Adaptation to space conditions of novel bacterial species isolated from the International Space Station revealed by functional gene annotations and comparative genome analysis

Lukasz M. Szydlowski[1,5], Alper A. Bulbul[2], Anna C. Simpson[3], Deniz E. Kaya[3], Nitin K. Singh[3], Ugur O. Sezerman[2], Paweł P. Łabaj[1], Tomasz Kosciolek[1,4,5*], Kasthuri J. Venkateswaran[3*]

[1]Malopolska Centre of Biotechnology, Jagiellonian University, Gronostajowa 7A, Krakow, 30-387, Malopolska, Poland.
[2]Biostatistics and Medical Informatics Department, Acibadem University, İçerenköy, Kayışdağı Cd.32, Istanbul, 34752, Turkey.
[3]Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, California, USA.
[4]Department of Data Science and Engineering, Silesian University of Technology, Akademicka 2A, Gliwice, 44-100, Slaskie, Poland.
[5]Sano Centre for Computational Personalized Medicine, Czarnowiejska 36, Krakow, 30-054, Malopolskie, Poland.

*Corresponding author(s). E-mail(s): t.kosciolek@sanoscience.org; kjvenkat1955@gmail.com;

Abstract

Background: The extreme environment of the International Space Station (ISS) puts selective pressure on microorganisms unintentionally introduced during its 20+ years of service as a low-orbit science platform and human habitat. Such pressure leads to the development of new features not found in the Earth-bound relatives, which enable them to adapt to unfavorable conditions.

Results: In this study, we generated the functional annotation of the genomes of five newly identified species of Gram-positive bacteria, four of which are

non-spore-forming and one spore-forming, all isolated from the ISS. Using a deep-learning based tool - deepFRI - we were able to functionally annotate close to 100% of protein-coding genes in all studied species, overcoming other annotation tools. Our comparative genomic analysis highlights common characteristics across all five species and specific genetic traits that appear unique to these ISS microorganisms. Proteome analysis mirrored these genomic patterns, revealing similar traits. The collective annotations suggest adaptations to life in space, including the management of hypoosmotic stress related to microgravity via mechanosensitive channel proteins, increased DNA repair activity to counteract heightened radiation exposure, and the presence of mobile genetic elements enhancing metabolism. In addition, our findings suggest the evolution of certain genetic traits indicative of potential pathogenic capabilities, such as small molecule and peptide synthesis and ATP-dependent transporters. These traits, exclusive to the ISS microorganisms, further substantiate previous reports explaining why microbes exposed to space conditions demonstrate enhanced antibiotic resistance and pathogenicity.

Conclusion: Our findings indicate that the microorganisms isolated from ISS we studied have adapted to life in space. Evidence such as mechanosensitive channel proteins, increased DNA repair activity, as well as metallopeptidases and novel S-layer oxidoreductases suggest a convergent adaptation among these diverse microorganisms, potentially complementing one another within the context of the microbiome. The common genes that facilitate adaptation to the ISS environment may enable bioproduction of essential biomolecules need during future space missions, or serve as potential drug targets, if these microorganisms pose health risks.

Keywords: functional annotation, International Space Station, extremophiles

1 Introduction

The environment of outer space, characterized by intense radiation and high vacuum, presents harsh conditions for any form of life. Despite this, microorganisms have demonstrated a remarkable adaptability to these extreme space environments. Organisms known as extremophiles, such as bacteria (e.g. *Bacillus*, *Deinococcus* species) [1, 2], fungi [3], tardigrades [4], bdelloid rotifers [5], and many others, are well-documented for their ability to survive in space conditions. As such, it is virtually impossible to prevent the transfer of microbial life to exploratory space vessels or planets[6, 7]. In addition to microorganisms potentially hitching a ride with robotic components, the microbiomes of humans also travel to space. This not only includes the microbiomes of space crew members, but also free-living microorganisms associated with cargo, science instruments, and test animals, including potential pathogens which should be considered for their potential to adapt to space conditions [8–13].

The exposure of microbes to outer space, or simulated versions of these conditions, has been extensively studied to test survival capabilities and assess the panspermia hypothesis. Over several decades, various microbial samples have been exposed to

space conditions via balloons, rockets, and spacecraft as part of the pioneering experiments in astrobiology (reviewed in [14]). For example, after 18 months of exposure, *Bacillus pumilus* SAFR-032, a bacterium isolated from the Jet Propulsion Laboratory's Spacecraft Assembly Facility (JPL-SAF), exhibited 10-40% survivability when UV was masked [1]. However, when these SAFR-032 spores were exposed to UV light (above 110 nm) and vacuum in space for 18 months, except for a few spores, there was a roughly 7-log reduction in viability. Similarly, dried *Deinococcus* cell pellets, particularly those 500 μm in thickness, remained viable after three years of space exposure due to a shadowing effect. Moreover, *D. radiodurans* after being subjected to 1-year exposure to outer space during Tanpopo mission[15], showed a collection of molecular alterations e.g. DNA damage and oxidative stress response, accumulation of metal transporters etc. In a recent study, a survival of *D. radiodurans* exposed to near space conditions was correlated with a growth condition prior to the exposure[16]. A comparative proteomics analysis of *B. pumilus* SAFR-032 revealed that proteins conferring resistance traits, such as superoxide dismutase, were present in higher concentrations the vegetative cells derived from spores exposed to space compared to ground controls. Furthermore, the first-generation spores of *B. pumilus* SAFR-032 resulting from space-exposed samples demonstrated increased UVC resistance (around 4,000 J/m^2) compared to their ground control counterparts[1].

The International Space Station (ISS) is a closed environment that houses a diverse range of microorganisms, including potential pathogens [17]. As we prepare for long-term space explorations, understanding the molecular mechanisms that allow microorganisms to survive and adapt during spaceflight is becoming increasingly important [10, 15]. Monitoring the microbial population onboard the ISS is crucial for safeguarding astronaut health and preventing contamination of both the spacecraft and its equipment [7]. Whole-genome sequencing (WGS) and phenotypic characterization of any novel microbial species discovered on the ISS can help identify potential pathogens and comprehend their impact on the closed habitat and crew health [7]. However, to accurately identify potential microbial threats, it is essential to maintain a current and comprehensive database of microbial genomes, detailing both their species identities and phenotypic traits. This database, continuously updated through WGS of pure cultures, is crucial for gathering genetic data and assisting in the identification of microbial threats via shotgun metagenomic sequencing. Moreover, the active initiatives to characterize the cultivable microbiome of previously not sequenced space habitats contribute to decreasing the unidentified microbial 'dark matter' in metagenomic sequencing outcomes.

Research conducted in space or using ground-based simulators has shown that microgravity conditions can alter various biological processes, including growth, morphology, gene expression, virulence, drug resistance, biofilm formation, and secondary metabolism [8–10, 12, 18]. Studies focusing on the microbiome have reported an increase in antibiotic resistance markers in astronauts' saliva [19], although these changes were temporary and reversed upon returning to Earth. Other research has indicated that microbial isolates from space environments are differentiating themselves from their Earth-bound counterparts, showing signs of adaptation to their new

environment [17, 20–23]. Interestingly, under spaceflight conditions, one species, *Bacillus safensis* JPL-MERTA-8-2, exhibited a 60% improvement in growth compared to its Earth-based growth [24].

This study focused on the genomic peculiarities of novel species isolated from the ISS and compared WGS of the closest relatives found in Earth environments. A thorough analysis was conducted on genes associated with DNA repair, radiation resistance, microgravity adaptation, stress responses, and metabolic changes (Fig. 1). These genes were compared to their terrestrial counterparts, and potential variations induced by space conditions were observed. The protein-coding sequences from each species pair were compared using a set of similarity metrics, unearthing potentially novel genes and clusters that might have been evolved in space. Furthermore, mutations within these sequences were rigorously analyzed, and their protein structures were predicted using *in silico* models. This approach illuminates the genetic adaptations undertaken in stringent conditions of the space environment. In essence, by comparing space and Earth genomes and identifying mutation patterns, a clearer insight into genetic evolution influenced by space conditions has been obtained.

2 Results

2.1 Comparison of different annotations and between the isolates and within pangenome context

Five novel species of Gram-positive bacteria that were isolated from the ISS were analyzed. Their generic features and the results of other molecular analyses are presented in Table 1 and Supplementary Figure S1. These bacteria were obtained from various flights, locations, and time periods, and are associated with different phylogenetic groups. The strain F6_3S_P_1C, which belongs to the *Paenibacillus* genus, has been identified as a spore-former, while the other four species were identified as non-spore-forming Actinobacteria. Through ANI and AAI analysis, we established closest Earth relatives (Table 1 and Supplementary Table S1). Additionally, we performed synteny analysis [25] using all top ANI hits for each of the five organisms (Supplementary Table S1), but yielded no results (data not shown), thus indicating all ISS isolates are distinct species.

2.1.1 Annotations common for space and Earth species (closest relative, see Supplementary Figure S1 for more comparisons)

When the annotated genomes were compared with several methods (PGAP, eggNOG), DeepFRI analysis showed highest annotation coverage and provided a confidence scale (based on the score value), to filter medium quality (MQ) and high-quality (HQ) annotations and the results are shown in Figure 2 (see Methods for details). In total, the genomes of the novel species encompassed 1290 and 1184 shared MQ and HQ annotations (1778 all annotations, Fig. 2ab). The majority of shared annotations were related to metabolism, followed by cellular processing and signaling and information storage and processing (Fig. 2c; all common annotations (GO terms) are detailed in Supplementary Table S12 in the Supplementary material).

