Supplementary Information

Supplementary figures

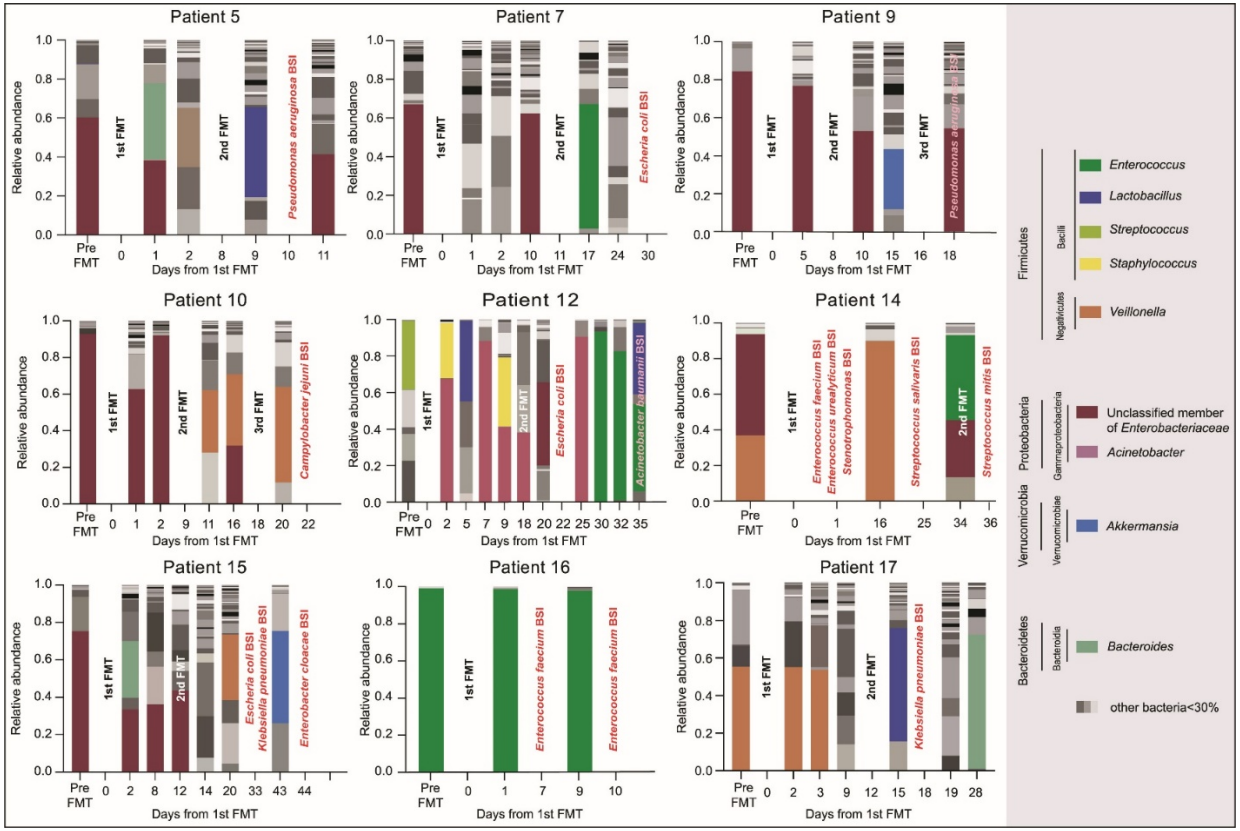

Figure S1. Bacterial monodomination among FMT recipients who developed BSIs. Temporal dynamics of the gut microbiota (genus level) during FMT sessions. BSI type and timing are designated in red. Monodominating taxa are coded by color. Taxa with a relative abundance of less than 30% are summarized as "other". FMT – fecal microbiota transplantation; BSI – bloodstream infection

Table S1. Characteristics of patients experiencing and not experiencing bloodstream infections after FMT

