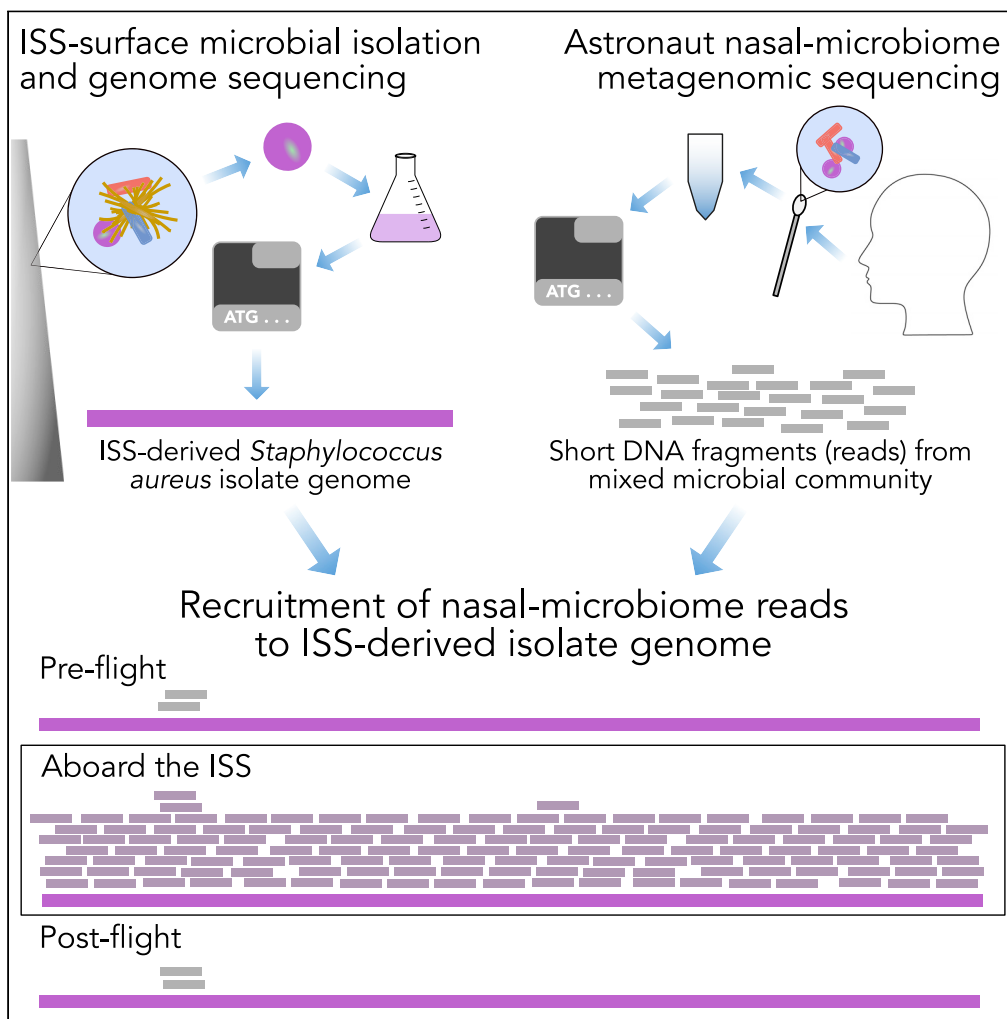## Article

Reference-guided metagenomics reveals genome-level evidence of potential microbial transmission from the ISS environment to an astronaut's microbiome

ISS-surface microbial isolation and genome sequencing

Astronaut nasal-microbiome metagenomic sequencing

ATG . . .

ATG . . .

ISS-derived *Staphylococcus aureus* isolate genome

Short DNA fragments (reads) from mixed microbial community

Recruitment of nasal-microbiome reads to ISS-derived isolate genome

Pre-flight

Aboard the ISS

Post-flight

Michael D. Lee, Aubrie O'Rourke, Hernan Lorenzi, Brad M. Bebout, Chris L. Dupont, R. Craig Everroad

Mike.Lee@nasa.gov

HIGHLIGHTS

Understanding built-environment microbiomes is critical for human space exploration

Reference-guided metagenomics is a powerful approach for monitoring microbiomes

We show potential microbial colonization of an astronaut microbiome while aboard the ISS

CellPress
OPEN ACCESS

Article

Reference-guided metagenomics reveals genome-level evidence of potential microbial transmission from the ISS environment to an astronaut's microbiome

Michael D. Lee,[1,2,6,7,*] Aubrie O'Rourke,[3] Hernan Lorenzi,[4] Brad M. Bebout,[1] Chris L. Dupont,[5] and R. Craig Everroad[1]

SUMMARY

Monitoring microbial communities aboard the International Space Station (ISS) is essential to maintaining astronaut health and the integrity of life-support systems. Using assembled genomes of ISS-derived microbial isolates as references, recruiting metagenomic reads from an astronaut's nasal microbiome revealed no recruitment to a *Staphylococcus aureus* isolate from samples before launch, yet systematic recruitment across the genome when sampled after 3 months aboard the ISS, with a median percent identity of 100%. This suggests that either a highly similar *S. aureus* population colonized the astronaut's nasal microbiome while the astronaut was aboard the ISS or that it may have been below detection before spaceflight, instead supporting a shift in community composition. This work highlights the value in generating genomic libraries of microbes from built-environments such as the ISS and demonstrates one way such data can be integrated with metagenomics to facilitate the tracking and monitoring of astronaut microbiomes and health.

INTRODUCTION

There is a growing awareness and consideration as to how the microbiome of built-environments affects human health (e.g., National Academies of Sciences and Medicine, 2017). The International Space Station (ISS) represents a particularly unique built-environment given its physical isolation, relatively low rate of occupant quantity, and turnover and impacts on immunity and health in general due to environmental factors such as microgravity and radiation. Tracking and monitoring of microbial communities within space-based built-environments such as the ISS is essential to maintaining astronaut health and preserving the integrity of life-support components such as potable-water and food-production systems. Particularly as we begin to look toward longer-duration human-spaceflight missions (Voorhies and Lorenzi, 2016), we will need to continue, and expand, our practices in gathering and utilizing information on the microbial populations inhabiting these closed environments that are traveling with us.

Microbial surveillance of the ISS environment by NASA has been underway in different respects for some time. These efforts have included culture-based approaches that have allowed the characterization of isolates recovered from surfaces (e.g., Knox et al., 2016; Sielaff et al., 2017; Romsdahl et al., 2018; Blachowicz et al., 2019) and the ISS potable-water system (O'Rourke et al., 2020). Culture-independent methods have also been employed, including targeted-amplicon sequencing (e.g., 16S ribosomal RNA, e.g., Checinska et al., 2015; Sielaff et al., 2019; Voorhies et al., 2019) and metagenomic sequencing (e.g., Be et al., 2017; Singh et al., 2018; Avila-Herrera et al., 2020) of surfaces and particles. Some of the primary findings from these have included demonstrating that the ISS microbiome differs from spacecraft-assembly clean rooms (Be et al., 2017), that different ISS surface areas appear to harbor different microbiomes (Sielaff et al., 2019), and that human skin-associated microorganisms seem to be a primary source for the microbiomes of ISS surfaces (Checinska et al., 2015; Voorhies et al., 2019). A few studies incorporating targeted-amplicon sequencing focusing on astronaut microbiomes have also been published. A study on astronaut salivary microbiomes revealed an increase in alpha-diversity during spaceflight (Urbaniak et al., 2020). Additional work focusing on different astronaut microbiome sources detected varied alpha-diversity responses during spaceflight for feces (increased), skin (mixed), and nasal (decreased) microbiomes, in addition to identifying

