Fecal microbiota transplantation influences procarcinogenic Escherichia coli in recipient recurrent Clostridioides difficile patients

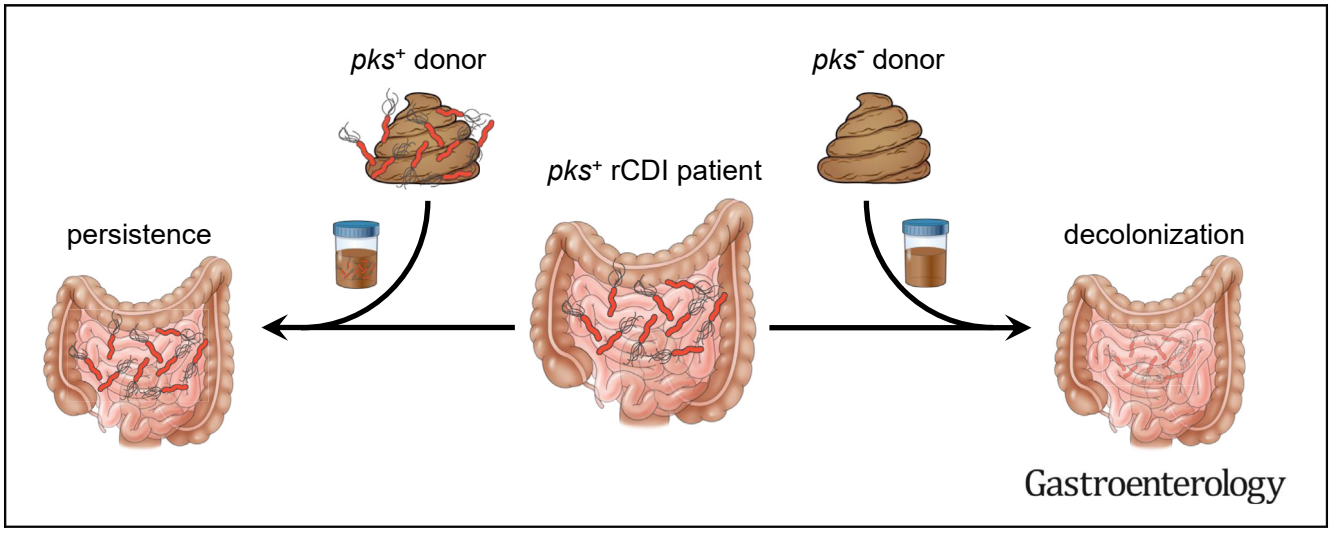
Nooij, S.; Ducarmon, Q.R.; Laros, J.F.J.; Zwittink, R.D.; Norman, J.M.; Smits, W.K.; ... ; Netherlands Donor Feces Bank

Gastroenterology 2021;161:1218–1228

Fecal Microbiota Transplantation Influences Procarcinogenic *Escherichia coli* in Recipient Recurrent *Clostridioides difficile* Patients

Sam Nooij,[1,2,3] Quinten R. Ducarmon,[1,3] Jeroen F. J. Laros,[3,4,5,6] Romy D. Zwittink,[1,3] Jason M. Norman,[7] Wiep Klaas Smits,[1,3] Hein W. Verspaget,[8,9] Josbert J. Keller,[8,10] Elisabeth M. Terveer,[1,2,3] and Ed J. Kuijper,[1,2,3,6] on behalf of the Working Group of the Netherlands Donor Feces Bank

[1]*Experimental Bacteriology, Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands;* [2]*Netherlands Donor Feces Bank, Leiden, the Netherlands;* [3]*Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, the Netherlands;* [4]*Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands;* [5]*Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands;* [6]*National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands;* [7]*Vedanta Biosciences Inc, Cambridge, Massachusetts;* [8]*Department of Gastroenterology, Leiden University Medical Center, Leiden, the Netherlands;* [9]*Department of Biobanking, Leiden University Medical Center, Leiden, the Netherlands; and* [10]*Department of Gastroenterology, Haaglanden Medical Center, The Hague, the Netherlands*



See Covering the Cover synopsis on page 1077;
See editorial on page 1103.

BACKGROUND & AIMS: Patients with multiple recurrent *Clostridioides difficile* infection (rCDI) have a disturbed gut microbiota that can be restored by fecal microbiota transplantation (FMT). Despite extensive screening, healthy feces donors may carry bacteria in their intestinal tract that could have long-term health effects, such as potentially procarcinogenic polyketide synthase-positive (*pks*$^+$) *Escherichia coli*. Here, we aim to determine whether the *pks* abundance and persistence of *pks*$^+$ *E coli* is influenced by *pks* status of the donor feces. **METHODS:** In a cohort of 49 patients with rCDI treated with FMT and matching donor samples—the largest cohort of its kind, to our knowledge—we retrospectively screened fecal metagenomes for *pks*$^+$ *E coli* and compared the presence of *pks* in patients before and after treatment and to their respective donors. **RESULTS:** The *pks* island was more prevalent ($P = .026$) and abundant ($P < .001$) in patients with rCDI (pre-FMT, 27 of 49 [55%]; median, 0.46 reads per kilobase per million [RPKM] *pks*) than in healthy donors (3 of 8 donors [37.5%], 11 of 38 samples [29%]; median, 0.01 RPKM *pks*). The *pks* status of patients post-FMT depended on the *pks* status of the donor suspension with which the patient was treated ($P = .046$). Particularly, persistence (8 of 9 cases) or clearance (13 of 18) of *pks*$^+$ *E coli* in *pks*$^+$ patients was correlated to *pks* in the donor ($P = .004$). **CONCLUSIONS:** We conclude that FMT contributes to *pks*$^+$ *E coli* persistence or eradication in patients with rCDI but that donor-to-patient transmission of *pks*$^+$ *E coli* is unlikely.

Keywords: Colibactin; Colorectal Cancer; Genotoxin; Microbiome.

Multiple recurrent *Clostridioides difficile* infection (rCDI) may occur as a complication after *C difficile* infection (CDI) and is associated with a disturbance of the colonic microbiota.[1–3] Patients with rCDI can be treated with fecal microbiota transplantation (FMT), which has a cure rate of up to 89%.[1,4] The mechanism of action appears to be restoration of a healthy microbiome.[5] The donor fecal microbiota derives from carefully screened, healthy donors.[4,6,7] Donor screening focuses on the prevention of transfer of pathogens and the potential transmission of microbiota-associated disorders. In recent years, there has been increasing attention for the causative role of gut bacteria in the development of colorectal cancer, which may pose a long-term risk.[8]