	Overall (n = 22)	BSI (n = 15)	No-BSI (n = 7)	P-value*
Transplant center, *n* (%)				1.0000
Chaim Sheba Medical Center	17 (77.3)	12 (80.0)	5 (71.4)	
Rambam Health Care	5 (22.7)	3 (20.0)	2 (28.6)	
Age at transplantation (med. [IQR])	56.0 [40.2, 64.5]	52.0 [38.5, 62.5]	62.0 [43.5, 65.0]	0.4583
Patient sex, *n* (%)				0.0743
Female	10 (45.5)	9 (60.0)	1 (14.3)	
Male	12 (54.5)	6 (40.0)	6 (85.7)	
Diagnosis, *n* (%)				0.6482
Acute lymphoblastic leukemia	1 (4.5)	1 (6.7)	0 (0.0)	
Acute myeloid leukemia	7 (31.8)	3 (20.0)	4 (57.1)	
Lymphoma	4 (18.1)	3 (20.0)	1 (14.3)	
Myelodysplastic syndrome	6 (27.3)	4 (26.7)	2 (28.6)	
Multiple myeloma	1 (4.5)	1 (6.7)	0 (0.0)	
Myeloproliferative neoplasms	3 (13.6)	3 (20.0)	0 (0.0)	
Remission status at HSCT, *n* (%)				1.0000
in remission	11 (50.0)	7 (46.7)	4 (57.1)	
relapsed/refractory disease	11 (50.0)	8 (53.3)	3 (42.9)	
Donor type, *n* (%)				0.7765
Matched sibling donor	11 (50.0)	7 (46.7)	4 (57.1)	
10/10 HLA match unrelated donor	7 (31.8)	4 (26.7)	3 (42.9)	
9/10 HLA match unrelated donor	3 (13.6)	3 (20.0)	0 (0.0)	
Haploidentical donor	1 (4.5)	1 (6.7)	0 (0.0)	
Conditioning intensity, *n* (%)				0.3762
Myeloablative	13 (59.1)	10 (66.7)	3 (42.9)	
Reduced intensity	9 (40.9)	5 (33.3)	4 (57.1)	
GvHD prophylaxis, *n* (%)				0.6749
CSA + MTX	11 (50.0)	6 (40.0)	5 (71.4)	
CSA + MTX + MMF	1 (4.5)	1 (6.7)	0 (0.0)	
CSA + MMF	8 (36.4)	6 (40.0)	2 (28.6)	
PTCy + MMF	2 (9.1)	2 (13.3)	0 (0.0)	
In-vivo T-cell depletion (ATG), *n* (%)				0.3246
received	14 (63.6)	11 (73.3)	3 (42.9)	
not received	8 (36.4)	4 (26.7)	4 (57.1)	
HSCT to GvHD onset, days (med. [IQR])	87.0 [34.2, 142.5]	38 [31.5, 120.0]	96.0 [67.5, 178.0]	0.2902
Stage of GI GvHD at onset, *n* (%)				0.1854
II	1 (4.5)	0 (0.0)	1 (14.3)	
III	18 (81.8)	12 (80.0)	6 (85.7)	
IV	3 (13.6)	3 (20.0)	0 (0.0)	
IBMTR Stage of GvHD at onset, *n* (%)				0.0513
C	15 (68.2)	8 (53.3)	7 (100.0)	
D	7 (31.8)	7 (46.7)	0 (0.0)	
Indication for FMT, *n* (%)				0.3182
Steroid Resistant	21 (95.5)	15 (100.0)	6 (85.7)	
Steroid Dependent	1 (4.5)	0 (0.0)	1 (14.3)	
GvHD onset to FMT, days (med. [IQR])	14.5 [10.0, 26.5]	11.0 [85., 15.0]	22.0 [18.5, 45.0]	0.0149
Steroid dose prior to FMT, mg/kg (med. [IQR])	1.6 [1.0, 2.0]	1.8 [1.4, 2.0]	1.0 [1.0, 1.5]	0.0540

* p-value for the comparison of bacteremic to non-bacteremic patients. Fisher's Exact Test for categorical variables; Wilcoxon Test for continuous variables.

BSI - bloodstream infection; med. - median; IQR - intraquartile range; FMT - fecal microbiota transplantation; HSCT - hematopoietic stem cell transplantation; HLA - human leukocyte antigen; ATG - antithymocyte globulin; GvHD - graft-versus-host-disease; CSA - cyclosporine A; MTX - methotrexate; MMF - mycophenolate mofetil; PTCy - post-transplant cyclophosphamide; GI - gastrointestinal; IBMTR - interational bone marrow transplantation registry; TPN - total parenteral nutrition.

