

RESEARCH ARTICLE

# Long-term taxonomic and functional divergence from donor bacterial strains following fecal microbiota transplantation in immunocompromised patients

Eli L. Moss<sup>1</sup>✉, Shannon B. Falconer<sup>2</sup>✉, Ekaterina Tkachenko<sup>1</sup>, Mingjie Wang<sup>1</sup>, Hannah Systrom<sup>4</sup>, Jasmin Mahabamunuge<sup>4</sup>, David A. Relman<sup>2,3</sup>, Elizabeth L. Hohmann<sup>4\*</sup>, Ami S. Bhatt<sup>1\*</sup>

**1** Departments of Genetics, and Medicine, Stanford University, Stanford, California, United States of America, **2** Departments of Microbiology and Immunology, and Medicine, Stanford University, Stanford, California, United States of America, **3** Veterans Affairs Palo Alto Health Care System, Palo Alto, California, United States of America, **4** Division of Infectious Diseases, Massachusetts General Hospital; Harvard Medical School, Boston, Massachusetts, United States of America

✉ These authors contributed equally to this work.

\* [ehohmann@partners.org](mailto:ehohmann@partners.org) (ELH); [asbhatt@stanford.edu](mailto:asbhatt@stanford.edu) (ASB)



**OPEN ACCESS**

**Citation:** Moss EL, Falconer SB, Tkachenko E, Wang M, Systrom H, Mahabamunuge J, et al. (2017) Long-term taxonomic and functional divergence from donor bacterial strains following fecal microbiota transplantation in immunocompromised patients. PLoS ONE 12(8): e0182585. <https://doi.org/10.1371/journal.pone.0182585>

**Editor:** Richard E. Isaacson, University of Minnesota, UNITED STATES

**Received:** May 2, 2017

**Accepted:** July 20, 2017

**Published:** August 21, 2017

**Copyright:** © 2017 Moss et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The datasets generated during the current study are available in the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under Bioproject accession PRJNA349197.

**Funding:** National Science Foundation Graduate Research Fellowship ([nsfgrfp.org](https://www.nsfgrfp.org)) DGE-114747 awarded to ELM; American Society of Hematology ([hematology.org](https://www.hematology.org)) Scholar and Amy Strelzer

## Abstract

Immunocompromised individuals are at high risk of developing *Clostridium difficile*-associated disease (CDAD). Fecal microbiota transplantation (FMT) is a highly effective therapy for refractory or recurrent CDAD and, despite safety concerns, has recently been offered to immunocompromised patients. We investigated the genomics of bacterial composition following FMT in immunocompromised patients over a 1-year period. Metagenomic, strain and gene-level bacterial dynamics were characterized in two CDAD-affected hematopoietic stem cell (HCT) recipients following FMT. We found alterations in gene content, including loss of virulence and antibiotic resistance genes. These alterations were accompanied by long-term bacterial divergence at the species and strain levels. Our findings suggest limited durability of the specific bacterial consortium introduced with FMT and indicate that alterations of the functional potential of the microbiome are more complex than can be inferred by taxonomic information alone. Our observation that FMT alone cannot induce long-term donor-like alterations of the microbiota of HCT recipients suggests that FMT cannot indefinitely supersede environmental and/or host factors in shaping bacterial composition.

## Introduction

Fecal microbiota transplantation (FMT) is a remarkably safe and effective therapy for resolving recurrent *Clostridium difficile*-associated disease (CDAD) in immunocompetent individuals [1], and it is increasingly being used to treat CDAD in immunocompromised individuals [2]. Among immunocompromised populations at greatest risk of developing CDAD are hematopoietic stem cell transplantation (HCT) recipients, who experience attack rates as high as 25%,

Manasevit (National Marrow Donor Program (bethematch.org)) award to ASB; National Cancer Institute (cancer.gov) K08 CA184420 to ASB., and the Thomas C. and Joan M. Merigan Endowment at Stanford University to DAR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

and who are up to nine times more likely to suffer from CDAD than immunocompetent individuals [3–5]. A small number of recent studies have reported that FMT can safely resolve CDAD in post-HCT recipients [2,6–9]. However, persisting concerns regarding the safety of administering FMT therapy to immunocompromised patients continue to limit use of the therapy in post-HCT individuals. Better understanding the effects of FMT through investigations of post-treatment gut bacterial community structure in immunocompromised patients is needed.

Several studies leveraging 16S rRNA gene sequencing have begun to reveal the taxonomic underpinnings of FMT therapy by informing on genus- and species-level bacterial dynamics. These studies have demonstrated that FMT induces taxonomic compositional changes in the recipient gut microbiota, resulting in a more donor-like state [10–15]. These analyses have provided a foundation for metagenomic investigation of FMT: there are good reasons for considering whole genomes and strain-level diversity, since bacterial strains with identical 16S rRNA sequences can exhibit wide variability in terms of phenotype and pathogenicity resulting from nucleotide level DNA sequence differences or differing gene content. For example, commensal *Escherichia coli* are abundant facultative anaerobes in the human gut [16], yet enterohemorrhagic *Escherichia coli* are important pathogens. A recent study of patients undergoing FMT for treatment of metabolic syndrome was the first to report strain-level dynamics after FMT. In this study of 10 patients, the investigators found that strains of donor bacteria are capable of both replacing and coexisting with same-species recipient strains at three months post-FMT [17]. These findings are intriguing, but questions remain regarding the generalizability of this finding to patients undergoing FMT for treatment of CDAD where, unlike the metabolic syndrome cohort, patients receive antibiotics prior to FMT and have an underlying dysbiosis. In the setting of prior antibiotic therapy, it is presumed that the total amount and diversity of microorganisms present in the FMT recipient is lower, and thus the gut may represent a relatively barren ecological niche for repopulation with donor-derived microbes. Additionally, no studies have yet reported on strain level engraftment at time points greater than three months after FMT.

To date, the collective understanding of FMT impact on the taxonomic composition of the gut microbiota has been defined by studies of immunocompetent patients. However, as immunocompromised patients are at greatest risk of developing CDAD, there is much to gain from an investigation of FMT in this patient population, in which the impaired host immune system may have diminished influence on the gut microbiota community. Further, although the recent studies relating the success of FMT in immunocompromised individuals have begun to assuage concerns of safety [2,6–9], those concerns are not unfounded [18]. Characterizing the taxonomic and functional capacity of the immunocompromised patient microbiota post-FMT carries potential for both improving clinical outcomes of CDAD in patients lacking a robust immune system and advancing our understanding of the human microbiota landscape during transitions between states of health and disease.

We present eight HCT patients with three or more microbiologically-documented CDAD episodes, all of whom were successfully treated with FMT. For two of these patients, we were able to perform a longitudinal study of microbiota dynamics, employing shotgun metagenomic sequencing of serial stool samples obtained before, immediately after and over one year following FMT. We show that recipient stool assumed donor-like taxonomic and functional composition immediately following FMT, but that functional and taxonomic concordance were diminished after one year. These findings suggest that environmental, lifestyle and other host-factors are important determinants in long-term shaping of the microbiota. Furthermore, we show that despite the divergence of overall taxonomic and functional similarity at >1 year post-FMT, potentially important specific changes to the gene repertoire were retained,

including the loss of pathogenicity genes from the pan-genome of *Escherichia coli* and a reduction of community-wide antibiotic-resistance genes.