In 2019, Drewes et al[9] studied transmission and clearance of putative procarcinogenic bacteria, including colibactin-encoding *Escherichia coli*, by FMT in 11 pediatric patients with rCDI.[9] Colibactin-producing *E coli*, also known as polyketide synthase-positive (*pks*$^+$) *E coli*, has gained much attention lately, because it can be present in healthy and diseased people, has been used as a probiotic (strain Nissle 1917), and is now suspected to contribute to colorectal carcinogenesis.[10–12] The *pks*$^+$ *E coli* carries a *pks* gene island of approximately 54 kilobases long, with 19 genes that encode the machinery to produce the nonribosomal peptide-polyketide hybrid genotoxin colibactin.[13–15] Colibactin can induce double-strand DNA breaks, which cause specific mutational signatures found in colorectal carcinomas.[10,13–15]

Here, we determine the effect of FMT from healthy donors on *pks*$^+$ *E coli* in patients with rCDI by retrospectively screening 49 matching triplets of deep-sequenced fecal metagenomes. Our results show that patients carry higher levels of *pks*$^+$ *E coli* than donors and that the potential procarcinogen is more likely to persist when *pks*$^+$ donor feces is used for FMT. When donor material with nondetectable levels of *pks* is used, FMT can contribute to clearance of *pks*$^+$ *E coli*. Thus, FMT is unlikely to lead to transmission of *pks*$^+$ *E coli* but can reduce the levels of this procarcinogenic bacterium in patients with rCDI.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Fecal microbiota transplantation from a healthy donor may influence potentially procarcinogenic *pks*$^+$ *Escherichia coli* in the gut of patients with multiple recurrent *Clostridioides difficile* infection.

NEW FINDINGS

Procarcinogenic *pks*$^+$ *E coli* is more prevalent and abundant in patients with multiple recurrent *C difficile* infection than in healthy donors. Fecal microbiota transplantation contributes to persistence or clearance but not transmission of *pks*$^+$ *E coli* in patients with multiple recurrent *C difficile* infection.

LIMITATIONS

It is currently unknown how long colonization by *pks*$^+$ *E coli* persists and what level and duration of exposure is needed to contribute to the development of colorectal cancer.

IMPACT

Better insight into the influence of fecal microbiota transplantation on procarcinogenic *pks*$^+$ *E coli* can aid in donor selection and prevention of colorectal cancer development in patients.

Materials and Methods

Fecal Microbiota Transplantation Treatment

After careful screening by a multidisciplinary FMT expert panel, 49 patients with multiple rCDIs were treated with FMT with a fecal suspension provided by the Netherlands Donor Feces Bank. The complete procedure of donor recruitment, selection and screening, fecal suspension processing, evaluation of FMT requests, FMT treatment, and patient follow-up has been described before.[4,16] In short, these patients received a vancomycin treatment for a minimum of 4 days (median, 12 days) until 24 hours before FMT. The antibiotics course in 11 patients was shorter or of an unrecorded duration. Treatment in 8 patients was with fidaxomicin or an unrecorded antibiotic.

FMT was administered as a single treatment by a nasoduodenal tube, except in 2 patients who received a single infusion by colonoscopy. Differences in antibiotic pretreatment regimen and mode of FMT were not found to influence the study results. The 198-mL fecal suspensions contained 60 g of donor feces in saline solution with 10% volume/volume glycerol and were stored until use at −80°C. All donors were healthy individuals between the ages of 18 and 60 and were extensively screened for disorders associated with microbiota and potential transmissible diseases.

Ethics Approval

Patients provided informed consent for collection of stool samples and outcome data of FMT for research purposes, which was approved by the Leiden University Medical Center Medical Ethics Committee (P15.145).

Sample Preparation and Sequencing

Stool samples of patients were collected for sequencing approximately 1 day before FMT, after antibiotic pretreatment,

Abbreviations used in this paper: **ANI, average nucleotide identity; CDI, *Clostridioides difficile* infection; D, donor; rCDI, recurrent *Clostridioides difficile* infection; FMT, fecal microbiota transplantation; MLST, Multi-Locus Sequence Typing; *pks*, polyketide synthase; P, patient; RPKM, reads per kilobase per million; ST, sequence type.**

Most current article

0016-5085
https://doi.org/10.1053/j.gastro.2021.06.009

and 3 weeks after FMT. Stool samples of donors used for FMT treatment were also used for sequencing. Stool samples were stored at −80°C. DNA extraction and sequencing was provided by DNA Genotek Inc. In short, DNA was extracted using the automated KingFisher Flex platform (Thermo Fisher Scientific Inc) and MoBio PowerMag +ClearMag microbiome RNA/DNA isolation kit (QIAGEN Inc). Sequencing libraries were prepared using Illumina's NexteraXT protocol, and libraries were sequenced on the Illumina NextSeq platform (Illumina) to a target depth of 10 million reads per sample (2 × 150 base pair paired-end).

No matching donor metagenome was available for patients (P) 2 and P5, so instead they were matched to donor metagenomic data from samples taken 19 and 7 days later than those used for FMT, respectively.

Sequence Analysis

Raw metagenomic reads were randomly subsampled to 10 million read pairs to facilitate de novo assembly using BBtools reformat.sh version 38.79 (parameters: "samplereadstarget = 10000000, sampleseed = 20"; http://sourceforge.net/projects/bbmap/). The subsampled data sets were then analyzed with the assembly-based workflow Jovian version 0.9.6,[17] using the National Center for Biotechnology Information BLAST nt database from March 13, 2020 (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). In short, Jovian trims raw reads (trimmomatic, 0.38[18]; parameters "PE SLIDINGWINDOW: 5:30, MINLEN: 50"), removes human reads by mapping to the human genome (bowtie2 version 2.3.4.3,[19] parameters "–local," samtools version 1.9 and bedtools version 2.27.1[20,21]; reference genome "GRCh38.p7" [https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.33/]), assembles filtered reads into scaffolds (SPAdes version 3.11.0,[22] parameters "–meta—only-assembler -k 21,33,55,77"), and filters scaffolds to a minimum length of 500 nucleotides, classifies scaffolds taxonomically using a lowest common ancestor approach (BLAST version 2.9.0 parameters "-evalue 0.05 -max_target_seqs 250 -max_hsps 1",[23] and MGKit version 0.3.4 using quantile threshold .97 and bitscore threshold 100),[24] and quantifies scaffolds by mapping filtered reads back to the scaffolds (BWA-mem version 0.7.17,[25] default parameters).

From Jovian, we used (1) the quality-trimmed reads from which human reads are removed and (2) assembled scaffolds and (3) taxonomic classifications and quantifications of scaffolds. These reads and scaffolds were screened for *E coli* and the *pks* island (19 *clb* genes from GenBank accession ID AM229678) using the custom workflow "Jovian screener" version 0.1.[26] This screening extension to Jovian works in 3 ways:

1. It extracts the species of interest from the taxonomic profile results table (based on the "tax_name" column).
2. It looks for the (gene) sequences of interest in all scaffolds using BLAST (version 2.9.0, parameters "-task blastn -perc_identity 75 -qcov_hsp_perc 0 -evalue 1E-20 -culling_limit 1 -max_hsps 1"), which are then filtered to include only matches that cover at least half (50%) of the sequences of interest.
3. It maps trimmed and filtered reads to the sequences of interest (BWA version 0.7.17, samtools version 1.10), after which mapped reads are deduplicated using Picard MarkDuplicates version 2.23.3 (https://broadinstitute.github.io/picard/) and counted (samtools version 1.10).