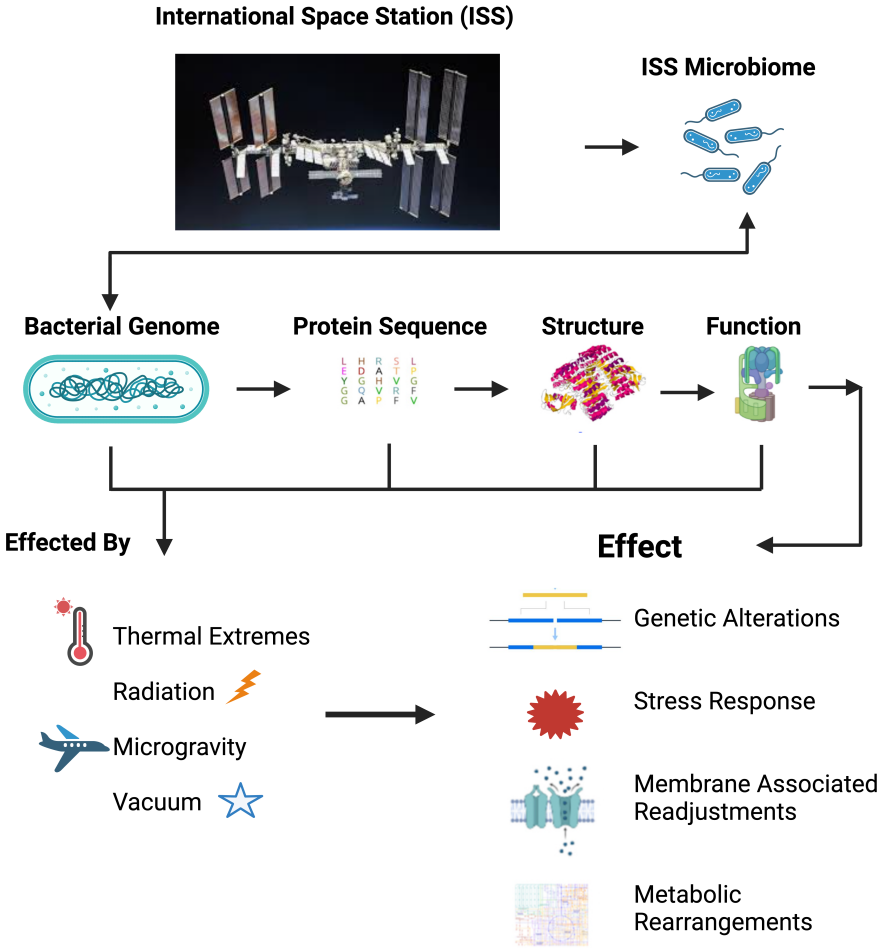
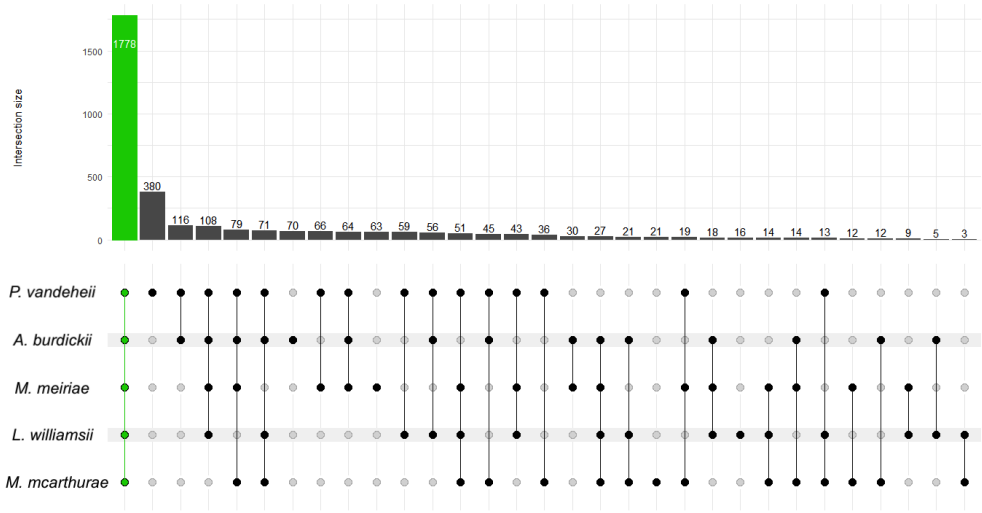


Fig. 1 Overview of the project.

2.1.2 Annotations unique to ISS bacterial species

All curated National Center for Biotechnology Information (NCBI) genomes were used to find the closest terrestrial relatives for the novel ISS bacterial species and the genome that exhibited the highest average nucleotide identity (ANI) value was further compared with genomes of ISS strains to characterize unique annotation. When compared with their closest terrestrial relatives all ISS species have different numbers of unique annotations (Fig. 3a), which corresponds to their ANI (the higher the ANI, the

All	MQ	HQ	
1778	1290	1184	**Total**
852	712	678	**BP**
121	54	47	**CC**
352	167	139	**EC**
452	357	320	**MF**



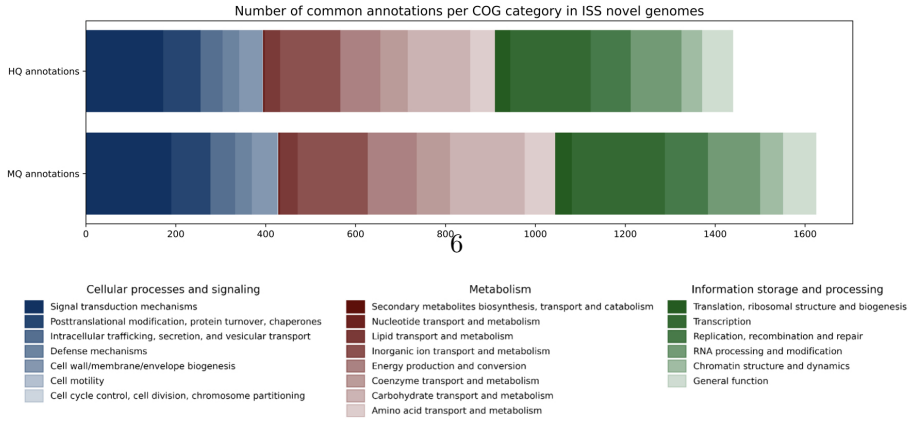


Fig. 2 Overview of the annotations common in all ISS species. A) Number of unique, common annotations found in all species (BP - biological process, CC - cell component, EC - enzyme commission, MF - molecular function, MQ - medium quality, HQ - high quality) B) Upset plot showing common annotations (derived from DeepFRI) with different groups of interests C) Clustering all isolates' common annotations in the form of clusters of orthologous genes (COG) categories. (See Methods for more details.

Table 1 General features of the isolates used in this study. **ANI** - average nucleotide identity.

Strain	Location	Scientific name	Genome ID
IF8SW-P5	ISS-Crew quarters	*Microbacterium mcarthurae*	JAROCE000000000
F6_8S_P_2B	ISS-Crew quarters	*Microbacterium meiriae*	CP133984
F6_3S_P_1C	ISS-ARED A	*Paenibacillus vandeheii*	JAROCD000000000
IIF3SC_B10	ISS-ARED	*Arthrobacter burdickii*	JAROCG000000000
F6_8S_P_1B	ISS-Crew quarters	*Leifsonia williamsii*	JAROCF000000000

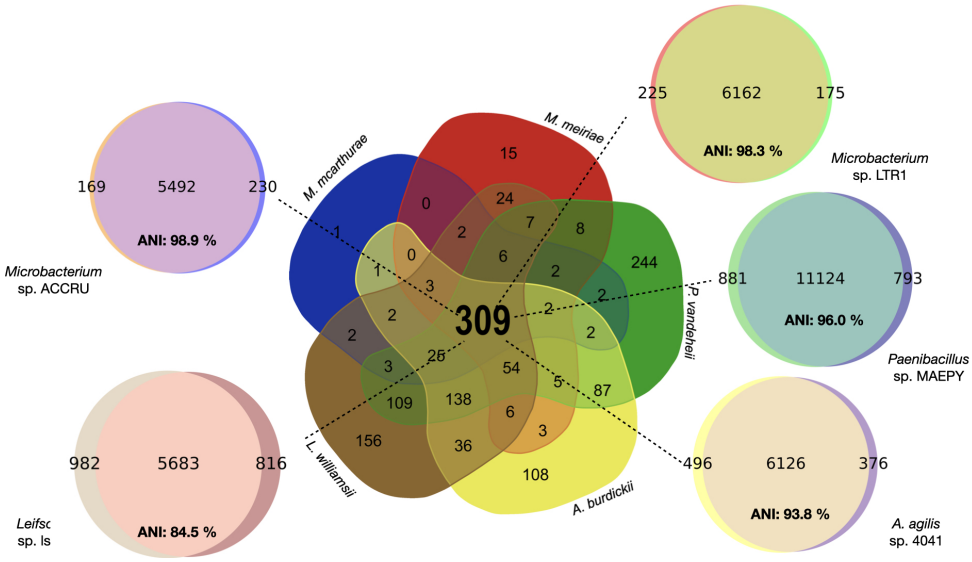
Spore-forming?	Number of genes	PGAP coverage	eggNOG coverage
No	2977	81.9%	81.9%
No	3310	80.9%	81.2%
Yes	6223	81.3%	82.4%
No	3557	80.3%	82.3%
No	3649	79.5%	78.6%

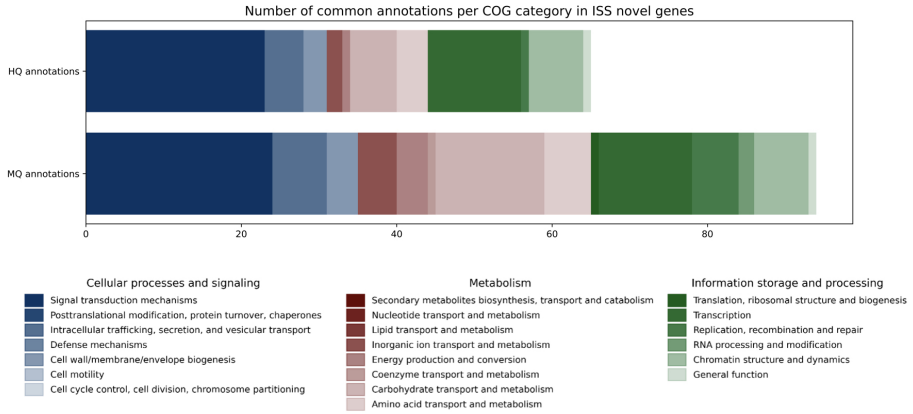
DeepFRI coverage	Closest Earth strain	Assembly ID	ANI (AAI)
100%	*Microbacterium* sp. ACCRU	GCF_022346945.1	98.9% (>90%)
99.9%	*Microbacterium* sp. LTR1	GCF_026805445.1	98.3% (>90%)
100%	*Paenibacillus* sp. MAEPY	GCA_000499205.1	96.0% (>90%)
99.9%	*Arthrobacter agilis* 4041	GCF_002157505.1	93.8% (>90%)
100%	*Leifsonia* sp. ku-ls	GCA_003367665.1	84.5% (75.91%)

fewer unique annotations). Additionally, two-dimensional Conserved Domain (CD)-hit clustering was performed between each isolate and its closest relative, using a 40% sequence similarity threshold. Thus, for each isolate a subgroup of unique genes (with no counterparts in their closest relative) was found (Supplementary Tables S2-S11). A comparison between these genes revealed the presence of 74 and 49 MQ and HQ common annotations, respectively (309 total annotations, Fig. 3b, Supplementary Table S13).

When aligning the novel genes in the genome context, we found the presence of potential novel biosynthetic gene clusters (BGCs) in all species. These BGCs encode secondary metabolites such as terpenes, carotenes, betalactones, lanthipeptides, lassopeptides, non-ribosomal metallophores, furans, non-alpha poly-amino acids (NAPAA), type 3 polyketide synthase (T3PKS), ribosomally synthesized and post-translationally modified peptides (RiPP), proteusins (Supplementary Fig. S1), Our findings of BGCs are complimentary to the previously reported BGCs in *A. burdickii*, *L. williamsii* and *P. vandeheii* [28]. We also expand the BGCs classification of the studied ISS species into the "novel" category, which are not found in the genomes of their closest terrestrial relatives. In this study, we report small molecule biosynthetic pathways, metal ion transporters and other transmembrane transporters. Furthermore, transcription factors associated with environmental stress have been found (Supplementary Table S14).