[1]Exobiology Branch, NASA Ames Research Center, Mountain View, CA, USA

[2]Blue Marble Space Institute of Science, Seattle, WA, USA

[3]Space Life Sciences, NASA Kennedy Space Center, Merritt Island, FL, USA

[4]Department of Infectious Diseases, J. Craig Venter Institute, Rockville, MD, USA

[5]J. Craig Venter Institute, San Diego, CA, USA

[6]Twitter: @AstrobioMike

[7]Lead contact

*Correspondence:
Mike.Lee@nasa.gov

https://doi.org/10.1016/j.isci.2021.102114

Table 1. Overview of metagenomic samples

Subject ID	# Samples	# Reads* per sample (mean $\pm$ 1 SD)	Megabases (mean $\pm$ 1 SD)
AstB	10	448,828 $\pm$ 60,542	67.3 $\pm$ 9.08
AstC	9	786,374 $\pm$ 559,220	118 $\pm$ 83.9
AstE	9	1,110,925 $\pm$ 567,810	167 $\pm$ 85.2
AstG	9	828,184 $\pm$ 251,201	124 $\pm$ 37.7
AstH	10	702,696 $\pm$ 210,988	105 $\pm$ 31.6

*Count is of total reads remaining after human-read removal.

consistent trends such as an overall decrease in Gammaproteobacteria sequences recovered from skin samples and an overall increase in *Staphylococcus* spp. in nasal samples (Voorhies et al., 2019).

The current work combines newly sequenced and assembled genomes of *Staphylococcus* microbial isolates recovered from the ISS with astronaut nasal microbiome metagenomic data sampled before, during, and after their time aboard the ISS. This is an observational, *ex post facto* integration and exploration of these datasets demonstrating one avenue of leveraging different aspects of NASA's efforts toward tracking microbial communities aboard the ISS.

RESULTS AND DISCUSSION

Fifty-three *Staphylococcus* isolates were recovered from ISS surfaces between 2006 and 2015, sequenced, and their genomes assembled (see Transparent Methods). These included 47 *Staphylococcus epidermidis*, 3 *S. aureus* (all lacking the *mecA* gene indicative of methicillin resistance; Wielders et al., 2002), and 1 each of *Staphylococcus auricularis*, *Staphylococcus lugdunensis*, and *Staphylococcus saprophyticus* as classified by NCBI (Figure S1; Table S1). Metagenomic sequencing was performed on DNA extracted from nasal swabs collected from 5 astronauts at multiple time points before (n = 3 or 4), during (n = 3), and after (n = 3) their time aboard the ISS for 6-month missions (Voorhies et al., 2019; Tables S2; S3).

Although we have the dates of isolation for the isolates, information is not publicly available for dates on specific astronaut missions, leaving us only able to discuss the metagenomic microbiome datasets in terms of relative time (i.e., days before, during, and after flight). Following removal of human reads, the average number of reads per metagenomic sample was 766,906 $\pm$ 421,561 (mean $\pm$ 1 SD; Table 1), which incorporating the read size equates to ~115 $\pm$ 63.2 Mbp. This relatively low coverage precludes doing extensive analyses on any microbial members present (see limitations of the study section), but as presented later can still enable presence or absence detection at the genome level.

We dereplicated the genomes (i.e., chose single representatives for clusters of highly similar genomes) before recruiting the nasal microbiome metagenomic reads (see Transparent Methods). This reduced the total number of ISS-derived reference genomes being used to recruit to from 53 to 7 (Table S4). Read-mapping was carried out for each metagenomic sample to each of the seven dereplicated isolate genomes (Table S5). A minimum detection (proportion of the genome that recruited reads) of at least 20% was employed to filter out sample-to-isolate mappings that were potentially due to non-specific read-recruitment. Two *S. epidermidis* isolates (s29 and s32) and two *S. aureus* isolates (s9 and s42) surpassed this threshold in at least 1 sample in all 5 astronaut datasets (Table S5).

S. epidermidis read-recruitment was relatively consistent across all samples

All 5 astronaut datasets recruited to the s29 and s32 *S. epidermidis* reference genomes relatively consistently across all samples (e.g., Figure 1A and s32 *S. epidermidis* on left side; Figure S2; Table S5). Read-mapping necessarily allows for some variation, however, and based upon the percent identities of the reads recruited to the isolate genomes, all datasets held *S. epidermidis* populations that were more similar to the s32 reference (median percent identity of 98.68%) than the s29 reference (97.97%; Welch t test p = 6 $\times$ 10^{-4}). With the most similar ISS-derived isolate genome being around 98.7% identical to the reads recovered from the astronauts before, during, and after their time aboard the ISS, this serves as an example of consistently present *S. epidermidis* populations in all astronaut nasal microbiomes that are simply similar enough to recruit reads to our reference genomes.
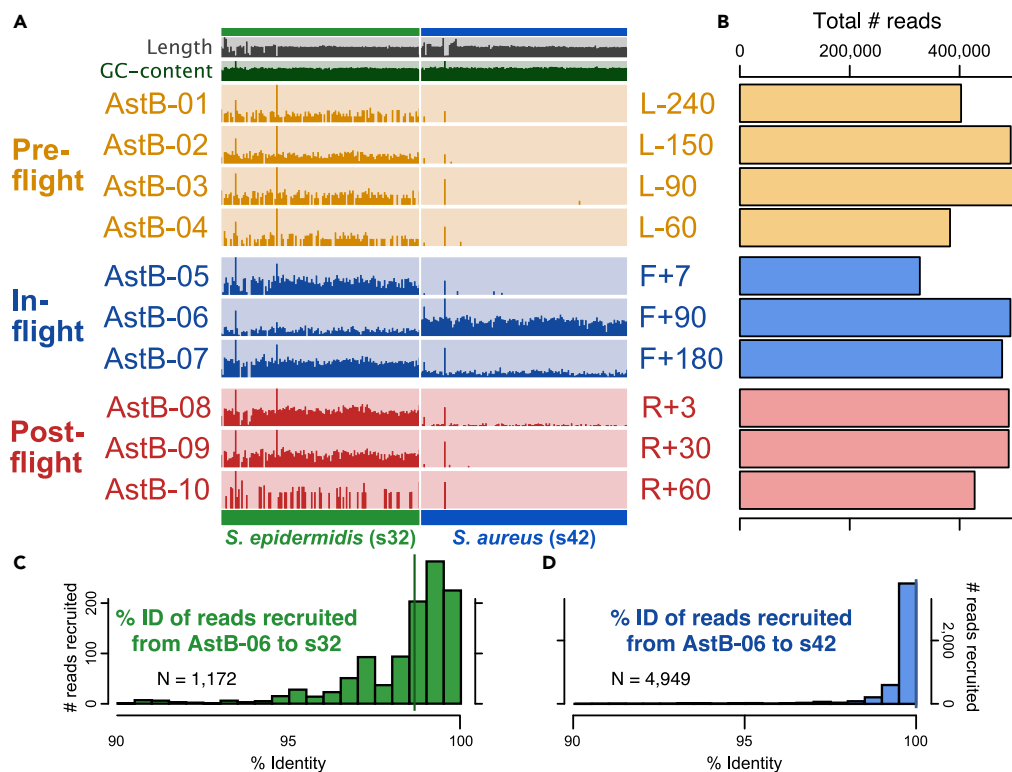
Figure 1. Visualization of metagenomic read-recruitment from astronaut AstB's nasal microbiome through time to genomes of two microbial isolates recovered from surfaces of the ISS

(A) Stretched across the x axis at the bottom are 2 genomes of microbial isolates recovered from the ISS, and the small, vertical columns above them represent information about ~20,000-bp fragments of the genome. The top 2 rows depict length and GC-content of those fragments across the genomes. Following those, each row represents a different time point of astronaut AstB's nasal microbiome, with metagenomic read-recruitment to the ISS-derived isolate genome presented with each fragment's mean coverage normalized to the overall mean coverage for that sample. The green *S. epidermidis* on the left represents what consistent read recruitment from almost all time points looks like. In contrast, reads are recruited consistently across genome *S. aureus* s42 (blue, on right) after 90 and 180 days aboard the ISS, and 3 days after returning to Earth, but not at any other time points. Lone high recruitment peaks are due to highly conserved regions beyond *Staphylococcus* (see Transparent Methods, "metagenomic read-recruitment, analysis, and visualization).