## Results

### Treatment outcomes of FMT in eight HCT patients

FMT was delivered as an oral, encapsulated therapy using stool from a healthy unrelated donor (Table 1). All eight of the HCT patients to receive FMT experienced resolution of CDAD at 8 weeks, a standard time frame for determining cure. Subjects G and H died of underlying CNS disease and intracranial hemorrhage at 208 and 101 days following FMT, respectively. Both subjects were negative for CDAD at the time of death. Subjects A, B, C, E and F were free of CDAD at time of final stool sample collection at 408, 384, 456, 448 and 410 days post-FMT, respectively. Subject D experienced a recurrence of CDAD, and tested positive for CDAD at time of final stool sample collection, 179 days post-FMT. FMT is typically performed in patients with CDAD after significant clinical improvement is achieved with antimicrobial therapy. Subject B had active diarrhea at the time of FMT; the remaining five subjects were symptomatically quiescent. FMT was well tolerated in all subjects, and the subject with active diarrhea experienced symptom improvement within days of treatment. No attributable side effects or exacerbation of graft-versus-host disease were observed.

### Short-term concordance and long-term divergence between recipient and donor microbiota following FMT

To better understand the taxonomic impact of FMT therapy in HCT recipients, we employed shotgun metagenomic sequencing of serial stool samples—both pre- and post-FMT—from

Table 1.

Subject	M/F	Age at FMT	Underlying illness	Type of HCT	CDAD history	Date of FMT relative to HCT (days post-HCT)	Rx	Symptoms at time of FMT	Donor #	Recurrence of CDAD post FMT at time of latest sample collection
A	F	56	T-PLL	dUCB-HCT	1,2,4 and 5 m after HCT	196	V	BM character declining off V; no frank diarrhea	29	No
B	F	63	DLBCL with CNS relapse	Auto-HCT	1,2 m before and 1 m after HCT	56	V,Fi	4–5 loose BM/d	29	No
C	F	56	Relapsed AML	Allo-HCT (MRD)	Months after second HCT with 8+ recurrences	1160	M,V, Culturelle	1 loose BM/d	29	No
D	F	39	AML	dUCB-HCT	1 m before and 1,6 m after HCT	577	V	No significant symptoms	28	No
E	M	57	CNS NHL	Auto-HCT	4 m after HCT, and 6+ recurrences	411	V, probiotics	No significant symptoms	29	No
F	F	38	AML (FLT3 ITD+)	dUCB-HCT	3 episodes starting 26 m after HCT	926	M,V	Frequent, formed BM	29	No
G	M	50	DLBCL	Allo-HCT	5 m before, 2 w and 3 m after HCT	83	V	1–3 BM/d	32	N/A
H	M	71	ALL	Allo-HCT (MUD)	4 m after HCT	129	V	3–4 loose BM/d	34	N/A

**List of abbreviations:** FMT (fecal microbiota transplantation); CDAD (*Clostridium difficile*-associated disease); F (female); M (male); T-PLL (T prolymphocytic leukemia); DLBCL (diffuse large B-cell lymphoma); CNS (central nervous system); AML (acute myelogenous leukemia); NHL (non-Hodgkin lymphoma); FLT3 ITD (FLT3 internal tandem duplication); ALL (acute lymphoblastic leukemia); dUCB (double umbilical cord blood); HCT (stem cell transplantation); Auto (autogenic); Allo (allogenic); V (vancomycin); Fi (fidaxomicin); M (metronidazole); BM (bowel movement); d (day); w (week). N/A is indicated for patients who were deceased prior to the time of final long term timepoint stool sample collection.

<https://doi.org/10.1371/journal.pone.0182585.t001>

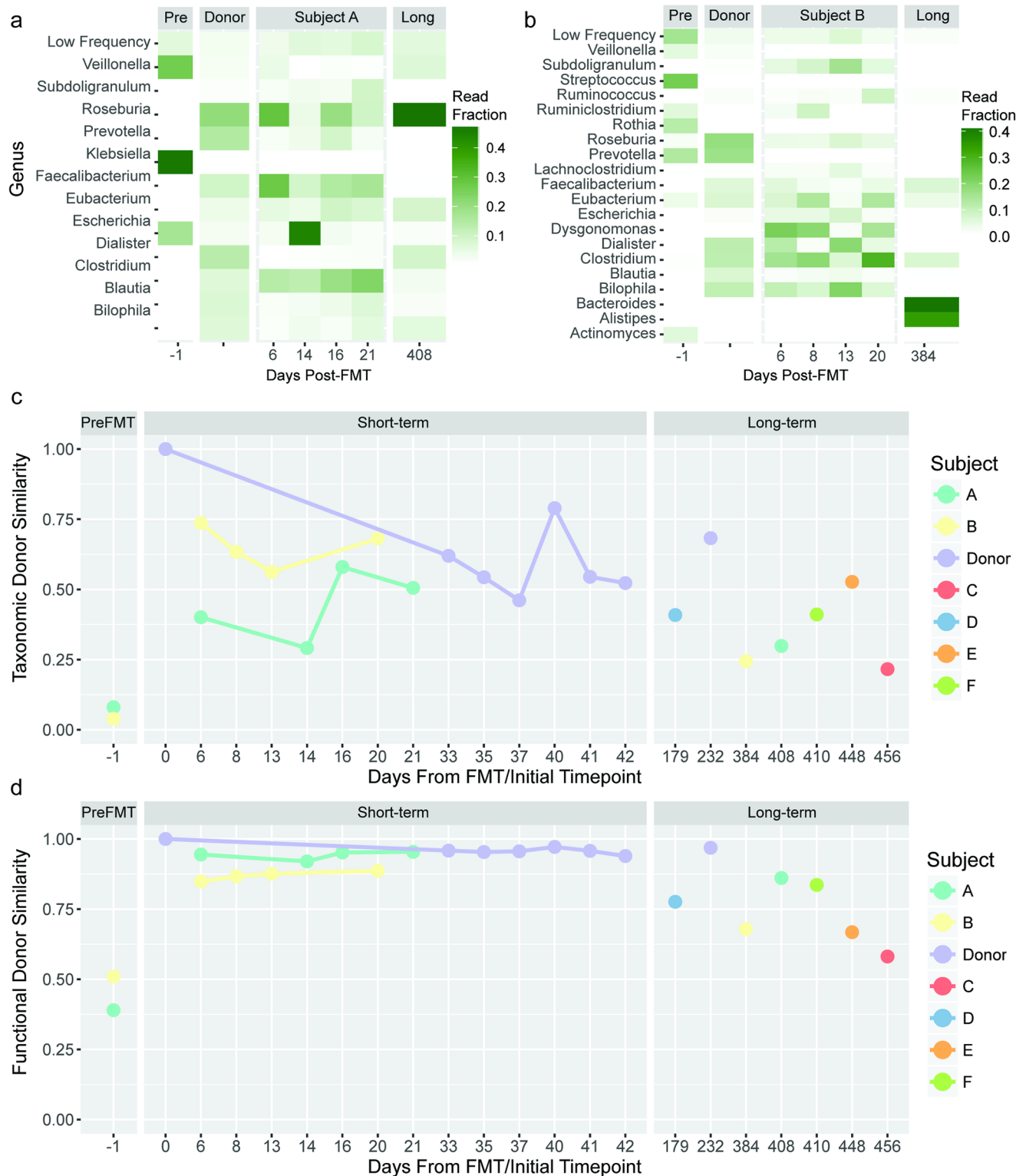
two HCT recipients, and investigated bacterial genus, strain and gene-level changes over a period of 408 and 384 days for Subjects A and B, respectively. Sample collection for Subjects A and B was performed as follows: a single stool sample was obtained immediately before FMT; four stool samples were collected within a short-term period following FMT (on days 6, 14, 16 and 21 for Subject A; and days 6, 8, 13 and 20 for Subject B); and one stool sample was obtained at a long-term timepoint post FMT (days 408 and 384 for Subjects A and B, respectively). We also performed shotgun metagenomic sequencing of one long-term post-FMT sample from subjects C, D, E and F; pre-FMT samples were not available for these individuals. Long-term samples for subjects G and H were not collected as subjects were deceased at time of collection. Samples immediately pre-FMT and within 21 days post-FMT were not available for Subjects C, D, E, F, G and H.