Given the difference in ANI for each ISS-species:closest-relative pair, a different number of novel gens was obtained in each case, proportional to the phylogenetic distance (the lower the ANI, the more novel genes were found). We therefore focused on the two *Microbacteria* species, as their ANI to the closest relative was higher than


225
6162
175
ANI: 98.3 %
Microbacterium
sp. LTR1
169
5492
230
ANI: 98.9 %
Microbacterium
sp. ACCRU
M. mcarthurae
M. meiriae
P. vandeheii
L. williamsii
A. burdickii
309
881
11124
793
ANI: 96.0 %
Paenibacillus
sp. MAEPY
982
5683
816
ANI: 84.5 %
496
6126
376
ANI: 93.8 %
A. agilis
sp. 4041


Number of common annotations per COG category in ISS novel genes
HQ annotations
MQ annotations
0
20
40
60
80
Cellular processes and signaling
Signal transduction mechanisms
Posttranslational modification, protein turnover, chaperones
Intracellular trafficking, secretion, and vesicular transport
Defense mechanisms
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Metabolism
Secondary metabolites biosynthesis, transport and catabolism
Nucleotide transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism
Energy production and conversion
Coenzyme transport and metabolism
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Information storage and processing
Translation, ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
RNA processing and modification
Chromatin structure and dynamics
General function

Mma_002827: magenta (pLDDT: 83.7, pTM: 0.708)
Mme_000934: orange (pLDDT: 87.6, pTM: 0.74)
TM aligned: 0.81

98%. Using Muscle, we aligned sequences of novel genes from each species together. We found an example of novel genes (Mma_002827 and Mme_000934) being present in both species with high (52.8%) sequence similarity. Given that initial annotation ("heavy metal associated domain", "response to copper ion", see Supplementary Tables S2 and S6), was sequence-based (see Methods for details), another approach was assayed, with structure prediction, followed by a functional annotation. Superpositions of predicted structures (both with pTM>0.7 and pLDDT>80) are almost identical (Fig. 3c). The structure-based deepFRI annotations reveal that these proteins have oxidoreductase activity (acting on CH-NH2 group of donors) and that are located in the extracellular layer of the cell (Supplementary Table S15). When compared to the proteome of *D. radiodurans*, weak structural homology was found to extracellular protein WP_027479552.1 (TM-score 0.34, see Supplementary Fig. S2).

Mobile genetic elements (MGEs) such as transposases were also found in all species except *L. williamsii*. There is also evidence of plasmid in some of the species (*M. mcarthurae*, *M. meiriae*), due to the presence of relaxosome encoding genes, such as MobC/L family relaxase (Figure 5, Supplementary Table S16).

2.2 Comparison of proteome-level mutations in the ISS species and within pangenome context

Several groups of proteins, important for adaptation to space environmental conditions were analyzed between ISS species and their closest Earth relatives. In metallopeptidases, the mutation rate of hydrophobic amino acids and glycine are higher than 1 in this protein group, suggesting an enrichment of these amino acids in the metallopeptidase proteins of ISS bacteria. These mutations might enhance the stability of these proteins under space environmental conditions (Figure 4A). Glycine enrichment is a measure of how often the amino acid glycine appears in a protein sequence compared to what would be expected by chance (see Methods for more details). Significant glycine enrichment in a protein can suggest evolutionary pressures and adaptive responses to make proteins more flexible via mutation to glycine[29]. The metallopeptidase activity stands out as it exhibits consistent glycine enrichment across all strains, suggesting it may play a critical role in universal adaptive mechanisms (Fig. 4B). Other molecular functions, such as "ATP-dependent peptidase activity" and "2-C-methyl-D-erythritol-4-phosphate cytidylyltransferase activity", show high glycine enrichment, but only in both *Microbaterium* species, indicating potential taxonomy-based adaptations (Fig. 4B).

Certain molecular functions, such as "5 formyltetrahydrofolate cyclo ligase activity", "ATP dependent peptidase activity", and "ATP phosphoribosyltransferase activity", display a high degree of hydrophobic amino acid enrichment, but only in *M. meiriae* and *L. williamsii*. This could suggest convergent adaptations involving changes in hydrophobicity of aforementioned proteins. Interestingly, "metallopeptidase activity" shows hydrophobic amino acid enrichment across all ISS species, albeit at a lower level than glycine enrichment. This could suggest a universal, but less dominant, adaptive strategy involving changes in protein hydrophobicity. These patterns underscore the complexity of potential adaptive strategies, highlighting the interplay between different types of amino acid changes in response to ISS environments (Figure

4B). Metallopeptidases play role in celluar stress response [15], biofilm formation [30], as well as virulence [31] and hence are important factors in microbial adaptations to space conditions. Therefore, a comparative analysis of sequences and structures of selected metallopeptidases from *M. meiriae* and its closest relative will be further discussed in the section below.

2.3 Case studies of adaptations

The functional annotations were further investigated to search for genes that would allow adaptation to life in space. The genes of interest were summarized in Figure 5.

2.3.1 Energy metabolism

As illustrated in Fig.1, metabolic rearrangements are an essential part of adaptations to life in space, which was also reported earlier [8–10, 12, 13, 18]. Common annotations show the presence of genes encoding proteins related to carbohydrate, cofactors and coenzymes metabolism (Fig 2c, Fig. 3b). Apart from genomic annotation (Supplementary Table S12), metabolic tests conducted on the ISS bacterial species reveal their ability to grow on a broad range of carbon sources, including cyclic compounds, oligo- and polysaccharides (Supplementary Fig. S3a-b). All species are capable of growing on 27 substrates being the sole carbon source, including carbohydrates (pectin, dextrin, mannose, fucose, gentibiose, lactose, melibiose, glucose, galactose, fructose as well as their branched and glycosylated derivatives), fatty acids (acetic acid, L-malic acid, sodium lactate and butyrate), amino acids (L-glutamic acid), alcohols (glycerol) and antibiotics (vancomycin and aztreonam). They can also grow at 1% NaCl concentration and pH6. Apart from the spore-forming *P. vandeheii*, the other bacteria can also utilize tween 40, gelatin, D-malic acid, L-lactic acid, propionic acid, D/L-serine, L-alanine, sorbitol, lincomycin, minocycline, nalidixic acid, rifamycin, as well as grow in the presence of lithium chloride. On the other hand, only *P. vandeheii* is able to grow in the presence of sodium bromate. Additionally to the species described by Simpson and colleagues [28], two species of *Microbacterium* were characterized for the first time. These species are able to utilize 72 carbon sources altogether (80 and 73 substrates for *M. mcarthurae* and *M. meiriae*, respectively), grow at 8% NaCl concentration, as well as in the presence of lithium chloride and potassium tellurite. Unlike other species, either *M. mcarthurae* or *M. meiriae* can utilize α-keto-glutaric acid and citric acid.

2.3.2 Mechanosensitive ion channels (Msc) and osmoprotectants

Except for *A. burdickii*, each of the studied organisms has at least one copy of *msc*S gene annotated. Msc proteins are known to play a crucial role in adapting cells to the osmotic pressure [32]. Overall, two variants of Msc proteins exist: large conductance, (3.6 nS) MscL and small conductance (1 nS) MscS. Large conductance channels are generally conserved and present in most bacteria as well as eukaryotes, although they are not found in deep sea species, where no pressure changes occur. Conversely, MscS are diverse structurally and functionally and occur in those organisms, which exhibit hypo-osmotic stress periodically (e.g. desiccation). MscS proteins have a common core architecture but show different gating mechanisms and conductive properties. Because

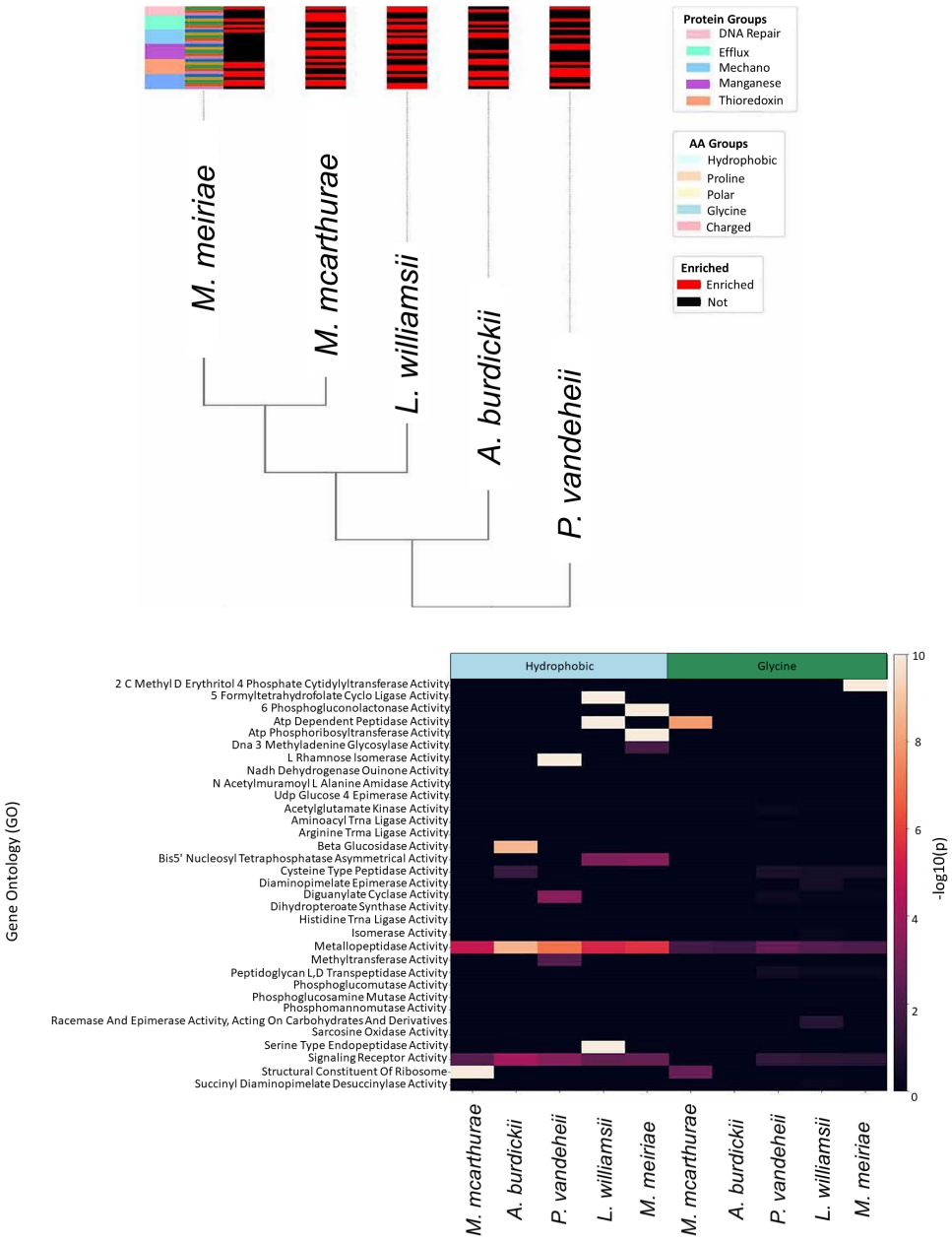


Fig. 4 Comparison of mutation enrichment in ISS species across different protein families. A) Clustering dendrogram representing enrichment in protein mutations (wrt closest relative) related to the protein group and type of mutation B) Mutation enrichment of selected proteins (wrt Uniprot database) with their significance scores.

