Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Culture-benchmarked metagenomics maps microbial strain sharing in long-term acute care

Michael Woodworth

mwoodwo@emory.edu

Emory University School of Medicine

Victoria Chu

UCSF https://orcid.org/0000-0002-7480-965X

Abigail Glascock

UCSF

Ahmed Babiker

Emory University

Deepti Suchindran

Emory University https://orcid.org/0009-0002-5948-0515

Amanda Ashley

Emory University

Danielle Barrios Steed

UCSD

Chris Bower

Emory

Alex Page

Emory

Sujan Reddy

US CDC

Alison Laufer Halpin

Centers for Disease Control and Prevention https://orcid.org/0000-0003-1643-1617

Maroya Walters

Centers for Disease Control and Prevention

Julia Van Riel

Emory University

Eileen Burd

Emory University School of Medicine

Sarah Lohsen

Emory University

Sarah Satola

Emory Antibiotic Resistance Center

Charles Langelier

University of California, San Francisco https://orcid.org/0000-0002-6708-4646

Article

Keywords:

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11209504/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Culture-benchmarked metagenomics maps microbial strain sharing in long-term acute care

Authors: Victoria T. Chu[1]*, Abigail L. Glascock[2,3]*, Ahmed Babiker[4], Deepti Suchindran[4], Amanda Ashley[4], Danielle Barrios Steed[5], Chris W. Bower[4], Alex Page[4], Sujan C. Reddy[6,7], Alison Laufer Halpin[6,7], Maroya Spalding Walters[6,7], Julia Van Riel[4], Eileen M. Burd[4,8], Sarah Lohsen[4], Sarah W. Satola[4], Charles R. Langelier[2,3]*, Michael H. Woodworth[4]*†

*These authors contributed equally to this work

†Corresponding author: Michael H. Woodworth. Email: mwoodwo@emory.edu

Author Affiliations:

[1]Department of Pediatrics, Division of Infectious Diseases and Global Health, University of California, San Francisco, San Francisco, CA, United States

[2]Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, San Francisco, CA, United States

[3]Chan Zuckerberg Biohub, San Francisco, CA, United States

[4]Department of Medicine, Division of Infectious Diseases, Emory University School of Medicine, Atlanta, GA, United States

[5]Department of Medicine, Division of Infectious Diseases and Global Public Health, University of California San Diego, La Jolla, CA United States

[6]Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention (CDC), Atlanta, GA, United States

[7]U.S. Public Health Service, Department of Health and Human Services, Rockville, MD, United States

[8]Department of Pathology, Emory University School of Medicine, Atlanta, GA, United States

Abstract

Long-term acute care hospitals (LTACHs) are major reservoirs for multidrug-resistant organisms (MDROs), yet conventional surveillance methods based on culture and targeted sequencing incompletely characterize microbial transmission. Metagenomic next-generation sequencing (mNGS) offers the potential to capture strain-level transmission dynamics across the broader microbiome, providing a more comprehensive view of healthcare-associated microbial spread. We conducted a prospective observational cohort study of 49 LTACH patients from April to July 2023. All LTACH patients provided perirectal samples on Weeks 1 and 6; a subset of patients who received fecal microbiota transplantation (FMT) submitted enhanced weekly sampling for a six-week period at the time of and after FMT administration. Samples underwent mNGS and were selectively cultured for MRDOs, with cultured isolates undergoing whole genome sequencing (WGS). We developed an open-source computational pipeline, MAG-based Strain Profiling and Identification (MAGSPI), which reconstructs metagenome-assembled genomes (MAGs), collapses them into species-level genome bins (SGBs), and enables strain-level comparisons using inStrain. WGS-defined strain assignments served as the reference standard for optimization of strain-sharing thresholds. Among 93 perirectal mNGS samples and 22 donor stool samples, MAGSPI recovered 3,509 MAGs and 710 unique SGBs spanning 319 genera and 10 bacterial phyla. InStrain identification thresholds were optimized using WGS-defined strain assignments as the reference standards. Across all pairwise comparisons, 62,316 strain-sharing events involving 661 SGBs were identified, revealing extensive microbial sharing throughout the facility. Samples from the same individual demonstrated significantly higher strain-sharing rates than samples from different individuals, supporting longitudinal strain persistence. Network analyses identified widespread interpatient strain sharing, including evidence of donor-derived strain engraftment and subsequent dissemination beyond FMT recipients. In longitudinal risk-factor analyses, increased interpatient staff crossings of nurses and respiratory therapists were associated with higher strain-sharing rates between patients over time. We demonstrate that strain-resolved mNGS surveillance provides a culture-independent approach to capture transmission events not readily identified by traditional culture-based methods. Integration of strain-sharing networks with healthcare workflow data can identify potential transmission pathways and modifiable risk factors. These findings support the use of metagenomic surveillance as a complementary tool for infection prevention, outbreak detection, and targeted interventions in LTACHs and other high-risk healthcare environments.

Introduction

Multidrug-resistant organisms (MDROs) are a major public health threat, contributing to substantial morbidity, mortality, and healthcare costs[1,2]. Healthcare settings are associated with increased risk of acquiring and transmitting multidrug-resistant organisms (MDROs), which can be spread via contaminated hands[3,4], medical equipment, and environmental surfaces[5,6]. Patients in long-term acute care hospitals (LTACHs) are particularly vulnerable to MDRO acquisition, with prevalence reaching 80% in some studies[7]. This vulnerability is driven by multiple risk factors including prolonged hospitalization, extensive antimicrobial exposure, immunocompromising conditions, active wounds, and the presence of invasive devices such as endotracheal tubes or indwelling urinary catheters[8-10]. Consequently, LTACHs serve as key reservoirs of MDRO transmission both within individual healthcare facilities and across healthcare networks through frequent interfacility patient transfers[10,11].

Fecal microbiota transplantation (FMT) is a proven microbe-based treatment approach for recurrent *C. difficile* infections[12,13]. Recent work has also demonstrated the potential of FMT to reduce MDRO colonization and subsequent infection risk in high-risk populations[14-16]. Existing studies, however, have focused almost exclusively on individual patient outcomes, and it remains unclear whether FMT can serve as a tool for controlling MDRO transmission within healthcare settings.

An important barrier to understanding MDRO transmission dynamics in healthcare settings is the limitation of traditional surveillance methods, which incompletely capture asymptomatic carriage of resistant organisms. Culture-based approaches and targeted PCR assays, even when supplemented by whole genome sequencing (WGS), are restricted to select organisms, underestimate overall colonization burden, and fail to fully capture transmission events[17,18]. These approaches also fail to capture broader ecological changes within the microbiome that

may influence MDRO colonization resistance and thus strain acquisition and dissemination within healthcare settings. Recent advances in metagenomic next-generation sequencing (mNGS), particularly metagenome assembly, may help overcome these limitations by enabling culture-independent, strain-level assessment of community composition[19,20], antimicrobial resistance, and genomic relatedness within a single workflow[21].

Recent studies have demonstrated the utility of metagenome-assembled genome (MAG) and species-level genome bin (SGB) approaches for tracking bacterial transmission within households[19] and following FMT[22,23]. However, few studies have evaluated these methods for longitudinal surveillance of transmission dynamics within healthcare environments, and none have focused on LTACHs. To address this gap, we performed multimodal epidemiologic, WGS, and metagenomic analyses of microbial colonization among LTACH patients over time to assess the utility of mNGS for characterizing transmission dynamics in a healthcare setting.

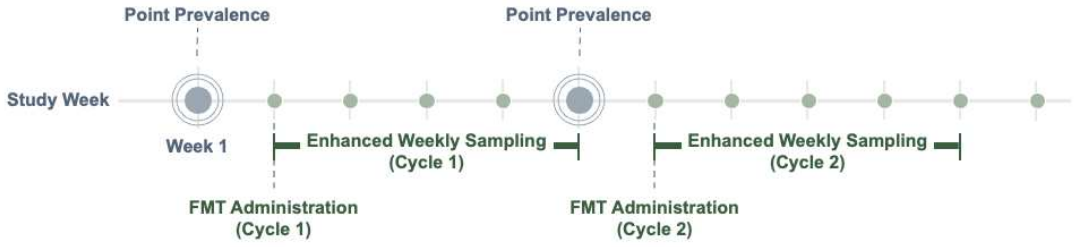
Results

Study enrollment

We carried out a prospective observational study of 49 LTACH patients from April -July 2023. The median age of study participants was 63 years (IQR: 54-70 years old) and 27 (55%) were male; 27 (55%) were African American or Black, 11 (22%) were White, 4 (8%) other, and 7 (14%) were an unknown race; 1 (2%) was of Hispanic ethnicity.

Point prevalence surveys consisting of swab samples were collected from three anatomical sites (inguinal, perirectal, or stool) and one environmental site (high-touch room surface) for all participants were conducted during Weeks 1 and 6 of the study period (Figure 1). Ten participants with MDRO-positive cultures additionally received FMT during either Cycle 1 (Week

2) or Cycle 2 (Week 7) and underwent longitudinal sampling before FMT and then weekly for the subsequent five weeks, enabling evaluation of donor strain engraftment and intra-facility transmission. One participant received FMT during both cycles because of persistent MDRO colonization. Minor deviations from the sampling schedule included one delayed Week 11 collection (performed during Week 12) and one participant who received FMT during Week 8 rather than Week 7, extending follow-up through Week 12. At Week 1, 97 samples were collected from 30 participants; at Week 6, 114 samples were collected from 35 participants. Over the 12-week study period, 313 samples were collected among the 49 participants: 102 (33%) high-touch surfaces, 108 (34%) perirectal, 65 (21%) inguinal, and 38 (12%) stool samples. All swabs were selectively cultured to identify MDROs; of the 108 perirectal samples, 93 (86%) were available and additionally underwent mNGS[23].