Table S2. Summary of assembly statistics and information for the genomes recovered from the blood culture samples

Sample	Patient	Species	Days from last FMT	Assembly size	# scafs	Completeness (%)	Contamination (%)
1	9	*Pseudomonas aeruginosa*	2	6,761,798	55	99.68	0.49
2*	8	*Escherichia coli*	24	5,001,133	120	99.32	0.22
3*	8	*Escherichia coli*	5	5,171,391	124	99.01	0.35
4B	8	*Brevibacterium frigoritolerans*	16	5,091,565	90	97.81	1.39
4E	8	*Enterococcus faecium*	16	2,655,734	106	98.5	0
5	12	*Escherichia coli*	4	4,792,771	76	99.67	0.04
6	12	*Acinetobacter baumannii*	17	3,576,220	75	95.86	0.66
7	13	*Enterococcus faecium*	10	2,565,989	109	98.5	0.12
8	13	*Mycolicibacterium mucogenicum*	1	6,195,761	158	99.37	0.31
9	10	*Campylobacter jejuni*	4	1,616,004	29	99.85	0.85
10**	1	*Enterococcus faecium*	1	2,586,232	29	99.63	0.06
11**	1	*Enterococcus faecium*	8	2,618,694	32	99.63	0.06
12	1	*Enterococcus faecium*	64	2,658,282	118	99.63	0.00
13	R01	*Escherichia coli*	2	4,693,495	60	99.32	0.1

*The genomes recovered from samples 2 and 3, both for patient 8, were found to represent the same strain based on genome comparison analysis. **Samples 10 and 11 (patient 1) contain the genome of the same strain.

FMT – fecal microbiota transplantation

Table S3. **BSI occurring up to 30 days from the last FMT**

BSI type	No. of events
Enterococcus faecium	10
Escherichia coli	6
Klebsiella pneumoniae	3
Pseudomonas aeruginosa	3
Campylobacter jejuni	2
Acinetobacter baumannii	2
Mycobacterium mucogenicum	1
Staphylococcus epidermidis	1
Enterococcus urealyticum	1
Stenotrophomonas	1
Streptococcus salivarius	1
Streptococcus mitis	1
Enterobacter cloacae	1

BSI – bloodstream infection; FMT – fecal microbiota transplantation

Table S4. Information for the metagenomics samples. Day/Month/Year refers to the collection date. Npairs - number of paired-end reads. Nbps - number of basepairs. No-human columns refer to the data after the removal of human reads. FMT – fecal microbiota transplantation