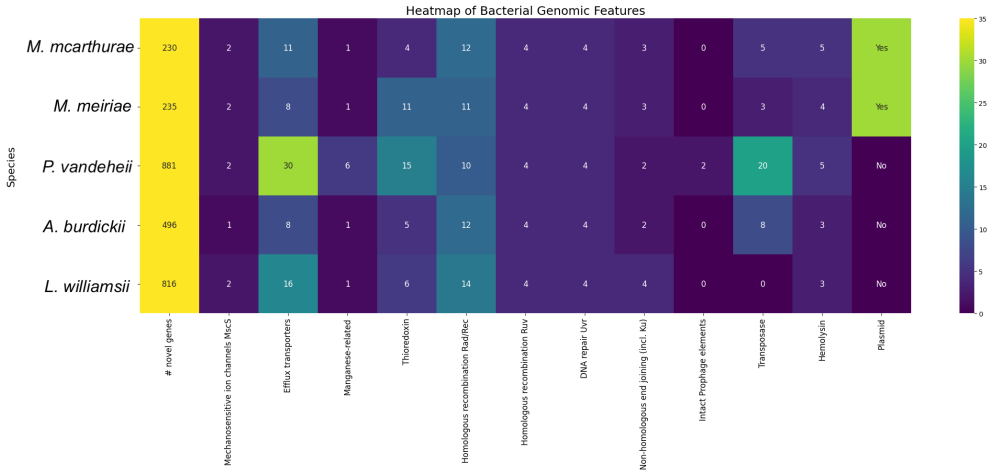


Fig. 5 Heatmap representing abundance of annotations of interest in the studied species.

low gravity conditions can be viewed as decrease in osmotic pressure, leading to the hypo-osmotic stress, their presence is therefore crucial for survival in ISS. All terrestrial relatives have a copy of mscS gene with over 90% sequence similarity to their homologous gene from the ISS species (Supplementary Fig. S4), except for *L. williamsii*, where there is an additional copy of the mscS with no homology in the genome of *Leifsonia* sp. ku-ls. Nevertheless, the presence of an additional copy of mscS gene could be due to the larger phylogenetic distance, as indicated by ANI (Fig. 3A). As for the osmoprotectants, there is no commonly shared metabolite, as genes related to betaine, choline, ectoine and trehalose synthesis are not present in all five species.

2.3.3 Redox stress

Changes in osmotic pressure, as well as exposure to ionizing radiation may lead to the formation of reactive oxygen species (ROS) that have detrimental effects on cells. From the studies on *Deinococcus* it was found that it has elevated intracellular concentrations of Mn as well as Fe[33]. All studied ISS strains show an elevated number of metalloproteins compared to their closest terrestrial counterparts (Supplementary Table S12). In the case of *M. mcarthurae* and *M. meiriae* there are 27 and 25 unique metalloproteins, respectively (Supplementary Table S14), which include metallopeptidases, thioredoxins, metalloregulated transcription factors and heavy metal associated P-type ATPases. Metallopeptidases were of our special interest, as they showed certain adaptive mutations in all studied species (as described in previous section). Therefore, we decided to further analyze the metallopeptidase pair from *M. meiriae* and its terrestrial relative *Microbacterium* sp. LTR1. We found aminopeptidases in two of these species (Mme_001931 and WP_268776344.1 from *M. meiriae* and its terrestrial relative *Microbacterium* sp. LTR1, respectively), which share only 44% protein sequence similarity (58.4% nucleotide similarity, Supplementary Fig. S5). When structural models were calculated for these proteins (both with pTM-score $\gg 0.9$), their superposition

is almost identical (Fig. 5A), despite the low sequence similarity. When analyzing enrichment of the amino acids in Mme_001931 with WP_268776344.1 as a baseline, we observed enrichment of arginine and methionine, which were over 1.5 times more abundant compared to terrestrial protein. Opposite case was observed for lysine, which was over 2 times less abundant in ISS organism (Fig. 5B). When superimposed to deinococcal aminopeptidase (WP_0516188611), Mme_001931 exhibited similarly high TM-score (0.83), although the latter's size is 1.81 times bigger than WP_0516188611 (Supplementary Fig. S6). In terms of amino acid enrichment, WP_051618861 also showed an elevated amount of arginines (8.3% wrt 5.5%) with subsequent decrease of lysines (1.1% wrt 2.6%) and increase in methionine with decrease of negatively charged amino acids, but also increase in glycines, histidines and prolines. Nonetheless, there is almost two times the size difference and further phylogenetic distance than in the case of *M. meiriae*.

2.3.4 DNA protection, repair and recombination

Another feature associated with ISS is the elevated rate of DNA-damaging due to radiation. *Deinococcus radiodurans* was extensively tested for its ability to protect and repair damaged DNA, which was due to its homologous recombination (HR) kit. All ISS species encode 10-14 genes related to HR, compared to 10 genes identified in *D. radiodurans* (Fig. 5). Similarly, all of the genomes share the set of *rec*, *ruv*, *rad* and *uvr* genes. In terms of non-homologous end-joining (NHEJ), all ISS strains share at least one gene. When compared with the Rec protein family from *D. radiodurans* and *B. pumilus* (Supplementary Fig. S7), clusters reflect phylogenetic distances. Moreover, all ISS strains share the LexA repressor (EC 3.4.21.88), which is part of the SOS response.

2.3.5 Phage elements

Phage elements were found in all genomes of ISS strains, with at least two phage-related genes found in each strain. Holin is a short protein containing few transmembrane regions. Although holin is usually accompanied by endolysin, there is no evidence of phage-related endolysin in any of the genome of ISS strains, suggesting that the phage elements could have been domesticated or removed. When tested with PHASTER software, prophage elements were detected only in the genome of *P. vandeheii*, as well as in its closest Earth relative (Supplementary Table S16).

2.3.6 Potential pathogenicity and antibiotic resistance

Common annotations (Supplementary Table S2) indicate the presence of genes with virulence potential, such as metallophores and hemolysins (including the group III family, which forms oligomeric complexes), which are present in all ISS strains. However, ISS strains often exhibit enhanced expression of these virulence genes due to the unique selective pressures of the space environment, such as microgravity and increased radiation. In contrast, Earth analogous strains show a greater diversity in these genes, driven by varied environmental challenges and human influences like antibiotic use and pollution ([34]. Genomes of all ISS bacterial species contain genes involved in biofilm formation, potentially posing a pathogenic risk to humans under

Mme_001931: magenta (pLDDT: 95.0 pTM: 0.93)
WP_268776344.1: cyan (pLDDT: 96.8, pTM: 0.95)
pTM aligned: 0.95

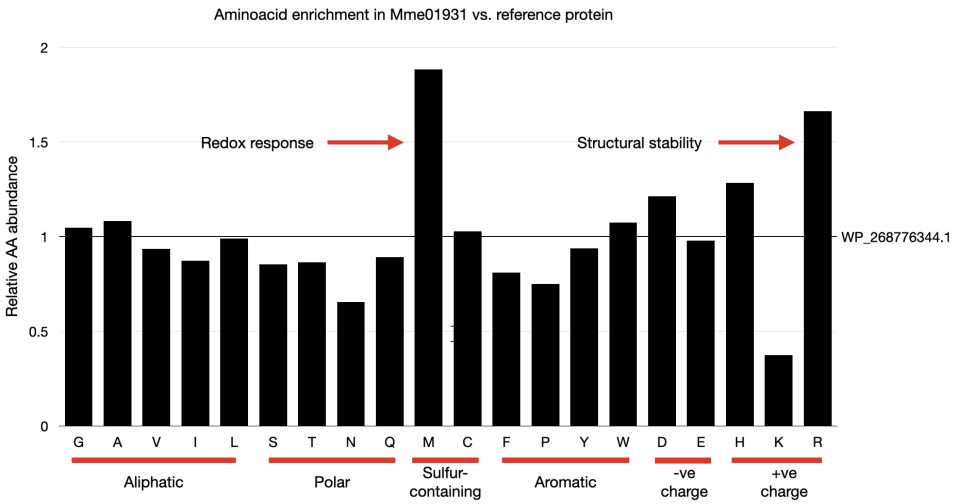


Fig. 6 Structure alignment of selected metallopeptidases from *Microbacterium* species, with amino acid enrichment in space isolate. amino acids were grouped by their side chains properties. **pLDDT** - predicted local-distance difference test, **pTM** - predicted template modelling, **TM** - template modelling, see [26, 27]

specific conditions, even though these genera and their closest terrestrial relatives are not typically known to be infectious agents. Comparatively, Earth counterparts also form biofilms, but the selective pressures driving this ability differ. On Earth, biofilm formation is often a response to environmental challenges such as antibiotic exposure and surface colonization in diverse habitats. In contrast, ISS bacteria have shown adaptations favoring biofilm formation and surface interactions due to the unique conditions of the space environment, such as microgravity and limited nutrient availability [35]. Additionally, as Mora *et al.* highlighted in their work [36], the genomic and physiological features selected by ISS conditions, though not directly pertinent to human health, exhibit considerable adaptations for biofilm formation and surface interactions. Their research underscores the importance of minimizing local moisture to prevent the formation of potentially harmful biofilms. Metabolic assays indicate that all ISS bacterial species have the ability to degrade at least two antimicrobial compounds (Supplementary Fig. S3ab), showcasing their resilience in a controlled, high-stress environment. These bacteria also harbor various P-type translocating transportases and metallopeptidases within their pool of novel genes (Supplementary Table S12), suggesting advanced mechanisms for survival and resistance. Similarly, Earth analogs have evolved to degrade antimicrobial compounds, especially in environments heavily influenced by human activity, such as hospitals and agricultural sites [31].