Study Week	1	2	3	4	5	6	7	8	9	10	11	12	Total
Participants	32	6	6	5	5	35	4	5	4	4	4	2	49
Samples Collected	97	17	15	14	10	114	8	12	8	8	7	3	313
Anatomical Samples	67	11	10	9	7	81	4	7	4	4	5	2	211
Environmental Samples	30	6	5	5	3	33	4	5	4	4	2	1	102
Samples sent for mNGS	23	6	6	5	4	28	4	5	4	3	3	2	93

Figure 1. Study design and sampling time points. Point-prevalence surveys were conducted during study weeks 1 and 6 (April–July 2023) among all verbally consented patients (n=49) admitted to a single academically affiliated long-term acute care hospital (LTACH) in Atlanta, Georgia, U.S. Ten patients who received FMT (administered as a single dose during either Week 2 or Week 7) provided samples before FMT and then underwent enhanced weekly sampling for the subsequent five weeks. Participant numbers varied over time because of patient admissions and discharges throughout the study period.

Whole genome sequencing of MDRO isolates

WGS was performed on 344 MDRO isolates from the collected swabs. The most prevalent cultured MDROs included *Enterococcus faecium* (n=99), *Pseudomonas aeruginosa* (n=51), *Enterobacter cloacae* complex (n=41), *Klebsiella pneumoniae* (n=43), and *Clostridioides difficile* (n=29) (Supplementary Table 1). Among the 344 isolates, 107 distinct bacterial strains representing 15 bacterial species were identified (Supplementary Figure 1). Using SNVPhyl-defined strain assignments of MDRO culture isolates, we identified 37 possible transmission events across 28 bacterial strain clusters, involving 28 unique patients and assuming a single facility introduction per cluster.

MAGSPI enables strain-level metagenomic MDRO surveillance

We performed mNGS sequencing on 115 samples from 51 individuals; 93 perirectal samples from 49 LTACH patients (1-7 samples per patient), and 22 stool samples from 2 donors used for FMT administration (donor 1: 12 samples, donor 2: 10 samples). mNGS was not performed on point prevalence perirectal samples from patients who received FMT; instead, baseline microbiome data were derived from samples collected during the week of FMT administration. Across samples with mNGS data, the median total reads per sample was 48.0 million (IQR: 38.3–83.7).

To enable strain-level tracking using metagenomic data, we developed an open-source computational pipeline called MAGSPI (MAG-based Strain Profiling and Identification) which integrates and harmonizes several existing, distinct genomic tools (Figure 2a)[24]. MAGSPI takes metagenomic sequencing data as input, generates high-quality MAGs using a tripartite algorithmic approach, collapses these MAGs to species-level genome bins (SGBs) and then uses the resulting high-quality SGB database to align and compare strains across population-level data. We applied our MAGSPI pipeline to our 115 peri-rectal and donor stool metagenomic

samples. In total, 3,509 MAGs were recovered; 3,389 (96.6%) were classified as bacteria, 70 (2%) were unclassified bacteria, 36 (1%) were classified as archaea, 1 (0.03%) was unclassified archaea and 13 (0.4%) were unclassified after mapping to the Genome Taxonomy Database (GTDB).

A median of 28 distinct MAGs was recovered per sample (IQR: 12–50). Of the 3,509 MAGs, 2,329 (66%) were ≥90% complete and 2,927 (83%) had <500 contiguous sequences (contigs) (Supplementary Figure 2a and 2b), and a median size of 2.4 Mb (IQR: 2.0–3.1) (Supplementary Table 2). We found that the total number of MAGs recovered per sample (richness) were independently correlated with sequencing depth (total reads) and microbiome alpha diversity (Shannon Diversity Index, SDI). Each one-unit increase in SDI was associated with a 40% increase in expected MAG richness, while sequencing yield showed a more modest but significant effect of a 0.8% increase per additional 10 million reads (Supplementary Figure 2c).

Among the 3,509 MAGs, genes encoding multidrug efflux pumps were the most common class of ARGs detected (n=1,321 ARGs), followed distantly by lipopeptide (n=357) and aminoglycoside (n=271) resistance genes (Supplementary Figure 3a). Among these ARGs, we noted several of distinct public health importance, including carbapenemase genes (bla_{KPC}), extended spectrum beta-lactamase genes ($bla_{CTX\text{-}M}$, bla_{SHV}), vancomycin resistance genes (*vanA*, *vanB*), and plasmid-mediated fluoroquinolone genes (*qnr*) (Supplementary Figure 3b).

Next, we dereplicated the high-quality MAGs at 97% average nucleotide identity (ANI) to define species-level genome bins (SGBs). Dereplication is a process that clusters nearly identical genomes together and reduces them down to a single, high-quality representative genome. Across the 710 unique SGBs detected, 609 unique species were identified from 319 different genera and 10 different bacterial phyla; 6 SGBs lacked taxonomic classification (Supplementary

Figure 4). Per-sample SGB richness increased linearly with MAG richness (Figure 2b). Most samples had a ≥80% MAG to SGB recovery rate (percentage of assembled MAGs meeting quality thresholds for species-level classification), with a median recovery rate of 89% (IQR: 83–94%) from all samples (Figure 2c). Recovery rates were comparable between stool and perirectal samples (Wilcoxon rank-sum p=0.41, Figure 2d). The majority of SGBs were only

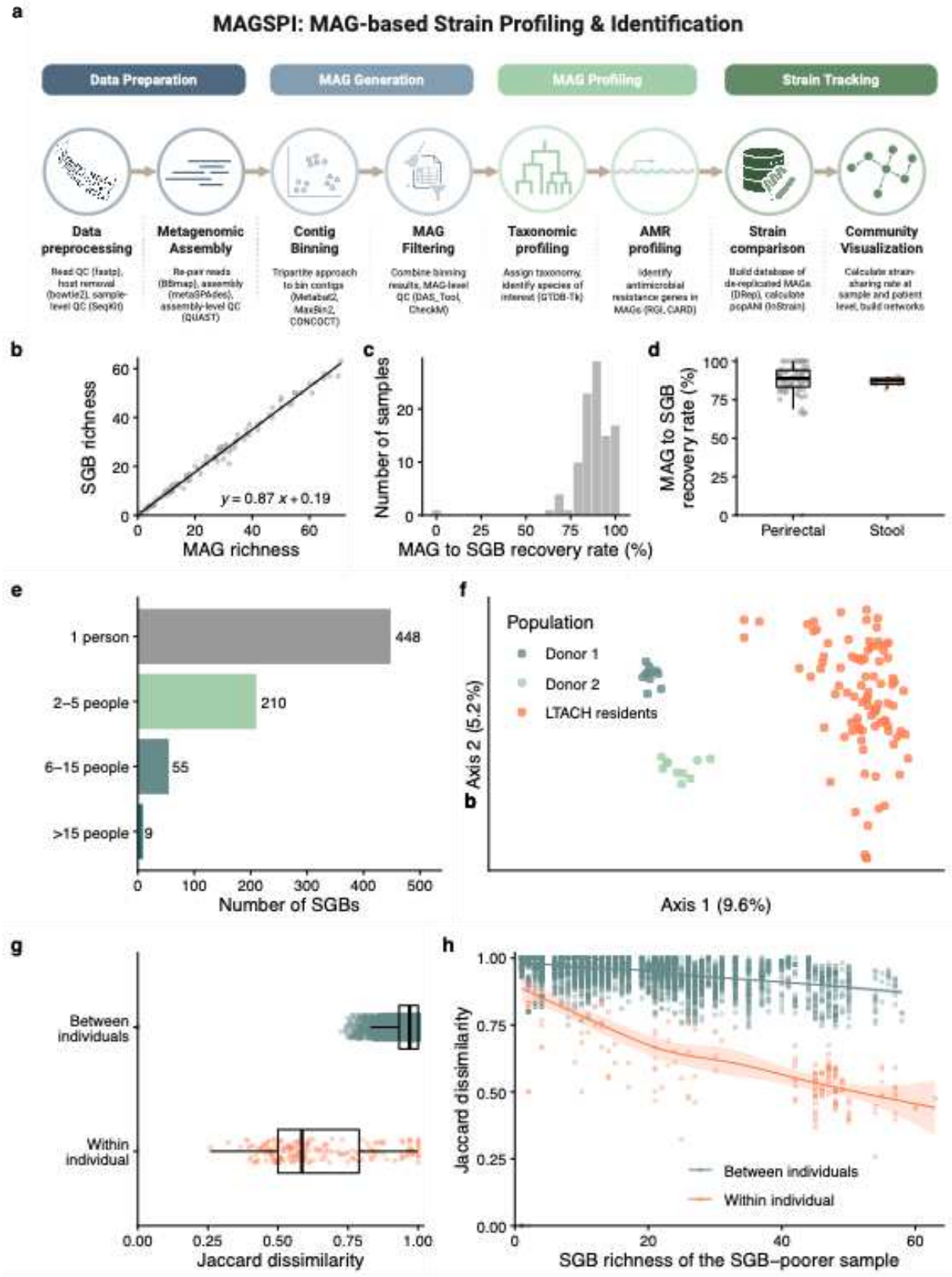

a
MAGSPI: MAG-based Strain Profiling & Identification
Data Preparation
MAG Generation
MAG Profiling
Strain Tracking
Data preprocessing
Read QC (fastp), host removal (bowtie2), sample-level QC (SeqKit)
Metagenomic Assembly
Re-pair reads (BBmap), assembly (metaSPAdes), assembly-level QC (QUAST)
Contig Binning
Tripartite approach to bin contigs (Metabat2, MaxBin2, CONCOCT)
MAG Filtering
Combine binning results, MAG-level QC (DAS_Tool, CheckM)
Taxonomic profiling
Assign taxonomy, identify species of interest (GTDB-Tk)
AMR profiling
Identify antimicrobial resistance genes in MAGs (RGI, CARD)
Strain comparison
Build database of de-replicated MAGs (DRep), calculate popANI (inStrain)
Community Visualization
Calculate strain-sharing rate at sample and patient level, build networks
b
SGB richness
MAG richness
y = 0.87 x + 0.19
c
Number of samples
MAG to SGB recovery rate (%)
d
MAG to SGB recovery rate (%)
Perirectal
Stool
e
1 person
2–5 people
6–15 people
>15 people
448
210
55
9
Number of SGBs
f
Population
Donor 1
Donor 2
LTACH residents
Axis 2 (5.2%)
Axis 1 (9.6%)
g
Between individuals
Within individual
Jaccard dissimilarity
h
Jaccard dissimilarity
Between individuals
Within individual
SGB richness of the SGB–poorer sample