Patient/ donor number	Day	Month	Year	Sample type	Sample ID	Npairs	Nbps	Npairs/ no-human	Nbps/ no-human	Pcnt/ no-human
R001	NA	NA	NA	Patient's stool	10	19,479,657	2,941,428,207	1,695,269	255,985,619	8.70
9	21	10	2018	Patient's stool	13	3,163,800	955,467,600	3,066,880	926,197,760	96.94
8	25	9	2018	Patient's stool	14	3,040,735	918,301,970	2,390,646	721,975,092	78.62
9	2	10	2018	Patient's stool	15	8,884,621	2,683,155,542	8,860,761	2,675,949,822	99.73
9	10	10	2018	Patient's stool	16	6,884,304	2,079,059,808	6,787,780	2,049,909,560	98.60
9	13	10	2018	Patient's stool	17	3,066,262	926,011,124	3,033,361	916,075,022	98.93
8	21	10	2018	Patient's stool	18	5,627,248	1,699,428,896	5,581,479	1,685,606,658	99.19
8	1	10	2018	Patient's stool	19	2,405,281	726,394,862	2,107,476	636,457,752	87.62
8	16	9	2018	Patient's stool	20	11,800,852	3,563,857,304	2,240,537	676,642,174	18.99
8	17	9	2018	Patient's stool	21	3,910,127	1,180,858,354	370,897	112,010,894	9.49
12	21	3	2019	Patient's stool	27	3,910,817	1,181,066,734	3,503,716	1,058,122,232	89.59
12	17	2	2019	Patient's stool	28	13,433,187	4,056,822,474	1,040,129	314,118,958	7.74
13	15	3	2019	Patient's stool	29	4,540,809	1,371,324,318	976,665	294,952,830	21.51
12	24	2	2019	Patient's stool	30	4,821,669	1,456,144,038	511,746	154,547,292	10.61
12	9	3	2019	Patient's stool	31	3,503,362	1,058,015,324	2,783,716	840,682,232	79.46
10	8	1	2019	Patient's stool	32	13,731,810	4,147,006,620	4,090,319	1,235,276,338	29.79
13	1	3	2019	Patient's stool	34	4,816,162	1,454,480,924	959,267	289,698,634	19.92
13	28	2	2019	Patient's stool	35	15,929,915	4,810,834,330	1,357,449	409,949,598	8.52
10	11	1	2019	Patient's stool	36	4,279,935	1,292,540,370	575,792	173,889,184	13.45
10	25	12	2018	Patient's stool	37	2,523,110	761,979,220	1,945,823	587,638,546	77.12
10	23	12	2018	Patient's stool	38	20,727,769	6,259,786,238	14,443,740	4,362,009,480	69.68
1	9	7	2017	Patient's stool	51	1,447,523	726,656,546	619,049	310,762,598	42.77
1	20	7	2017	Patient's stool	54	1,649,242	827,919,484	468,595	235,234,690	28.41
1	24	7	2017	Patient's stool	55	1,405,249	705,434,998	1,311,023	658,133,546	93.29
1	26	7	2017	Patient's stool	56	1,061,451	532,848,402	928,889	466,302,278	87.51
1	17	8	2017	Patient's stool	59	1,536,619	771,382,738	1,535,635	770,888,770	99.94
ELK6				Donor capsule/R	22	47,753,847	14,421,661,794	47,749,144	14,420,241,488	99.99
ELK6				Donor capsule	22	10,955,722	3,308,628,044	10,898,573	3,291,369,046	99.48
ALM01				Donor capsule /R	23	44,423,901	13,416,018,102	44,314,452	13,382,964,504	99.75
ALM01				Donor capsule	23	8,190,871	2,473,643,042	8,168,428	2,466,865,256	99.73
ALM05				Donor capsule	24	10,132,889	3,060,132,478	10,101,270	3,050,583,540	99.69
ALM07				Donor capsule /R	25	48,556,184	14,663,967,568	48,376,314	14,609,646,828	99.63
ALM07				Donor capsule	25	13,777,535	4,160,815,570	13,717,516	4,142,689,832	99.56
ARK05				Donor capsule	26	4,307,738	1,300,936,876	4,256,814	1,285,557,828	98.82
EZ01				Donor capsule	64	1,260,720	632,881,440	1,260,456	632,748,912	99.98
ELK07				Donor capsule	65	1,582,327	794,328,154	1,581,338	793,831,676	99.94
CR001				Donor capsule	12	17,496,725	5,284,010,950	17,427,893	5,263,223,686	99.61

Supplementary Methods

Patients and study design

This was an extension of a single-arm, non-randomized, pilot study (NCT# 03214289) evaluating the safety and efficacy of fecal microbiota transplantation (FMT) for patients with steroid-resistant or steroid-dependent lower gastrointestinal (LGI) acute graft-versus-host disease (aGVHD). A total of 22 patients from two centers (Chaim Sheba Medical Center, Tal Hashomer, Ramat Gan, Israel [n=17] and Rambam Health Care Campus, Haifa, Israel [n=5]) were enrolled.

Patients included met the following criteria:

1) Age ≥ 16 years (initially 18 years, amended during the study) and ≤ 75 years

2) Presence of active LGI aGVHD. The diagnosis of aGVHD could be made using standard clinical criteriawhere in doubt, a biopsy was performed.[1-3]

3) Meet criteria for steroid-resistant or steroid-dependent LGI aGVHD. Steroid-resistant LGI aGVHD was defined as intestinal GVHD that did not improve within seven days after initial steroid therapy (≥1 mg/kg of methylprednisolone) or had progressed after five days of treatment. Steroid-dependent LGI aGVHD was defined as repeated recurrence of LGI aGVHD during steroid tapering.[4]

Donors recruitment and FMT procedure

Donors were unrelated healthy, non-pregnant adults aged 18 to 50 years, taking no medications, with normal body mass index. Volunteerswere excluded for any significant medical history or for the use of any antibiotics in the six months preceeding stool collection. Candidates werescreened with the American Association of Blood Banks donor questionnaire and underwent a physical

examination and general laboratory screening tests as described previously, in accordance with the guidance of the Israeli Ministry of Health for FMT.[5] Additional testing for capsules administered in this immunocompromised population included polymerase chain reaction testing for cytomegalovirus and screening of stool aliquots for vancomycin-resistant enterococci, methicillin-resistant staphylococcus aureus, and carbapenemase-resistant Enterobacteriaceae (CRE).