3 Discussion

The thorough examination of genomes of five novel ISS bacterial species, when compared with their closest terrestrial counterparts, unveiled clear adaptations to the rigors of ISS space environment. Notably, significant shifts in cell physiology and metabolism were seen. The extent to which genes lacked homologs in their terrestrial relatives was found to be associated with their phylogenetic distance. A specific set of genes was discovered to be present in all isolates but absent in their closest terrestrial relatives. This observation is consistent with earlier findings on the ISS-specific evolutionary patterns [37, 38]. Mannose is a known substrate of glycosylation in prokaryotes[39] and thus may play a role in cell wall modifications, cell adhesion and biofilm formation. Moreover, genes related to beta-mannosidase and mannose activities were up-regulated in high salinity media in the brine shrimp *Artemia salina*[40], suggesting that mannose and its derivatives may be accumulated as organic osmolytes. Clusters associated with small molecule and peptide biosynthesis in fungi were found in abundance, reaffirming earlier studies into molecular responses to microgravity [11]. The pool of novel genes unique to the ISS species (Supplementary Table S13) showed an abundance of genes related to peptide and small molecule synthesis, transmembrane secretion/transport, and mechanosensitive ion channels. This suggests they might have a defense mechanism against hypoosmotic shocks. Furthermore, novel surface oxidoreductases have been identified in *M. mcarthurae* and *M. meiriae* which are not present in their terrestrial analogs. In light of these findings, a deeper investigation using comparative transcriptomics, proteomics, and metabolomics is recommended. To perform meta-omic analysis while cultivating these microorganisms on ISS would be very challenging. Nevertheless, such studies were conducted before, e.g. Tanpopo mission [15],

although this study focused on meta-omic analyses post-exposure. Conducting *in vivo* experiments on microogranisms could be performed via simulating some of the conditions experienced in ISS, e.g. increased radiation, vacuum or microgravity. Ground facilities to recreate these conditions exist and has been utilized in such experiments before ([41–43], also reviewed in [10]). It would therefore be possible to compare gene expression profile under these conditions, protein activity (e.g. extracellular metallopeptidases) or even assay secretion of certain metabolites (e.g. chelators, antibiotics) on indicative media.

Efficient DNA repair and its impact on survival in the extreme conditions has long been studied in many organisms [41, 44, 45]. LexA repressor-mediated response is known to be crucial to DNA protection and repair process [46]. Studies on extremophilic bdelloid rotifers highlighted the linkage of efficient DNA repair mechanisms with the presence of high rates of HGT [47, 48]. The detection of MGEs underscores a significant genetic factor hinting at the selective pressures faced by species isolated from the ISS. Interestingly, the quantity of annotated transposases seems to be in proportion to genome size, with *P. vandeheii* genome showcasing the most substantial presence of MGEs. On the other hand, transposable elements were not detected in all ISS species, thus suggesting that TE-mediated genome rearrangements are not a common route for genomic adaptations to space environment. Given that microgravity and hypoosmotic shock may induce dessication stress, the adaptation mechanisms may be similar to dessication-tolerant organisms, e.g. bdelloid rotifers, where TEs are highly diverged [49] and containing unique ORFs, thus indicating a domestication. Hence, it is possible that in the species described in this study the TEs are significantly diverged. Extremophilic lifestyle provides selective pressure on maintaining genome integrity. Therefore, TEs disruptive activity is expected to be quenched, at least in the non-spore forming species. On the other hand, *P. vandeheii* may have a different adaptation system, being a spore-former such as *D. radiodurans*, which is known to contain TEs that are a part of stress-response[50]. Notably, sequences tied to plasmid-specific genes, such as *mob*C and relaxase, were discerned in both *Microbacterium* species, a feature absent in their terrestrial analogs. This emphasizes the potential unique genetic adaptations of these space-borne bacteria.

Building upon established research[33, 47], our analysis of proteomic alterations points to a significant enrichment across various protein classes. A particularly noteworthy observation is the global glycine enrichment, a phenomenon that aligns with findings from earlier studies on *Acinetobacter* sp., a desiccation-tolerant organism[17, 37, 51]. Historically, these organisms have not only displayed enhanced hydrophilicity but have also shown a fortified resistance to ionizing radiation, a stress factor intrinsically tied to redox reactions. The emphasis on these genetic mechanisms holds paramount importance, especially in the context of space-related stress factors. As organisms venture into space, they confront an array of unique and potent stressors, including elevated radiation, microgravity, and vacuum conditions. The genetic adaptability and resilience exhibited by certain organisms, as demonstrated by their proteomic profiles and specific enrichments, provide invaluable insights. These genetic mechanisms might well be the key to understanding how life can persist, and perhaps thrive, in the harsh environs of space. Such knowledge not only expands our

understanding of extremophile biology but also informs potential biotechnological applications and space colonization strategies in the future.

When looking at the singular case of metallopeptidase (aminopeptidase N) from *M. meiriae* and its closely-related Earth counterpart *Microbacterium* sp. LTR1, we observed almost perfect structure alignment, despite 43% sequence similarity. The changes in amino acid composition reflect the adaptations to the lifestyle of *M. meiriae*, with enrichment of arginine(s) and methionine(s), with the subsequent reduction of lysine and asparagine. Arginine is known to provide more stability to the structure than lysine [52–54]. Methionine in proteins plays an important role in cellular antioxidant activity, protein structure stabilization, and sequence-independent recognition of protein surfaces. It can also act as a regulatory switch through reversible oxidation and reduction (reviewed in [55]). This could hence explain its presence in the catalytic site (HEXXH motif) of aminopeptidase N.

Drawing from foundational work such as that by Daly et al.[33], metalloproteins stand out as a consistently enriched protein class in strains adapted to space conditions. This enrichment pattern mirrors the remarkable adaptations seen in organisms like *D. radiodurans*, which has gained attention for its extraordinary radiation resistance[41]. The prominence of these metal-dependent enzymes in space-adapted strains is not merely a coincidence. These enzymes play a pivotal role in efficiently managing the redox stress that such microorganisms face in space. Their enrichment suggests an enhanced capability for early stress detection, with these enzymes acting as both environmental sensors and transcription factors. This is a vital adaptation, as space presents a plethora of unique stressors, with redox imbalances being a major concern. The ability of certain strains to bolster their defense mechanisms, as evidenced by the enrichment of specific proteins, reveals the intricate genetic adaptations that enable survival in such hostile settings. These findings not only deepen our grasp of astrobiology but also have profound implications for potential biotechnological applications, hinting at ways to engineer organisms or systems for better resilience in space or other extreme conditions. The extreme ISS environment explains the need for metals scavenging to balance the redox stress e.g. in the form of metalloproteins. The study of metallopeptidases is essential in the context of microbial adaptation to space conditions due to their significant roles in protein degradation, regulation of biological functions, and contribution to microbial virulence and biofilm formation[30]. This research is vital for ensuring the health and safety of astronauts, maintaining the integrity of space missions, and exploring new biotechnological applications. On the other hand, the prevalence of genes such as hemolysins, ferrochelatases, as well as protein glycosylation (mannose), antibiotic resistance and biofilm pathways indicates the risk of potential pathogenicity of the ISS-inhabiting microorganisms. These genes also appear in the genomes of their closest terrestrial relatives, but more studies needed to validate whether microgravity enhance the antibiotic resistance and biofilm production[56]. Conversely, these observations highlight the importance of metal uptake systems, which could lead to the development of antimicrobial agents utilized in the space habitat. Hemolysins are known to be repressed upon high Fe concentrations, as they are tools to scavenge it from the environment. On the other hand,

LexA-mediated stress response is known to initiate bacteriocin synthesis [57]. Thiopeptide antibiotic called TP-1161 has been detected in the ISS organism *A. burdickii*[28] and the presence of BGCs encoding potentially novel antimicrobial molecules in the genomes of remaining four ISS species has been confirmed. Therefore, these organisms may serve as a source of potential new antimicrobial agents available upon spaceflight.

In summary, the present study underscored the benefits of deep learning-based annotations, here using DeepFRI[58], which are able to go beyond the limitations of homology-based tools like PGAP and eggNOG[59]. In fact, DeepFRI enabled the construction of a comprehensive and cohesive picture of adaptations to space conditions. Nevertheless, combined use of annotations derived from all tools was applied to e.g. mechanosensitive ion channel proteins (see Section 2.3.2). While traditional tools often relegated these genes to the ambiguous "hypothetical" category, DeepFRI was adept at providing precise annotations. This allowed for the discernment of novel BGCs, a finding that was further corroborated using specialized tools such as antiSMASH[28, 60]. In addition, a significant portion of these novel annotations pertained to membrane-associated proteins. This resonates with prior observations highlighting membrane-related adjustments. The challenge of studying the membrane proteome cannot be understated. Yet, with the recent augmentation of databases like AlphaFold[27] and ESM atlas[61], the scientific community has been endowed with a richer repository of membrane-bound protein structures. This surge in data availability paves the way for more refined structure-based annotations of novel genomes.

This deep dive into the genetic underpinnings bears profound significance, especially when contextualized within space-related stress factors. As organisms confront the multifaceted challenges of space, their genetic machinery undergoes intricate recalibrations. By harnessing sophisticated annotation tools, we gain unparalleled insights into these genetic adaptations, potentially unlocking the secrets of space resilience. Such revelations not only propel our understanding of astrobiology forward but also spotlight the potential of genetic mechanisms in tailoring organisms for optimal performance in extraterrestrial environments.

4 Methods

4.1 Bacteria isolation, genome assembly and annotation

Sample collection and bacterial isolation from ISS were performed as described in [23, 62]. Biochemical tests were performed using a Gram-positive identification card (Vitek 2 GP ID, bioMérieux) and phenotypic fingerprint was generated through GNIII MicroPlate according to BioLog's protocol as described elsewhere [28]. We assembled the genomes of five novel bacterial species isolated from various surfaces of the ISS (specifically strains IF8SW-P5, IIF3SC_B10, F6_8S_P_1B, F6_8S_P_2B, F6_3S_P_1C), using the SPAdes workflow[63]. Phylogenetic affiliations are as outlined in [28] and presented in Table 1.

4.2 DNA extraction

The ISS isolates (n=5) were grown on R2A medium and incubated at 30°C, for 2 to 7 days. For WGS, DNA was isolated with the ZymoBIOMICS DNA MagBead kit, and WGS libraries prepared using the Illumina Nextera DNA Flex kit. Sequencing on the NovaSeq 6000 platform involved paired-end, 2 x 150 bp reads. FastQC (https://github.com/s-andrews/FastQC) and fastp [64] ensured data quality. WGS reads were assembled using SPAdes [63], with genome statistics from QUAST [65] and quality checks from CheckM [66]. Species verification through fastANI [67] and fastAAI (https://github.com/cruizperez/FastAAI) confirmed identities with over 95% identity, and digital DNA-DNA hybridization (dDDH) was estimated with GGDC using Formula 2[68]. For genome synteny analysis, we used SynTracker [25].

4.3 Genome annotation

Following assembly, the contigs were subjected to the NCBI Prokaryotic Genome Annotation Pipeline (PGAP)[69], available at https://github.com/ncbi/pgap/wiki, in order to obtain draft genomes and proteomes. Genomes were further investigated for their functions, where all identified genes (bearing PGAP descriptions in their names) underwent homology-based annotation using eggNOG v5.0, accessible at https://github.com/eggnogdb/eggnog-mapper. Genomes of closest relatives of the ISS species (highest average nucleotide identity [ANI]) were also subjected to PGAP and eggNOG annotation.