In order to draw parallels between our results and the majority of FMT metagenomic studies to date, we began with an investigation of bacterial composition following FMT at the level of genus, a level of resolution typically targeted with the common approach of 16S rRNA gene amplicon analysis. For both Subjects A (Fig 1a) and B (Fig 1b) we observed a rapid shift towards a more donor-like state for all bacterial members within the most abundant genera. Although donor-concordant patterns of engraftment were found to be sustained for all short-term samples (<21 days), compositional similarity was largely diminished in the long-term (>1 year) timepoints for both Subjects A and B.

To quantitatively assess parity between donor and recipient gut communities, we calculated whole-community similarity of all timepoints for Subjects A and B and long-term timepoints for Subjects C, D, E and F to corresponding donor timepoints at both the levels of genus (Fig 1c, S1 Table) and gene (Fig 1d, S2 Table). We observed a large increase in community-wide donor similarity at both the level of genus and gene in subjects A and B with FMT, with maintenance of elevated similarity during the three week short-term follow-up period. However, at >1 year post-FMT, donor similarity was reduced beyond the range of variability observed in donor self-similarity. As a point of comparison, we also calculated donor self-similarity using multiple stool samples taken over 180 days from donor #29, who provided fecal samples for all subjects aside from D, G and H. Donor self-similarity was calculated as similarity across the donor time series to the initial timepoint. At the taxonomic level, donor self-similarity fluctuated by as much as 33%, though at the level of gene, variations in donor self-similarity were minimal (within 1%). Although stool sampled immediately pre-FMT and within a month post-FMT were not collected for Subjects C, D, E, F, G or H, we were able to obtain samples for Subjects C, E and F at >1 year post-FMT, and for Subject D at ~6 months following FMT. We noted low donor similarity indices in these samples for both genus and gene (Fig 1c and 1d) (S1 and S2 Tables).

### Strain-level genomic remodeling accompanies FMT-induced community-level turnover

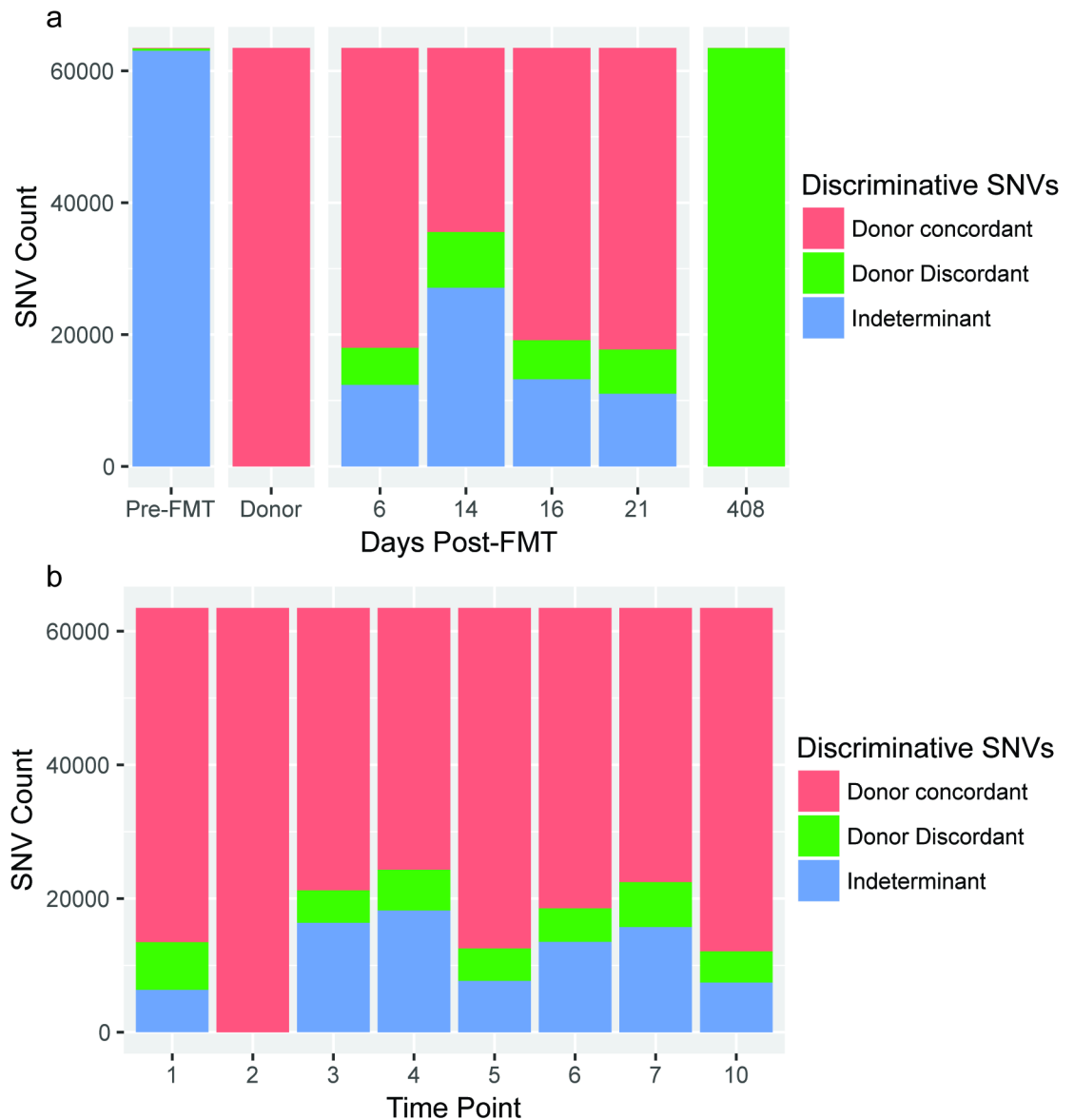
To achieve a higher-resolution perspective of the metagenomic events surrounding FMT, we performed strain-level analysis on a subset of high-abundance taxa. Since the presence of genus *Roseburia* was found to be negligible in Subject A pre-FMT, yet abundant in both the donor and recipient post-FMT (Fig 1a), we sought to investigate the concordance of *Roseburia* strains between donor and Subject A over both the short-term (within 21 days of FMT) and the long-term (>1 year). Using nucleotide polymorphisms to distinguish between strains within the *Roseburia* genus, we observed the ratio of concordant to discordant single nucleotide variants (SNVs) to be similar between Subject A short-term samples post-FMT (Fig 2a), and those from the donor sampled from within a period of a month (Fig 2b). At >1 year post-



**Fig 1. Genus-level taxonomic composition reveals extensive short-term donor microbial engraftment with FMT, followed by long-term reduction in donor similarity relative to donor self-similarity over a similar time period.** Genera present in the donor appear in (a) Subject A and (b) Subject B recipients with inconsistent long-term residence. For visual clarity, genera represented by >5% of assigned reads from at least one timepoint are shown. (c) Whole-community Bray-Curtis donor similarity for Subjects A and B, and an additional four patients (Subjects C, D, E and F) for which long-term and donor samples were obtained, is shown for genus-level composition. Donor similarity is calculated for

each recipient timepoint with the corresponding donor sample used for FMT. As a point of reference, we show multiple timepoints for Donor #29, where similarity is measured against the first timepoint. Microbial sequencing reads were classified using Kraken [29] in conjunction with a sequence database collected from NCBI Refseq and Genbank microbial genome references. (d) Whole-community donor Chao similarity [33] in gene content is shown. Gene abundances were measured by alignment to the Uniref50 functionally annotated protein sequence database [32].