Capsules were produced at the Center for Microbiome Research at Shamir Medical Center, Israel, following a previously described protocol.[6] Briefly, a fecal suspension was generated in normal saline. Materials were sequentially sieved to remove particulate material. The final slurry was concentrated by centrifugation and re-suspended in saline at 1/10 volume of the initial sample with 20% glycerol added as a bacterial cryoprotectant. Fecal matter solution was pipetted into size 0 capsules (650μL), which were closed and then secondarily sealed in size 00 capsules (DRCaps, Capsugel, Cambridge, MA, USA). Each inoculum was prepared from the feces of a single donor, and a dose of 15 capsules contained sieved, concentrated material derived from a mean of 18 grams of fecal matter. Capsules were stored frozen at -80^0C before use.

Participants received a course of oral FMT, which consisted of 15 capsules per day for two consecutive days (total of 30 capsules). In patients with no or partial response to the FMT, an FMT could be repeated from the same or a different donor at the treating physician discretion. Participants were asked to fast for 4 hours before and 1 hour following capsule intake. Capsules were taken with sips of water. There was a recommendation to stop any form of anti-bacterial antibiotic within 48 hours before and after FMT administration. Subjects were followed for six months post-FMT.

Bacterial DNA extraction, 16S amplification, and sequencing.

DNA was extracted from 100 mg of stool using PureLink Microbiome DNA Purification Kit (Invitrogen, Carlsbad, California, USA) according to the manufacturer's instructions, following a 2-minute bead-beating step (BioSpec). Following DNA extraction, the V4 region of the bacterial 16S gene was amplified by Polymerase chain reaction (PCR), using the barcoded 515F (10 μM) and 806R (10 μM) primers. PCR reactions were performed with PrimeStar max PCR mix (Takara, Mountain View, California, USA), for 30 cycles of denaturation (98°C for 10sec), annealing (55°C for 5sec), and extension (72°C for 20sec), with final elongation at 72°C (1min). Amplicons were purified (AMPure, Beckman Coulter, Brea, California, USA), quantified by the PicoGreen assay (Molecular Probes, Eugene, OR), and pooled at equal concentrations (30 ng/μL). Finally, pooled samples were purified using 2% E-Gel (Invitrogen, Carlsbad, CA) and sequenced using the Illumina MiSeq platform at the Genomic Center, Azrieli Faculty of Medicine, Bar-Ilan University, Israel.

Library Prep and shotgun Sequencing

DNA libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina) and Nextera Index Kit (Illumina) with total DNA input of 1ng. Genomic DNA was fragmented using a proportional amount of Illumina Nextera XT fragmentation enzyme. Dual indexes were added to each sample followed by 12 cycles of PCR to construct libraries. DNA libraries were purified using AMpure magnetic Beads (Beckman Coulter) and eluted in QIAGEN EB buffer. Finally, library quantity was assessed with Qubit (ThermoFisher) and sequenced on an Illumina HiSeq platform 2x150bp.

16S rRNA gene sequence analysis

Data analysis was performed using the Quantitative Insights Into Microbial Ecology software (QIIME2).[7] Paired-end sequences were demultiplexed by per-sample barcodes and error-corrected by Divisive Amplicon Denoising Algorithm (DADA2).[8] Final feature sequences were aligned against Greengenes database[9] with 99% confidence for taxonomic annotation. We filtered out features with low abundance (<0.001%) as well as mitochondria and chloroplast sequences.

Genome assembly of blood culture samples

Samples were assembled using Spades v3.10.1 with default parameters[10]. To remove potential foreign DNA sequences, we discarded all scaffolds whose coverage was out of the range of [0.6, 1.4]*(median coverage of the assembly). Genes and proteins were predicted using Prodigal v2.6.3 with default parameters[11]. Genome completeness and percent contamination were determined using Checkm[12]. To assign taxonomy to the genomes, we aligned all the genome scaffolds against the NCBI's nt database using blastn[13] and searched for the species whose genomes covered the highest number of sequences. In all cases, the best hit's alignment quality was sufficient for determining taxonomy at the species level.

Identification of strains in the metagenomics samples

We applied a similar approach to the one described in references 14-15.[14,15] The approach is based on searching species and strain genes identified in the reference genome in the metagenomics samples. In addition, we added a step designed to assist the identification at low coverage.