The resulting tables were analyzed and visualized using R 4.0.2 (R Foundation for Statistical Computing) software (Rmarkdown scripts online at https://git.lumc.nl/snooij/pks-in-rcdi-metagenomes-analysis).[27]

We renamed the "hypothetical protein (clb)" from AM229678 "*clbS*," because it is identical to accession number KX683217 (*E* coli *clbS* gene, complete coding sequences). Reads per kilobase per million (RPKM) values were calculated by dividing the number of mapped reads by the length of the respective gene (coding sequence), multiplied by 1000, divided by the number of trimmed and filtered reads, multiplied by 1 million.

Pilot Screening for Other Procarcinogenic Bacteria

Next to *pks* from *E coli*, we also screened metagenomes for the presence of *Fusobacterium nucleatum* adhesin A (*fadA*; accession ID AVQ22939), *Bacteroides fragilis* toxin (*bft1*, *bft2, bft3*; accession IDs AB026624, AB026625, and AB026626), and *Campylobacter jejuni* cytolethal distending toxin (*cdtA*, *cdtB*, *cdtC*; accession ID AB274791). The *fadA* and *cdt* genes were not detected, and *bft* was detected in patient 7 pre- and post-FMT and in patient 42 post-FMT. Because these genes were almost completely absent from the current data set and absent from donors, we did not analyze these further.

Improving the Identification of pks by Combining Data From All 19 clb Genes

To improve the sensitivity of screening for the *pks* island in metagenomics data compared with using a single marker gene only (eg, *clbB*), we assessed whether the genes always co-occur and may be considered as 1 unit. In assembled scaffolds, the *clb* genes are mostly assembled on 1 or a few scaffolds when depth is sufficiently high, suggesting consistent close genomic proximity (Supplementary File 1). For the mapping approach, we compared the RPKM values of *clb* genes in each metagenomics sample with one another and calculated Spearman's correlations (Supplementary Figures 1–3). Correlations between *clb* genes are high ($r = 0.54$–0.99, $P < .01$) and are also higher than correlations with flanking regions and more distant *E coli* marker genes. Therefore, we conclude that the *pks* island is present in genomes as a complete island, and for its detection and quantification, we combine all hits to *clb* genes to calculate average *pks* values.

Selecting Representative Escherichia coli Marker Genes

To calculate the abundance of *E coli* in a similar way as the *pks* genes to facilitate the estimation of *pks* island copies per genome, we also mapped the reads to a number of putative single-copy marker genes. The putative marker genes are *dxs*, *rodA*, and *uidA* (accession IDs AF035440, M22857, and S69414, respectively),[28,29] MultiLocus Sequence Typing (MLST) genes *adk*, *fumC*, *gyrB icd*, *mdh*, *purA*, and *recA* from Enterobase,[30] and

adk, *gcl*, *zwf*, *mdh*, *metA* and *ppk* by Adiri et al.[31] Based on the depth of coverage, we found that the MLST genes sometimes showed uneven coverage, especially in metagenomes with more non-*E coli* Enterobacteriaceae. This indicates that these genes may have regions that are homologous to genes from other species, so the MLST genes were excluded.

Next, we compared RPKM values for *dxs*, *rodA*, and *uidA* to the relative abundance of *E coli* and Enterobacteriaceae in all analyzed metagenomes. The *uidA* gene has the best correlation with *E coli* abundance (Spearman's correlation $r = 0.97$, $P < 2.2^{-16}$; Supplementary Figure 4*A*). The *dxs* and *rodA* genes also appear in metagenomes with no *E coli* (Supplementary Figure 4*B* and *C*) and show better correlations with Enterobacteriaceae (Spearman's correlations $r = 0.96$ and $r = 0.97$, respectively; Supplementary Figure 4*E* and *F*). Even though *uidA* is the most specific marker gene, it is not sufficiently present in all metagenomes. Therefore, we calculated the average RPKM of *dxs*, *rodA*, and *uidA*, whichever are present, to represent the quantity of *E coli* genomes in the metagenomes, which was then used to calculate the ratio of the average *pks* island copies per *E coli* genome.

Comparing Escherichia coli *Assembled Scaffolds Between Individuals*

We assessed whether our data allows us to distinguish within-species variants of *E coli* that could be used to conclusively demonstrate transfer or persistence. By using average nucleotide identity (ANI) (pyANI version 0.2.10, parameter "-m ANIb")[32] we compared *E coli* scaffolds from patients with *pks*$^+$ *E coli* before and after FMT and their donors, and *E coli* from patients who had no *pks* before FMT but were *pks*$^+$ after FMT and the corresponding donors. ANIs between *E coli* scaffolds from different patients were as high as within patients pre- and post-FMT, so they could not be used to support persistence. Furthermore, donor scaffolds were too short to make a reasonable argument for transmission based on ANI. Where possible, *E coli* assemblies were typed in silico using the MLST software "mlst" (https://www.github.com/tseemann/mlst) and the PubMLST database (downloaded April 14, 2021).[33] MLST was successful only for 20 patients and for 2 patients provided a result for both before and after FMT (Supplementary Table 1). These 2 patients (P21 and P36), had the same sequence type before and after FMT, suggesting persistence of the same dominant variant. Both patients lost *pks* after FMT, so these sequence types do not carry *pks*.

Data Availability

All raw metagenomics data, from which human-derived reads have been excluded by mapping to a human reference genome (bowtie2 version 2.4.2, using parameter "-very-sensitive-local" to the GRCh38.p7 assembly), have been deposited in the European Nucleotide Archive under accession number PRJEB44737.

Results

Study Characteristics

We screened the fecal metagenomes of 49 patients with rCDI before and after FMT and 38 corresponding metagenomes from feces of 8 donors of the Netherlands Donor Feces Bank for *pks*$^+$ *E coli* via read mapping against a *pks* island reference (GenBank accession ID AM229678) and *E coli* single-copy marker genes (*dxs*, AF035440; *rodA*, M22857; *uidA*, S69414).

Suspensions from 11 donors were used to treat 2 patients each, thus yielding 49 sample triplets. The patients, 31 women and 18 men, were a median age of 73.5 years (range, 27–92 years). Only patients were included who had been treated with a single FMT and for which pre- and post-FMT samples were available. The donors, 4 women and 4 men, were a median age of 33 years (range, 24–46 years). Samples were collected between April 2016 and March 2018. For patients P2 and P5, no donor metagenomic data were available from the date of FMT but were from 19 and 7 days later, respectively. For other patients, corresponding donor data were available from the same sample as used for FMT. A rCDI cure was achieved in all patients, except 3 patients (P12, P25, and P26), who had a recurrence within 3 weeks after FMT.