(B) Depicts the total number of reads available from each sample following human-read removal.

(C and D) Histograms depicting the percent identities of the reads recruited from AstB-06 to the s32 (median percent ID of recruited reads of 98.68%) and s42 (median percent ID of 100%) genomes. Vertical bars represent medians.

S. aureus read-recruitment was exclusive to AstB and AstE

S. aureus ISS-derived genomes were only detected in the nasal microbiome datasets AstB and AstE (Table S5). AstE recruited reads across both s9 and s42 *S. aureus* genomes for all time points before, during, and after flight with about the same percent identity (median of 98.67% identity to each; Figures S3–S5; Table S5). As with the aforementioned consistent *S. epidermidis* recruitment, this reveals that the AstE nasal microbiome held a population of *S. aureus* before, during, and after flight that was similar to the ISS-derived *S. aureus* genomes, and we do not see any change during their time aboard the ISS. For AstB, on the other hand, *S. aureus* ISS-derived genomes were only detected during their time aboard the ISS.

AstB recruits systematically to *S. aureus* In-Flight only

Recruitment of metagenomic data from AstB's nasal microbiome through time revealed no recruitment to *S. aureus* in 5 samples spanning from 240 days before launch up to being aboard the ISS for 7 days. However, there was systematic read-recruitment across the entire *S. aureus* s9 and s42 genome when sampled at 90 and 180 days aboard the ISS that remained detectable at 3 days following the astronaut's return to Earth (R+3), but not at 30 or 60 days after returning to Earth (Figure 1). This suggests either that the astronaut's nasal microbiome was colonized by a population similar to the ISS-derived *S. aureus* isolates

while the astronaut was aboard the ISS, followed by subsequent carriage back to Earth, or that its relative abundance was below detection before and after spaceflight.

AstB's In-Flight *S. aureus* population is 100% identical to the ISS-derived isolate s42 *S. aureus* genome

In the case of all nasal microbiomes recruiting to *S. epidermidis*, and all of AstE's samples' reads recruiting to *S. aureus* isolates, the microbial populations that were consistently present before, during, and after flight were about 98.5% similar to the ISS-derived isolate genomes based on percent identity of recruited reads (Figures S2–S5; Tables S5 and S6). Figures 2A and 2B depict what this looks like for AstE's 90-day In-Flight time point ("F+90"), where the percent identity of the reads recovered from the astronaut's nasal *S. aureus* population has a long-tail distribution and median of 98.67% to both *S. aureus* ISS-derived isolate genomes s9 (Figure 2A) and s42 (Figure 2B)—the same is true for all AstE's time points (Figures S4 and S5; Tables S5 and S6). This type of recruitment can be thought of as "off-target" in the sense that there were indeed microbial populations of *S. aureus* present, but the reads recovered from them were only roughly 98.5% identical to the ISS-derived isolate genomes. This can also be seen for the recruitment of AstB to ISS-derived *S. aureus* s9, having a median of 98.7% ID of reads recruited and a similar long-tail distribution (Figure 2C). However, the same sample recruited to ISS-derived *S. aureus* s42 shows the reads recovered have a median percent identity of 100%, without the long-tail distribution (Figure 2D). Based on a phylogenomic tree with NCBI RefSeq complete genomes built with 119 single-copy genes specific to the Firmicutes phylum (Lee, 2019; Data S1), s42 is most closely related to *S. aureus* strain CFSAN082783 (GCF_008330585.1), and the average nucleotide identity between the two is 99.8% (Jain et al., 2018). Multi-locus sequence-typing based on the seven genes in the PubMLST (Jolley and Maiden, 2010) *S. aureus* database categorized both the nearest relative strain, CFSAN082783, and ISS-derived isolate s42 as sequence-type (ST) 45, clonal complex 45.

Conclusion

Here we present evidence at the genomic level of either colonization or enrichment of a microbial population at of *S. aureus* in the nasal microbiome of an astronaut while aboard the ISS—a population that based on read-recruitment is virtually identical to an ISS-derived isolate. These results build on our previously reported amplicon-based characterization of these same nasal microbiome samples that noted an increase in recovered *Staphylococcus* sequences (Voorhies et al., 2019). Upper respiratory symptoms including congestion, rhinitis, and sneezing were some of the most reported incidences in a study of astronauts from 46 long-duration ISS missions (Crucian et al., 2016), and greater abundances of *S. aureus* in the upper respiratory tract have been shown to be associated with several respiratory diseases (e.g., Feazel et al., 2012; Muluk et al., 2018; Ramakrishnan et al., 2013). It is plausible that the increased relative abundances of *S. aureus* in the nasal microbiomes of astronauts while aboard the ISS is a factor in causing some of these symptoms and is worth investigating further.

This is the first report demonstrating colonization or enrichment of an ISS microbial population at the genomic level within an astronaut's nasal microbiome. The lack of detection of the population at 30 days after returning to Earth suggests that this may have been a transient event, although we cannot rule out it may have simply been below detection. This work highlights the value in generating genomic libraries of microbes from built-environments such as the ISS and demonstrates one way that genomic and metagenomic data can be integrated to facilitate efforts to track and monitor astronaut microbiomes—efforts that will become increasingly important as we begin undertaking longer-duration human-spaceflight missions.

Limitations of the study

The average number of metagenomic reads per sample following removal of human-derived reads was $766,906 \pm 421,561$ (mean $\pm$ 1 SD; Table 1). Incorporating read-size this equated to ~115 $\pm$ 63.2 Mbp of possible coverage per sample. To put this in context of thinking about a mixed microbial community, even if there were only one microbial organism with a 5-Mbp genome present in the sampled microbiomes, this would only leave just over ~20× coverage for that one genome. Although genome-level presence/absence detection was still attainable as presented earlier, this relatively low coverage precluded the ability to do extensive analyses on any microbial members present.
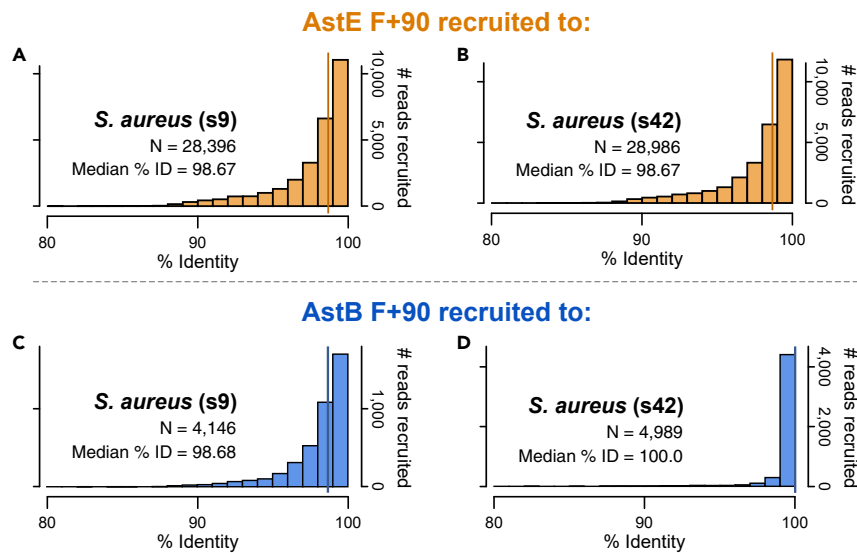
AstE F+90 recruited to:

AstB F+90 recruited to:

Figure 2. Histograms of percent identities of reads recruited to ISS-derived *S. aureus* isolate genomes

(A–D) (A and B) AstE's In-Flight day-90 nasal metagenome reads recruited to the 2 *S. aureus* isolates, showing a median percent identity of 98.67% to both. (C and D) AstB's same time point recruited to the same 2 *S. aureus* isolate genomes as (A and B), showing that unlike the population in AstE's nasal microbiome, the recruited reads have a median percent identity of 100% to one of the isolate references, *S. aureus* s42 (D).

