8SW-P3), *Bacillus amyloliquefaciens* (strains IIF7SW-P4 and IIF7SC-B1), *Pseudomonas fulva* strain F8_7S_9B, and *Staphylococcus saprophyticus* (strains F5_7S_P13 and F6_7S_P5) were grown on tryptic soy agar plates. After 72 h the plates were scraped and biomass was suspended in 1 mL of buffer (100 mM Tris HCl, 50 mM NaCl, pH 7.0), and genomic DNA was extracted via a lysozyme, proteinase K treatment followed by phenol-chloroform extraction and ethanol precipitation as outlined in⁶¹. DNA quality was assessed by fluorometric quantification (Qubit, Invitrogen, Waltham, MA, USA) and UV-Vis microvolume spectrophotometry (Nanodrop, Thermo Fischer Scientific, Waltham, MA, USA). *Paenibacillus polymyxa* and *Bacillus amyloliquefaciens*, as well as *Staphylococcus saprophyticus* F5_7S_P13, were excluded from the sequencing run due to biofilm polysaccharides preventing extraction of gDNA with sufficient purity for MinION sequencing. All gDNA manipulations were performed using wide-bore pipet tips to prevent genomic DNA (gDNA) shearing. Native gDNA was sequenced on a MinION R9.4 flowcell (Oxford Nanopore Technologies (ONT), Oxford, UK). The sequencing library was prepared using the ONT Rapid Barcoding Sequencing kit (SQK-RBK004, Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's protocol. Sequencing Fast5 files were base-called using the ONT Guppy base-caller (v6.2.1) using the super high accuracy model. Genomes were assembled using Flye [v2.9.1⁶²] and polished with the ONT tool Medaka (v1.7.0, <https://github.com/nanoporetech/medaka>) using default parameters for both. Closed contigs were checked for prophage regions using the PHASTER web-based interface as outlined above.

Functional annotation of prophage regions

Open reading frames encoded within the prophage nucleotide sequences from the PHASTER output were determined using the NCBI ORFfinder¹⁸. Open reading frames were annotated using Microbe-Annotator [2.0.5,⁶³], using DIAMOND search for homology determination⁵⁸. The output file was combined with the E values for the homology scores and any E value higher than 1×10^{-40} was flagged for manual curation. Proteins flagged for manual curation were compiled into a single file in fasta format and searched against the NCBI RefSeq database. The results were curated for accuracy and confidence and the annotation added to the summary file. All annotations in the summary file were then manually assigned to one of 21 categories based on their functional annotation.

All data visualization was performed using in Python (v3.9) using Jupyter Notebooks [v1.0.0,⁶⁴] and the packages Matplotlib [v3.5.1,⁶⁵] and Seaborn [v0.11.2,⁶⁶]. Additional packages used for analysis included NumPy [v1.22.3,⁶⁷], Pandas [v1.4.2,^{68,69}], and Biopython [1.79,⁷⁰]. No custom code, algorithms, or scripts central to the research were generated.

Reporting summary

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

Data availability

The long-read assemblies and raw fast5 files generated in this study have been deposited in the NASA Open Science Data Repository (<https://osdr.nasa.gov/>) under accession code OSD-582 (<https://osdr.nasa.gov/bio/repo/data/studies/OSD-582>, <https://doi.org/10.26030/v6s6-w170>) and in the NCBI databases Assembly under accession codes GCF_017167745.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_017167745.1) and GCF_017166245.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_017166245.1), BioProject under accession code PRJNA994947 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA994947>), BioSamples under accession codes SAMN36453224 (<https://www.ncbi.nlm.nih.gov/biosample/36453224>) and SAMN36453224 (<https://www.ncbi.nlm.nih.gov/biosample/36453224>), and SRA under accession codes SRR25281509 ([https://www.ncbi.nlm.nih.gov/sra/SRX21026417\[accn\]](https://www.ncbi.nlm.nih.gov/sra/SRX21026417[accn])) and SRR25281510 ([https://www.ncbi.nlm.nih.gov/sra/SRX21026416\[accn\]](https://www.ncbi.nlm.nih.gov/sra/SRX21026416[accn])). All data for figures have been uploaded to Figshare (<https://doi.org/10.6084/m9.figshare.23290226>). Source data are provided with this paper.

References

1. Lang, J. M. et al. A microbial survey of the International Space Station (ISS). *PeerJ* **5**, e4029 (2017).
2. Avila-Herrera, A. et al. Crewmember microbiome may influence microbial composition of ISS habitable surfaces. *PLoS ONE* **15**, e0231838 (2020).
3. Mora, M. et al. Space Station conditions are selective but do not alter microbial characteristics relevant to human health. *Nat. Commun.* **10**, 3990 (2019).
4. Madrigal, P. et al. Machine learning algorithm to characterize antimicrobial resistance associated with the International Space Station surface microbiome. *Microbiome* **10**, 134 (2022).
5. Ramisetty, B. C. M. & Sudhakari, P. A. Bacterial ‘Grounded’ Prophages: Hotspots for Genetic Renovation and Innovation. *Front. Genet.* **10**, 65 (2019).
6. Koskella, B. & Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**, 916–931 (2014).
7. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* **11**, 1511–1520 (2017).