Proteomes with the PGAP identifiers were used as input for DeepFRI analysis [58], available on github: (https://github.com/bioinf-mcb/Metagenomic-DeepFRI/), which generated query contact map using results from mmseqs2 target database[70] search for similar protein sequences with known structures and with v 1.0 model weights. For reference database, we combined structures from Swissprot (https://alphafold.ebi.ac.uk/download), microbiome immunity project (MIP) [71] and protein databank (PDB, https://www.rcsb.org). Next, the contact map alignment was performed to use it as input to DeepFRI's Graph Convolutional Network (GCN) and annotations without the alignment to known structures were processed with Convolutional Neural Network (CNN). For medium quality (MQ) and high quality (HQ) annotation, scores higher than or equal to 0.3 (50% of prediction confidence; MQ) and 0.5 (90% of prediction confidence; HQ) were used respectively (Figs. 2-3). All annotations for ISS strains and their terrestrial relatives are provided in Supplementary Tables S2-S11. Phylogenetically close relatives from Earth for each tested species were identified using ANI (Table 1). Their genomes were annotated using DeepFRI as described above. Protein sequences from the aforementioned five species were aligned and conducted two-dimensional clustering using CD-HIT[72], available at https://github.com/weizhongli/cdhit, employing a 40% similarity threshold. Genes below that threshold were considered "novel" and subjected to analysis shown in Fig.3A-B.

Annotations (GO terms) were then compared between ISS strains (Supplementary Table S12) and their terrestrial counterparts (Supplementary Table S13). Common annotations were aggregated according to clusters of orthologous genes (COG)

categories[73] using go2cog mapper (https://github.com/Tomasz-Lab/go2cog). GO terms that could be assigned to more than 3 COG categories were grouped as category R - "general function prediction" (Figures 2 and 3). Heatmaps were generated using seaborn (https://github.com/mwaskom/seaborn). Venn diagrams and Upsetplot were generated using matplotlib (https://github.com/matplotlib/matplotlib). Using a knowledge-based approach, a set of genes was selected for further analysis. Protein structures were modeled using *in silico* structure prediction tool – AlphaFold Colab[27] – and hence obtained protein databank (PDB) format structures were aligned using Geneious Prime 2023.1.1 (Biomatters). Structures were aligned using TM-align tool [26]. Based on the predicted structure, novel functional annotation was performed using DeepFRI. All the annotation data and scripts are placed in the repository on github (https://github.com/lmszydlowski/Adaptation-of-novel-bacterial-species-isolated-from-the-International-Space-Station).

4.4 Sequence alignments, functional annotations for plasmids, phage elements and biosynthetic gene clusters in the genome

Sequences of selected proteins were aligned using Muscle 5.1 [74]. Clustering dendrograms were built using PhyML 3.3,[75] using LG amino acid substitution model and approximate likelihood ratio test (aLRT) for branch support. All contigs generated for each genome were scanned for potential plasmids using PLASme software (https://github.com/HubertTang/PLASMe), phage-related elements using PHASTER software[76, 77] and results provided in the Supplementary Tables S15-S16. Genome sequences were also processed through the AntiSMASH 7.0 software [60] to look for biosynthetic gene clusters (BGCs) (Supplementary Table S14).

4.5 Protein mutation and enrichment analysis

To investigate the roles and processes of proteins highly enriched in bacteria from the ISS, we specifically focused on the enrichment of certain amino acids, especially proline and glycine, which belong to unique amino acid change groups. The amino acid compositions of proteins from the ISS strains were compared with those proteins found in the closest terrestrial species through global alignment methods[78]. This facilitates the identification and quantification of any shifts in the amino acid composition. These shifts were evaluated by calculating the mutation rate, which is achieved by dividing the number of amino acid substitutions in the ISS bacteria by the number of amino acid substitutions in their closest terrestrial relatives. Proteins amino acid mutation rate normalized and Z-score above 1 .96 or below -1.96 were classified as enriched. To investigate the functions and pathways of these proteins, we utilize the g:Profiler enrichment tool (https://biit.cs.ut.ee/gprofiler/gost[79]). Custom gene sets for bacteria were created using UniProt's[80] predicted Gene Ontology (GO) term associations[81] for bacterial proteins. The tool was configured to include a minimum gene set size of 5 and a maximum of 10,000, focusing on molecular function GO terms. Shared gene sets between samples were filtered using a false discovery rate (FDR) cutoff of 0.05 in 1424 molecular function GO terms. Data visualization was enhanced

through ete3 (https://doi.org/10.1093/molbev/msw046) for constructing phylogenetic trees and seaborn (https://github.com/mwaskom/seaborn) for generating heatmaps. The phylogenetic distance between bacterial species was calculated using NCBI's distance calculations, with the closest terrestrial bacteria as a reference.

Supplementary information.
Supplementary information contains 7 figures and 16 Tables.

Supplementary Figure 1. Assembled contigs with regions corresponding to biosynthetic gene clusters marked. Full description of BGC from each organism is provided in the Supplementary Table S14.

Supplementary Figure 2. Alignment of novel S-layer osidoreductases with their closest sequenctial homolog from *D. radiodurans* (WP_027479552.1, red). Mma - *M. mcarthurae* (magenta), Mme - *M. meiriae*, orange. **pLDDT** - predicted local-distance difference test, **pTM** - predicted template modelling, **TM** - template modelling, see [26, 27].

Supplementary Figure 3. A) Phenotypic assays of studied organisms with GNIII MicroPlate (Biolog). **B)** Biochemical assays of studied organisms Gram-positive identification card Vitek 2 GP ID (bioMérieux). Blue - absent, red - present. Assays were performed according to manufacturers' protocol.

Supplementary Figure 4. Phylogenetic tree based on alignment of MscS protein sequences from ISS organisms and their closest terrestrial relatives. Branch labels indicate substitutions per site. Ab - *A. burdickii* (blue), Asp - *A. agilis* 4041 (blue), Mma - *M. mcarthurae* (burgundy), Msp - *Microbacterium* sp. ACCRU (burgundy), Lw - *L. williamsii* (yellow), Lsp - *Leifsonia* sp.ku-ls (yellow), Mme - *M. meiriae* (orange), Mp - *Microbacterium* sp. LTR1 (orange), Pv - *P. vandeheii*, (green), Psp - *Paenibacillus* sp. MAEPY (green).

Supplementary Figure 5. Alignment of DNA and amino acid sequence of aminopeptidases Mme_001931 and WP_268776344. Green boxes indicate sequence identity.

Supplementary Figure 6. Structure alignment of aminopeptidase Mme_001931 (cyan) and WP_0516188611 (red). **pLDDT** - predicted local-distance difference test, **pTM** - predicted template modelling, **TM** - template modelling, see [26, 27].

Supplementary Figure 7. Phylogenetic tree based on alignment of DNA-repair protein sequences from ISS organisms and *D. radiodurans* and *B. pumilus* SAFR-032. Branch labels indicate substitutions per site. Ab - *A. burdickii* (blue), Bp - *B. pumilus* (black), Dr - *D. radiodurans* (red), Mma - *M. mcarthurae* (burgundy), Lw - *L. williamsii* (yellow), Mme - *M. meiriae* (orange), Pv - *P. vandeheii*, (green).

Supplementary Table 1. ANI results (top 8) for all ISS species.

Supplementary Table 2. Annotations of *M. mcarthurae* genes.

Supplementary Table 3. Annotations of *A. burdickii* genes.

Supplementary Table 4. Annotations of *P. vandeheii* genes.

Supplementary Table 5. Annotations of *L. williamsii* genes.

Supplementary Table 6. Annotations of *M. meiriae* genes.

Supplementary Table 7. Annotations of *Microbacterium* sp. ACCRU genes.

Supplementary Table 8. Annotations of *A. agilis* 4041 genes.

Supplementary Table 9. Annotations of *Paenibacillus* sp. MAEPY2 genes.

Supplementary Table 10. Annotations of *Leifsonia* sp. ls-ku genes.
Supplementary Table 11. Annotations of *Microbacterium* sp. LTR1 genes.
Supplementary Table 12. Annotations of genes common in all ISS species.
Supplementary Table 13. Annotations of unique (not present in closest relative) genes shared by all ISS species.
Supplementary Table 14. DeepFRI annotations of unique (not present in closest relative) genes and antiSMASH predictions of BGCs.
Supplementary Table 15. Structure - based DeepFRI annotations of Mma_002827 and Mme_000934.
Supplementary Table 16. Predictions for plasmids and phage elements in the genomes of ISS species.

Acknowledgements

We acknowledge the Jet Propulsion Laboratory supercomputing facility staff, notably Narendra J. Patel (Jimmy) and Edward Villanueva, for their continuous support in providing the best possible infrastructure for BIGDATA analysis. © 2023 California Institute of Technology, Government sponsorship acknowledged.

Declarations

Funding

Part of the research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with National Aeronautics and Space Administration. This research was funded by a 2012 Space Biology NNH12ZTT001N grant no. 19-12829-26 under Task Order NNN13D111T award to KV. The funders had no role in study design, data collection and interpretation, the writing of the manuscript, or the decision to submit the work for publication. TK and LS have been funded by the National Science Centre, Poland grant 2019/35/D/NZ2/04353. PL has been partially funded by the National Science Centre, Poland grant 2020/38/E/NZ2/00598.

Ethics approval

Not applicable

Consent to participate

Not applicable

Consent for publication

All authors have agreed to be co-authors and have approved the final version of this manuscript.

Availability of data and materials

The draft genome sequences of strains IIF3SC_B10, F6_3S_P_1C and F6_8S_P_1B were deposited in NCBI under BioProject PRJNA935338 (see [28]. Strains F6_8S_P_2B and IF8SW-P5 were deposited in BioProjects PRJNA869446 and PRJNA224116, respectively. The WGS accession numbers are given in Table 1 and the genome versions described in this paper are the first versions. Moreover, genomes of all ISS strains and their terrestrial relatives are available on Kbase (narrative 107942).

Code availability

All bioinformatic tools used in this study are available on GitHub: https://github.com/lmszydlowski/Adaptation-of-novel-bacterial-species-isolated-from-the-International-Space-Station.