Figure 2. Overview of MAG-based Strain Profiling and Identification (MAGSPI) pipeline (a) and species-level genome bins (SGBs) detection and composition (b-h) from metagenomic next-generation sequencing (mNGS) data. The analysis included 22 stool samples from 2 fecal microbiota transplantation (FMT) donors and 93 perirectal samples from 49 patients at a single long-term acute care hospital (LTACH) in Atlanta, Georgia between April and July 2023. **a** Overview of the MAGSPI pipeline. **b** Linear regression (black line) fit of the relationship between MAG richness to SGB richness. **c** Distribution and frequency of MAG to SGB recovery rate (percentage of assembled MAGs meeting quality thresholds for species-level classification) per sample **d** Distribution and frequency of MAG to SGB recovery rate by sample type. Boxplot elements include a center line (median), box limits (upper and lower quartiles), whiskers (1.5x interquartile range). **e** Distribution of SGBs shared across participants. The bars indicate the different levels of SGB sharing: no transmission (1 person), low (2-5 people), moderate (6-15 people), or high (>15 people) transmission. Numbers on the right indicate the number of SGBs in each category. **f** SGB microbiome composition overview of all samples from the 49 LTACH patients and the 2 donors (principal coordinate analysis of Jaccard dissimilarity. n=110 samples). **g** Pairwise comparison of SGB microbiome composition differences, measured by Jaccard dissimilarity, from samples between and within individuals. Boxplot elements include a center line (median), box limits (upper and lower quartiles), whiskers (1.5x interquartile range). **h** Pairwise comparison of SGB microbiome composition differences, measured by Jaccard dissimilarity, as a distribution by the number of SGBs detected in the SGB-poorer sample of the pair. Smoothed trend lines (LOESS) with 95% confidence intervals show the relationship between SGB richness and microbiome composition differences for each comparison group.

detected in a single individual (n=436, 61%), while an additional 30% (n=210) were detected in 2 to 5 individuals (Figure 2e). Among the 64 SGBs detected in ≥6 individuals, the majority were non-pathogenic gut commensals; however, bacterial species with pathogenic potential were also identified (e.g., *E. faecalis*, *E. faecium*, *E. coli*, *K. pneumoniae*, *P. mirabilis*, *P. aeruginosa*) (Supplementary Figure 5).

We noted compositional divergence of SGBs between the LTACH patients, and FMT donor 1 and donor 2 (Figure 2f). Samples from the same donor clustered together and most individuals with longitudinal samples demonstrated a compositional shift over time, although there were some exceptions (e.g., P04 and P32) (Supplementary Figure 6). SGB microbial composition demonstrated greater variation (measured by Jaccard Index) between individuals than within individuals (Figure 2g) and also demonstrated greater variation with increasing SGB richness (Figure 2h). Notably, within individual comparisons consistently exhibited higher Jaccard similarity than between individual comparisons across all richness levels, demonstrating that individuals maintain distinctive SGB microbiome signatures.

Determining strain thresholds using SGBs

We next used inStrain compare[25] to perform pairwise comparisons of all WGS and mNGS data. For each comparison, we calculated population average nucleotide identity (popANI), which accounts for minor alleles within each sample and considers a nucleotide position as divergent only when no shared allele exists. In principle, this attribute of popANI affords the sensitivity needed to distinguish shared strains from closely related but distinct ones. Of the pairwise comparisons that met the defined minimum criteria (median popANI ≥0.99995, ≥0.50 breadth, ≥0.80 shared scaffold proportion, and ≥10 shared scaffolds), 99% had a median popANI of 1.00, while the distribution of breadth and shared scaffold proportion were more varied (breadth: median 98, IQR 92–100; shared scaffold proportion: median 93%, IQR 86-100) (Supplementary Figure 7a, 7b). We initially found that this approach, however, frequently combined multiple different WGS-confirmed strains (based on SNVPhyl[26]) together, labeling them incorrectly as a single strain (Supplementary Figure 7c).

We thus evaluated the impact of varying inStrain thresholds (popANI, breadth and shared scaffold proportion) on sensitivity and specificity. Sensitivity was defined by WGS strain consistency (inStrain cluster contains all established WGS strains) and longitudinal consistency (inStrain cluster contains all WGS collected longitudinally from the same patient). Specificity was defined by WGS strain purity (inStrain cluster contains sequences from only a single SNVPhyl-defined strain). WGS strain consistency was 100% for all inStrain threshold combinations except when the shared scaffold proportion was 1.00; thus, we relied mainly on longitudinal consistency and WGS strain purity as the markers of threshold performance (Supplementary Table 3, Figure 3a). As this study was intended to evaluate transmission pathways and it is feasible that an individual could acquire new strains over time, we prioritized specificity over sensitivity in the final strain thresholds. Optimal thresholds were ultimately defined as median popANI ≥0.99995, ≥0.50 breadth, and ≥0.95 shared scaffold proportion, and

required ≥10 shared scaffolds (Figure 3b). We then applied these thresholds to define final network transmission maps for *C. difficile*, *E. hormaechei*, *E. faecalis*, *E. faecium*, *E. coli*, and *K. pnuemoniae* (Supplementary Figure 8).

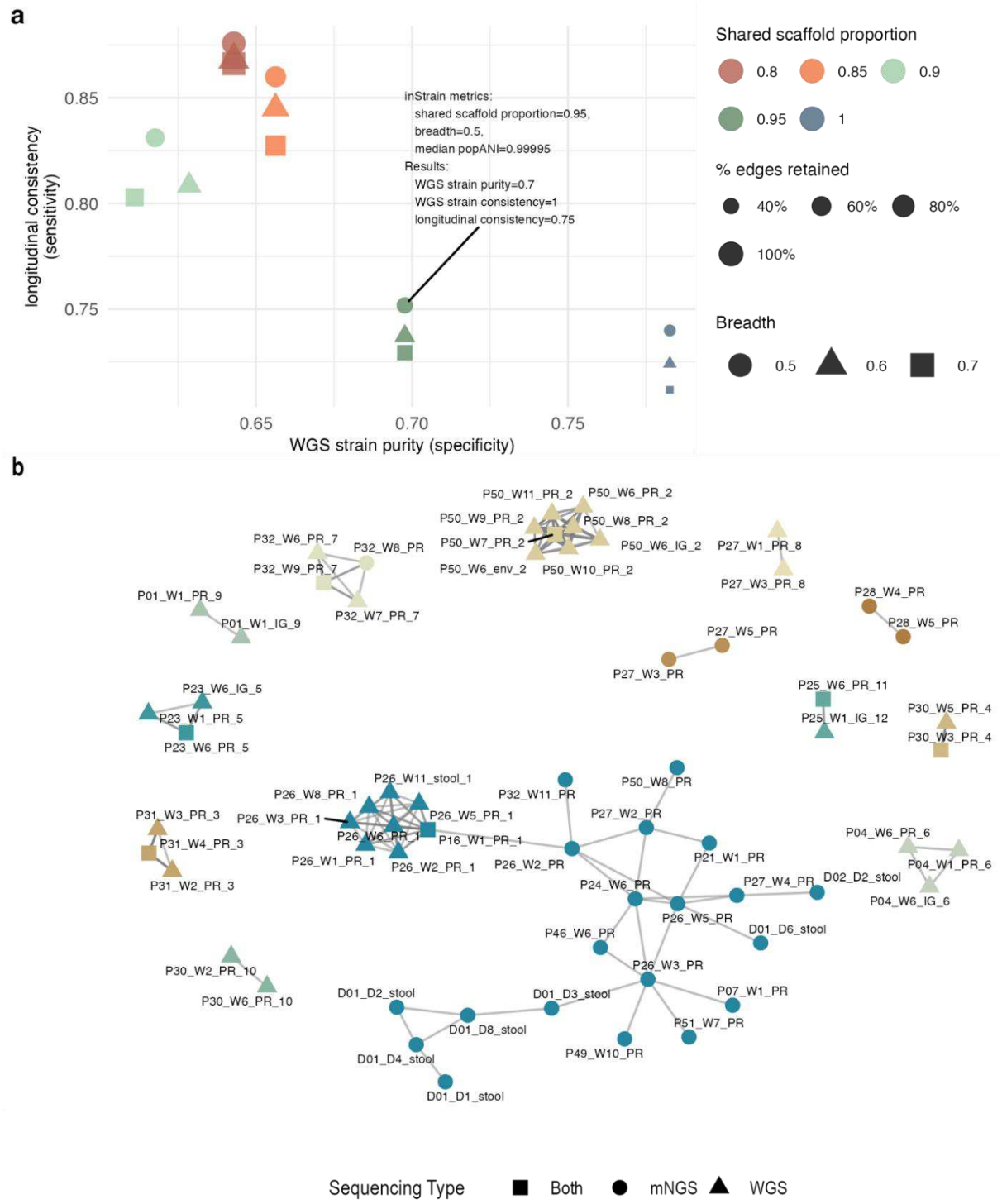


Figure 3. Threshold optimization for inStrain-based strain clustering. The analysis included metagenomic next-generation sequencing (mNGS) data from 93 samples and the whole-genome sequencing data from the 344 cultured isolates collected from 49 patients at a single long-term acute care hospital (LTACH) in Atlanta, Georgia between April and July 2023. **a** Sensitivity (longitudinal consistency) and specificity (SNVPhyl-defined WGS strain purity) across threshold combinations for inStrain-based strain clustering. Thresholds shown include breadth (shape) and shared scaffold proportion (color). Point size reflect the percentage of edges retained compared with the minimum threshold requirements. Results were stable across the popANI thresholds tested; median popANI ≥0.99999 is shown for clarity. **b** *Pseudomonas aeruginosa* strain clustering network generated from inStrain compare results using optimal thresholds (median popANI ≥0.99995, ≥0.50 breadth, ≥0.95 shared scaffold proportion). Nodes represent individual WGS data or mNGS samples, with the shape indicating sequencing type (mNGS, WGS, or both) and color indicating strain cluster assignment. Node labels indicate in order the patient ID (P#), timepoint week (W#), and sample type (PR: perirectal, IG: inguinal, stool, env: environmental), and where WGS data was available, the SNVPhyl-defined strain number. Edges connect samples that met the inStrain compare thresholds for strain-level clustering.