The pks *Island Is Absent in Most Donors or Detected in Very Low Quantities*

To quantify *pks* levels in metagenomes, we averaged normalized RPKM values of 19 *clb* genes (colibactin-encoding, AM229678). Individuals with 0 reads mapped to *clb* genes in their metagenome are considered *pks*-negative (*pks*$^-$) and individuals with 1 or more reads mapped to 1 or more *clb* genes are *pks*$^+$. The metagenomes in 5 of 8 healthy donors were always negative for the *pks* island (Figure 1). One donor (D5) had multiple samples with *pks*$^+$ metagenomes (see below), another donor (D2) provided only 1 sample, which was *pks*$^+$, and a third donor (D6) had 1 *pks*$^+$ sample of 9 used for transplantation. Levels of *pks* among positive donors ranged from 0.0013 to 0.0498 RPKM (median, 0.010 RPKM). In total, 11 of 38 donor samples (29%) were *pks*$^+$.

Healthy Donors May Persistently Carry pks

To study the temporal dynamics of *pks* colonization of donors, we evaluated longitudinal samples from 1 donor (D5) that were consistently *pks*$^+$. Of 15 metagenomes total from D5, of which 10 were used for transplantation, 13 had traces of the *pks* island (Figure 2*A*). The first and last samples were both positive, suggesting that this donor was persistently colonized with a *pks*$^+$ bacterium during the donation period of 6 months. Levels of *pks* ranged from 0 to .05 RPKM and correlated loosely with the relative abundance of *E coli* (Spearman's $r = 0.43$, $P = .11$) or Enterobacteriaceae (Spearman's $r = 0.1$, $P = .72$). *E coli* was not always detected in the metagenomes of donor D5 (Figure 2*B* and *C*). It seems most likely that due to the low abundance of *E coli*, the sequencing depth in the metagenome was low, yielding only short genomic regions that are shared between *E coli* and other Enterobacteriaceae, resulting in these being classified as Enterobacteriaceae.

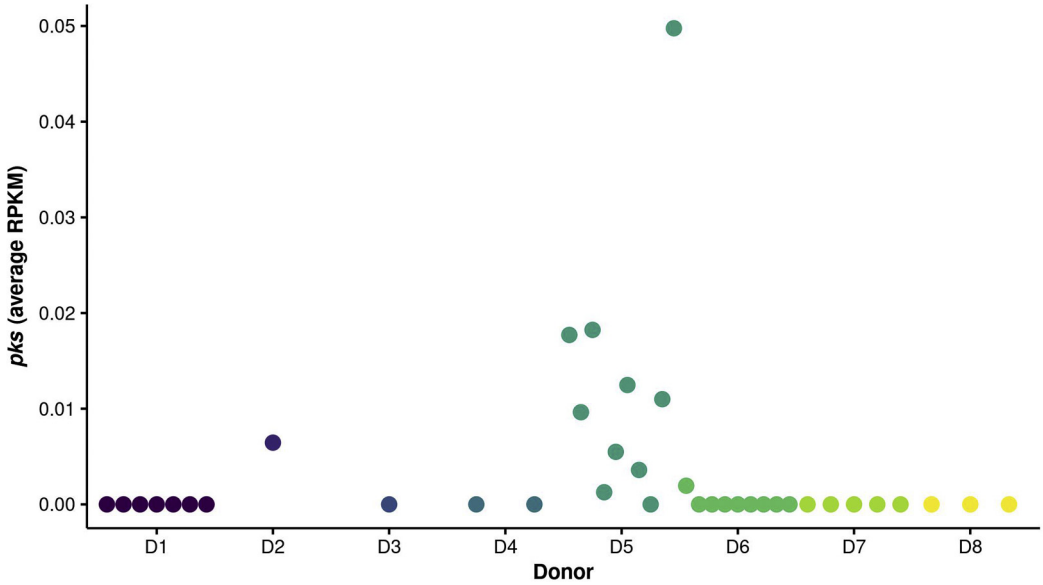


Figure 1. Levels of *pks* in donor metagenomes. Reads from donor metagenomes were mapped to the *pks* island (AM227896) to quantify the *pks* island per sample. Values are normalized by gene length and library size (RPKM), and the average RPKM value of the 19 *clb* genes is calculated as the average *pks*. Of 8 donors, 3 had pks^+ samples, of whom only 1 had multiple pks^+ samples. Of 38 donor samples included, 11 were pks^+.

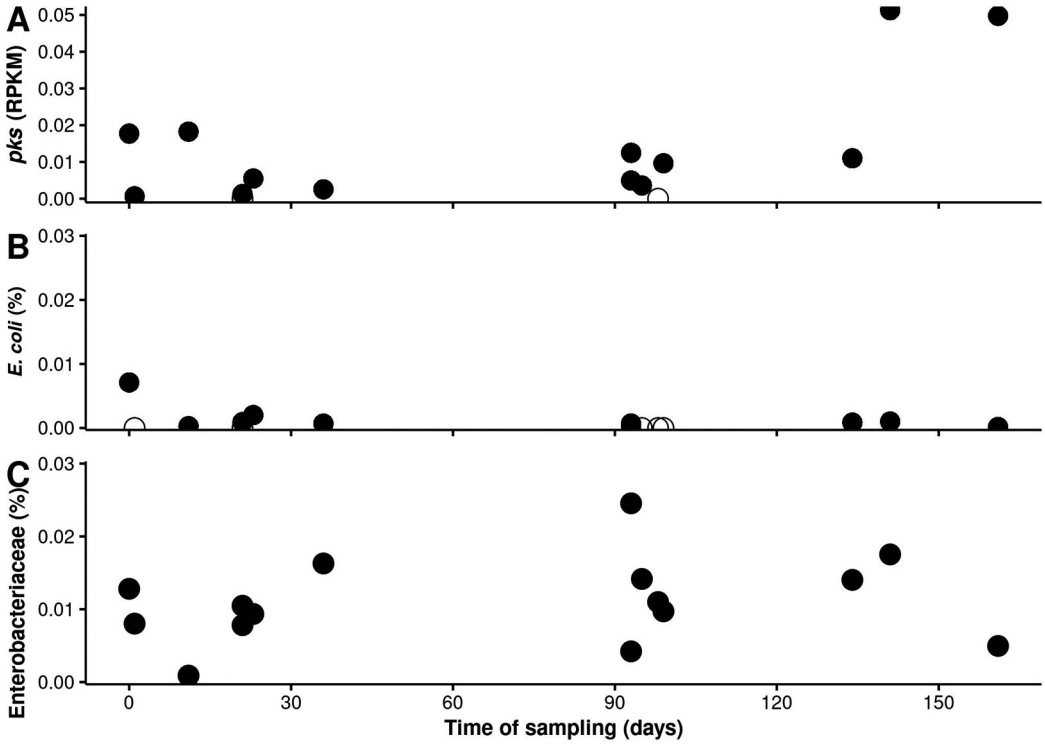


Figure 2. Levels of *pks* and relative abundances of *E coli* and Enterobacteriaceae over time in 1 healthy donor (D5). To gain more insight in the duration of colonization by pks^+ *E coli* in healthy donors, we quantified average *pks* levels of all available metagenomes of D5 by read mapping and determined the relative abundances of *E coli* and Enterobacteriaceae for each metagenome using Jovian. (*A*) Average normalized values of *pks* island (*clb* genes) in D5 over approximately 6 months' time. Samples from days 21 and 93 are duplicates; both are shown. Levels of *pks* are generally very low, near the limit of detection. D5 seems to have been persistently colonized by a pks^+ species for at least half a year, even though 2 samples in the first and fourth months were pks^-. Relative abundances of (*B*) *E coli* and (*C*) Enterobacteriaceae (including *E coli*) in each metagenomic sample of D5. No clear correlation is found between abundance and the level of *pks*. The *empty circles* denote that no *pks* or *E coli* was detected at that time.