Resource availability

Lead contact

Further information should be directed to and will be fulfilled by the Lead Contact, Michael D. Lee (Mike. Lee@nasa.gov).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Due to IRB considerations, the Astronaut metagenomic data is available upon request from NASA's Life Sciences Data Archive (LSDA) through experiment 1836 (lsda.jsc.nasa.gov/Experiment/exper/1836). The ISS-derived *Staphylococcus* isolate genomes are available through NCBI at BioProject: PRJNA486830 (Table S1) and at our Open-Science Framework repository (OSF; Foster and Deardorff, 2017), project "mr582," which also holds walkthroughs and annotated code for the processing and analyses that were performed (see osf.io/mr582/wiki/).

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2021.102114.

ACKNOWLEDGMENTS

REFERENCES

Avila-Herrera, A., Thissen, J., Urbaniak, C., Be, N.A., Smith, D.J., Karouia, F., Mehta, S., Venkateswaran, K., and Jaing, C. (2020). Crewmember microbiome may influence microbial composition of ISS habitable surfaces. PLoS One 15, 1–20.

Be, N.A., Avila-Herrera, A., Allen, J.E., Singh, N., Checinska Sielaff, A., Jaing, C., and Venkateswaran, K. (2017). Whole metagenome profiles of particulates collected from the International Space Station. Microbiome 5, 81.

Blachowicz, A., Chiang, A.J., Romsdahl, J., Kalkum, M., Wang, C.C.C., and Venkateswaran, K. (2019). Proteomic characterization of Aspergillus fumigatus isolated from air and surfaces of the International Space Station. Fungal Genet. Biol. 124, 39–46.

Checinska, A., Probst, A.J., Vaishampayan, P., White, J.R., Kumar, D., Stepanov, V.G., Fox, G.E., Nilsson, H.R., Pierson, D.L., Perry, J., et al. (2015). Microbiomes of the dust particles collected from the international space station and spacecraft assembly facilities. Microbiome 3, 50.

Crucian, B., Babiak-Vazquez, A., Johnston, S., Pierson, D.L., Ott, C.M., and Sams, C. (2016). Incidence of clinical symptoms during long-duration orbital spaceflight. Int. J. Gen. Med. 9, 383–391.

Feazel, L.M., Robertson, C.E., Ramakrishnan, V.R., and Frank, D.N. (2012). Microbiome complexity and Staphylococcus aureus in chronic rhinosinusitis. Laryngoscope 122, 467–472.

Foster, E.D., and Deardorff, A. (2017). Open science Framework (OSF). J. Med. Libr. Assoc. 105, 203–206.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 1–8.

Jolley, K.A., and Maiden, M.C.J. (2010). BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11, 595.

Knox, B.P., Blachowicz, A., Palmer, J.M., Romsdahl, J., Huttenlocher, A., Wang, C.C.C., Keller, N.P., and Venkateswaran, K. (2016). Characterization of Aspergillus fumigatus isolates from air and surfaces of the international space station. MSphere 1, e00227–16.

Lee, M.D. (2019). GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35, 1–3.

Muluk, N.B., Altın, F., and Cingi, C. (2018). Role of superantigens in allergic inflammation: their relationship to allergic rhinitis, chronic rhinosinusitis, asthma, and atopic dermatitis. Am. J. Rhinol. Allerg. 32, 502–517.

National Academies of Sciences, E., and Medicine, A. (2017). Microbiomes of the Built Environment: A Research Agenda for Indoor Microbiology, Human Health, and Buildings (The National Academies Press).

O'Rourke, A., Lee, M.D., Nierman, W.C., Everroad, R.C., and Dupont, C.L. (2020). Genomic and phenotypic characterization of Burkholderia isolates from the potable water system of the International Space Station. PLoS One 15, e0227152.

Ramakrishnan, V.R., Feazel, L.M., Abrass, L.J., and Frank, D.N. (2013). Prevalence and abundance of Staphylococcus aureus in the middle meatus of patients with chronic rhinosinusitis, nasal polyps, and asthma. Int. Forum Allergy Rhinol. 3, 267–271.

Romsdahl, J., Blachowicz, A., Chiang, A.J., Singh, N., Stajich, J.E., Kalkum, M., Venkateswaran, K., and Wang, C.C.C. (2018). Characterization of Aspergillus niger isolated from the international space station. MSystems 3, 1–13.

Sielaff, A.C., Kumar, R.M., Pal, D., Mayilraj, S., and Venkateswaran, K. (2017). Solibacillus kalamii sp. nov., isolated from a high-efficiency particulate arrestance filter system used in the international space station. Int. J. Syst. Evol. Microbiol. 67, 896–901.

Sielaff, A.C., Urbaniak, C., Babu, G., Mohan, M., Stepanov, V.G., Tran, Q., Wood, J.M., Minich, J., Mcdonald, D., Mayer, T., et al. (2019). Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. Microbiome 7, 50.

Singh, N.K., Wood, J.M., Karouia, F., and Venkateswaran, K. (2018). Succession and persistence of microbial communities and antimicrobial resistance genes associated with International Space Station environmental surfaces. Microbiome 6, 1–23.

Urbaniak, C., Lorenzi, H., Thissen, J., Jaing, C., Crucian, B., Sams, C., Pierson, D., Venkateswaran, K., and Mehta, S. (2020). The influence of spaceflight on the astronaut salivary microbiome and the search for a microbiome biomarker for viral reactivation. Microbiome 8, 1–14.

Voorhies, A.A., and Lorenzi, H.A. (2016). The challenge of maintaining a healthy microbiome during long-duration space missions. Front. Astron. Sp. Sci. 3, 1–7.

Voorhies, A.A., Mark Ott, C., Mehta, S., Pierson, D.L., Crucian, B.E., Feiveson, A., Oubre, C.M., Torralba, M., Moncera, K., Zhang, Y., et al. (2019). Study of the impact of long-duration space missions at the International Space Station on the astronaut microbiome. Sci. Rep. 9, 1–17.

Wielders, C.L.C., Fluit, A.C., Brisse, S., Verhoef, J., and Schmitz, F.J. (2002). mecA gene is widely disseminated in Staphylococcus aureus population. J. Clin. Microbiol. 40, 3970–3975.