Finally, we evaluated whether the application of the Louvain clustering algorithm to the optimal thresholds would improve the sensitivity and specificity of strain determination. Louvain clustering did not impact the WGS strain purity (70% to 71%) or the WGS strain consistency (100%) but led to a substantial loss in longitudinal consistency (75% to 65%) (Supplementary Figure 7d). As a result, Louvain clustering was not applied for strain determination.

inStrain compare-based MDRO detection from mNGS data

Having defined optimal thresholds for MAG-based strain identification, we next assessed diagnostic performance versus culture-based methods. Compared with the gold standard of culture-based MDRO detection, mNGS had an overall sensitivity of 42% (95% CI: 34-50%), specificity of 92% (89-94%), positive predictive value (PPV) of 62% (53-71%), and negative predictive value (NPV) of 83% (80-86%) (Supplementary Figure 9a). When stratified by species, diagnostic performance varied significantly. For the bacterial species *C. difficile* and *P. aeruginosa,* culture had a higher detection rate than mNGS (GEE p <0.01 for both); all other bacterial species found no difference in detection rates (Supplementary Figure 9b). Notably for *E. hormaechei*, while detection rates were comparable, there was little concordance between detection by mNGS and culture.

Strain sharing between patients and environmental samples in the LTACH

To investigate strain transmission within the LTACH using SGBs, we used the above described thresholds to identify shared strains. Across all pairwise metagenomic sample comparisons, we identified 62,316 instances of strain sharing involving 661 of the 710 (93.1%) SGBs, demonstrating extensive strain sharing between samples. 12,864 (20.6%) of the shared strains were observed in same patient pairs, 11,431 (18.3%) were observed in donor:recipient pairs, and 38,021 (61.0%) were observed in non-connected pairs. We identified 20,275 possible transmission events, assuming a single introduction into the facility for each unique bacterial strain. We calculated a strain sharing rate for each sample pair by dividing the total number of shared strains by the total number of shared SGB detections between samples. Strain-sharing rates varied substantially between sample pairs (range= 0.00-0.93), with most exhibiting low levels of strain sharing (median= 0.07, mean= 0.11), and a smaller number demonstrating high strain-sharing rates (Figure 4a). Samples collected from the same individual exhibited significantly higher strain-sharing rates than samples collected from different individuals ($P < 2.2 \times 10^{-16}$; Figure 4b), consistent with longitudinal persistence of host-associated strains. Several clinically relevant species and potential MDROs displayed elevated rates of between-sample strain sharing, including *Clostridioides difficile* (0.23), *Enterobacter hormaechei* (0.21), *Pseudomonas aeruginosa* (0.20), *Enterococcus faecium* (0.19), and *Proteus mirabilis* (0.13) (Figure 4c).

A network of sample-level strain sharing revealed a highly interconnected structure spanning much of the cohort, with the greatest sharing observed between FMT donor:recipient sample pairs and samples collected from the same individual over time, as expected (Figure 4d). Numerous strain-sharing links were observed between samples without a FMT donor:recipient or intra-patient connection, however, suggesting that strain sharing was not restricted to these

cases. The average strain sharing rate was 0.22 between FMT donor:recipient sample pairs, 0.61 between intra-patient sample pairs, and 0.09 between unconnected sample pairs.

To characterize strain sharing at the individual level, shared strains were aggregated across samples belonging to the same participant or donor. The average rate of strain sharing between individuals was 0.075, with 0.26 being the average between donor:recipient pairs and 0.074 being the average between unconnected individuals. To more closely observe the strain sharing between FMT donors and recipients, we built a patient-level network of donor specific strain-sharing. We observed strain-sharing rates of up to 0.46 between FMT donors and recipients (Figure 4e). Interestingly, we observed strain sharing between donors and non-recipients, suggesting that donor-acquired strains from FMT recipients were transmitted to other patients within the LTACH.

To further investigate this, we leveraged the longitudinal structure of the data to build time-series networks across the 12 weeks of the study. Specifically, we evaluated FMT Donor 1 strain-sharing with all patients in the LTACH. We found that FMT Donor 1-associated strains successfully engrafted following transplantation and contributed to stable connections between the donor and multiple recipients over time. Additionally, non-recipients shared strains above baseline-level sharing with FMT Donor 1, consistent with secondary dissemination of FMT donor-derived strains within the LTACH. There were 40 individuals who did not receive FMT from Donor 1 (39 Non-FMT recipients, 1 FMT Donor 2 recipient). Of these, 13 (32.5%) were found to have evidence of strain sharing above baseline, including the single FMT Donor 2 recipient.

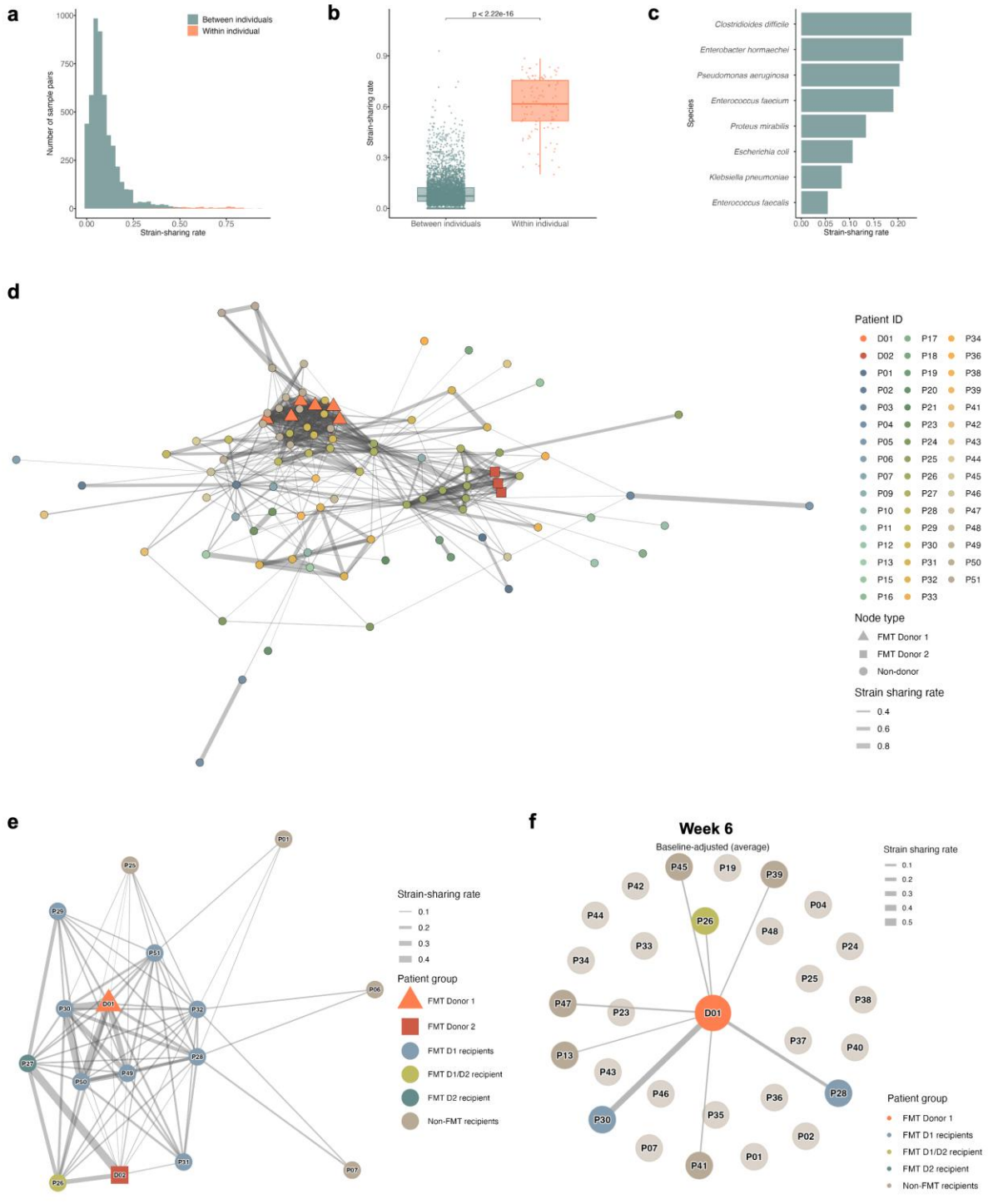

a
Between individuals
Within individual
Number of sample pairs
Strain-sharing rate
b
p < 2.22e-16
Strain-sharing rate
Between individuals
Within individual
c
Clostridioides difficile
Enterobacter hormaechei
Pseudomonas aeruginosa
Enterococcus faecium
Proteus mirabilis
Escherichia coli
Klebsiella pneumoniae
Enterococcus faecalis
Species
Strain-sharing rate
d
Patient ID
Node type
FMT Donor 1
FMT Donor 2
Non-donor
Strain sharing rate
e
Strain-sharing rate
Patient group
FMT Donor 1
FMT Donor 2
FMT D1 recipients
FMT D1/D2 recipient
FMT D2 recipient
Non-FMT recipients
f
Week 6
Baseline-adjusted (average)
Strain sharing rate
Patient group
FMT Donor 1
FMT D1 recipients
FMT D1/D2 recipient
FMT D2 recipient
Non-FMT recipients

Figure 4. Strain sharing networks reveal widespread transmission within the long-term acute care hospital (LTACH) and persistence of donor-associated strains following fecal microbiota transplantation (FMT). Networks were created from longitudinal samples (n=93) obtained from patients (n=49) at a single long-term acute care hospital (LTACH) in Atlanta, Georgia between April and July 2023. **a** Distribution of strain-sharing rates across all pairwise sample comparisons. Comparisons from the same participant (within-patient) are shown in green and comparisons between different participants (between-patient) are shown in orange. **b** Strain-sharing rates were significantly higher among samples collected from the same participant than among samples collected from different participants (Wilcoxon rank-sum test, $P < 2.2 \times 10^{-16}$). **c** Strain-sharing rates of MDROs of interest. **d** Sample-level strain-sharing network where each node is a sample, colored by participant/donor and edge width corresponds to strain-sharing rate. **e** Individual-level strain-sharing network where each node is a participant/donor colored according to study group and edge width corresponds to aggregated strain-sharing rate. **f** Longitudinal Donor 1 strain-sharing network. Network is shown for week 6 of the study. Each node is a patient/donor, colored by study group as before. Edge widths correspond to baseline-adjusted strain-sharing rates. Only connections between Donor 1 and other nodes are shown. All non-donor nodes shown had sampling and mNGS sequencing performed at week 6.