<https://doi.org/10.1371/journal.pone.0182585.g001>



**Fig 2. Dominant Roseburia strain(s) are different in the short versus long term timepoint after FMT.** (a) A high number of single nucleotide variants (SNVs) distinguish the genus Roseburia strains present in the donor FMT sample (day 33 in donor time series) and Subject A recipient shortly after FMT from those in Subject A after one year following FMT. (b) Roseburia strains demonstrate stable, low-level diversity in the donor across eight time points taken over ~8 months, contrasting with the divergence observed in the recipient. Sequences were assembled from short read data obtained at the long-term time-point, and resultant contigs were classified taxonomically (see Methods). Contigs belonging to genus Roseburia were used as a reference for alignment of read data from all timepoints, followed by SNV calling and variable site selection. All polymorphisms distinguishing the donor from the long-term sample were selected, and presence of these discriminant polymorphisms in intermediate timepoints was recorded (adapted from a previously described approach [17]).

<https://doi.org/10.1371/journal.pone.0182585.g002>

FMT, we found the strains present within the genus *Roseburia* in Subject A to be genetically different from those of the donor.

As organisms from the genus *Escherichia* were observed in both donor and Subject A pre-FMT (Fig 1a), we focused on the representative species, *Escherichia coli*, and investigated whether strains post-FMT showed greater concordance with recipient pre-FMT or donor strains of *E. coli*. Using PanPhlAn (see [Materials and methods](#)) to calculate individual gene abundances within the *E. coli* pan-genome, we observed large-scale changes in the gene complement of *E. coli* in the recipient following FMT (Fig 3). Hierarchical clustering identified 719 genes present prior to FMT that disappear within the first two timepoints post-FMT and remain absent through the >1 year timepoint (Fig 3, S1 File). This group is significantly enriched for known pathogenicity genes (Fig 3, S2 File). While there was a loss of many notable *E. coli*-specific virulence genes over time, there was an overall increase in the relative abundance of *E. coli* in the broader microbial community in Subject A post-FMT.

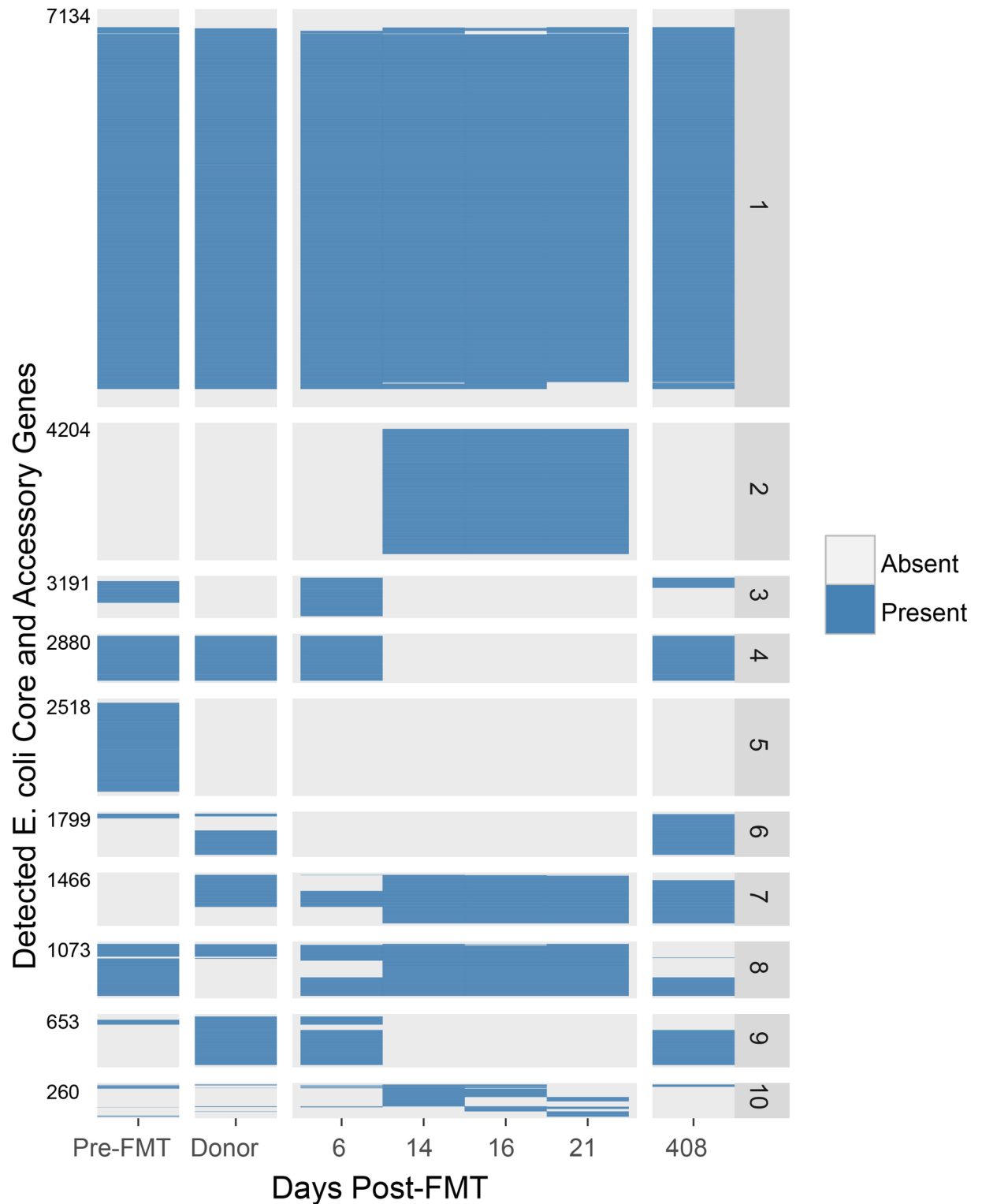
To investigate the effects of FMT on the antibiotic resistome, we measured the abundance of antibiotic resistance genes from donor stool and Subject A both pre- and post-FMT (Fig 4). Subject B was excluded from this analysis due to a large concentration of human DNA pre-FMT, reducing the available microbial data volume. We performed alignment of shotgun data to the Comprehensive Antibiotic Resistance Database [19] to obtain antibiotic resistance gene counts, normalized by total per-sample coverage and aggregated by resistance phenotype. Following FMT, we observed an immediate decrease in the abundance of antibiotic resistance genes by 2.4-fold which remained stable beyond 1 year. In addition, we noted a sustained increase in Chao similarity of unaggregated gene abundance profiles between the resistome profile of Subject A and the donor (S3 Table).

