## Supplementary Information

### Supplementary Results: Patient Demographics

Table S1: Summary of demographics for 22 patients included in efficacy statistics

| SupplementaryTablesS1-8.xlsx |
| --- |

### Supplementary Results: Adverse Events

Table S2: Summary of adverse events for all patients enrolled. Numbers indicate number of patients experiencing one or more events of each type. Where events are on a three point scale, higher values mean more severe events.

| SupplementaryTablesS1-8.xlsx |
| --- |

### Supplementary Tables: Samples and Microbiome Profiles Collected by Patient

Table S3: List of fecal samples collected.

| SupplementaryTablesS1-8.xlsx |
| --- |

Table S4: List of 16S rRNA gene libraries sequenced.

| SupplementaryTablesS1-8.xlsx |
| --- |

Table S5: List of shotgun metagenomic libraries sequenced.

| SupplementaryTablesS1-8.xlsx |
| --- |

Table S6: List of metabolomic profiles collected.

| SupplementaryTablesS1-8.xlsx |
| --- |

### Supplementary Results: Patients Deviating from Initial Study Design

Table S7: Summary of relevant deviations from main protocol.

| SupplementaryTablesS1-8.xlsx |
| --- |

### Supplementary Data: Ordinations of microbiome profiles colored by patient

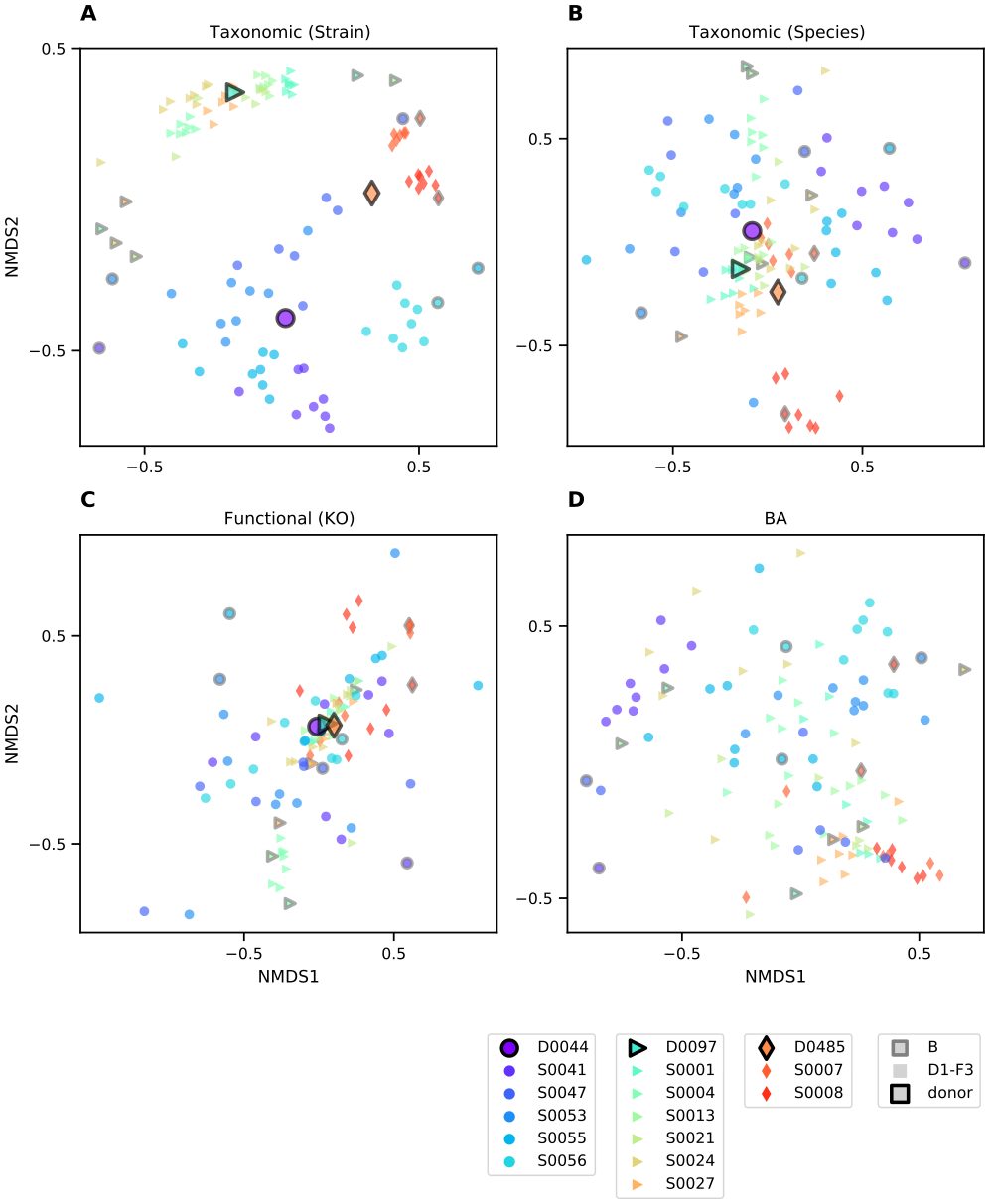


Figure S1: Clustering of taxonomic, functional, and BA profiles across time points. Ordinations are calculated and plotted as in Fig. 2. For **(A-C)** datatypes where donor samples were also profiled, larger points with black outlines represent the mean of all samples from that donor. Samples