We next evaluated whether epidemiological exposures between sampling timepoints (timepoint 1 [T1] to timepoint 2 [T2]) were associated with increases in strain-sharing rates between patient pairs. In univariable models adjusting for baseline sharing rate at T1 and the number of MAGs detected at T2, an increased daily rate of interpatient crossings (defined as the number of times per day that a staff member moved between a given patient pair) was significantly associated with increased strain-sharing rates over time (OR 1.04, 95% CI 1.00–1.07, p=0.03; Figure 5). When stratified by staff role, nurse (OR 1.15, 95% CI 1.01–1.30) and respiratory therapist (OR 1.44, 95% CI 1.24–1.67) interpatient crossings were significantly associated with increased strain-sharing rates over time, whereas technician (i.e., nursing technicians and patient care assistants), support services staff (i.e., environmental services, facilities, and food services), therapist (i.e., physical and occupational therapists), and physician interpatient crossings were not. The estimate for physician interpatient crossings was accompanied by a wide confidence interval, reflecting the sparse physician data available, as most (n=71/89, 80%) patient pairs had no physician interpatient crossings during the observation interval. Although both hand hygiene non-compliance (OR 6.8, 95% CI 0.7–62.3, p=0.09) and spending a majority of days on the same floor (OR 1.4, 95% CI 1.0–2.1, p=0.08) were associated with increased strain-sharing rate, neither association was statistically significant.

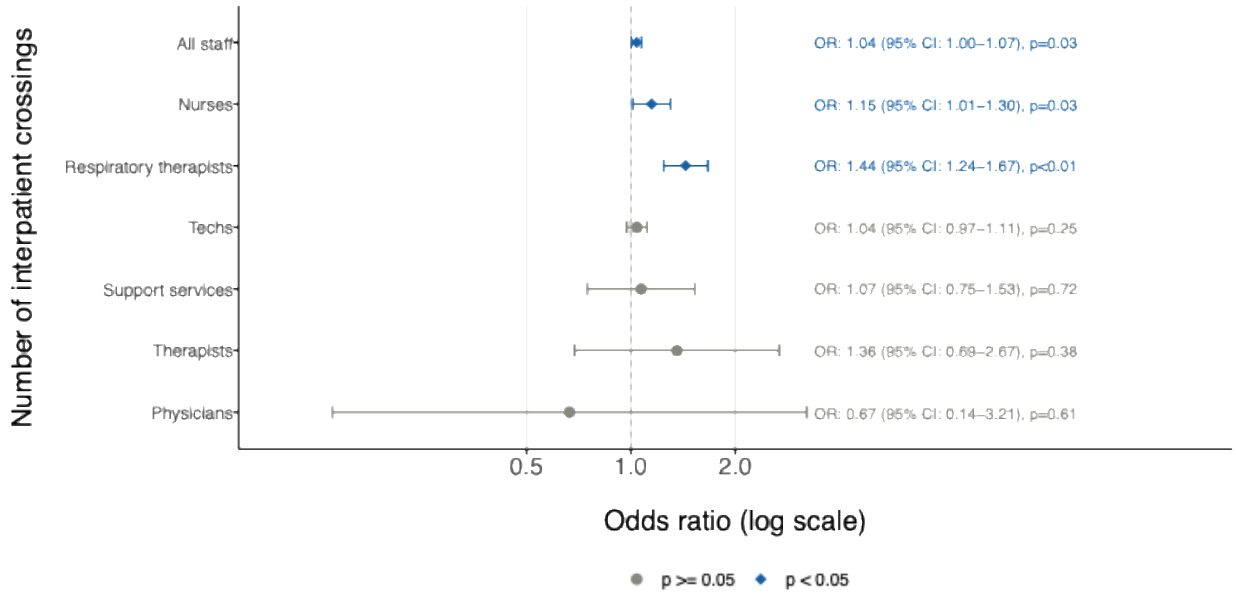


Figure 5. Forest plot of regression models evaluating association of shared staff exposures, by staff roles, with microbial transmission within the long-term acute care hospital (LTACH) over time (n=89 patient dyads-intervals). Associations between each number of interpatient crossings by shared staff within the time interval and strain sharing rate at each timepoint were estimated using univariable zero-inflated beta mixed-effects regression, adjusted for baseline sharing at the immediate prior timepoint, sequencing depth, and patient-level clustering. Colors differentiate between statistical significance inferred from Wald z-test p-values; $p<0.05$ was considered statistically significant.

Discussion

Leveraging data from a deeply characterized longitudinal LTACH cohort, we developed MAGSPI, an open-source pipeline for culture-independent mNGS surveillance of pathogen transmission in healthcare facilities. Using this approach, we identified evidence of ongoing, occult transmission of both pathogenic and commensal organisms that was not captured by standard WGS surveillance alone, including transmission patterns linked to otherwise missed epidemiologic risk factors such as healthcare staff. We further used concomitant FMTs performed within the LTACH as a benchmark for intentional strain transmission, enabling us to optimize strain-level inference and sharing parameters, characterize longitudinal dynamics of in-host microbial colonization, and identify off-target sharing of FMT-derived strains among patients in the facility.

Our study focused on transmission within LTACHs, an understudied healthcare setting with a uniquely elevated risk of MDRO transmission. The availability of culture-based WGS data enabled development of more stringent thresholds for strain identification than those traditionally used[27]. We chose to prioritize specificity over sensitivity to reduce false strain-sharing detections that may lead to inaccurate conclusions of potential transmission networks[28]. We adopted an approach using the strain-sharing rate, a normalized measure of microbial overlap between samples[19,20,29] as a framework to understand changes in baseline similarities between patient pairs and infer transmission. Network analysis identified both patient clusters in which strain sharing increased over time, and uncovered evidence of FMT donor strain colonization in non-FMT recipients, offering a higher-resolution view of transmission dynamics than prior approaches.

Our findings highlight the deeper resolution of strain-sharing events and transmission networks afforded by SGB-based mNGS surveillance, which can provide fresh insight into microbial dissemination in clinical settings. Culture-based WGS identified only 37 transmission events, limited by focused MDRO selection due to laboratory and personnel limitations, compared with the 20,275 strain-sharing events identified by the single assay, SGB-based mNGS. As such, this method may uncover additional epidemiologic factors that could contribute to healthcare-associated transmission and may be particularly useful during non-outbreak situations. One example was the discovery that transmission of the FMT Donor 1-associated strains to non-FMT recipients within the LTACH facility. While engraftment of donor strains has been well documented[22,30,31], to our knowledge, no previous studies have demonstrated facility-level spread of FMT donor strains beyond FMT recipients. While FMT has been shown to reduce MDRO colonization and infection risk in high-risk populations at the individual level[14-16], our

findings raise the possibility that administering FMT to patients colonized with MDROs in an LTACH could have broader, and potentially beneficial, effects on facility-level microbial ecology. Whether such individual-level treatments translate into community-level protection, however, remains an important question for further study.

Beyond identifying previously unrecognized transmission events, the SGB-based mNGS transmission network also uncovered epidemiologic factors associated with strain sharing. Regression models demonstrated that higher strain-sharing rates were associated with increased number of times respiratory therapists or nurses went back and forth between a given patient pair. Interestingly, poor hand hygiene compliance and shared floor were not significantly associated with increased strain-sharing rates over time. The data were likely too homogenous to effectively evaluate poor hand hygiene, as evidenced by the wide confidence intervals, and were underpowered to evaluate geographic proximity of shared floors, as the lower bound of the 95% confidence interval was close to 1.0. Both are well-established factors in healthcare-associated transmission and outbreaks[3,4,32,33]. The relationship between increased shared provider crossings and strain sharing is less straightforward. Increased crossings by respiratory therapists and nurses could simply be correlated with other risk factors for microbial colonization (e.g., higher acuity patient, greater exposure to antibiotics, or a greater likelihood of having indwelling lines and other invasive medical devices), or they may by causal, representing direct patient-to-staff-to-patient contact-mediated transmission. Ultimately, intervention studies of infection prevention and control measures targeting these staff roles would be needed to confirm causality. Taken together, these findings demonstrate that integrating SGB-based mNGS data with epidemiologic data can reveal both unexpected patterns of microbial dissemination and potential drivers of healthcare-associated transmission, providing insights that may guide future mechanistic studies and targeted infection prevention interventions.

Although SGB-based mNGS may be valuable for ecological characterization of transmission dynamics, our data reinforce that culture-based WGS or targeted PCR assays remain preferable for outbreak investigations or surveillance of specific pathogens[34]. Compared with culture-based WGS, SGB-based mNGS demonstrated high overall specificity (92%) but limited sensitivity (42%) for MDRO detection, with performance varying by species. Additionally, although MAG-based mNGS successfully detected high-consequence ARGs and attributed them to specific bacterial species, discordance with culture-based WGS data highlights the risk of ARG misattribution due to incomplete binning or chimeric MAG assembly, a known challenge specific to ARGs in MAG assembly[35], which is further exacerbated for plasmid-associated ARGs[36]. Some of these plasmid ARG-specific misbinning challenges may be resolved or improved upon by use of other methodologies such as long-read sequencing. However, these limitations must be taken in context of the large increase in additional strain-sharing events that the SGB-based MAGSPI pipeline has uncovered compared with culture-based WGS. SGB-based mNGS offers a powerful and scalable approach to healthcare surveillance capturing transmission dynamics, resistome complexity, and microbial community structure in a single assay.

Our work builds on other recent studies have explored the use of SGB-based mNGS data to track bacterial strain movements following donor engraftment of FMT recipients[22,37-39] and in hospital settings[40]. In one study conducted at a hospital in Pakistan[41], mNGS data from three pooled environmental surface samples identified nine high-quality MAGs (>50% completeness and <5% contamination) and suggested potential transmission events between hospital environments. Similarly, another study used strain-level MAGs to trace skin colonization and transmission of *C. auris* and bacterial pathogens among 42 nursing home residents[42], demonstrating network clusters of shared bacterial strains and widespread colonization with *C.*

auris and multidrug-resistant bacterial pathogens. Together, these studies support the feasibility of applying strain-resolved mNGS approaches to healthcare-associated transmission surveillance.