Supplemental Information

Reference-guided metagenomics reveals genome-level evidence of potential microbial transmission from the ISS environment to an astronaut's microbiome

Michael D. Lee, Aubrie O'Rourke, Hernan Lorenzi, Brad M. Bebout, Chris L. Dupont, and R. Craig Everroad
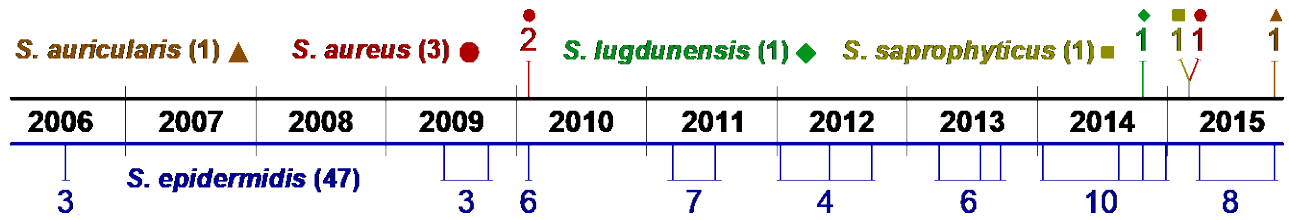
Supplemental Data Items

Figure S1. Isolation timeline, Related to first paragraph of Results and Discussion, and Table S1. The 3 *S. aureus* isolates were isolated only at one sampling timepoint in 2010 (2 of them, s8 and s9), and then one timepoint in 2015 (s42). Data in Table S1.
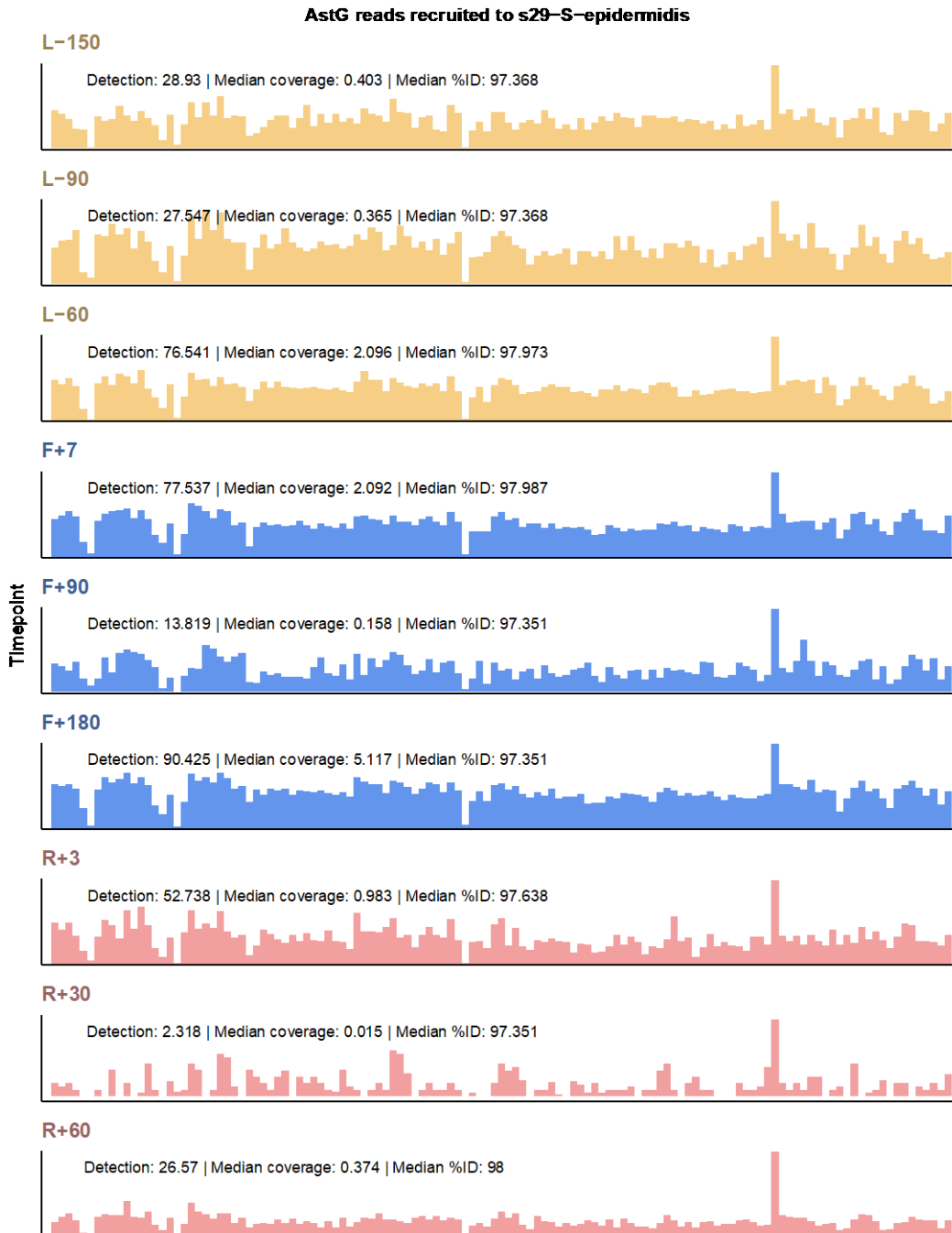
AstG reads recruited to s29−S−epidermidis

L−150
Detection: 28.93 | Median coverage: 0.403 | Median %ID: 97.368

L−90
Detection: 27.547 | Median coverage: 0.365 | Median %ID: 97.368

L−60
Detection: 76.541 | Median coverage: 2.096 | Median %ID: 97.973

F+7
Detection: 77.537 | Median coverage: 2.092 | Median %ID: 97.987

F+90
Detection: 13.819 | Median coverage: 0.158 | Median %ID: 97.351

F+180
Detection: 90.425 | Median coverage: 5.117 | Median %ID: 97.351

R+3
Detection: 52.738 | Median coverage: 0.983 | Median %ID: 97.638

R+30
Detection: 2.318 | Median coverage: 0.015 | Median %ID: 97.351

R+60
Detection: 26.57 | Median coverage: 0.374 | Median %ID: 98

Timepoint

Figure S2. Read recruitment from all AstG timepoints to _S. epidermidis_ s29 isolate genome, Related to Figure 1A. Bars represent 20,000 bp fragments of the genome, so the whole genome spans the x-axis. Proportion of the genome covered ("Detection"), median coverage, and median percent ID of all aligned reads based on BLAST are overlain in text on each. All 5 nasal microbiome datasets looked similar (Table S5).