The pks *Island Is Frequent and Abundant in Metagenomes of Patients With Recurrent* Clostridioides difficile *Infection Before Fecal Microbiota Transplant*

Genes from the *pks* island were detected in 27 of 49 patients (55%) before FMT treatment and in 19 of 49 patients (39%) after FMT. Overall, *pks* was more prevalent in metagenomes of patients with rCDI pre-FMT than in donor metagenomes (odds ratio, 2.246; 95% confidence interval, 1.010–4.589; χ^2 test, $P = .026$). Patient metagenomes reached *pks* levels up to 59.7 RPKM, corresponding to roughly 0.3% of all analyzed reads. The median *pks* levels of pks^+ patient metagenomes were 0.461 RPKM for pre-FMT and 1.348 RPKM for post-FMT. The *pks* level in pks^+ patient metagenomes was significantly higher than in donors (pairwise Wilcoxon rank sum exact test, $P < .001$ for both pre- and post-FMT), but was not different between pre- and post-FMT ($P = .66$; Figure 3*A*). When the *pks*-negative metagenomes were included in the comparison, the differences were less apparent, and *pks* levels of donors only differed from patients before FMT (Supplementary Figure 5).

Fecal Microbiota Transplant Changes pks *Levels in* pks^+ *Patients With Recurrent* Clostridioides difficile *Infection*

Because *pks* levels—when detected—are much higher in patients with rCDI than in healthy donors (Figure 3*A*), we determined whether treatment by FMT leads to a reduction in *pks* levels in the subset of pks^+ patients with rCDI. Overall, there was no significant difference between *pks* levels pre- and post-FMT in patients who had *pks* before FMT (paired Wilcoxon's signed rank exact test, $P = .2901$). However, when stratified by *pks* status of the donor, we find that the change of *pks* level in the patient with rCDI after FMT depends on the *pks* status of the donor (Fisher's exact test, $P = .0006$).

The metagenomes of 27 patients were pks^+ before FMT treatment, and *pks* levels varied more than 1000-fold. After FMT, 20 (74%) had decreased levels of *pks*, of which 14 (52%) were reduced to levels below our detection limit (Figure 3*B*). Of these 20, 18 (90%) were treated with a pks^- donor suspension (Table 1, Supplementary Figure 6). The level of *pks* increased in 7, 6 (86%) of whom were treated with a pks^+ donor suspension.

Patients Who Were pks^- *Before Fecal Microbiota Transplant Mostly Remained Negative*

We assessed the effect of FMT on patients who had no detectable levels of *pks* before FMT to evaluate the possibility of transmission from a pks^+ donor. Overall, as expected with a conversion of *pks* status in a pks^- population, the *pks* level increased after FMT (paired Wilcoxon's signed rank test, $P = .03603$). Before FMT, 22 patients had no detectable *pks*. After FMT, 16 patients (73%) remained negative and 6 acquired *pks* (Table 1). Interestingly, from the 6 patients who became pks^+ after FMT, 5 (83%) were treated with a pks^- donor suspension derived from 3 different donors (Figure 3*C*, Supplementary Figure 6). Of the 16 patients who remained negative, 12 (75%) were treated with pks^- donor suspensions. The average *pks* level of the metagenomes of the 6 patients who acquired *pks* after FMT was 0.420 RPKM (median, 0.245 RPKM).

The *pks* levels in patients who acquired *pks* varied more than 10-fold and were generally low compared with *pks* levels in patients with persistent *pks* (median, 2.874 RPKM), although the difference was not significant (Wilcoxon's rank sum test, $P = .08742$). Importantly, however, whether a patient in this group of patients with rCDI who was pks^- demonstrated increased *pks* levels after FMT was not significantly correlated with the *pks* status of the donor (Fisher's exact test, $P = 1$), suggesting that an increase in *pks* is as likely to derive from the patient as from the donor.

Persistence or Clearance of pks *in Patients After Fecal Microbiota Transplant Depends on the Presence of* pks *in Donors*

In addition to testing the effect of FMT on the quantity of *pks* in patients, we assessed the likelihood of *pks* persistence, clearance, and transmission based on the presence of *pks* in patients and their donors. Of 35 patients with rCDI who were treated with a pks^- donor suspension, 25 (71%) were pks^- after FMT (Table 1), and 4 (11%) had reduced *pks* (Supplementary Figure 6). Among 14 patients who were treated with a pks^+ donor suspension, 9 (64%) were pks^+ after FMT. Based on these numbers, we find a direct dependence between the post-FMT *pks* status and *pks* in the donor suspension (χ^2 test with Yates' continuity correction, $P = .046$).

The effect of FMT on post-FMT *pks* status was strongest in patients who were already colonized with a pks^+ *E coli* before FMT (Fisher's exact test, $P = .004$). Of 27 patients who were pks^+ pre-FMT, 9 were treated with a pks^+ donor suspension, of whom 8 remained pks^+ post-FMT. Of the 18 patients who were treated with a *pks*-free donor suspension, 5 contained *pks* post-FMT. No such effect was found in patients in whom we detected no *pks* pre-FMT (Fisher's exact test, $P = 1$). Although our metagenomic analyses could not demonstrate persistence or transmission using within-species variants (see Materials and Methods), taken together, these results indicate that FMT contributes to persistence or clearance rather than donor-to-patient transmission of *pks* in patients with rCDI.