This study has several strengths, including its deeply phenotyped longitudinal LTACH cohort, paired WGS and mNGS profiling, rich healthcare workflow and site-level metadata, and development of MAGSPI, a novel open-source pipeline for culture-independent strain-sharing analysis. In addition, the study leverages concomitant FMTs as intentional strain-transmission events to benchmark strain-level inference, enabling what is, to our knowledge, the first evaluation of culture-independent bacterial transmission dynamics in the LTACH setting.

This study also has limitations. First, evaluation of diagnostic performance between SGB-based mNGS and culture-based WGS was limited because cultures were performed only for MDROs. Second, SGBs represent partial reconstructions of microbial genomes; although we applied stringent thresholds to improve confidence in species-level identification, SGBs lack the resolution of isolate WGS and may misclassify closely related strains. Third, healthcare workers and environmental surfaces were not directly sampled for mNGS, and therefore inferred transmission pathways are based on predictive modeling rather than direct microbiologic confirmation. Finally, as a single-center study, further work is need to understand the generalizability of our findings to other LTACHs and different patient populations.

Taken together, we report the development of MAGSPI and demonstrate that SGB-based mNGS analysis of surveillance swabs provides a minimally invasive strategy for longitudinal surveillance of microbial transmission dynamics in the LTACH setting. By enabling simultaneous

strain-level assessment across the broader gut microbiome, this approach offers complementary insights to traditional culture-based surveillance and may improve detection of transmission events that would otherwise remain unrecognized. Integrating mNGS-derived strain-sharing networks with epidemiologic and healthcare workflow data may also help identify modifiable transmission pathways and inform more targeted infection prevention interventions. As the field continues to mature, SGB-based mNGS holds considerable promise as a complement to, and potentially a partial replacement for, more resource-intensive culture-based WGS strategies.

Methods

Study Design

We conducted a cohort study of patients admitted to a single academic-affiliated LTACH in Atlanta, Georgia from April to July 2023. The facility is a 50-bed facility with a typical daily census of 30–40 patients. Patients are housed in private, single-occupancy rooms across two floors. Participants included all admitted patients who provided verbal consent (or for whom verbal consent was provided by a legally-authorized representative) for a perirectal swab point-prevalence sample to identify patients with MDRO colonization under an IRB-reviewed and approved protocol (STUDY00004199). Participants with positive MDRO cultures were further screened for participation in the Sentinel REACT trial[43]. Sentinel REACT was a pilot trial of fecal microbiota transplantation (FMT) conducted at the LTACH under the IRB-approved protocol STUDY00005467.

Participant demographic characteristics (e.g. sex, race, age, and medical history) were abstracted from the medical record. Samples were collected from all participants with verbal consent admitted on Week 1 and Week 6 of the study period. Ten participants with MDRO-positive perirectal cultures consented to FMT administration and underwent enhanced weekly sampling on the day of FMT administration and 7, 14, 21, and 28 days post-FMT administration. FMT doses were produced with material from just two feces donors who met screening criteria, including stool tests that were negative for MDROs. Samples at each time point included up to three anatomic (inguinal, perirectal, or stool) culture swabs and one environmental (high-touch room surface) culture swabs.

Culture and Whole Genome Sequencing (WGS)

Participant ESwab (COPAN Diagnostics, Murrieta, CA, USA) eluates were plated on differential and selective culture media (Remel Spectra VRE, Thermo Fisher Scientific, Lenexa, KS, USA;

HardyCHROM ESBL, Hardy Diagnostics, Santa Maria, CA, USA, cycloserine cefoxitin mannitol broth with taurocholate lysozyme cysteine [CCMB-TAL], Anaerobe Systems, Morgan Hill, CA, USA) for qualitative MDRO detection of target MDRO categories. Target MDROs included extended-spectrum beta-lactamase (ESBL) producing Enterobacterales, carbapenem-resistant Enterobacterales (CRE), multidrug-resistant *Pseudomonas* (MDRP), vancomycin-resistant *Enterococcus* spp. (VRE), and toxigenic *C. difficile*. Suspected *C. difficile* isolates were sub-cultured in anaerobic conditions and underwent PCR assays for detection of the pan-*C. difficile* *rplP* gene and *tcdB* toxin gene targets as previously described to further classify isolates as toxigenic or non-toxigenic (Supplementary Table 4). For non-*C. difficile* positive cultures, pure-growth colonies were identified by MALDI-TOF mass spectrometry (bioMérieux, Marcy-l'Étoile, France) with susceptibility testing performed by Vitek2 microdilution (bioMérieux, Marcy-l'Étoile, France) or Oxoid™ 30µg vancomycin disc diffusion (Thermo Fisher Scientific) (to classify *Enterococcus* spp. as VRE or vancomycin-susceptible *Enterococcus* spp.) as per Clinical Laboratory Standards Institute cutoffs[44]. Participants with MDRO-positive perirectal cultures consented to FMT administration and underwent enhanced weekly sampling (anatomic and environment) on the day of FMT administration and up to four weeks post-FMT administration; one participant received 2 doses of FMT administered 5 weeks apart.

Morphologically distinct MDRO isolates were selected to undergo WGS. For each participant timepoint, when identical isolates, as determined by morphologic inspection and phenotypic antimicrobial susceptibility testing, were recovered from multiple culture plates from the same sample, only one cultured isolate was sequenced. DNA was extracted from pure-growth plates using the MagMax Viral/Pathogen Nucleic Acid Isolation kit (Applied Biosystems A42352) with the KingFisher Apex instrument (Thermo Fisher Scientific). Libraries were prepared using the Nextera XT kit (Illumina), and after quality control, pooled DNA libraries were sequenced on an Illumina NovaSeq 6000 with a target of 600,000 minimum PE150 reads per isolate. Nineteen

samples yielded genomes with apparent contamination or low-quality assemblies were re-cultured and re-sequenced. SNVPhyl[26] was used to align sequences and call SNVs between genomes. Genome pairs differing by ≤10 SNVs were considered shared strains. For *C. difficile* genomes, a more stringent cutoff of ≤2 SNVs was applied, consistent with previously published studies[45-47]. Strain assignments using the above SNV thresholds for cultured isolates were considered gold standard and used to assess the strain assignment performance of inStrain[25]. The minimum number of inferred transmission events was calculated for each SNVPhyl-defined strain cluster as the number of unique patients in each strain cluster minus one, assuming a single introduction into the facility.

Metagenomic Next-Generation Sequencing (mNGS)

mNGS was performed on 93 perirectal samples. These comprised (1) point-prevalence samples from participants without enhanced weekly sampling, and (2) weekly longitudinal samples from participants enrolled in the enhanced sampling protocol (Figure 1a). For most perirectal swabs, DNA was extracted with the ZymoBIOMICS MagBead DNA kit (Zymo Research), libraries were prepared with the Nextera XT kit (Illumina), and after quality control, pooled libraries were sequenced on an Illumina NovaSeq 6000. Negative template water controls were included with each extraction batch and sequenced. Adapter trimming, low-quality and low-complexity read filtering were performed using fastp[48]. High-quality reads were human host-filtered using Bowtie2[49] and the Hg38 human genome reference[50]. Sample-level sequencing quality was assessed using SeqKit[51].

Species-level Genome Bins (SGB) Pipeline

High-quality non-host metagenomic reads were re-paired using BBmap[52] and assembled into contigs using metaSPAdes[53]. Assembly quality was assessed using QUAST[54]. The assembled contigs were binned into metagenome-assembled genomes (MAGs) using three tools with distinct algorithmic approaches: MaxBin2[55], MetaBAT2[56], and CONCOCT[57]. DAS Tool was used to compare MAGs across the three approaches on a per sample basis and generate high-quality non-redundant bins[58]. MAG quality was assessed using CheckM[59]; high-quality MAGs as designated by DAS Tool with a completeness score ≥50% and a contamination level ≤10% were retained for downstream analysis. Taxonomic assignment of MAGs was performed using GTDB-Tk[60] and antimicrobial resistance gene (ARG) profiling was performed using RGI and the CARD database[61]. MAGs were de-replicated at 97% similarity using dRep[62] to generate a database of species-level genome bins (SGBs). This SGB database was used as reference to identify MAGs in metagenomic samples and trace their transmission between samples using inStrain profile and inStrain compare[25], respectively.

Detected MAGs were summarized using metrics including recovered MAG richness, sequencing yield (total reads), microbial diversity (SDI) contig counts, and completeness. Separate univariable negative binomial regression models were used to assess the associations microbial diversity and sequencing yield with recovered MAG richness. MAG-to-SGB recovery rates were calculated overall and stratified by recovered MAG richness and sample type.

inStrain Bacterial Strain Assignment

To identify shared bacterial strains in mNGS samples, we performed pairwise comparisons among all WGS and mNGS sequences using inStrain profile and inStrain compare to generate pairwise population average nucleotide identity (popANI) values for each SGB scaffold. These

values were then averaged across all scaffolds to generate a median popANI value between pairwise comparisons.

inStrain thresholds to determine strain assignments were chosen by systematically evaluating combinations of the following metrics: median popANI (range: 0.99995-0.999999), coverage breadth across the SGB reference genome (range: 0.50-0.70), and proportion of shared scaffolds meeting the high-coverage criteria of ≥0.99995 popANI in both samples (range: 0.80-1.00). We set the lower bounds of the range tested as median popANI ≥0.99995, ≥0.50 breadth, ≥0.80 shared scaffold proportion, and required ≥10 shared scaffolds. Using strain assignments defined by SNVPhyl from WGS culture isolates as the gold standard, the different threshold combinations of inStrain metrics were evaluated for optimization of the sensitivity and specificity for strain assignment. Two measures of sensitivity were used: (1) WGS strain consistency, defined as the ability of inStrain compare to classify all WGS sequences belonging to the same SNVPhyl-defined strain as the same strain; and (2) longitudinal consistency, defined as the ability of inStrain compare to classify repeated samples from the same individuals as the same strain. Specificity was defined as WGS strain purity, where the inStrain-defined strains contained sequences from only a single SNVPhyl-defined strain. Only SGBs representing species with ≥25 cultured isolates with WGS data were included in this threshold determination analysis; these SGBs included *E. coli* (n=2), *E. faecalis* (n=2), *C. difficile* (n=1), *E. faecium* (n=1), *E. hormaechei* (n=1), *K. pneumoniae* (n=1), and *P. aeruginosa* (n=1).