## Discussion

Our study provides one of the first metagenomic analysis of strain- and gene-level bacterial engraftment following FMT, and the first such analysis in patients being treated for CDAD. The success of FMT in resolving CDAD in HCT patients has been demonstrated previously [2,6–9], however a detailed characterization of this “black box” therapy has been lacking. We used a shotgun metagenomic approach to examine both short- and long-term bacterial engraftment following FMT to treat CDAD in HCT patients. We show that although concordance was observed between microbiota composition of the donor and recipient in the short-term, concordance was reduced after 1 year. Previous studies investigating long-term donor similarity have been confined to immunocompetent individuals and have either (1) been limited to 16S rRNA sequencing analysis [13–15] or (2) observed strain-level changes, but only through to three months following FMT. [17]

In the case of the 16S rRNA studies, Broecker et al. [13] investigated family donor FMT in one subject, and reported long-term (>4.5 years) donor similarity but limited short-term (<7 months) concordance. Conversely, we observed short-term similarity and long-term donor discordance. For all eight HCT patients in our study, recipients received stool from a healthy, unrelated donor, in contrast to the method employed by Broecker et al, where fecal material was donated by the sister of the FMT recipient. Our findings that unrelated donors and recipients display dissimilar microbiota profiles at long-term timepoints post-FMT suggest that FMT engraftment is broadly impacted by external effects such as diet, host genetics, host immune surveillance, and subsequent antibiotic exposure. These factors may supersede FMT composition in determining long-term FMT durability.

Also employing 16S rRNA sequencing were studies by 1) Jalanka et al., who used FMT from “universal donors” to treat CDAD in 14 immunocompetent patients, and report on



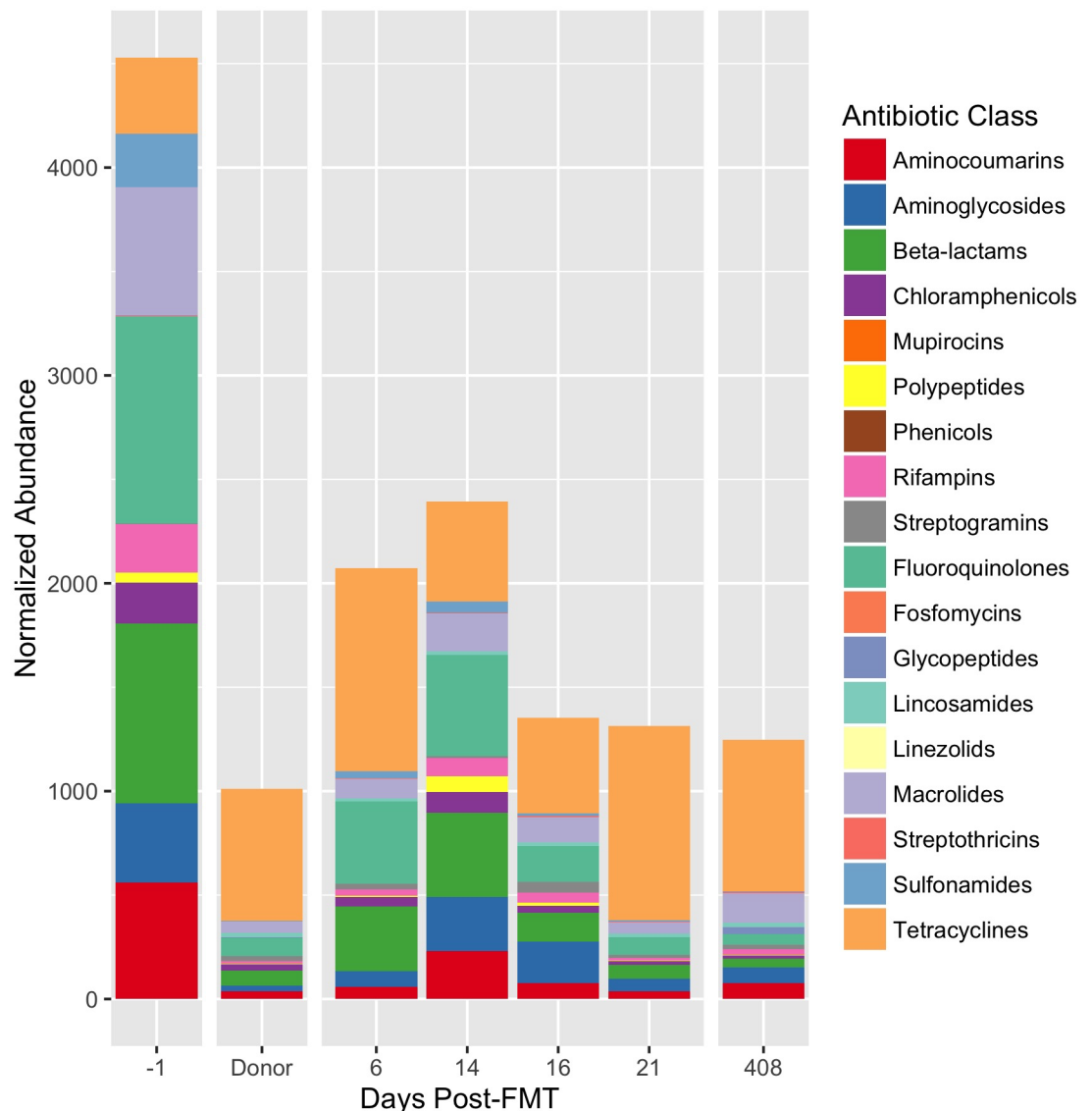
**Fig 3. The total gene complement of *E. coli* in Subject A experiences large-scale remodeling in association with FMT, resulting in broad reductions in genes significantly enriched for virulence factors.** PanPhlAn [37] was used to identify the presence of individual genes within the pan-genome of *E. coli*. Reads were aligned to a nonredundant collection of genes present in phylogenetically diverse *E. coli* isolate genome sequences. Those genes occurring at an abundance consistent with *E. coli*-specific occurrence were chosen and designated present or absent, as described previously [37]. Gene occurrence profiles were



hierarchically clustered and divided into 10 groups with similar occurrence over the time series. Cumulative gene counts across groups are indicated on the left margin. Genes associated with virulence were identified by alignment to the Virulence Factor Database [39]. Over- or underrepresentation of virulence genes in cluster groups was determined using Fisher's exact test to detect significant departure from random occurrence. Significant ( $p < 0.05$ ) underrepresentation was found in group 1, and significant overrepresentation was found in groups 5 and 9. KEGG annotations of genes within each group are given in S1 File. Virulence factors found within each group are given in S2 File.

<https://doi.org/10.1371/journal.pone.0182585.g003>

bacterial composition through to 1 year following FMT [14] Weingarden et al., who followed taxonomic changes for 151 days post-FMT from 4 immunocompetent CDAD-afflicted subjects treated with stool from a single donor. [15] Like Broecker et al., Jalanka et al. found the unrelated donor bacterial profile to be retained in recipients at 1 year post-FMT. However,



**Fig 4. Antibiotic resistome profile of Subject A is rapidly remodeled immediately after FMT.** Antibiotic resistance gene abundances were measured by alignment to the CARD antibiotic resistance database [19] and normalizing by per-sample coverage. Individual gene abundances are aggregated by antibiotic class. Genes conferring multiple antibiotic class resistance phenotypes are counted toward each antibiotic class.