from each subject are differentiated by color and shape as indicated in the legend (bottom left). Patients’ baseline samples are outlined in black. The same four ordinations colored by assigned donors are available as Fig. 2A-D.

### Supplementary Data: Shared ASVs/Species/Strains across Donors

Comparing taxa across donors presents one way to evaluate the effect of taxonomic resolution on the sensitivity of tracking transfers between individuals. That very few strains are found to be shared across donor communities (see Supplementary Fig. S2) indicates that false positives—non-transfer of a donor strain incorrectly inferred to be a transfer—are infrequent.

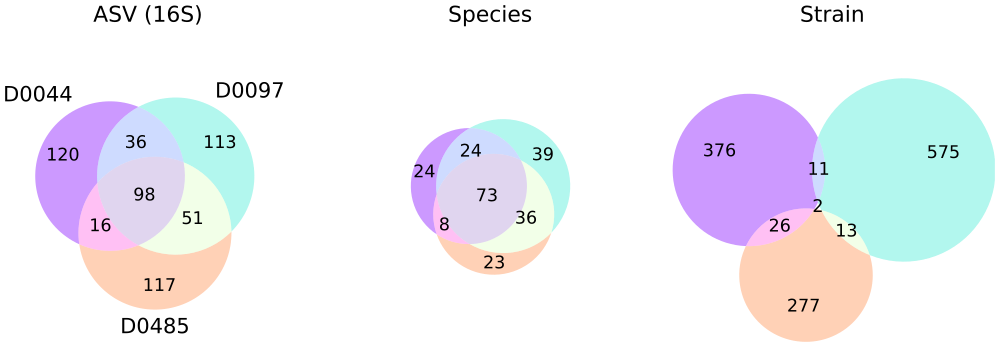


Figure S2: Overlap between taxa detected in three donors shows that strain-level taxonomic resolution increases sensitivity and specificity of engraftment detection. Venn diagrams depict relationships between sets of taxa detected in donor samples, and numbers indicate the size of the respective sets. Taxa were considered to be detected in a donor if their mean relative abundance across that donor’s samples was greater than 0.01%. Circles are colored by donor as in Fig. 2. Species composition was estimated from metagenomes based on the mean coverage reported by GT-PRO[36](#ref-Shi2020), and was further partitioned into strain composition based on haplotype deconvolution with Strain Finder[22](#ref-Smillie2018).

### Supplementary Methods: Stool Sampling Instructions to Patients

Instructions to patients for stool self-sampling.