Identifying the sets of core and strain genes

Given the reference genome derived from the blood culture samples and its species affiliation, we first collected a set of genomes from the same species from NCBI. Fifty genomes were chosen from this group based on average nucleotide identity (ANI) scores such that the ANI score between the NCBI genomes and the reference will be at least 95, but the score between any two NCBI genomes will be 99 at most. ANI scores were calculated using fastANI with default parameters.[16] Next, we clustered the genes of the NCBI + reference genomes using usearch (80% identity).[17] All clusters that contained representatives from at least 85% of the genomes but with no more than 10% redundancy were defined as core species genes. Clusters that included representatives from no more than 20% of the genomes with at least one gene from the NCBI set and one from the reference genome were considered as strain genes.

Calculating the fraction of species and strain genes present in the metagenomics sample

We aligned each metagenomics sample's reads against the reference genome using bowtie2 v2.3.4.1 with parameters -X 1000 -sensitive.[18] Each gene to which at least one read was aligned is considered to be present. The fraction of genes from the species/strain sets that were detected in a metagenomics sample was used to determine whether the species and strain are present in the sample as described next.

Determining species presence in a metagenomics sample

The presence of a species in a metagenomics sample was determined by comparing the fraction of species genes found in the sample to the expected fraction for the average coverage of the species genes in the sample. To compute the expected fraction, we created simulated datasets of reads of varying sizes by subsampling the sample from which the reference genome was recovered. Next,

we applied the aforementioned procedure to calculate the fraction of species and strain genes in each simulated set. A species was said to be present in a sample if its fraction of species genes fell within the expected range of fractions for its coverage. Refer to Supplementary File S1 for complete information for all samples analyzed.

Determining strain presence in a metagenomics sample

We applied two criteria: (i) the fraction of strain genes detected vs. the fraction of species genes, and (ii) the ratio between the number of genes detected when considering reads that mapped to the genome at edit distance ≤ 5 and the number of genes detected using edit distance ≤1. This criterion is useful for low-coverage cases. In both cases, we used simulated data to generate a range of expected values.

We ignored cases where less than 30% of the core genes were identified (coverage <~0.2x). Cases in which a genome passed the criteria for same species and same strain with the fraction of core genes ranging between 30% to 70% (~0.2x<coverage <~0.4x) were labeled inconclusive because analyses using close strains showed that cases falling in this range are likely to be correct (results not shown). Samples in which the reference genome passed the criteria for the same species and same strain and the fraction of core genes are higher than 70% (coverage >~0.4x) were determined to contain the reference strain.

Calculating relative abundance of blood culture species in metagenomic samples

Community composition in the metagenomics samples was evaluated using metaphlan2 v2.7.5.[19] For samples in which blood culture strains were detected, we recovered the relative abundance of their species.

Inconclusive identification of target stains

We applied our detection pipeline to the FMT capsules and patient stool samples using the assembled bacterial genomes. Overall, blood culture strains were conclusively detected in 29 metagenomics samples. In four fecal and three capsule samples, we could not conclusively determine whether the target strains were present (Supplementary file S1). For the three capsule samples (Table S4), we re-sequenced the respective metagenomic samples at ~3.5-5.5 times the original sequencing depth and repeated the analysis. This time the strains were clearly absent from the samples. We attribute the initial inconclusive results to cross-contamination between samples that sometimes occur in Illumina multiplex sequencing.[20] Nevertheless, this demonstrates the sensitivity and accuracy of our bioinformatics pipeline.

Limitations of the strain identification method

Here we evaluate the limitations of our strain-detection approach. We consider two main factors that may affect the method's sensitivity and accuracy: sequencing depth and strain variation.