8. Fortier, L.-C. & Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **4**, 354–365 (2013).
9. Brüssow, H., Canchaya, C. & Hardt, W.-D. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol. Mol. Biol. Rev.* **68**, 560–602 (2004).
10. Braga, L. P. P. et al. Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios. *Microbiome* **8**, 52 (2020).
11. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
12. Bao, H. et al. Alterations in the diversity and composition of mice gut microbiota by lytic or temperate gut phage treatment. *Appl. Microbiol. Biotechnol.* **102**, 10219–10230 (2018).
13. Voorhies, A. A. et al. Study of the impact of long-duration space missions at the International Space Station on the astronaut microbiome. *Sci. Rep.* **9**, 9911 (2019).
14. Bryan, N. C. et al. Genomic and Functional Characterization of Enterococcus faecalis Isolates Recovered From the International Space Station and Their Potential for Pathogenicity. *Front. Microbiol.* **11**, 515319 (2020).
15. Singh, N. K., Wood, J. M., Karouia, F. & Venkateswaran, K. Succession and persistence of microbial communities and antimicrobial resistance genes associated with International Space Station environmental surfaces. *Microbiome* **6**, 204 (2018).
16. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. & Wishart, D. S. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res.* **39**, W347–W352 (2011).
17. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
18. Sayers, E. W. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **50**, D20–D26 (2021).
19. Akhter, S., Aziz, R. K. & Edwards, R. A. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* **40**, e126 (2012).
20. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
21. O’Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
22. Camargo, A. P. et al. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* **51**, D733–D743 (2023).
23. Turner, D., Kropinski, A. M. & Adriaenssens, E. M. A Roadmap for Genome-Based Phage Taxonomy. *Viruses* **13**, 506 (2021).
24. Olishevskaya, S., Nickzad, A. & Déziel, E. Bacillus and Paenibacillus secreted polyketides and peptides involved in controlling human and plant pathogens. *Appl. Microbiol. Biotechnol.* **103**, 1189–1215 (2019).
25. Tierney, B. T. et al. Multidrug-resistant *Acinetobacter pittii* is adapting to and exhibiting potential succession aboard the International Space Station. *Microbiome* **10**, 210 (2022).
26. Morgan, G. J., Hatfull, G. F., Casjens, S. & Hendrix, R. W. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J. Mol. Biol.* **317**, 337–359 (2002).
27. Vigil-Stenman, T., Innbergs, K., Bergman, B. & Ekman, M. High abundance and expression of transposases in bacteria from the Baltic Sea. *ISME J.* **11**, 2611–2623 (2017).
28. Waller, A. S. et al. Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J.* **8**, 1391–1402 (2014).
29. Liang, G. & Bushman, F. D. The human virome: assembly, composition and host interactions. *Nat. Rev. Microbiol.* **19**, 514–527 (2021).
30. Dávila-Ramos, S. et al. A Review on Viral Metagenomics in Extreme Environments. *Front. Microbiol.* **10**, 2403 (2019).
31. Kondo, K., Kawano, M. & Sugai, M. Distribution of Antimicrobial Resistance and Virulence Genes within the Prophage-Associated Regions in Nosocomial Pathogens. *mSphere* **6**, e00452–21 (2021).
32. Botelho, J., Cazares, A. & Schulenburg, H. The ESKAPE mobilome contributes to the spread of antimicrobial resistance and CRISPR-mediated conflict between mobile genetic elements. *Nucleic Acids Res.* **51**, 236–252 (2023).
33. Wendling, C. C., Refardt, D. & Hall, A. R. Fitness benefits to bacteria of carrying prophages and prophage-encoded antibiotic-resistance genes peak in different environments. *Evolution* **75**, 515–528 (2021).
34. Tixador, R. et al. Study of minimal inhibitory concentration of antibiotics on bacteria cultivated in vitro in space (Cytos 2 experiment). *Aviat. Space Environ. Med.* **56**, 748–751 (1985).
35. Hammond, T. G. et al. Effects of microgravity on the virulence of *Listeria monocytogenes*, *Enterococcus faecalis*, *Candida albicans*, and methicillin-resistant *Staphylococcus aureus*. *Astrobiology* **13**, 1081–1090 (2013).
36. Aunins, T. R. et al. Spaceflight Modifies *Escherichia coli* Gene Expression in Response to Antibiotic Exposure and Reveals Role of Oxidative Stress Response. *Front. Microbiol.* **9**, 310 (2018).
37. Bondy-Denomy, J. et al. Prophages mediate defense against phage infection through diverse mechanisms. *ISME J.* **10**, 2854–2866 (2016).
38. Owen, S. V. et al. Prophages encode phage-defense systems with cognate self-immunity. *Cell Host Microbe* **29**, 1620–1633.e8 (2021).
39. Mobberley, J. et al. Lysis and Sporulation in *Bacillus* Isolates from the Gulf of Mexico. *Appl. Environ. Microbiol.* **76**, 829–842 (2010).
40. Abe, K. et al. Developmentally-Regulated Excision of the SP β Prophage Reconstitutes a Gene Required for Spore Envelope Maturation in *Bacillus subtilis*. *PLOS Genet.* **10**, e1004636 (2014).
41. Kodaira, S. et al. Space Radiation Dosimetry at the Exposure Facility of the International Space Station for the Tanpopo Mission. *Astrobiology* **21**, 1473–1478 (2021).
42. Bharadwaj, A. R. et al. Draft Genome Sequences of Enterobacteriales Strains Isolated from the International Space Station. *Microbiol. Resour. Announc.* **9**, e00817–e00820 (2020).
43. Daudu, R. et al. Draft Genome Sequences of *Bacillaceae* Strains Isolated from the International Space Station. *Microbiol. Resour. Announc.* **9**, e00701–e00720 (2020).
44. Simpson, A. C. et al. Draft Genome Sequences of Various Bacterial Phyla Isolated from the International Space Station. *Microbiol. Resour. Announc.* **10**, e00214–e00221 (2021).
45. Venkateswaran, K. Microbial Observatory (ISS-MO): Indoor microbiome study of the International Space Station surfaces. <https://doi.org/10.26030/cary-q426> (2016).
46. Feldgarden, M. et al. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob. Agents Chemother.* **63**, e00483–19 (2019).
47. Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
48. Zankari, E. et al. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).
49. Gupta, S. K. et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58**, 212–220 (2014).
50. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
51. Ingle, D. J. et al. In silico serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H