SelfSampling.pdf

### Supplementary Methods: Clinical Laboratory Assessments

Table S8: List of clinical laboratory assessments

| SupplementaryTablesS1-8.xlsx |
| --- |

### Supplementary Methods: Sample Processing

#### Sample collection and DNA extraction

Patients were instructed to collect a stool sample into provided sample vials (Sarstedt faeces tubes cat. no. 80.734.311) before each study visit and either bring it refrigerated to our clinic in-person or send it with a frozen cool pack via overnight courier. Upon delivery, stool specimens were stored at -80 °C until their analysis.

DNA extraction from fecal samples was performed using the modified cetyltrimethylammonium bromide (CTAB) as previously described[64](#ref-Piceno2020). In brief, ~0.3 g aliquot was taken from each frozen stool sample and suspended in 500 µL CTAB extraction buffer in a Lysing Matrix E tube (MP Biomedicals) by vortexing, followed by incubation for 15 minutes at 65 °C. After adding 500 µL phenol:chloroform:isoamyl alcohol (25:24:1), the solution underwent bead-beating (5.5 m/s for 30 seconds), followed by centrifugation (16,000 g for 5 minutes at 4 °C). The resulting aqueous phase (approximately 400 µL) was transferred to a new 2 mL 96-well plate. An additional 500 µL CTAB extraction buffer was added to the fecal aliquot and both the bead-beating and centrifugation steps were repeated, resulting in approximately 800 µL. An equal volume of chloroform was added and the solution was mixed and centrifuged (3,000 g for 10 minutes). The aqueous phase (approximately 600 µl) was transferred to another 2 mL 96-well plate, combined with 2-volume polyethylene glycol and stored at 4 °C overnight to precipitate DNA. Samples were then centrifuged (3,000 g for 60 minutes) to pellet DNA, washed twice with cold 70% EtOH, resuspended in sterile water and diluted to 10 ng DNA/µL (Qubit dsDNA BR Assay Kit; ThermoFisher Scientific, MA)

#### Sequencing

For 16S rRNA gene profiling, The V4 region was amplified as previously described[64](#ref-Piceno2020). PCR reactions were performed with 0.625 U Hot Start ExTaq and 1x buffer, 200 µM dNTPs, 0.56 µL/µL BSA, 0.4 µM each forward (F515) and reverse (R806) primers in triplicate 25 µL reactions containing 10ng of template gDNA. Thermal cycling was set at: 98 °C for 2 minutes, 30 rounds of 98 °C for 20 sec, 50 °C for 30 sec, 72 °C for 45 sec, and a final extension at 72 °C for 10 minutes. Amplicons were normalized (SequalPrep Normalization Plate Kit; ThermoFisher Scientific, MA), quantified (Qubit dsDNA BR Assay Kit; ThermoFisher Scientific, MA), pooled in equimolar concentrations, purified (Agencourt AMPure XP System; Beckman-Coulter), quantified (KAPA Library Quantification Kit; KAPA Biosystems), and diluted to 2 nM. Equimolar PhiX spike-in control was added at 40% final volume, and the samples were sequenced on an Illumina NextSeq 500 Platform.

For metagenomic sequencing, an aliquot of the extracted DNA was sent to QB3 at the University of California, Berkeley <<https://qb3.berkeley.edu/>> for sequencing on the NovaSeq 6000 platform using the 150PE Flow Cell S4 format.

#### Metabolomics

For metabolomics profiling, 112 fecal samples (200 mg) were provided to Metabolon (Durham, NC) who performed Ultrahigh Performance Liquid Chromotography/Tandem Mass Spectrometry (UPLC-MS/MS) and Gas Chromatography-Mass Spectrometry (GC-MS) based on standardized published protocol <<http://www.metabolon.com/>>. Detected molecules were identified through Metabolon’s library of purified standards, which encompasses >3,300 commercially available compounds. This yielded 1050 distinct metabolites among the samples, from which the 51 metabolites identified as primary and secondary bile acids were analyzed for this study. For each identified BA, peak intensities were normalized by the root mean squared intensity of that peak across samples before downstream analysis.