<https://doi.org/10.1371/journal.pone.0182585.g004>

Weingarden et al. report the highest similarity between donor and recipient to be 1 day following FMT, with considerable divergence at all later time points. Such findings are more consistent with our observations of short-term similarity and long-term dissimilarity between donor and recipient gut microbiota composition. We attribute the respective discrepancy between the findings of Broecker and Jalanka with those of Weingarden and ourselves to differences in sensitivity between 16S and metagenomic methodologies, and to possible involvement of the host immune status and potential subsequent therapies in influencing the long-term fate of donor microbiota.

Our analysis of nucleotide-level similarity between donor and recipient, adapted from Li et al [17], showed donor concordance over short-term timepoints consistent with that group's findings, however after 1 year post-FMT we report high discordance between donor and recipient genus *Roseburia* SNVs (Fig 2). We focused on this genus since it appeared to be absent in the recipient pre-FMT, abundant in both the donor and the recipient at all timepoints post-FMT, and showed the highest abundance of reads at >1 year post-FMT. Given the consistent presence of *Roseburia* across all timepoints, it was surprising that the genomic sequence(s) of *Roseburia* at >1 year post-FMT was extensively dissimilar to that of the *Roseburia* organism (s) that had initially engrafted in the recipient. It is possible that intra-species competition, neutral drift, or other mechanisms of nucleotide divergence drove strain turnover within the ecological niche harbouring *Roseburia*; however, more studies are needed to understand the importance of sub-species interactions and how such fluctuations might affect gut microbiome behavior and ultimately human health.

To gain a better appreciation for how gene-level differences between same-species strains of bacteria affect engraftment following FMT, we chose to investigate the gene complement of *E. coli*, a species that exhibits wide variability between strains, ranging from dominant commensal in the human gut to highly virulent pathogen [16]. We chose this organism both because it occurs at high abundance through the time series obtained from Subject A, and because there is an extensive genome collection available for *E. coli*. At the taxonomic level, the presence of *E. coli* was consistent in Subject A throughout all samples; however, at the gene level, we observed massive losses and gains in the total *E. coli* accessory genome between all measured timepoints, with the most pronounced differences occurring pre- and post-FMT (Fig 3). Whether this change is due to strain replacement or genomic remodeling within a single resident strain is unconfirmed, although the rapid timeframe in which it occurs suggests the former.

Multi-drug-resistant (MDR) bacteria are becoming increasingly common in patients with hematological malignancies, such as HCT recipients [20]. It has been previously shown that FMT is capable of reducing antibiotic resistance genes in immunocompetent patients being treated for CDAD [21]. Due presumably to both the requirement that antibiotic treatment be administered for at least 3 CDAD episodes before FMT will be attempted, as well as the use of antibiotics in conjunction with HCT therapy, recipients have been found to have higher numbers of antibiotic resistance genes than those of healthy donors. We sought to determine (1) whether immunocompromised patients carry an increased number of antibiotic resistance genes compared to a healthy donor, and (2) if FMT is similarly capable of reducing the abundance and variety of antibiotic resistance genes in HCT patients. We demonstrate that, for the immunocompromised subjects analyzed, there was an increased presence of antibiotic resistance genes in the recipient compared to the donor, and that abundance of the antibiotic resistome was reduced by >50% followed FMT (Fig 4). Furthermore, this reduction persisted more than one year after FMT. This is consistent with findings in immunocompetent recipients, and bodes well for the potential reduction of antibiotic resistance via FMT in HCT patients. Given the lean toolkit of antibiotics that remain effective in the face of MDR pathogens, identifying

strategies to mitigate antibiotic resistance is of tremendous importance, especially for immunocompromised individuals. Reducing antibiotic resistance may prove to be yet another potential application of FMT.

In conclusion, our work establishes a foundation for future studies employing a shotgun metagenomic approach to better understand the evolution of microbial communities after FMT in various disease states. We demonstrate that granular shotgun metagenomic analyses show high initial donor similarity and long-term donor divergence at the community-wide taxonomic and functional levels, as well as in single-species gene composition and nucleotide variation. While further studies involving a larger cohort of patients and more detailed statistical analysis are necessary to validate our early findings, the findings reported herein lend support to the safe use of FMT for CDAD and reveal preliminary evidence for short and long-term FMT dynamics in immunocompromised HCT patients. In addition, our results strongly support the value of detailed analysis of metagenomic sequence data, and suggest a holistic model of FMT engraftment durability informed by more than donor or recipient microbiome taxonomic composition at the species, genus or phylum level.

## Methods

### Stool sample collection

Bone marrow transplant recipient fecal samples were collected by patients at home using collection kits provided. Fecal samples were shipped to the laboratory on pre-frozen freezer packs and stored at  $-80^{\circ}\text{C}$  within 24 hours of production. Upon defecation, pre-FMT and early (<21 days) timepoint fecal samples were aliquoted in 95% EtOH. Donor and all late timepoint (>6 months) fecal samples were stored in the absence of EtOH.

### Fecal microbiota transplantation

Donors were screened and fecal samples were collected, and processed into FMT capsules (Capsugel hypromellose capsules) as described [22–24]. FMT was performed as previously reported [22]. Briefly, each patient received 15 frozen FMT capsules with water on each of two consecutive days followed by monitoring for any adverse effects. Subjects took nothing by mouth for four hours before and one hour after administration. Subjects were instructed to stop oral vancomycin 24–48 hours prior to FMT. Subject A vomited more than 12 hours after the first dose but did not bring up any capsules. There were no serious adverse events attributable to FMT.

### DNA extraction

Stool samples were stored at  $-80^{\circ}\text{C}$  until use. DNA extractions were performed on all samples using the QIAamp DNA Stool Mini Kit (QIAGEN<sup>®</sup>) as per manufacturer instructions, with an added bead-beating step using the Mini-Beadbeater-16 (BioSpec Products) and 1 mm diameter Zirconia/Silica beads (BioSpec Products) consisting of 7 rounds of alternating 30 second bead-beating bursts followed by cooling on ice. DNA concentration and quality estimations were performed using Qubit<sup>®</sup> Fluorometric Quantitation (Life Technologies) and Bioanalyzer 2100 (Agilent). DNA libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina<sup>®</sup>). Prepared libraries were subjected to 100 base pair, paired-end sequencing on a HiSeq 2500 (Illumina<sup>®</sup>).

## Metagenomic shotgun sequencing and analysis

Data from individual sequencing runs were first aggregated by clinical sample using an in-house script. Paired-end raw reads from shotgun sequencing were trimmed using cutAdapt v1.7.1 [25] using a minimum length 80bp, maximum N count of 0, minimum terminal base score of 30, and the Illumina Nextera transposase and adapter sequences. In addition, reads were trimmed by 16bp at the 5' end to remove low quality bases and deduplicated with PrinSeq v0.20.4 [26]. Sequencing was performed to asymptotic saturation of available taxonomic diversity, assessed by calculating rarefaction curves using the vegan v2.4 [27] package in the R statistical computing language [28] and defining an asymptotic threshold at less than one new genus discovered per each additional million reads (S1 Fig).