- serotype combinations within and between pathogenic lineages. *Micro. Genom.* **2**, e000064 (2016).
52. Doster, E. et al. MEGARes 2.0: a database for classification of anti-microbial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res.* **48**, D561–D569 (2020).
 53. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).
 54. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
 55. Girgis, H. Z. MeShClust v3.0: high-quality clustering of DNA sequences using the mean shift algorithm and alignment-free identity scores. *BMC Genom.* **23**, 423 (2022).
 56. Berrios, D. C., Galazka, J., Grigorev, K., Gebre, S. & Costes, S. V. NASA GeneLab: interfaces for the exploration of space omics data. *Nucleic Acids Res.* **49**, D1515–D1522 (2021).
 57. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
 58. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
 59. Shaiber, A. et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol.* **21**, 292 (2020).
 60. Eren, A. M. et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* **6**, 3–6 (2021).
 61. Kiang, N. Y. et al. Discovery of Chlorophyll d: Isolation and Characterization of a Far-Red Cyanobacterium from the Original Site of Manning and Strain (1943) at Moss Beach, California. *Microorganisms* **10**, 819 (2022).
 62. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
 63. Ruiz-Perez, C. A., Conrad, R. E. & Konstantinidis, K. T. Microbe-Annotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinforma.* **22**, 11 (2021).
 64. Granger, B. E. & Pérez, F. Jupyter: Thinking and Storytelling With Code and Data. *Comput. Sci. Eng.* **23**, 7–14 (2021).
 65. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
 66. Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
 67. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
 68. team, T. pandas development. pandas-dev/pandas: Pandas. <https://doi.org/10.5281/zenodo.3509134> (2020).
 69. McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference* (eds. Walt, S. van der & Millman, J.) 56–61 <https://doi.org/10.25080/Majora-92bf1922-00a> (2010).
 70. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

Acknowledgements

The authors would like to acknowledge Dr. Kasthuri Venkateswaran and Dr. Anna Simpson at the NASA Jet Propulsion Laboratory for providing the ISS flight isolates and Dr. Braden Tierney at Weill-Cornell Medical College for his critical review of the paper. This work was supported by the NASA Space Biology Space Life Science Training Program and the NASA Human Research Program Model Translation and Space Biology Integration project.

Author contributions

J.T.B. and I.I. conceived and designed the study. J.T.B. performed the pan-genome analysis, mitomycin C and UV experiments, MinION sequencing and assembly, and curation of prophage-encoded protein functions. I.I. built and executed the pipeline for prophage identification, comparison, and classification, and the presence of antibiotic resistance markers. Both authors wrote and edited the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-42104-w>.

Correspondence and requests for materials should be addressed to Jared T. Brodrick.

Peer review information *Nature Communications* thanks Louis-Charles Fortier and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024