AstE reads recruited to s42−S−aureus

L−150

Detection: 36.467 | Median coverage: 0.689 | Median %ID: 98.675

L−90

Detection: 26.548 | Median coverage: 0.358 | Median %ID: 98.667

L−60

Detection: 66.332 | Median coverage: 1.383 | Median %ID: 98.675

F+7

Detection: 35.236 | Median coverage: 0.628 | Median %ID: 98.013

F+90

Detection: 53.019 | Median coverage: 1.483 | Median %ID: 98.675

F+180

Detection: 90.294 | Median coverage: 6.122 | Median %ID: 98.675

R+3

Detection: 31.861 | Median coverage: 0.497 | Median %ID: 98.425

R+30

Detection: 15.22 | Median coverage: 0.181 | Median %ID: 98.667

R+60

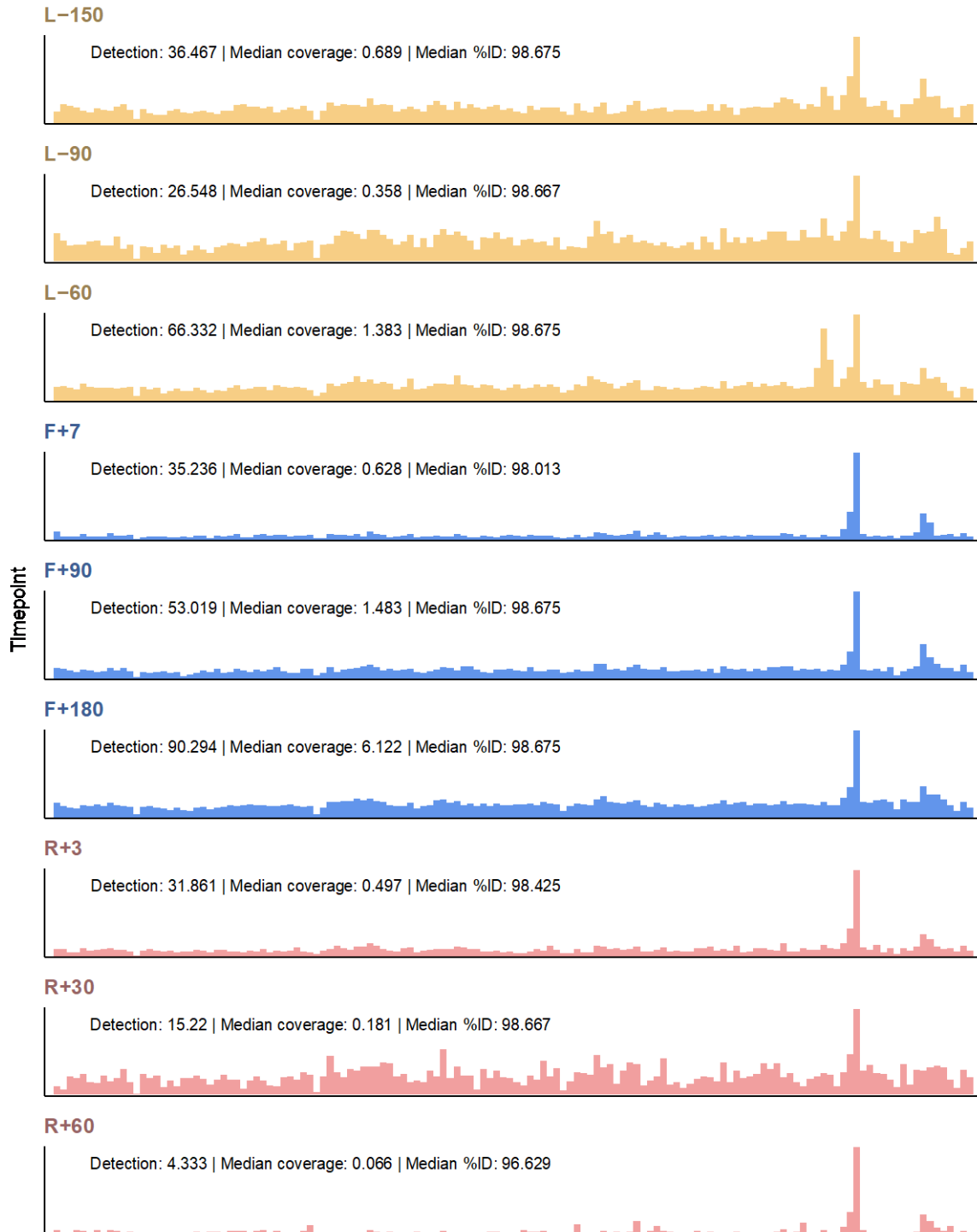Detection: 4.333 | Median coverage: 0.066 | Median %ID: 96.629

Timepoint

Figure S3. Read recruitment from all AstE timepoints to *S. aureus* s42 isolate genome, Related to sections '*S. aureus* read-recruitment was exclusive to AstB and AstE' and '*AstB's In-Flight S. aureus population is 100% identical to the ISS-derived isolate s42 S. aureus genome'.Bars represent 20,000 bp fragments of the genome, so the whole genome spans the x-axis. Proportion of the genome covered ("Detection"), median coverage, and median percent ID of all aligned reads based on BLAST are overlain in text on each. Recruitment of AstE to *S. aureus* s9 looked similar (Table S6).
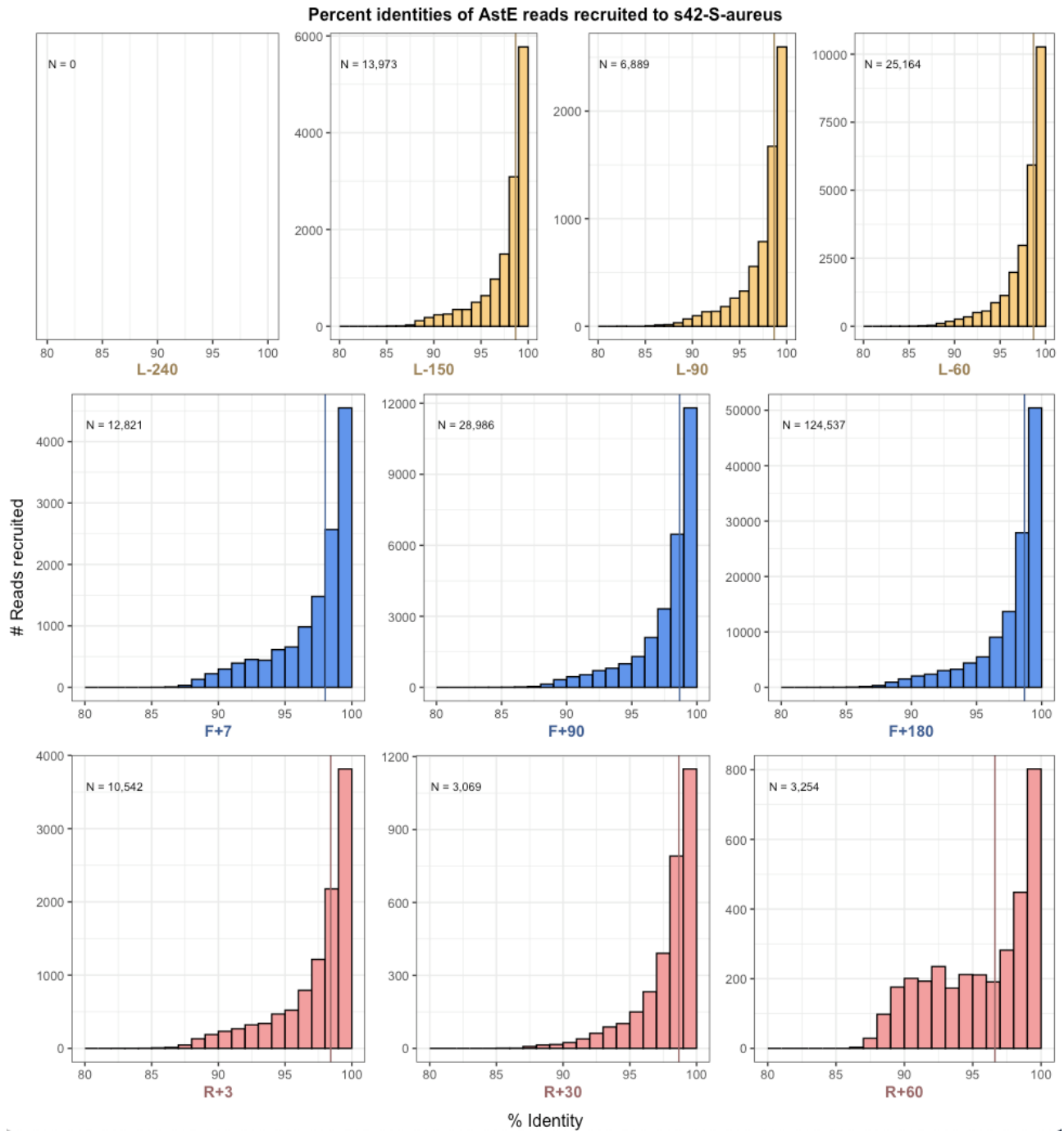
Figure S4. Percent identities of reads recruited from all AstE timepoints to *S. aureus* s42 isolate genome, Related to sections '*S. aureus* read-recruitment was exclusive to AstB and AstE' and '*AstB's In-Flight S. aureus population is 100% identical to the ISS-derived isolate s42 S. aureus genome'. Percent ID is based on BLAST of aligned reads to the genome. AstE to *S. aureus* s9 looked similar (Figure S5; Tables S5 and S6).
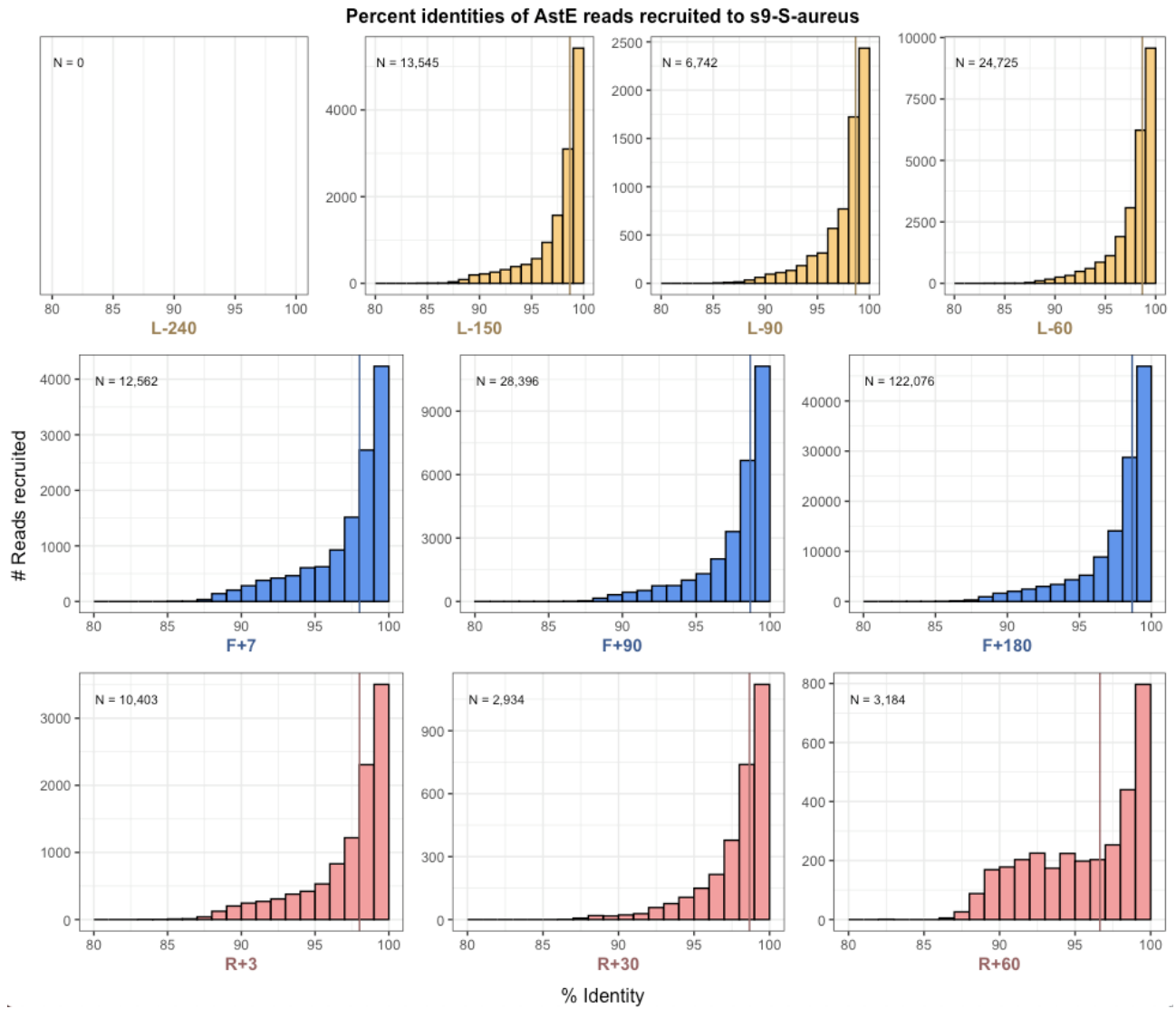
Figure S5. Percent identities of reads recruited from all AstE timepoints to *S. aureus* s9 isolate genome. '*AstB's In-Flight S. aureus population is 100% identical to the ISS-derived isolate s42 S. aureus genome'*. Percent ID is based on BLAST of aligned reads to the genome. AstE to *S. aureus* s42 looked similar (Figure S4; Tables S5 and S6).