Insufficient sequencing depth may result in false negatives. We estimate that coverage of >0.4x is required for the target genome to be reliably detected because at this coverage, roughly ~50% of the core genes and ~40-50% of the strain genes are typically detected if the target strain is present in the sample (see Supplementary file 1, core genes vs. coverage and core genes vs. Strain genes figures). We considered coverage between 0.2x to 0.4x to be marginal; samples in which such coverage was detected were resequenced to achieve higher coverage. Very few strain genes are typically detected for coverage below 0.2x, which makes the results unreliable. We estimate that in human samples, a coverage of 0.2x translates into a relative abundance of <0.1% of the community for a sequencing depth of ~1Gbp as follows: assuming an average microbial genome

size of ~3Mbp in stool communities, a coverage of 0.2x yields 600Kbp. For a sample sequencing depth of 1Gbp, this corresponds to 0.06% of the community. Considering that some of the sequenced DNA may originate from phages and human DNA (typically <25% of the DNA in our samples), the actual relative abundance is closer to 0.1%. Detection can be improved by increasing the coverage (e.g., 10Gbp should provide a 0.01% detection sensitivity). To conclude, in the current study, our method could typically detect the target genome if its relative abundance in the sample was 0.1% or higher.

Strain variation may lead to false positives. When multiple strains of the target genome's species are present, their combined gene pool may include most of the target strain's strain-specific genes and result in false detection. Strain genes selected in this study are present in <20% of the strains of the target genome's species. Therefore, assuming a random strain includes <20% of the strain-specific genes, multiple random strains may include most of the strain-specific genes and lead to false detection of the target genome. However, at least seven strains with >0.4x coverage are required to cover ~80% of the complete set of strain genes in a sample ($1-0.8^7 = 0.21$). Given that a single strain in gut metagenomes typically represents most species,[21] these cases are expected to be very rare. Moreover, this issue can be addressed by identifying cases in which the average coverage of the core genes is significantly higher than the coverage of the strain genes or when the variation in coverage of the strain genes is very high.

Data availability

All 16S rRNA data and metagenomic raw reads are currently under submission to EBI and NCBI-SRA, respectively.

References

.1 Dignan FL, Clark A, Amrolia P, et al. Diagnosis and management of acute graft-versus-host disease. *Br J Haematol*. 2012;158(1):30-45.

.2 Firoz BF, Lee SJ, Nghiem P, Qureshi AA. Role of skin biopsy to confirm suspected acute graft-vs-host disease: results of decision analysis. *Arch Dermatol*. 2006;142(2):175-182.

.3 Weisdorf DJ, Hurd D, Carter S, et al. Prospective grading of graft-versus-host disease after unrelated donor marrow transplantation: a grading algorithm versus blinded expert panel review. *Biol Blood Marrow Transplant*. 2003;9(8):512-518.

.4 Ruutu T, Gratwohl A, de Witte T, et al. Prophylaxis and treatment of GVHD: EBMT-ELN working group recommendations for a standardized practice. *Bone Marrow Transplant*. 2014;49(2):168-173.

.5 Sayer HG, Kroger M, Beyer J, et al. Reduced intensity conditioning for allogeneic hematopoietic stem cell transplantation in patients with acute myeloid leukemia: disease status by marrow blasts is the strongest prognostic factor. *Bone Marrow Transplant*. 2003;31(12):1089-.1095

.6 Youngster I, Russell GH, Pindar C, Ziv-Baran T, Sauk J, Hohmann EL. Oral, capsulized, frozen fecal microbiota transplantation for relapsing Clostridium difficile infection. *JAMA*. 2014;312(17):1772-1778.

.7 Bolyen E, Rideout JR, Dillon MR, et al .Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature biotechnology*. 2019;37(8):852-857.

.8 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods*. 2016;13(7):581-583.

.9 DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*. 2006;72(7):5069-5072.

.10 Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*. 2012;19(5):455-477.

.11 Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*. 2010;11(1):1-11.

.12 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 research*. 2015;25(7):1043-1055.

.13 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990;215(3):403-410.

.14 Suez J, Zmora N, Zilberman-Schapira G, et al. Post-antibiotic gut mucosal microbiome reconstitution is impaired by probiotics and improved by autologous FMT. *Cell*. 2018;174(6):1406-1423. e1416.

.15 Zmora N, Zilberman-Schapira G, Suez J, et al. Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. *Cell*. 2018;174(6):1388-1405. e1321.

.16 Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications*. 2018;9(1):1-8.

.17 Chen C-Y, Pollack S, Hunter DJ, Hirschhorn JN, Kraft P, Price AL. Improved ancestry inference using weights from external reference panels. *Bioinformatics*. 2013;29(11):1399-1406.
.18 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357.
.19 Truong DT, Franzosa EA, Tickle TL, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature methods*. 2015;12(10):902-903.
.20 Kircher M, Sawyer S ,Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. 2012;40(1):e3.
.21 Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome research*. 2017;27(4):626-638.