Based on the optimization analysis, a shared strain was defined as a pairwise comparison between WGS data and mNGS samples meeting all the following thresholds: median popANI ≥0.99999, ≥0.50 breadth across the SGB reference genome, ≥10 compared scaffolds, and ≥0.95 shared scaffold proportion. Strain determinations were determined by network connectedness and then visualized per SGB using network plots.

Comparison of SGB inStrain compare-based mNGS with culture for MDRO detection

We assessed metagenomic detection against the gold standard of MDRO culture detection of toxigenic *C. difficile*, vancomycin-resistant *Enterococcus* (*E. faecalis* and *E. faecium*), and ESBL-producing or carbapenem-resistant Enterobacterales (*E. coli*, *E. hormaechei*, *K. pneumoniae*, and *P. aeruginosa*). To narrow detection to MDROs, we incorporated inStrain compare results; samples with metagenomic sequences that met the strain sharing thresholds with a WGS of a culture isolate were considered an MDRO.

We evaluated both diagnostic performance (i.e., sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]) as well as detection concordance of the 101 samples that were sent for mNGS and culture. To compare detection rates between inStrain compare-based mNGS and culture, Generalized Estimating Equations (GEE) with an independence working correlation structure and robust sandwich variance estimators were used, clustering samples by person to account for repeated samples (range 1-6 samples per person). GEE models were fit separately for each MDRO bacterial species; an additional pooled GEE adjusting for MDRO bacterial species was used to estimate the overall method effect across all seven species.

Finally, we performed a sensitivity analysis where samples with metagenomic sequences identified in the same network cluster as a WGS of an MDRO culture isolate, rather than requiring direct connection to a WGS were considered an MDRO.

Strain Sharing Rate Analysis

Strain sharing between samples was quantified using a strain-sharing rate metric. For each pairwise comparison, the number of shared strains was calculated using inStrain compare results and our previously optimized thresholds: median popANI ≥0.99999, ≥0.50 breadth across the SGB reference genome, ≥10 compared scaffolds, and ≥0.95 shared scaffold proportion. The number of shared strains was divided by the total number of shared SGBs between samples with at least 50% coverage breadth.

$$\text{Strain-sharing rate} = \frac{\sum \text{shared strains}}{\sum \text{SGBs detected in both samples}}$$

For patient-level analyses, the number of shared strains and number of shared SGBs were aggregated across all pairwise sample comparisons between a given pair of individuals. The total number of shared strains between individuals at all time points was divided by the total number of SGBs shared over those time points to create a single strain-sharing rate for each participant pair. To reduce noise, participant pairs with fewer than 20 shared SGBs were excluded from network analyses.

For longitudinal analyses of donor 1 strain-sharing across the LTACH, we analyzed all strain-sharing events between donor 1 and other patients, aggregated at the patient-level. To accurately capture acquired strain-sharing and exclude spurious low-level sharing between patients, we generated an average baseline strain-sharing rate from all inter-patient comparisons and subtracted this from all patient level strain sharing rates at later time points.

Automated Hand Hygiene Monitoring System

All inpatient rooms on the study units were equipped with hand hygiene sensors (Clean Hands - Safe Hands, Atlanta GA; now part of SwipeSense, Atlanta, GA) attached to the hand hygiene product dispenser at the entrance of each room. These sensors identified the unique Bluetooth beacon emitted by the badge worn by the care provider and detected motion into and out of patient rooms as well as the dispensing of alcohol-based hand rub or soap. All staff who provided direct patient care were asked to wear a badge, including nurses, nurse technicians, physicians, advanced practice providers and therapists, though badge wearing was not a mandatory requirement for employment. Hand hygiene adherence was recorded by the electronic system based on whether the healthcare worker dispensed hand hygiene product within a predetermined amount of time after being detected to have entered or exited the patient's room.

Transmission Risk Factor Analysis

Due to the sampling design and the skewed distribution of sharing rates, we used univariable zero-inflated beta mixed-effects models to evaluate whether epidemiological risk factors were associated with increases in strain-sharing rate between patient pairs. The outcome was the strain-sharing rate between two patients at a timepoint (T2). Separate models were fit for each epidemiological exposure during the interval between the prior patient pair timepoint (T1) and T2: hand hygiene non-compliance rate (defined as the proportion of room entries by a shared staff during which hand hygiene was not performed); majority days spent on the same floor; and daily interpatient crossing rate (defined as the number of times per day a staff member moved between patient pairs), evaluated overall and by staff role. All models included the strain-sharing rate at T1, an offset for the log number of MAGs detected at T2, and random intercepts for each patient in the patient pair.

The conditional component of the model was specified on the logit scale as:

$$\eta_{ij,T2} = \beta_0 + \beta_1 S_{ij,T1} + \beta_2 E_{ij} + \log(N_{ij,T2}) + u_i + u_j$$

where $\mu_{ij,T2}$ is the expected strain-sharing rate for patient pair ij at T2; β_0 is the fixed intercept; $S_{ij,T1}$ is the strain-sharing rate at T1; E_{ij} is the epidemiological exposure measured during the interval between T1 and T2; $N_{ij,T2}$ is the number of MAGs available for comparison at T2, with $\log(N_{ij,T2})$ included as an offset; and u_i and u_j are patient-specific random intercepts for the two patients comprising the dyad. An intercept-only zero-inflation component was included in all models.

Acknowledgments

We are grateful for the patients who made this study possible through their participation. We also thank Drs. Jesse Jacob and Lindsey Gottlieb for generously sharing the hand hygiene data utilized in the spatial data analysis.

Funding

Research reported in this publication was supported by the Emory Prevention Epicenter Program (PEACH) through the Centers for Disease Control and Prevention (CDC) [U54CK000601 to SL, SS, SFK, MHW], an Antibacterial Resistance Leadership Group Early Faculty Seedling Award (National Institute for Allergy and Infectious Diseases) [UM1AI104681 to AB], National Institute for Allergy and Infectious Diseases K32 awards [K23AI144036 to MHW and K23AI185326 to VTC], and the Chan Zuckerberg Biohub, San Francisco, CA [CRL and ALG]. CDC co-author roles included interpretation, preparation, review, and approval of the manuscript. These funding organizations had no other role in design and conduct of the study; collection, management, and analysis of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. This study was supported in

part by the Investigational Clinical Microbiology Core (ICMC), which was supported by the Department of Medicine, Division of Infectious Diseases, Emory University School of Medicine.

Disclaimer: This content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health or Centers for Disease Control and Prevention.

Author contributions:

Conceptualization: VTC, ALG, CRL, MHW

Methodology: VTC, ALG, CRL, MHW

Investigation: VTC, ALG, AB, DS, AA, DBS, CWB, AP, SCR, ALH, MSW, JVR, EMB, SL, SWS, CRL, MHW

Visualization: VTC, ALG, CRL, MHW

Funding acquisition: MHW, VTC, CRL

Project administration: VTC, ALG, CRL, MHW, CWB, DS, AA, SCR

Supervision: VTC, ALG, MHW, CWB, SCR

Writing – original draft: VTC, ALG, CRL, MHW

Writing – review & editing: VTC, ALG, AB, DS, AA, DBS, CWB, AP, SCR, ALH, MSW, JVR, EMB, SL, SWS, CRL, MHW

Ethics approval: This study was conducted in accordance with the Declaration of Helsinki and the U.S. federal regulations at 45 CFR part 46 and 21 CFR part 56. All study procedures were reviewed and approved by the Emory University Institutional Review Board (Atlanta, GA, United States). Perirectal and inguinal point prevalence sampling was performed under the Antimicrobial Resistance Point Prevalence Sampling (APPS) protocol, STUDY00004199.

Healthy donor consent and screening for FMT dose manufacture was performed under the Microbiome Enrichment Program stool donor protocol, 2025P011796. Fecal microbiota transplantation was performed under the Sentinel REACT protocol, STUDY00005467, which was reviewed as part of a U.S. Food and Drug Administration investigational new drug application and approved under 45 CFR part 46.101(c) and 21 CFR part 56 (ClinicalTrials.gov NCT05780801).

Consent to participate: Verbal informed consent for perirectal and inguinal sampling was obtained from every participant, or from a legally authorized representative where the participant was unable to provide consent, as approved by the Emory University Institutional Review Board. Healthy stool donors for FMT dose manufacturing provided written informed consent prior to any study activities. Participants who subsequently received fecal microbiota transplantation in the Sentinel REACT trial, or their legally authorized representatives, provided separate written informed consent prior to treatment.

Competing interests: A.B. reports personal fees from Beckman Coulter. M.H.W. has served on a clinical advisory board for CARB-X. All other authors declare that they have no competing interests.

Data Availability

Metagenomic data are available through the National Library of Medicine Database of Genotypes and Phenotypes (dbGaP) under accession ID phs004084.v1.p1. Isolate genomes are available through the National Center for Biotechnology Information Sequence Read Archive (SRA) under BioProject PRJNA1271147.

All code and de-identified source data can be found at https://github.com/infectiousdisease-langelier-lab/REACT. The authors are committed to sharing metagenomic data with approved investigators who complete dbGaP access requirements or with direct data use agreement executed with the corresponding author.

References

1. Murray CJL, Ikuta KS, Sharara F, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* 2022; **399**(10325): 629-55.

2. Jernigan JA, Hatfield KM, Wolford H, et al. Multidrug-Resistant Bacterial Infections in U.S. Hospitalized Patients, 2012-2017. *N Engl J Med* 2020; **382**(14): 1309-19.

3. Boyce JM. Hand and environmental hygiene: respective roles for MRSA, multi-resistant gram negatives, Clostridioides difficile, and Candida spp. *Antimicrob Resist Infect Control* 2024; **13**(1): 110.

4. Safety WP, Organization WH. WHO guidelines on hand hygiene in health care: World Health Organization, 2009.

5. Siegel JD, Rhinehart E, Jackson M, Chiarello L. 2007 Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Health Care Settings. *Am J Infect Control* 2007; **35**(10 Suppl 2): S65-164.