Variable Within-Species pks *Island Prevalence Suggests Persistence of* pks^+ Escherichia coli *in Patients*

Using single-copy marker genes *dxs*, *rodA*, and *uidA* (encoding D-1-deoxyxylulose 5-phosphate synthase, rod-shape determining protein, and β-glucuronidase, respectively) to quantify *E coli* genomes, we calculated the average number of *pks* island copies per *E coli* genome or prevalence of *pks* among *E coli* variant subpopulations. *E coli* was detected in less than half of the donor metagenomes (14 of

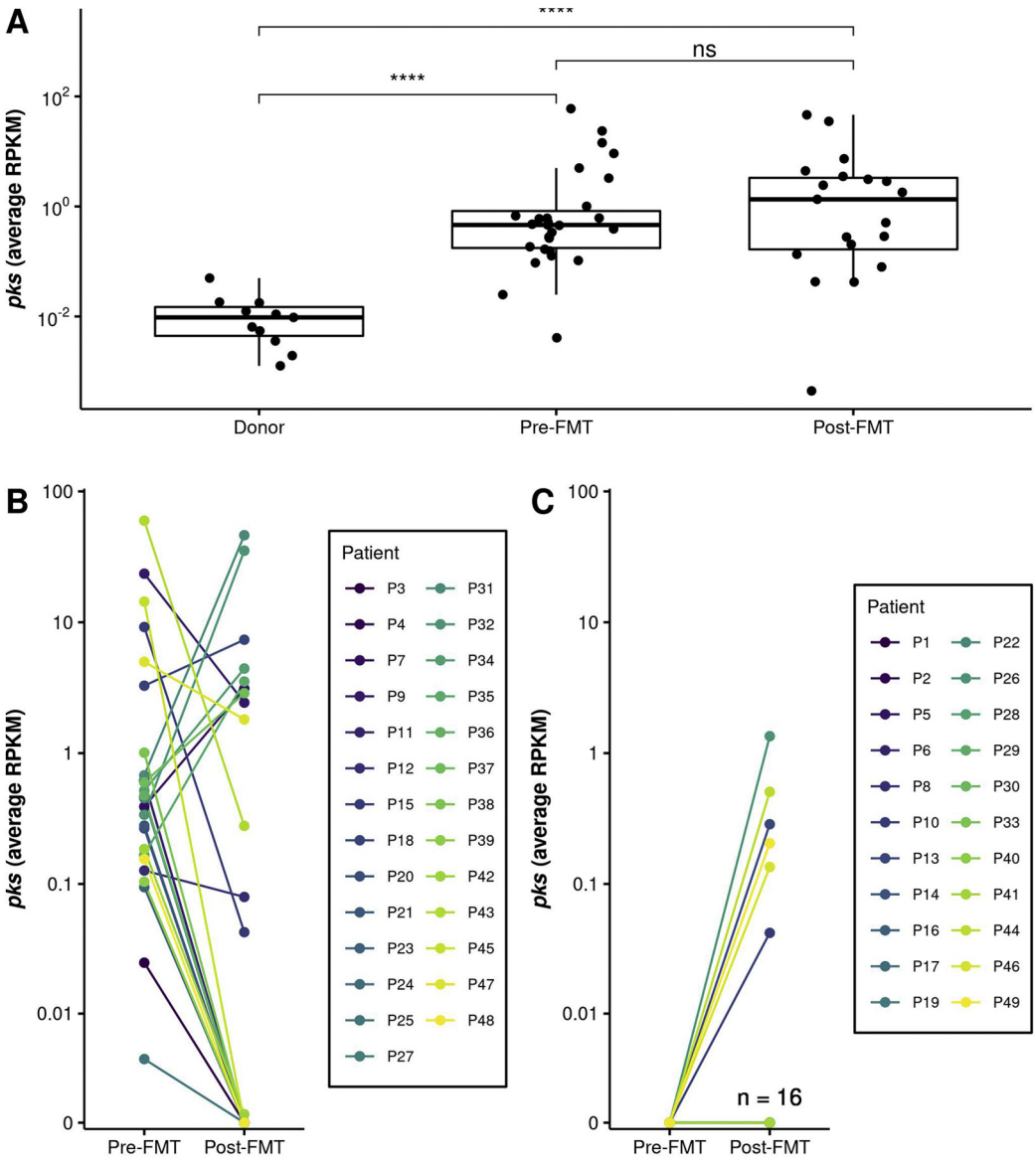


Figure 3. Levels of *pks* levels in *pks*$^+$ metagenomes of fecal donors and patients with rCDI before and after FMT treatment and changes in *pks* in patients after FMT. To assess the effect of FMT on the *pks* levels in patients, we mapped reads to a *pks* island reference and quantified *pks* by normalizing read counts by gene length and library size (RPKM), and averaged values over all 19 *clb* genes. The *y*-axes use a log10 scale. (*A*) Levels of *pks* in *pks*$^+$ donor samples were lower than those in patients pre- and post-FMT (Wilcoxon's rank sum test, $P < .001$). Levels of *pks* were not different between those found in patients pre-FMT and those post-FMT. The *horizontal line* in the middle of each *box* indicates the median; the *top and bottom borders* of the box mark the 75th and 25th percentiles, respectively, and the *vertical lines* mark minimum and maximum of all the data. ****$P < .0001$; ns, not significant. (*B*) Changes in *pks* levels in the 27 patients who were *pks*$^+$ before FMT. After FMT, 20 patients had lower *pks* levels, of whom 14 lost *pks*, and 7 had increased levels of *pks*. (*C*) Changes in FMT levels in the 22 patients who were *pks*$^-$ before FMT. After FMT, 6 of these patients acquired *pks*, and 16 remained *pks*$^-$. The number of patients with 0 *pks* after FMT is written in the figure to clarify the overlapping dots.

38; Figure 4*A*, *full circles*) and was detected in 5 of 11 *pks*$^+$ metagenomes from donors D2 and D5. *E coli* markers were present in 48 of 49 patients pre-FMT (Figure 4*B*), and *pks* per *E coli* ratios that are often near 0 indicate that *pks*$^+$ *E coli* variants are often a minority variant. After FMT, 26 of 49 patients still harbored *E coli* species (Figure 4*C*), and when *pks* is present the *pks* per *E coli* ratios are often higher.

Taken together, the results suggest that patients with rCDI treated with vancomycin or fidaxomicin before FMT may have high abundances of different *E coli* subpopulations or variants. When a patient is colonized with a *pks*$^+$ *E coli*, the *pks*$^+$ *E coli* is often a minority of *E coli* variants. After treatment by FMT using a donor suspension that contains little if any *E coli*, the number of different *E coli* variants decreases. And when a *pks*$^+$ variant is present post-FMT, this is often a larger fraction or the only *E coli* variant, possibly indicating that this variant is more persistent than other *E coli* variants.

Table 1. Summary of *pks* Status for Each Step in the Fecal Microbiota Transplant Treatments of the 49 Included Patients

Donor suspensions	Patients pre-FMT	Patients post-FMT
35 *pks*$^-$	17 *pks*$^-$	12 *pks*$^-$
		5 *pks*$^+$
	18 *pks*$^+$	13 *pks*$^-$
		5 *pks*$^+$
14 *pks*$^+$	5 *pks*$^-$	4 *pks*$^-$
		1 *pks*$^+$
	9 *pks*$^+$	1 *pks*$^-$
		8 *pks*$^+$

NOTE: Forty-nine patients, of which 22 were *pks*$^-$ and 27 were *pks*$^+$ before FMT, were treated with 35 *pks*− and 14 *pks*$^+$ donor suspensions. After FMT, 30 patients were *pks*$^-$ and 19 were *pks*$^+$.