### Supplementary Methods: Microbiome profiling

#### 16S rRNA gene amplicon analysis

Raw amplicon sequencing data was processed as in[64](#ref-Piceno2020). Briefly, sequencer output was converted to fastq with bcl2fastq v2.16.0.10 <<https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html>> and demultiplexed by barcode with QIIME[65](#ref-Caporaso2010). Quality-filtering via DADA2[66](#ref-Callahan2015) was performed using the recommended settings with the following adjustments: maximum expected errors allowed ≤ 3, truncation length of 150 bases for R1 and 140 for R2; chimeric sequences were found using minFoldParentOverAbundance = 8. Taxonomy was assigned with SILVA database V132[67](#ref-Quast2013),[68](#ref-Yilmaz2014) to amplicon sequence variants (ASV). Contaminant ASV were identified with the decontam[69](#ref-Davis2018) package for R. ASV with >2% of total read sums contributed by controls and those with < 0.001% total read count across all samples were excluded. A phylogenetic tree was constructed using the phangorn[70](#ref-Schliep2011),[71](#ref-Schliep2016), msa[72](#ref-Bodenhofer2015), and DECIPHER[73](#ref-Wright2016) packages for R. Amplicon sequencing resulted in an average of 3.4e5 paired reads per sample after processing.

#### Metagenomic reads pre-processing

Metagenomic sequences were deduplicated using FastUniq[74](#ref-Xu2012), any contaminating adapter sequence removed using Scythe[75](#ref-Buffalo2018), and then quality trimmed with Sickle[76](#ref-Joshi2011). Any cleaned reads that mapped to the human reference genome (GRCh38[77](#ref-Schneider2017)) using Bowtie2[78](#ref-Langmead2012) were removed. Shotgun metagenomic sequencing resulted in 4.9e7 paired reads per sample after processing.

#### Functional gene profiling

Reads were annotated with presumed functions by first identifying homology to the UHGP-50 (50% identity clusters,[59](#ref-Almeida2021)) reference database using DIAMOND[79](#ref-Buchfink2014), a fast implementation of the BLASTX algorithm[80](#ref-Altschul1990). All top hits to this database were tallied for each sample. UHGP-50 hits were subsequently annotated with KEGG Orthology (KO) numbers based on assignments previously generated for the UHGP using EggNOG-mapper[81](#ref-Huerta-Cepas2017). Tallies of reads mapping to these annotations were not corrected for gene length.

## References

64. Piceno, Y. M. *et al.* Reduced Mucosal Inflammation and Select Alterations in the Viable Mucosa-Associated Microbiota Observed in Clinically-Responsive Ulcerative Colitis Patients Following A Single Fecal Microbiota Transplantation in a Pilot Study Evaluating Safety and Efficacy. *J. Clin. Nutr. Food Sci.* **3**, 27 (2020).

65. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

66. Callahan, B. J. *et al.* DADA2 : High resolution sample inference from amplicon data. *bioRxiv* **13**, 0–14 (2015).

67. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

68. Yilmaz, P. *et al.* The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **42**, 643–648 (2014).

69. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).

70. Schliep, K. P. Phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).

71. Schliep, K., Potts, A. A., Morrison, D. A. & Grimm, G. W. *Intertwining phylogenetic trees and networks*. <https://peerj.com/preprints/2054v1> (2016) doi:[10.7287/peerj.preprints.2054v1](https://doi.org/10.7287/peerj.preprints.2054v1).

72. Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C. & Hochreiter, S. Msa: An R package for multiple sequence alignment. *Bioinformatics* **31**, 3997–3999 (2015).

73. Wright, E., S. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal* **8**, 352 (2016).

74. Xu, H. *et al.* FastUniq: A fast *de novo* duplicates removal tool for paired short reads. *PLoS ONE* **7**, 1–6 (2012).

75. Buffalo, V. *Scythe - A Bayesian adapter trimmer*. (2018).

76. Joshi, N. A. & Fass, J. N. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). <https://github.com/najoshi/sickle> (2011).

77. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).

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

79. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).

80. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

81. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution* **34**, 2115–2122 (2017).