Authors' contributions

- **LS**, manuscript writing, data analysis, figure preparation;
- **AB**, figure preparation, data analysis;
- **AS**, data analysis, manuscript reviewing;
- **NS**, data analysis, manuscript reviewing;
- **DEK**, figure preparation, manuscript reviewing;
- **PL**, manuscript reviewing, project supervision;
- **OUS**, manuscript reviewing, project supervision;
- **TK**, manuscript reviewing, data analysis, project supervision;
- **VK**, conceptualization, manuscript reviewing, project supervision;

References

[1] Vaishampayan, P.A., Rabbow, E., Horneck, G., Venkateswaran, K.J.: Survival of *Bacillus pumilus* spores for a prolonged period of time in real space conditions. Astrobiology **12**(5), 487–497 (2012) https://doi.org/10.1089/ast.2011.0738

[2] Cox, M.M., Battista, J.R.: *Deinococcus radiodurans* - The consummate survivor. Nature Reviews Microbiology **3**(11), 882–892 (2005) https://doi.org/10.1038/nrmicro1264

[3] Onofri, S., De Vera, J.P., Zucconi, L., Selbmann, L., Scalzi, G., Venkateswaran, K.J., Rabbow, E., De La Torre, R., Horneck, G.: Survival of Antarctic Cryptoendolithic Fungi in Simulated Martian Conditions on Board the International Space Station. Astrobiology **15**(12), 1052–1059 (2015) https://doi.org/10.1089/ast.2015.1324

[4] Rebecchi, L., Altiero, T., Cesari, M., Bertolani, R., Rizzo, A.M., Corsetto, P.a., Guidetti, R.: Resistance of the anhydrobiotic eutardigrade *Paramacrobiotus richtersi* to space flight (LIFE-TARSE mission on FOTON-M3). Journal

of Zoological Systematics and Evolutionary Research **49**, 98–103 (2011) https://doi.org/10.1111/j.1439-0469.2010.00606.x

[5] Ricci, C., Caprioli, M., Boschetti, C., Santo, N.: *Macrotrachela quadricornifera* featured in a space experiment. Hydrobiologia **534**(1-3), 239–244 (2005) https://doi.org/10.1007/s10750-004-1509-7

[6] Rummel, J.D.: Planetary exploration in the time of astrobiology: Protecting against biological contamination. Proceedings of the National Academy of Sciences of the United States of America **98**(5), 2128–2131 (2001) https://doi.org/10.1073/pnas.061021398

[7] Koehle, A.P., Brumwell, S.L., Seto, E.P., Lynch, A.M., Urbaniak, C.: Microbial applications for sustainable space exploration beyond low Earth orbit. npj Microgravity **9**(1) (2023) https://doi.org/10.1038/s41526-023-00285-0

[8] Prasad, B., Richter, P., Vadakedath, N., Haag, F.W.M., Strauch, S.M., Mancinelli, R., Schwarzwälder, A., Etcheparre, E., Gaume, N., Lebert, M.: How the space environment influences organisms: An astrobiological perspective and review. International Journal of Astrobiology **20**(2), 159–177 (2021) https://doi.org/10.1017/S1473550421000057

[9] Blaustein, R.A., Mcfarland, A.G., Maamar, B., Lopez, A., Castro-wallace, S., Hartmann, E.M.: Pangenomic Approach To Understanding Microbial Adaptations within a Model Built Environment , the International Space Station , Relative to Human Hosts and Soil. mSystems **4**(1) (2019) https://doi.org/10.1128/mSystems.00281-18

[10] Milojevic, T., Weckwerth, W.: Molecular Mechanisms of Microbial Survivability in Outer Space: A Systems Biology Approach. Frontiers in Microbiology **11**(May) (2020) https://doi.org/10.3389/fmicb.2020.00923

[11] Blachowicz, A., Romsdahl, J., Chiang, A.J., Masonjones, S., Kalkum, M., Stajich, J.E., Torok, T., Wang, C.C.C., Venkateswaran, K.: The International Space Station Environment Triggers Molecular Responses in Aspergillus niger. Frontiers in Microbiology **13**(June) (2022) https://doi.org/10.3389/fmicb.2022.893071

[12] Senatore, G., Mastroleo, F., Leys, N., Mauriello, G.: Effect of microgravity and space radiation on microbes. Future Microbiology **13**(7), 831–847 (2018) https://doi.org/10.2217/fmb-2017-0251

[13] Kumar, R.K., Singh, N.K., Balakrishnan, S., Parker, C.W., Raman, K., Venkateswaran, K.: Metabolic modeling of the International Space Station microbiome reveals key microbial interactions. Microbiome **10**(1), 1–16 (2022) https://doi.org/10.1186/s40168-022-01279-y

[14] Simões, M.F., Antunes, A.: Microbial pathogenicity in space. Pathogens **10**(4),

1–23 (2021) https://doi.org/10.3390/pathogens10040450

[15] Ott, E., Kawaguchi, Y., Kölbl, D., Rabbow, E., Rettberg, P., Mora, M., Moissl-Eichinger, C., Weckwerth, W., Yamagishi, A., Milojevic, T.: Molecular repertoire of deinococcus radiodurans after 1 year of exposure outside the international space station within the tanpopo mission. Microbiome **8** (2020) https://doi.org/10.1186/s40168-020-00927-5

[16] Chen, Y., Zhang, Q., Wang, D., Shu, Y.-G., Shi, H.: Memory Effect on the Survival of *Deinococcus radiodurans* after Exposure in Near Space. Microbiology Spectrum **11**(2) (2023) https://doi.org/10.1128/spectrum.03474-22

[17] Tierney, B.T., Singh, N.K., Simpson, A.C., Hujer, A.M., Bonomo, R.A., Mason, C.E., Venkateswaran, K.: Multidrug-resistant *Acinetobacter pittii* is adapting to and exhibiting potential succession aboard the International Space Station. Microbiome **10**(1), 1–14 (2022) https://doi.org/10.1186/s40168-022-01358-0

[18] Ott, E., Fuchs, F.M., Moeller, R., Hemmersbach, R., Kawaguchi, Y.: Molecular response of *Deinococcus radiodurans* to simulated microgravity explored by proteometabolomic approach, 1–13 (2019) https://doi.org/10.1038/s41598-019-54742-6

[19] Voorhies, A.A., Ott, C.M., Mehta, S., Pierson, D.L., Crucian, B.E., Feiveson, A., Oubre, C.M., Torralba, M., Moncera, K., Zhang, Y., Zurek, E., Lorenzi, H.A.: Study of the impact of long- duration space missions at the International Space Station on the astronaut microbiome. Scientific Reports **9**(9911), 1–17 (2019) https://doi.org/10.1038/s41598-019-46303-8

[20] Singh, N.K., Lavire, C., Nesme, J., Vial, L., Nesme, X., Mason, C.E., Lassalle, F., Venkateswaran, K.: Comparative Genomics of Novel *Agrobacterium* G3 Strains Isolated From the International Space Station and Description of *Agrobacterium tomkonis* sp. nov. Frontiers in Microbiology **12**(December), 1–15 (2021) https://doi.org/10.3389/fmicb.2021.765943

[21] Bijlani, S., Singh, N.K., Eedara, V.V.R., Podile, A.R., Mason, C.E., Wang, C.C.C., Venkateswaran, K.: *Methylobacterium ajmalii* sp. nov., Isolated From the International Space Station. Frontiers in Microbiology **12**(March), 1–14 (2021) https://doi.org/10.3389/fmicb.2021.639396

[22] Fajardo-Cavazos, P., Nicholson, W.L.: Cultivation of *Staphylococcus epidermidis* in the human spaceflight environment leads to alterations in the frequency and spectrum of spontaneous rifampicin-resistance mutations in the *rpo*B gene. Frontiers in Microbiology **7**(JUN), 1–10 (2016) https://doi.org/10.3389/fmicb.2016.00999

[23] Urbaniak, C., Morrison, M.D., Thissen, J.B., Karouia, F., Smith, D.J., Mehta, S., Jaing, C., Venkateswaran, K.: Microbial Tracking-2, a metagenomics analysis of

bacteria and fungi onboard the International Space Station. Microbiome **10**(1), 1–19 (2022) https://doi.org/10.1186/s40168-022-01293-0

[24] Coil, D.A., Neches, R.Y., Lang, J.M., Brown, W.E., Severance, M., Cavalier, D., Eisen, J.A.: Growth of 48 built environment bacterial isolates on board the International Space Station (ISS). PeerJ **2016**(3), 1–11 (2016) https://doi.org/10.7717/peerj.1842

[25] Enav, H., Paz, I., Ley, R.E.: Strain tracking in complex microbiomes using synteny analysis reveals per-species modes of evolution. Nature Biotechnology (2024) https://doi.org/10.1038/s41587-024-02276-2

[26] Zhang, Y., Skolnick, J.: TM-align: A protein structure alignment algorithm based on the TM-score. Nucleic Acids Research **33**(7), 2302–2309 (2005) https://doi.org/10.1093/nar/gki524

[27] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D.: Highly accurate protein structure prediction with AlphaFold. Nature **596**(7873), 583–589 (2021) https://doi.org/10.1038/s41586-021-03819-2

[28] Simpson, A.C., Sengupta, P., Zhang, F., Hameed, A., Parker, C.W., Singh, N.K., Miliotis, G., Rekha, P.D., Raman, K., Mason, C.E., Venkateswaran, K.: Phylogenomics, phenotypic, and functional traits of five novel (earth-derived) bacterial species isolated from the international space station and their prevalence in metagenomes. Scientific Reports **13** (2023) https://doi.org/10.1038/s41598-023-44172-w

[29] Högel, P., Götz, A., Kuhne, F., Ebert, M., Stelzer, W., Rand, K.D., Scharnagl, C., Langosch, D.: Glycine Perturbs Local and Global Conformational Flexibility of a Transmembrane Helix vol. 57, pp. 1326–1337 (2018). https://doi.org/10.1021/acs.biochem.7b01197

[30] Chavez, A., Topolski, C., Hicks, J., Villafania, M., Baez, N., Burke, M., Castillo, H.: Phenotypical changes in escherichia coli k12 after prolonged exposure to simulated microgravity. Frontiers in Space Technologies **4** (2024) https://doi.org/10.3389/frspt.2023.1282850

[31] Manyi-Loh, C., Mamphweli, S., Meyer, E., Okoh, A.: Antibiotic use in agriculture and its consequential resistance in environmental sources: Potential public health implications. Molecules **23** (2018) https://doi.org/10.3390/molecules23040795

[32] Fajardo-Cavazos, P., Nicholson, W.L.: Mechanotransduction in prokaryotes: A

possible mechanism of spaceflight adaptation. Life **11**(1), 1–11 (2021) https://doi.org/10.3390/life11010033

[33] Daly, M.J., Gaidamakova, E.K., Matrosova, V.Y., Vasilenko, A., Zhai, M., Leapman, R.D., Lai, B., Ravel, B., Li, S.M.W., Kemner, K.M., Fredrickson, J.K.: Protein oxidation implicated as the primary determinant of bacterial radioresistance. PLoS Biology **5**(4), 769–779 (2007) https://doi.org/10.1371/journal.pbio.0050092

[34] Bengtsson-Palme, J., Kristiansson, E., Larsson, D.G.J.: Environmental factors influencing the development and spread of antibiotic resistance. FEMS Microbiology Reviews (2018)

[35] Sharma, G., Curtis, P.D.: The impacts of microgravity on bacterial metabolism. Life (2022)