Transparent Methods

Organism isolation
Staphylococcus isolates were collected from surfaces of the ISS using the Surface Sampler Kit (https://lsda.jsc.nasa.gov/Hardware/hardw/670) between 2006–2015, isolated on tryptic soy agar (TSA) plates (see Figure S1 and Table S1 for dates of isolation), and stored at -80°C in 15% glycerol at NASA's Johnson Space Center – independent of the current study – as part of NASA's long-term efforts to recover microbial isolates from the ISS (see LSDA experiment 13823; https://lsda.jsc.nasa.gov/Experiment/exper/13823). The authors of the current study (specifically MDL and AO) received these isolates as slant cultures that were sent to the J. Craig Venter Institute (JCVI) in San Diego, California, USA. Isolates were plated on TSA places, incubated overnight at 35°C, and inoculated into 4ml of tryptic soy broth. 2ml were centrifuged to form pellets which were frozen at -80°C until DNA extractions were performed.

Nasal microbiome sampling
The astronaut sampling strategy is originally detailed in Voorhies et al., 2019, and is recapitulated here. Longitudinal samples were collected from 5 astronauts (indicated as AstB, AstC, AstE, AstG, AstH) at 9 or 10 timepoints (Table S2). Sample timepoints are reported as related to: 1) the astronaut's launch date to the ISS (being indicated by "L-" and the number of days prior to launch); 2) the astronaut's time aboard the ISS during their mission (being indicated by "F+" and the number of days aboard the ISS); and 3) the astronaut's return to Earth (being indicated by "R+" and the number of days after returning). Samples were self-collected by the astronauts via nasal swabs which were stored at -100°C aboard the ISS until returned to Earth.

DNA extraction
DNA was extracted from isolates using a phenol/chloroform protocol and quantified using a nanodrop and run on a gel to inspect quality. For the metagenomes, as described in Voorhies *et al.*, 2019, after sample collection, nasal swabs were resuspended in microcentrifuge tubes containing 1.2ml lysis buffer (20mM Tris-Cl, pH 8, 2mM EDTA, 1.2% Triton X-100). Following vortexing at maximum speed for one minute, 1ml of solution was removed and placed in a lysing Matrix B tube (MP Biomedicals CAT # 6911-500). These were vortexed at maximum speed for 5 minutes, then at 10,000 RPM for 1 minute. 700 µl of solution was then incubated for ten minutes at 75°C, allowed to cool, treated with 200mg/ml lysozyme and 20mg/ml Proteinase K, and then phenol/chloroform extractions were performed.

DNA library preparation and sequencing
Microbial-isolate DNA library preparation was performed with the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs) following the manufacturer's instructions, but using half the standard reaction volumes. Metagenomic libraries were prepared with the NexteraXT DNA Library Prep Kit (Illumina), following the manufacturer's specifications. Sequencing for both isolates and metagenomes was performed on Illumina's NextSeq 500 platform with 2x150 paired-end sequencing targeting roughly 1GB of data per sample. Human reads were removed from the metagenomic data using BMTagger (srprism v2.3.17; bmtool v0.0.0; ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/) with human assembly HG18 as the reference.