Discussion

In this cohort study of 49 FMT-treated adult patients with rCDI and their corresponding healthy feces donors, we screened fecal metagenomes for the presence of *pks*$^+$ *E coli* to assess its prevalence and the effect of FMT. To our knowledge, this is the first study to use metagenomic screening for potentially procarcinogenic bacteria in the largest FMT cohort to date. We found that *pks* was present in a minority of healthy donor samples, and when present, *pks* levels were low. Of 8 donors, 1 donor was likely persistently colonized with a *pks*$^+$ bacterium for 6 months. In contrast, patients with rCDI had high levels of *pks* and high abundances of *E coli* in their metagenomes. FMT resulted in changes in both status and levels of *pks* detected in the deep-sequenced metagenomes. In particular, we find that treatment of *pks*$^+$ patients with rCDI with a *pks*$^-$ donor suspension reduces *pks* levels, in some cases to below the detection limit, whereas treatment with a *pks*$^+$ donor suspension leads to persistence of *pks*$^+$ *E coli*.

A previous study elucidating the effects of FMT on the presence of intestinal procarcinogenic bacteria in 11 pediatric patients with rCDI reported similar findings.[9] Drewes et al[9] found that FMT supported decolonization depending on the microbiota of both patient and donor. We find that FMT supports decolonization of *pks*$^+$ bacteria in patients with rCDI treated with *pks*$^-$ donor suspensions. Drewes et al[9] reported transmission of putative procarcinogenic bacteria during FMT.

Although we cannot exclude the possibility of transmission on the basis of our data, we consider it more likely that *pks*$^+$ *E coli* variants persist in the patients than that they are transferred from the donor, because *E coli* is much lower in abundance in donor material than in patients, and engraftment of low-abundant bacteria in the donor is less likely.[34] A study specifically designed to address the question of transmission of procarcinogenic bacteria during FMT, using culturing and whole-genome sequencing, is needed to investigate this aspect further.

Other potentially procarcinogenic species were studied previously in the context of FMT, namely enterotoxigenic *Bacteroides fragilis* and *Fusobacterium nucleatum*.[9] We also screened the current set of fecal metagenomes for the presence of these and other putative procarcinogenic bacteria (the *bft* toxin gene from enterotoxigenic *Bacteroides fragilis*, the *fadA* adhesin gene from *Fusobacterium nucleatum*, and the *cdtA-C* genes from *Campylobacter jejuni*),[35] but we only detected *bft* in 2 metagenomes from 2 patients. Because copy numbers of the *bft*, *clbB*, and *fadA* genes were reported to be similar in colonized patients,[9] and we readily detected *clb* genes (from the *pks* island), we do not expect this to be a technical limitation of our method, but possibly reflects differences in the study population. Because the other procarcinogenic bacteria appeared to be absent and could not be analyzed further, we focused our analyses on *pks*$^+$ *E coli*.

In a few fecal metagenomes, especially from donors, we detected *pks* in absence of *E coli*. Possibly, *E coli* was present below the limit of detection or the *pks* gene island belonged to a different pathogenic species of Enterobacteriaceae that may also carry the *pks* island.[36] Although we detected *Klebsiella*, *Enterobacter*, and *Citrobacter* spp using the assembly based approach in 6 of 11 donor samples with *pks* but that were negative for *E coli* based on the 3 single-copy marker genes, pathogenic isolates of these species are unlikely to be present in healthy donors due to thorough screening. We therefore consider it more likely that *E coli* was present below the detection limit of our assay and is in fact the host of the *pks* island.

Earlier studies of procarcinogenic intestinal bacteria applied different technical approaches for detection and quantification. Detection of the *pks* island was often based on the longest gene (*clbB*; 9.6 kilobase) by polymerase chain reaction or by metagenomic screening.[9,11] Metagenomics-based assays require high sequencing depths to reliably detect single genes. We combined data of all 19 *clb* genes, which we showed are highly intercorrelated, to improve the sensitivity of our assay and allow us to detect very low levels of *pks*. Thereby we increased the number of *pks*$^+$ samples by 8 compared with detection based on only *clbB*.

The *pks* levels (as RPKM values) in our screened feces donors were slightly lower than previously reported for healthy controls (donors: 0 RPKM median, 0.05 maximum; healthy controls: 0.06 median), whereas *pks* levels in *pks*$^+$ patients with rCDI (pre-FMT: 0.46 median, post-FMT: 1.35 median) were higher than those found in patients with inflammatory bowel disease (0.18 RPKM median).[11] This suggests that the screened and approved donors of the National Donor Feces Bank may have lower *pks* levels than average healthy controls. In addition, these patients with rCDI seem to have been exposed to *pks* levels higher than those in patients with inflammatory bowel disease.

By comparing prevalence of *pks* in different groups we come to similar conclusions. The prevalence of *pks* among healthy donors (11 of 38 samples [28.9%]) is comparable to other studies that reported prevalence of *pks* among healthy