[36] Mora, M., Wink, L., Kögler, I., Mahnert, A., Rettberg, P., Schwendner, P., Demets, R., Cockell, C., Alekhova, T., Klingl, A., Krause, R., Zolotariof, A., Alexandrova, A., Moissl-Eichinger, C.: Space station conditions are selective but do not alter microbial characteristics relevant to human health. Nature Communications **10** (2019) https://doi.org/10.1038/s41467-019-11682-z

[37] Singh, N.K., Wood, J.M., Patane, J., Maria, L., Moura, S., Lombardino, J.: Characterization of Metagenome Assembled Genomes from the International Space Station. BMC Microbiome, 1–20 (2022) https://doi.org/10.21203/rs.3.rs-1934191/v1

[38] Bradbury, P., Wu, H., Choi, J.U., Rowan, A.E., Zhang, H., Poole, K., Lauko, J., Chou, J.: Modeling the Impact of Microgravity at the Cellular Level: Implications for Human Disease. Frontiers in Cell and Developmental Biology **8**(February), 1–8 (2020) https://doi.org/10.3389/fcell.2020.00096

[39] Nothaft, H., Szymanski, C.M.: Protein glycosylation in bacteria: Sweeter than ever. Nature Reviews Microbiology **8**(11), 765–778 (2010) https://doi.org/10.1038/nrmicro2383

[40] Lee, J.: Tolerance at the genetic level of the brine shrimp *Artemia salina* to a wide range of salinity. preprint, 1–21 (2020)

[41] Zahradka, K., Slade, D., Bailone, A., Sommer, S., Averbeck, D., Petranovic, M., Lindner, A.B., Radman, M.: Reassembly of shattered chromosomes in *Deinococcus radiodurans*. Nature **443**(7111), 569–573 (2006) https://doi.org/10.1038/nature05160

[42] Su, X., Guo, Y., Fang, T., Jiang, X., Wang, D., Li, D., Bai, P., Zhang, B., Wang, J., Liu, C.: Effects of simulated microgravity on the physiology of stenotrophomonas maltophilia and multiomic analysis. Frontiers in Microbiology **12** (2021) https:

//doi.org/10.3389/fmicb.2021.701265

[43] Su, X., Fang, T., Fang, L., Wang, D., Jiang, X., Liu, C., Zhang, H., Guo, R., Wang, J.: Effects of short-term exposure to simulated microgravity on the physiology of bacillus subtilis and multiomic analysis. Canadian Journal of Microbiology **69**, 464–478 (2023) https://doi.org/10.1139/cjm-2023-0061

[44] Jönsson, K.I., Harms-Ringdahl, M., Torudd, J.: Radiation tolerance in the eutardigrade *Richtersius coronifer*. International Journal of Radiation Biology **81**(9), 649–656 (2005) https://doi.org/10.1080/09553000500368453

[45] Gladyshev, E., Meselson, M.: Extreme resistance of bdelloid rotifers to ionizing radiation. Proceedings of the National Academy of Sciences of the United States of America **105**(13), 5139–44 (2008) https://doi.org/10.1073/pnas.0800966105

[46] Butala, M., Žgur-Bertok, D., Busby, S.J.W.: The bacterial LexA transcriptional repressor. Cellular and Molecular Life Sciences **66**(1), 82–93 (2009) https://doi.org/10.1007/s00018-008-8378-6

[47] Boschetti, C., Pouchkina-stantcheva, N., Hoffmann, P., Tunnacliffe, A.: Foreign genes and novel hydrophilic protein genes participate in the desiccation response of the bdelloid rotifer *Adineta ricciae*. The Journal of experimental biology **214**, 59–68 (2011) https://doi.org/10.1242/jeb.050328

[48] Eyres, I., Boschetti, C., Crisp, A., Smith, T.P., Fontaneto, D., Tunnacliffe, A., Barraclough, T.G.: Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. BMC Biology **13**(1) (2015) https://doi.org/10.1186/s12915-015-0202-9

[49] Gladyshev, E., Meselson, M., Arkhipova, I.: A deep-branching clade of retrovirus-like retrotransposons in bdelloid rotifers. Gene **390**, 136–145 (2007)

[50] Pasternak, C., Dulermo, R., Ton-Hoang, B., Debuchy, R., Siguier, P., Coste, G., Chandler, M., Sommer, S.: Isdra2 transposition in deinococcus radiodurans is downregulated by tnpb. Molecular Microbiology **88**, 443–455 (2013) https://doi.org/10.1111/mmi.12194

[51] Seuylemezian, A., Vaishampayan, P., Cooper, K., Venkateswaran, K.: Draft genome sequences of *Acinetobacter* and *Bacillus* strains isolated from spacecraft-associated surfaces. Genome Announcements **6**(6), 1–2 (2018) https://doi.org/10.1128/genomeA.01554-17

[52] Borders, C.L., Broadwater, J.A., Bekeny, P.A., Salmon, J.E., Lee, A.S., Eldridge, A.M., Pett, V.B.: A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. Protein Science **3**(4), 541–548 (1994) https://doi.org/10.1002/pro.5560030402

[53] Donald, J.E., Kulp, D.W., Degrado, W.F.: Salt Bridges: Geometrically Specific, Designable Interactions. Biochemistry **79**(3), 898–915 (2012) https://doi.org/10.1002/prot.22927.Salt

[54] Sokalingam, S., Raghunathan, G., Soundrarajan, N., Lee, S.G.: A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. PLoS ONE **7**(7) (2012) https://doi.org/10.1371/journal.pone.0040410

[55] Aledo, J.C.: Methionine in proteins: The Cinderella of the proteinogenic amino acids. Protein Science **28**(10), 1785–1796 (2019) https://doi.org/10.1002/pro.3698

[56] Kim, W., Tengra, F.K., Young, Z., Shong, J., Marchand, N., Chan, H.K., Pangule, R.C., Parra, M., Dordick, J.S., Plawsky, J.L., Collins, C.H.: Spaceflight Promotes Biofilm Formation by Pseudomonas aeruginosa. PLoS ONE **8**(4), 1–8 (2013) https://doi.org/10.1371/journal.pone.0062437

[57] Gillor, O., Vriezen, J.A.C., Riley, M.A.: The role of SOS boxes in enteric bacteriocin regulation. Microbiology **154**(6), 1783–1792 (2008) https://doi.org/10.1099/mic.0.2007/016139-0

[58] Gligorijević, V., Renfrew, P.D., Kosciolek, T., Leman, J.K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B.C., Fisk, I.M., Vlamakis, H., Xavier, R.J., Knight, R., Cho, K., Bonneau, R.: Structure-based protein function prediction using graph convolutional networks. Nature Communications **12**(1) (2021) https://doi.org/10.1038/s41467-021-23303-9

[59] Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., Von Mering, C., Bork, P.: EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research **47**(D1), 309–314 (2019) https://doi.org/10.1093/nar/gky1085

[60] Blin, K., Shaw, S., Augustijn, H.E., Reitz, Z.L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B.R., Metcalf, W.W., Helfrich, E.J.N., van Wezel, G.P., Medema, M.H., Weber, T.: antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Research **51**(May), 46–50 (2023) https://doi.org/10.1093/nar/gkad344

[61] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., Rives, A.: Evolutionary-scale prediction of atomic-level protein structure with a language model. Science **379**(6637), 1123–1130 (2023) https://doi.org/10.1126/science.ade2574

[62] Sielaff, A.C., Urbaniak, C., Babu, G., Mohan, M., Stepanov, V.G., Tran, Q.,

Wood, J.M., Minich, J., Mcdonald, D., Mayer, T., Knight, R., Karouia, F., Fox, G.E., Venkateswaran, K.: Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. Microbiome **7**, 50 (2019)

[63] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A.: Spades: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology **19**, 455–477 (2012) https://doi.org/10.1089/cmb.2012.0021

[64] Chen, S., Zhou, Y., Chen, Y., Gu, J.: Fastp: An ultra-fast all-in-one fastq preprocessor **34**, 884–890 (2018) https://doi.org/10.1093/bioinformatics/bty560

[65] Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G.: Quast: Quality assessment tool for genome assemblies. Bioinformatics **29**, 1072–1075 (2013) https://doi.org/10.1093/bioinformatics/btt086

[66] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W.: Checkm: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research **25**, 1043–1055 (2015) https://doi.org/10.1101/gr.186072.114

[67] Yoon, S.H., Ha, S., Lim, J., Kwon, S., Chun, J.: A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology **110**, 1281–1286 (2017) https://doi.org/10.1007/s10482-017-0844-4

[68] Meier-Kolthoff, J.P., Carbasse, J.S., Peinado-Olarte, R.L., Göker, M.: Tygs and lpsn: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Research **50**, 801–807 (2022) https://doi.org/10.1093/nar/gkab902

[69] Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J.: NCBI prokaryotic genome annotation pipeline. Nucleic Acids Research **44**(14), 6614–6624 (2016) https://doi.org/10.1093/nar/gkw569

[70] Steinegger, M., Söding, J.: MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nature Biotechnology **35**(11), 1026–1028 (2017) https://doi.org/10.1038/nbt.3988

[71] Koehler Leman, J., Szczerbiak, P., Renfrew, P.D., Gligorijevic, V., Berenberg, D., Vatanen, T., Taylor, B.C., Chandler, C., Janssen, S., Pataki, A., Carriero, N., Fisk, I., Xavier, R.J., Knight, R., Bonneau, R., Kosciolek, T.: Sequence-structure-function relationships in the microbial protein universe. Nature Communications

14(1) (2023) https://doi.org/10.1038/s41467-023-37896-w

[72] Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W.: CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics **28**(23), 3150–3152 (2012) https://doi.org/10.1093/bioinformatics/bts565

[73] Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Alvarez, R.V., Landsman, D., Koonin, E.V.: COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Research **49**(D1), 274–281 (2021) https://doi.org/10.1093/nar/gkaa1018

[74] Edgar, R.C.: Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nature Communications **13**(1), 1–9 (2022) https://doi.org/10.1038/s41467-022-34630-w

[75] Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O.: New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology **59**(3), 307–321 (2010) https://doi.org/10.1093/sysbio/syq010

[76] Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S.: PHAST: A Fast Phage Search Tool. Nucleic Acids Research **39**(SUPPL. 2), 347–352 (2011) https://doi.org/10.1093/nar/gkr485

[77] Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S.: PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research **44**(W1), 16–21 (2016) https://doi.org/10.1093/nar/gkw387

[78] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J.: Basic local alignment search tool. Journal of Molecular Biology **215**(3), 403–410 (1990) https://doi.org/10.1016/S0022-2836(05)80360-2

[79] Reimand, J., Kull, M., Peterson, H., Hansen, J., Vilo, J.: G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Research **35**(SUPPL.2), 193–200 (2007) https://doi.org/10.1093/nar/gkm226

[80] Consortium, U.: UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research **47**(D1), 506–515 (2019) https://doi.org/10.1093/nar/gky1049

[81] Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, Alan, P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene Ontology: tool for the unification of biology. Nature Genet. **25**(1), 25–29 (2000) https://doi.org/10.1038/75556.Gene