Isolate sequencing-data preprocessing, assembly, and dereplication
For all isolate sequence datasets, adapters were searched and removed with bcl2fastq2 Conversion Software v2.17 (Illumina). Quality profiles of read pairs were generated with FastQC v0.11.8 (Andrews, 2010), and trimming was performed with Trimmomatic v0.39 (Bolger et al., 2014) using a sliding window setting of 5:20 and filtering out reads shorter than 100 bp after trimming. Assemblies were performed using SPAdes v3.12.0 (Bankevich et al., 2012) with default settings. Assemblies were summarized with bit v1.8.06 (Lee, 2018) and estimated completion and redundancy were calculated with checkm v1.1.2 (Parks et al., 2015). Assembly summary information is presented in Table S1. The mecA gene was searched in *S. aureus* isolates via BLASTp (Altschul et al., 1990) using UniProtKB (Consortium, 2017) protein Q7DHH4. Isolate assemblies were dereplicated with dRep v2.6.2 (Olm et al., 2017) prior to initial mapping.

Metagenomic read-recruitment, analysis, and visualization
Metagenomic reads were recruited to individual isolate genomes with bowtie2 v2.4.1 (Langmead and Salzberg, 2012) with default settings. Sam files were converted to bam files and sorted and indexed with samtools v1.7 (Li et al., 2009), and bedtools v2.29.2 (Quinlan and Hall, 2010) was utilized to help calculate detection (portion of reference that recruited reads). Reads that successfully mapped were then aligned with

BLAST v2.5.0 (Altschul et al., 1990). Snakemake v5.19.1 (Köster and Rahmann, 2012) was used for much of the processing performed (see OSF osf.io/mr582/wiki/). Analyses were performed and some visualizations generated using R v3.6.3 (R Core Team, 2017) in RStudio v1.1.456 (RStudio Team, 2020), making heavy use of the tidyverse v1.3.0 (Wickham et al., 2019) and ggpubr v0.4.0 (Kassambara, 2020) packages. Read-recruitment across the genomes was also visualized with anvi'o v6.2 (Eren et al., 2015). For Figure 1, read-mapping was performed to a combined index of both *S. epidermidis* s32 and *S. aureus* s42.

The large peak in *S. aureus* s42 in Figure 1A is due to non-specific read recruitment to a highly conserved DNA sequence. The sequence responsible for its recruitment is:

>identical_sequence_between_s32_and_s42
AGCTATGCCGTTGGCACGACAACTGGTACACCAGAGGTATGTCCATCCCGGTCCTCTCGTACTAAGGAC
AGCTCCTCTCAAATTTCCTACGCCCACGACGGATAGGGACCGAACTGTCTCACGACGTTCTGAACCCAG
CTCGCGTACCGCTTTAATGGGCGAACAGCCCAACCCTTGGGACCGACTACAGCCCCAGGATGCGATGA
GCCGACATCGAGGTGCCAAACCTCCCCGTCGATGTGAACTCTTGGGGGAGATAAGCCTGTTATCCCCG
GGGTAGCTTTTATCCGTTGAGCGATGGCCCTTCCATGCGGAACCACCGGATCACTAAGTCCGTCTTTCG
ACCCTGCTCGACTTGTAGGTCTCGCAGTCAAGCTCCCTTATGCCTTTACACTCTATGAATGATTTCCAAC
CATTCTGAGGGAACCTTTGAGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAACTGCCCG
CCTGACACTGTCTCCCACC

This is identical to the central high peak of read recruitment seen in *S. epidermidis* s32 also. A BLASTN search yields many *Staphylococcus* with 100% identical, complete alignments, and if excluding the *Staphylococcus* genus, there are many different organisms with near 100% identity and full alignments. A BLASTX search yields hypothetical, uncharacterized, and conserved proteins, RNAases, and cell-wall associated hydrolases from many different organisms.

Multi-locus sequence-typing
mlst v2.19.0 (Seemann) was utilized to perform multi-locus sequence-typing of the *S. aureus* assemblies based on PubMLST (Jolley and Maiden, 2010).

Phylogenomic tree construction and average nucleotide identity calculation
GToTree v1.5.36 (Lee, 2019) was utilized to make a phylogenomic tree of ISS-derived isolates and those available through NCBI's RefSeq (Leary et al., 2016) using a single-copy gene-set of 119 target genes specific for bacteria of the phylum Firmicutes. GToTree by default downloads all reference genomes, identifies coding sequences as needed in all input genomes with prodigal v2.6.3 (Hyatt et al., 2012), scans amino-acid sequences for the target genes with HMMER3 v3.3.1 (Eddy, 2011), aligns each gene-set with Muscle v3.8.1551 (Edgar, 2004), trims the alignments with trimal v1.4.rev15 (Capella-Gutiérrez et al., 2009), concatenates all gene-set alignments, swaps labels for taxonomic information with TaxonKit v0.6.0 (Shen and Xiong, 2019), and then uses FastTree2 v2.1.10 (Price et al., 2010) for phylogenetic estimation. Average nucleotide identity was calculated with fastANI v1.32 (Jain et al., 2018).

General
Conda (Anaconda-Team, 2016) and Snakemake (Köster and Rahmann, 2012) were utilized throughout this work.

Data and code availability
Due to IRB considerations, the Astronaut metagenomic data is available upon request from NASA's Life Sciences Data Archive (LSDA) through experiment 1836 (lsda.jsc.nasa.gov/Experiment/exper/1836). The ISS-derived *Staphylococcus* isolate genomes are available through NCBI under project accession PRJNA486830 (Table S1) and at our Open-Science Framework repository (OSF; Foster and Deardorff, 2017), project "mr582", which also holds walkthroughs and annotated code for the processing and analyses that were performed (see osf.io/mr582/wiki/).

Supplemental References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool.

J. Mol. Biol. *215*, 403–410.

Anaconda-Team (2016). Anaconda Software Distribution.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data.

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., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

Consortium, T.U. (2017). UniProt : the universal protein knowledgebase. *45*, 158–169.

Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. *7*.

Edgar, R.C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 1–19.

Eren, A.M., Esen, C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015). Anvi ' o : an advanced analysis and visualization platform for ' omics data. 1–29.

Foster, E.D., and Deardorff, A. (2017). Open Science Framework (OSF). J. Med. Libr. Assoc. *105*, 203–206.

Hyatt, D., Locascio, P.F., Hauser, L.J., and Uberbacher, E.C. (2012). Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics *28*, 2223–2230.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. *9*, 1–8.

Jolley, K.A., and Maiden, M.C.J. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics *11*, 595.

Kassambara, A. (2020). ggpubr: "ggplot2" based publication-ready plots.

Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520–2522.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Leary, N.A.O., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., Mcveigh, R., Rajput, B., Robbertse, B., Smith-white, B., Ako-adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI : current status , taxonomic expansion , and functional annotation. *44*, 733–745.

Lee, M.D. (2018). Bioinformatics Tools (bit). doi:10.5281/zenodo.3383647

Lee, M.D. (2019). GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 1–3.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. *11*, 2864–2868.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM : assessing the quality of microbial genomes recovered from isolates , single cells , and metagenomes. 1043–1055.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *5*.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Seemann, T. mlst. https://github.com/tseemann/mlst

Shen, W., and Xiong, J. (2019). TaxonKit: a cross-platform and efficient NCBI taxonomy toolkit. BioRxiv 513523.

Team, R.C. (2017). R: A language and environment for statiscal computing.

Team, Rs. (2020). RStudio: Integrated development for R.

Voorhies, A.A., Mark Ott, C., Mehta, S., Pierson, D.L., Crucian, B.E., Feiveson, A., Oubre, C.M., Torralba, M., Moncera, K., Zhang, Y., et al. (2019). Study of the impact of long-duration space missions at the International Space Station on the astronaut microbiome. Sci. Rep. *9*, 1–17.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. J. Open Source Softw. *4*, 1686.