Reads were taxonomically assigned using Kraken v0.10.4 [29] based on k-mer resemblance to human, bacterial, viral and fungal k-mer profiles ( $k = 31$ ) generated from the NCBI Bacterial Refseq and Genbank collections (accessed November 14, 2015), filtered with the kraken-filter utility at a threshold of 0.2, and results visualized with in-house scripts. Similarity between samples based on taxonomic calls was calculated with the Bray-Curtis similarity metric [30].

Functional analysis was conducted on quality filtered reads using HUMAnN2 v0.2.1 [31] to obtain read alignments to the Uniref50 functionally annotated protein sequence database [32]. Relative gene family abundances were obtained by dividing gene family coverage by per-sample total coverage. The Chao similarity metric was chosen to assess functional similarity due to the high diversity present in the functional repertoire (~2 million distinct genes) and the consequent risk of negative bias due to undersampling, a problem affecting alternative choices of similarity index such as Bray-Curtis or Jaccard. [33]

Roseburia SNV diversity was calculated by first assembling reads from the subject 1 long-term time-point with Spades v3.6.1 [34], then aligning reads from all timepoints to the contigs so obtained using BWA [35], and finally calling SNVs using the GATK Unified Genotyper [36]. SNVs discriminating donor from long-term recipient timepoint were selected as previously described [17] and visualized with an in-house script.

PanPhlAn [37] was used to obtain gene repertoire membership information for *Escherichia coli* using the 2016 pan-genome index for this organism. Gene functions were annotated for all genes using the KEGG [38] database according to the method described in the PanPhlAn documentation. Those genes occurring at an abundance consistent with *E. coli*-specific occurrence were chosen and designated present or absent as described previously [37]. Genes associated with virulence were identified by alignment to the Virulence Factor Database [39]. Over- or underrepresentation of virulence genes in cluster groups was determined using Fisher's exact test to detect significant departure from random occurrence.

Antibiotic resistance profiles were obtained by using BWA [35] to align read data from all samples to the Comprehensive Antibiotic Resistance Database [19] (accessed 6/6/16) and collecting per-gene read counts and aggregating by resistance phenotypes using an in-house script.

Plots for all analyses were generated in the R statistical language [28] using the ggplot2 [40], reshape2 [41], and doBy [42] packages.

## Supporting information

**S1 Table. Bray-Curtis genus-level taxonomic donor similarity.** Taxonomic abundances were calculated with the Kraken classifier using a custom database constructed from the NCBI Genbank and Refseq reference collections. Similarity is expressed for each subject to the appropriate donor sample, and for each time point in the donor 29 time series to the initial

time point in that series.  
(DOCX)

**S2 Table. Chao functional donor similarity.** Gene family abundances were calculated by alignment to the UniRef50 database. Similarity is expressed for each subject to the appropriate donor sample, and for each time point in the donor 29 time series to the initial time point in that series.  
(DOCX)

**S3 Table. Chao antibiotic resistance gene-level donor similarity for subject A.** Antibiotic resistance gene abundances were measured by alignment to the CARD antibiotic resistance database and normalizing by per-sample coverage.  
(DOCX)

**S4 Table. Post-quality control sequencing read counts obtained for all samples.** These read files have been submitted to NCBI SRA under BioProject accession No. PRJNA349197.  
(DOCX)

**S1 Fig. Rarefaction curves for sequenced samples.** Rarefaction curves were calculated using the vegan library(Oksanen *et al*, 2007) in the R statistical computing language(Ihaka & Gentleman, 1996). Curves represent repeated random subsamplings of genus-level taxonomic calls taken at increasing sample sizes. Singleton calls were omitted from this analysis. We assess sufficient sampling when a curve has transitioned from the vertical phase to an asymptotic phase. Samples are labeled in the form “Subject\_Timepoint”, with ‘lt’ identifying long-term timepoints and donor samples prefixed with the letter D. Timepoints are numbered in chronological order for each subject. Note that labels D29\_11 and D29\_7 are coincident, as are D29\_9 and B\_2.  
(TIF)

**S1 File. KEGG annotations of clustered E. coli genes identified in Subject A.**  
(ZIP)

**S2 File. VFDB virulence annotations of clustered E. coli genes identified in Subject A.**  
(ZIP)

**S3 File. Taxonomic classification reports for all sequenced subject and donor timepoints.**  
(ZIP)

**S4 File. Script code for all analyses described in this manuscript.**  
(ZIP)

## Author Contributions

**Conceptualization:** Eli L. Moss, Elizabeth L. Hohmann, Ami S. Bhatt.

**Data curation:** Eli L. Moss.

**Formal analysis:** Eli L. Moss, Shannon B. Falconer, Mingjie Wang.

**Funding acquisition:** Elizabeth L. Hohmann, Ami S. Bhatt.

**Investigation:** Eli L. Moss, Shannon B. Falconer, Ekaterina Tkachenko, Hannah System, Jasmin Mahabamunuge, Elizabeth L. Hohmann, Ami S. Bhatt.

**Methodology:** Eli L. Moss, Shannon B. Falconer, Ekaterina Tkachenko, Elizabeth L. Hohmann, Ami S. Bhatt.

**Project administration:** Hannah Systrom, Jasmin Mahabamunuge, Elizabeth L. Hohmann, Ami S. Bhatt.

**Resources:** Elizabeth L. Hohmann.

**Software:** Eli L. Moss.

**Supervision:** David A. Relman, Elizabeth L. Hohmann, Ami S. Bhatt.

**Validation:** Eli L. Moss.

**Visualization:** Eli L. Moss.

**Writing – original draft:** Eli L. Moss, Shannon B. Falconer, Ami S. Bhatt.

**Writing – review & editing:** Eli L. Moss, Shannon B. Falconer, Ekaterina Tkachenko, David A. Relman, Elizabeth L. Hohmann, Ami S. Bhatt.