6. Tanner WD, Leecaster MK, Zhang Y, et al. Environmental Contamination of Contact Precaution and Non-Contact Precaution Patient Rooms in Six Acute Care Facilities. *Clin Infect Dis* 2021; **72**(Suppl 1): S8-s16.

7. McKinnell JA, Singh RD, Miller LG, et al. The SHIELD Orange County Project: Multidrug-resistant Organism Prevalence in 21 Nursing Homes and Long-term Acute Care Facilities in Southern California. *Clin Infect Dis* 2019; **69**(9): 1566-73.

8. Lim CJ, Cheng AC, Kennon J, et al. Prevalence of multidrug-resistant organisms and risk factors for carriage in long-term care facilities: a nested case-control study. *J Antimicrob Chemother* 2014; **69**(7): 1972-80.

9. Azzini AM, Be G, Naso L, et al. Risk factors for colonization with multidrug-resistant Gram-negative bacteria and Clostridioides difficile in Long Term Care Facilities (LTCFs) residents: the evidence from 27 facilities in a high endemic setting. *Front Cell Infect Microbiol* 2023; **13**: 1155320.

10. Han JH, Lapp Z, Bushman F, et al. Whole-Genome Sequencing To Identify Drivers of Carbapenem-Resistant Klebsiella pneumoniae Transmission within and between Regional Long-Term Acute-Care Hospitals. *Antimicrob Agents Chemother* 2019; **63**(11).

11. Cincotta SE, Walters MS, Ham DC, et al. Regional impact of multidrug-resistant organism prevention bundles implemented by facility type: A modeling study. *Infect Control Hosp Epidemiol* 2024; **45**(7): 856-63.

12. Lee JY, Kim Y, Kim J, Kim JK. Fecal Microbiota Transplantation: Indications, Methods, and Challenges. *J Microbiol* 2024; **62**(12): 1057-74.

13. Kelly CR, Fischer M, Allegretti JR, et al. ACG Clinical Guidelines: Prevention, Diagnosis, and Treatment of Clostridioides difficile Infections. *Official journal of the American College of Gastroenterology | ACG* 2021; **116**(6): 1124-47.

14. Saha S, Tariq R, Tosh PK, Pardi DS, Khanna S. Faecal microbiota transplantation for eradicating carriage of multidrug-resistant organisms: a systematic review. *Clin Microbiol Infect* 2019; **25**(8): 958-63.

15. Yoon YK, Suh JW, Kang EJ, Kim JY. Efficacy and safety of fecal microbiota transplantation for decolonization of intestinal multidrug-resistant microorganism carriage: beyond Clostridioides difficile infection. *Ann Med* 2019; **51**(7-8): 379-89.

16. Merrick B, Prossomariti D, Allen E, et al. Faecal microbiota transplant to ERadicate gastrointestinal carriage of Antibiotic-Resistant Organisms (FERARO): A feasibility randomised controlled trial. *J Infect* 2025; **91**(1): 106504.

17. Cummings LA, Hoogestraat DR, Rassoulian-Barrett SL, et al. Comprehensive evaluation of complex polymicrobial specimens using next generation sequencing and standard microbiological culture. *Sci Rep* 2020; **10**(1): 5446.

18. Yi Q, Zhang G, Wang T, et al. Comparative Analysis of Metagenomic Next-Generation Sequencing, Sanger Sequencing, and Conventional Culture for Detecting Common Pathogens Causing Lower Respiratory Tract Infections in Clinical Samples. *Microorganisms* 2025; **13**(3).

19. Valles-Colomer M, Blanco-Míguez A, Manghi P, et al. The person-to-person transmission landscape of the gut and oral microbiomes. *Nature* 2023; **614**(7946): 125-35.

20. Ricci L, Heidrich V, Punčochář M, et al. Baby-to-baby strain transmission shapes the developing gut microbiome. *Nature* 2026; **651**(8104): 191-200.

21. Lu D, Kalantar KL, Glascock AL, et al. Simultaneous detection of pathogens and antimicrobial resistance genes with the open source, cloud-based, CZ ID platform. *Genome Medicine* 2025; **17**(1): 46.

22. Ianiro G, Punčochář M, Karcher N, et al. Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. *Nat Med* 2022; **28**(9): 1913-23.

23. Fan Y, Ni M, Aggarwala V, et al. Long-read metagenomics for strain tracking after faecal microbiota transplant. *Nat Microbiol* 2025; **10**(12): 3258-71.

24. Glascock AL. MAGSPI. https://github.com/infectiousdisease-langelier-lab/MAGSPI-Nextflow: Github; 2026.

25. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nature Biotechnology* 2021; **39**(6): 727-36.

26. Petkau A, Mabon P, Sieffert C, et al. SNVPhyl: a single nucleotide variant phylogenomics pipeline for microbial genomic epidemiology. *Microb Genom* 2017; **3**(6): e000116.

27. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain enables population genomic analysis from metagenomic data and sensitive detection of shared microbial strains. *Nat Biotechnol* 2021; **39**(6): 727-36.

28. Hawken SE, Yelin RD, Lolans K, et al. Threshold-free genomic cluster detection to track transmission pathways in health-care settings: a genomic epidemiology analysis. *Lancet Microbe* 2022; **3**(9): e652-e62.

29. Beghini F, Pullman J, Alexander M, et al. Gut microbiome strain-sharing within isolated village social networks. *Nature* 2025; **637**(8044): 167-75.

30. Smillie CS, Sauk J, Gevers D, et al. Strain Tracking Reveals the Determinants of Bacterial Engraftment in the Human Gut Following Fecal Microbiota Transplantation. *Cell Host Microbe* 2018; **23**(2): 229-40.e5.

31. Porcari S, Benech N, Valles-Colomer M, et al. Key determinants of success in fecal microbiota transplantation: From microbiome to clinic. *Cell Host Microbe* 2023; **31**(5): 712-33.

32. King ME BD, Goodman RA, Rasmussen SA. The CDC Field Epidemiology Manuaal. In: Rasmussen SA, Goodman RA, eds.: New York: Oxford University Press; 2019.

33. World Health Organization (WHO). Outbreak Toolkit: Stages of an outbreak investigation. https://www.who.int/emergencies/outbreak-toolkit/investigating-outbreak-of-unknown-disease (accessed May 25, 2026).

34. Nutman A, Marchaim D. How to: molecular investigation of a hospital outbreak. *Clinical Microbiology and Infection* 2019; **25**(6): 688-95.

35. Abramova A, Karkman A, Bengtsson-Palme J. Metagenomic assemblies tend to break around antibiotic resistance genes. *BMC Genomics* 2024; **25**(1): 959.

36. Maguire F, Jia B, Gray KL, Lau WYV, Beiko RG, Brinkman FSL. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic Islands. *Microb Genom* 2020; **6**(10).

37. Hoops SL, Moutsoglou D, Vaughn BP, Khoruts A, Knights D. Metagenomic source tracking after microbiota transplant therapy. *Gut Microbes* 2025; **17**(1): 2487840.

38. Chen Q, Wu C, Xu J, et al. Donor-recipient intermicrobial interactions impact transfer of subspecies and fecal microbiota transplantation outcome. *Cell Host Microbe* 2024; **32**(3): 349-65.e4.

39. Schmidt TSB, Li SS, Maistrenko OM, et al. Drivers and determinants of strain dynamics following fecal microbiota transplantation. *Nat Med* 2022; **28**(9): 1902-12.

40. Siranosian BA, Brooks EF, Andermann T, et al. Rare transmission of commensal and pathogenic bacteria in the gut microbiome of hospitalized adults. *Nature Communications* 2022; **13**(1): 586.

41. Asghar MU, Zhai Y, Liu T, et al. A metagenomics-based approach to understanding the transmission of healthcare-associated antimicrobial resistance in Pakistan. *J Hazard Mater* 2025; **496**: 139384.

42. Proctor DM, Sansom SE, Deming C, et al. Clonal Candida auris and ESKAPE pathogens on the skin of residents of nursing homes. *Nature* 2025; **639**(8056): 1016-23.

43. Woodworth MH, Babiker A, Prakash-Asrani R, et al. Microbiota Transplantation Among Patients Receiving Long-Term Care: The Sentinel REACT Nonrandomized Clinical Trial. *JAMA Netw Open* 2025; **8**(7): e2522740.

44. Clinical Laboratory Standards Institute (CLSI). Performance Standards for Antimicrobial Susceptibility Testing, 2024.

45. Cho J, Cunningham S, Pu M, et al. Clostridioides difficile Whole-genome Sequencing Differentiates Relapse With the Same Strain From Reinfection With a New Strain. *Clin Infect Dis* 2021; **72**(5): 806-13.

46. Miles-Jay A, Young VB, Pamer EG, et al. A multisite genomic epidemiology study of Clostridioides difficile infections in the USA supports differential roles of healthcare versus community spread for two common strains. *Microb Genom* 2021; **7**(6).

47. Eyre DW, Cule ML, Wilson DJ, et al. Diverse sources of C. difficile infection identified on whole-genome sequencing. *N Engl J Med* 2013; **369**(13): 1195-205.

48. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018; **34**(17): i884-i90.

49. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**(4): 357-9.

50. Casper J, Speir ML, Raney BJ, et al. The UCSC Genome Browser database: 2026 update. *Nucleic Acids Res* 2026; **54**(D1): D1331-d5.

51. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* 2016; **11**(10): e0163962.

52. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. United States: Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US); 2014.

53. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017; **27**(5): 824-34.

54. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013; **29**(8): 1072-5.

55. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016; **32**(4): 605-7.

56. Kang DD, Li F, Kirton E, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019; **7**: e7359.

57. Alneberg J, Bjarnason BS, de Bruijn I, et al. Binning metagenomic contigs by coverage and composition. *Nature Methods* 2014; **11**(11): 1144-6.

58. Sieber CMK, Probst AJ, Sharrar A, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* 2018; **3**(7): 836-43.

59. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**(7): 1043-55.

60. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 2022; **38**(23): 5315-6.

61. Alcock BP, Huynh W, Chalil R, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* 2023; **51**(D1): D690-d9.

62. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *Isme j* 2017; **11**(12): 2864-8.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- REACTManuscriptSupplementary092926.docx