## References

1. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N Engl J Med*. 2013; 368: 407–415. <https://doi.org/10.1056/NEJMoa1205037> PMID: 23323867
2. Kelly CR, Ihunnah C, Fischer M, Khoruts A, Surawicz C, Afzali A, et al. Fecal microbiota transplant for treatment of *Clostridium difficile* infection in immunocompromised patients. *Am J Gastroenterol*. 2014; 109: 1065–1071. <https://doi.org/10.1038/ajg.2014.133> PMID: 24890442
3. Bruminhent J, Wang Z-X, Hu C, Wagner J, Sunday R, Bobik B, et al. *Clostridium difficile* colonization and disease in patients undergoing hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2014; 20: 1329–1334. <https://doi.org/10.1016/j.bbmt.2014.04.026> PMID: 24792871
4. Chopra T, Chandrasekar P, Salimnia H, Heilbrun LK, Smith D, Alangaden GJ. Recent epidemiology of *Clostridium difficile* infection during hematopoietic stem cell transplantation. *Clin Transplant*. 2011; 25: E82–7. <https://doi.org/10.1111/j.1399-0012.2010.01331.x> PMID: 20973823
5. de Castro CG, Ganc AJ, Ganc RL, Petrolli MS, Hamerschlack N. Fecal microbiota transplant after hematopoietic SCT: report of a successful case. *Bone Marrow Transplant*. 2015; 50: 145–145. <https://doi.org/10.1038/bmt.2014.212> PMID: 25265462
6. Mittal C, Miller N, Meighani A, Hart BR, John A, Ramesh M. Fecal microbiota transplant for recurrent *Clostridium difficile* infection after peripheral autologous stem cell transplant for diffuse large B-cell lymphoma. *Bone Marrow Transplant*. 2015; 50: 1010. <https://doi.org/10.1038/bmt.2015.85> PMID: 25893454
7. Neemann K, Eichele D d., Smith P w., Bociek R, Akhtari M, Freifeld A. Fecal microbiota transplantation for fulminant *Clostridium difficile* infection in an allogeneic stem cell transplant patient. *Transpl Infect Dis*. 2012; 14: E161–E165. <https://doi.org/10.1111/tid.12017> PMID: 23121625
8. Aroniadis OC, Brandt LJ, Greenberg A, Borody T, Kelly CR, Mellow M, et al. Long-term Follow-up Study of Fecal Microbiota Transplantation for Severe and/or Complicated *Clostridium difficile* Infection: A Multicenter Experience. *J Clin Gastroenterol*. 2016; 50: 398–402. PMID: 26125460
9. Webb BJ, Brunner A, Ford CD, Gazdik MA, Petersen FB, Hoda D. Fecal microbiota transplantation for recurrent *Clostridium difficile* infection in hematopoietic stem cell transplant recipients. *Transpl Infect Dis*. 2016; 18: 628–633. <https://doi.org/10.1111/tid.12550> PMID: 27214585
10. Seekatz AM, Aas J, Gessert CE, Rubin TA, Saman DM, Bakken JS, et al. Recovery of the gut microbiome following fecal microbiota transplantation. *MBio*. 2014; 5: e00893–14. <https://doi.org/10.1128/mBio.00893-14> PMID: 24939885
11. Shahinas D, Silverman M, Sittler T, Chiu C, Kim P, Allen-Vercoe E, et al. Toward an Understanding of Changes in Diversity Associated with Fecal Microbiome Transplantation Based on 16S rRNA Gene Deep Sequencing. *MBio*. 2012; 3: e00338–12. <https://doi.org/10.1128/mBio.00338-12> PMID: 23093385
12. Hamilton MJ, Weingarden AR, Unno T, Khoruts A, Sadowsky MJ. High-throughput DNA sequence analysis reveals stable engraftment of gut microbiota following transplantation of previously frozen fecal bacteria. *Gut Microbes*. 2013; 4: 125–135. <https://doi.org/10.4161/gmic.23571> PMID: 23333862
13. Broecker F, Klumpp J, Schuppler M, Russo G, Biedermann L, Hombach M, et al. Long-term changes of bacterial and viral compositions in the intestine of a recovered *Clostridium difficile* patient after fecal

- microbiota transplantation. *Cold Spring Harb Mol Case Stud.* 2016; 2: a000448. <https://doi.org/10.1101/mcs.a000448> PMID: 27148577
14. Jalanka J, Mattila E, Jouhten H, Hartman J, de Vos WM, Arkkila P, et al. Long-term effects on luminal and mucosal microbiota and commonly acquired taxa in faecal microbiota transplantation for recurrent *Clostridium difficile* infection. *BMC Med.* 2016; 14: 155. <https://doi.org/10.1186/s12916-016-0698-z> PMID: 27724956
  15. Weingarden A, González A, Vázquez-Baeza Y, Weiss S, Humphry G, Berg-Lyons D, et al. Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *Microbiome.* 2015; 3. <https://doi.org/10.1186/s40168-015-0070-0> PMID: 25825673
  16. Tenailon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol.* 2010; 8: 207–217. <https://doi.org/10.1038/nrmicro2298> PMID: 20157339
  17. Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, et al. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science.* 2016; 352: 586–589. <https://doi.org/10.1126/science.aad8852> PMID: 27126044
  18. Quera R, Espinoza R, Estay C, Rivera D. Bacteremia as an adverse event of fecal microbiota transplantation in a patient with Crohn's disease and recurrent *Clostridium difficile* infection. *J Crohns Colitis.* 2014; 8: 252–253. <https://doi.org/10.1016/j.crohns.2013.10.002> PMID: 24184170
  19. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother.* 2013; 57: 3348–3357. <https://doi.org/10.1128/AAC.00419-13> PMID: 23650175
  20. Baker TM, Satlin MJ. The growing threat of multidrug-resistant Gram-negative infections in patients with hematologic malignancies. *Leuk Lymphoma.* 2016; 57: 2245–2258. <https://doi.org/10.1080/10428194.2016.1193859> PMID: 27339405
  21. Millan B, Park H, Hotte N, Mathieu O, Burguiere P, Tompkins TA, et al. Fecal Microbial Transplants Reduce Antibiotic-resistant Genes in Patients With Recurrent *Clostridium difficile* Infection. *Clin Infect Dis.* 2016; 62: 1479–1486. <https://doi.org/10.1093/cid/ciw185> PMID: 27025836
  22. 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: 1772–1778. <https://doi.org/10.1001/jama.2014.13875> PMID: 25322359
  23. Youngster I, Mahabamunuge J, Systrom HK, Sauk J, Khalili H, Levin J, et al. Oral, frozen fecal microbiota transplant (FMT) capsules for recurrent *Clostridium difficile* infection. *BMC Med.* 2016; 14: 134. <https://doi.org/10.1186/s12916-016-0680-9> PMID: 27609178
  24. Hamilton MJ, Weingarden AR, Sadowsky MJ, Khoruts A. Standardized frozen preparation for transplantation of fecal microbiota for recurrent *Clostridium difficile* infection. *Am J Gastroenterol.* 2012; 107: 761–767. <https://doi.org/10.1038/ajg.2011.482> PMID: 22290405
  25. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal.* 2011; 17: 10–12.
  26. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011; btr026.
  27. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* Blackwell Publishing Ltd; 2003; 14: 927–930.
  28. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. *J Comput Graph Stat.* 1996; 5: 299–314.
  29. Wood D, Salzberg S. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014; 15: R46. <https://doi.org/10.1186/gb-2014-15-3-r46> PMID: 24580807
  30. Beals EW. Bray-Curtis Ordination: An Effective Strategy for Analysis of Multivariate Ecological Data. In: MacFadyen A. and Ford E.D., editor. *Advances in Ecological Research.* Academic Press; 1984. pp. 1–55.
  31. Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, et al. Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. *PLoS Comput Biol.* 2012; 8: e1002358. <https://doi.org/10.1371/journal.pcbi.1002358> PMID: 22719234
  32. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics.* 2007; 23: 1282–1288. <https://doi.org/10.1093/bioinformatics/btm098> PMID: 17379688
  33. Chao A, Chazdon RL, Colwell RK, Shen T-J. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol Lett.* Blackwell Science Ltd; 2005; 8: 148–159.

34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol.* 2012–5; 19: 455–477.
35. Li H, Durbin R. Fast and accurate short read alignment with Burrows—Wheeler transform. *Bioinformatics.* 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
36. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013; 43: 11.10.1–33.
37. Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, et al. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat Methods.* 2016; <https://doi.org/10.1038/nmeth.3802> PMID: 26999001
38. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000; 28: 27–30. PMID: 10592173
39. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 2005; 33: D325–8. <https://doi.org/10.1093/nar/gki008> PMID: 15608208
40. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer Science & Business Media; 2009.
41. Wickham H. *reshape2: Flexibly reshape data: a reboot of the reshape package.* R package version. 2012;1. <http://cran.ms.unimelb.edu.au/web/packages/reshape2/>
42. Højsgaard S, Højsgaard MS, Hmisc D. The doBy package. *The Newsletter of the R Project.* 2006; 6: 1.