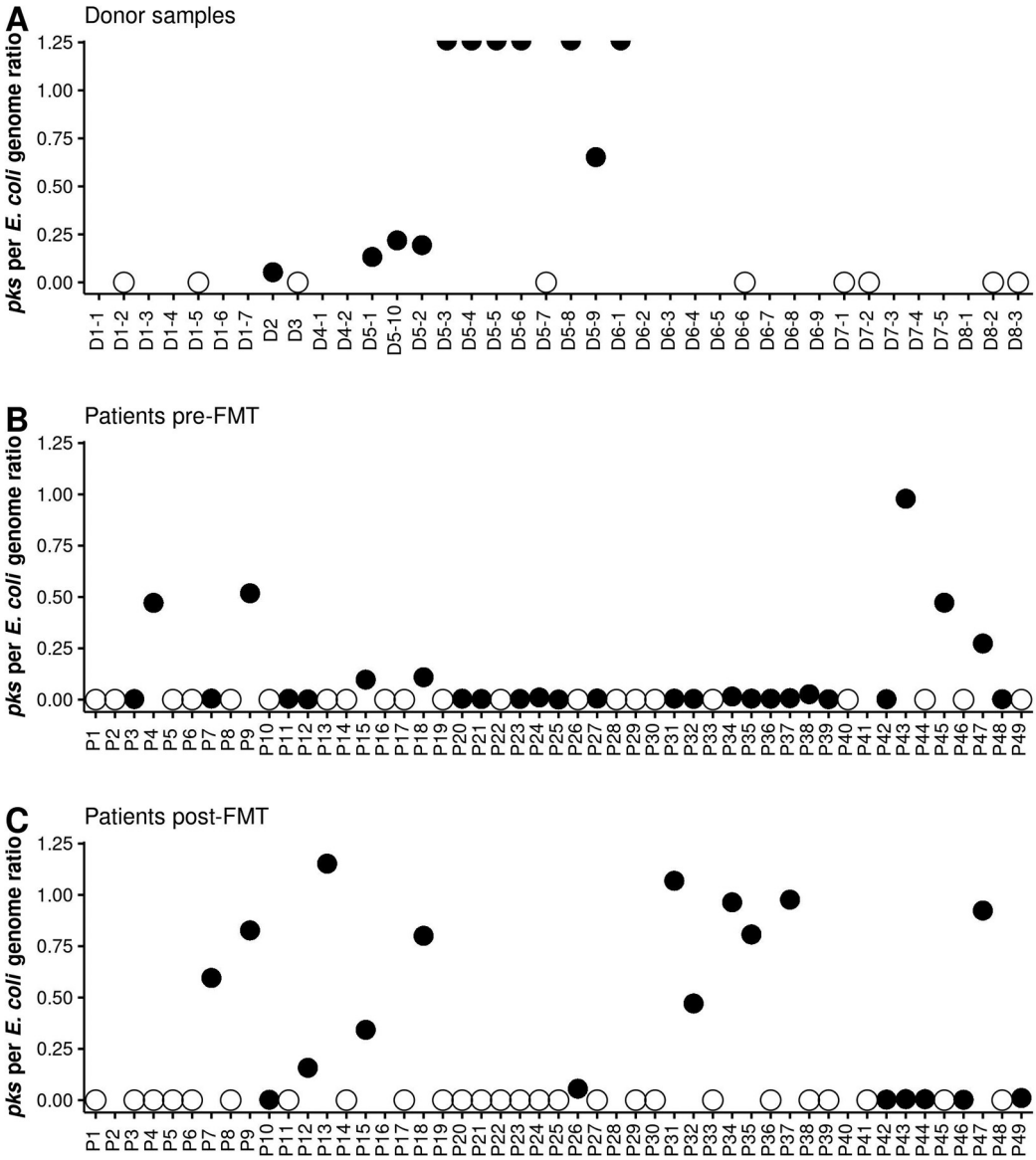


Figure 4. Average number of *pks* island copies per *E coli* genome, based on 3 marker genes. Average *pks* island abundances per sample are calculated as the average RPKM value of all 19 *clb* genes. *E coli* genome abundances are estimated based on the average RPKM value of the single-copy marker genes *dxs*, *rodA*, and *uidA*. The *empty circles* indicate presence of *E coli* and absence of *pks*, the *black semicircles* at the top of a figure indicate presence of *pks* and absence of *E coli*, and *missing circles* indicate neither *E coli* nor *pks* was detected in the metagenome. (*A*) *E coli* is detected in 14 donor samples, of which 5 samples from 2 different donors are pks^+. In these samples, there are between 0.05 and 0.65 *pks* island copies per *E coli* genome. The metagenomes of 6 donors have *pks* without *E coli* markers. (*B*) Nearly all patients have *E coli* in their fecal metagenome pre-FMT (all but P41). The *pks* per *E coli* ratios vary from 0 (*E coli* without the *pks* island) to 1 (all *E coli* genomes in the sample have the *pks* island). In most pks^+ cases, the *pks* island is present in a minority of *E coli* genomes. (*C*) After FMT, *E coli* is still detected in most patients, and the average number of *pks* copies per *E coli* genome varies from 0 to ~1. Fewer patients are pks^+ after FMT, and *pks* per *E coli* ratios are often higher: in 9 cases, the majority of *E coli* genomes has the *pks* island.

controls between 20% and 30%.[10,11,36] The prevalence of *pks* among patients with rCDI is higher (27 of 49 pre-FMT [55.1%] and 19 of 49 post-FMT [38.8%]) than in healthy donors. The numbers are between those reported for inflammatory bowel disease (~33%)[11] and those reported among patients with colorectal cancer (~60%).[10] However, the reported prevalence is often based on fewer than 50 patients and may therefore not accurately reflect the prevalence in the general population. Regardless, we find the *pks* levels and prevalence are approximately in the same range as described before and suggest that the relatively old patients with rCDI might be at an increased risk of developing colorectal cancer.

The age difference of approximately 40 years between our donors and patients is of note, because this may have been a confounding factor in the current study, assuming that age is correlated with pks^+ *E coli* colonization. To our knowledge, however, there has been no evidence indicating

that older age is associated with a higher incidence of *pks*$^{+}$ *E coli* colonization. Furthermore, Fukugaiti et al[37] suggested that rCDI is correlated to colorectal carcinogenesis. These hypotheses may be tested in a future study by screening a cohort of healthy people of similar age as the patients with rCDI and monitor them over time to compare the number of occurrences of colorectal cancer between these 2 groups.

Conclusion

We find that *pks* is prevalent and can be highly abundant in patients with rCDI. FMT with *pks*$^{-}$ donor suspensions generally decreases *pks* levels of patients, and *pks* persists in *pks*$^{+}$ patients when a *pks*$^{+}$ donor is used. We find no clear evidence of donor-to-patient transmission of *pks*$^{+}$ *E coli*, and if it were to occur, the impact may be negligible given the age and comorbidity of most patients in this particular cohort. Therefore, we think that current screening protocols for FMT donors are safe and that routine screening for *pks* is not required, at least for the treatment of rCDI. It is currently unknown how long *pks*$^{+}$ *E coli* persists in cured rCDI patients and on what time frame this may contribute to colorectal carcinogenesis. Further research addressing the long-term health effects is needed to evaluate the cancer risk due to *pks*$^{+}$ *E coli* in rCDI and other patients, because this might offer opportunities for early intervention in the development of colorectal cancer.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2021.06.009.

References

1. Smits WK, Lyras D, Lacy DB, et al. *Clostridium difficile* infection. Nat Rev Dis Primers 2016;2:1–20.
2. Chang JY, Antonopoulos DA, Kalra A, et al. Decreased diversity of the fecal microbiome in recurrent *Clostridium difficile*-associated diarrhea. J Infect Dis 2008;197:435–438.
3. Nood E van, Vrieze A, Nieuwdorp M, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. N Engl J Med 2013;368:407–415.
4. Terveer EM, Vendrik KE, Ooijevaar RE, et al. Faecal microbiota transplantation for *Clostridioides difficile* infection: four years' experience of the Netherlands Donor Feces Bank. United European Gastroenterol J 2020;8:1236–1247.
5. Khoruts A, Sadowsky MJ. Understanding the mechanisms of faecal microbiota transplantation. Nat Rev Gastroenterol Hepatol 2016;13:508–516.
6. Cammarota G, Ianiro G, Kelly CR, et al. International consensus conference on stool banking for faecal microbiota transplantation in clinical practice. Gut 2019;68:2111–2121.
7. Keller JJ, Ooijevaar RE, Hvas CL, et al. A standardised model for stool banking for faecal microbiota transplantation: a consensus report from a multidisciplinary UEG working group. United European Gastroenterol J 2021;9:229–247.
8. Scott AJ, Alexander JL, Merrifield CA, et al. International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. Gut 2019;68:1624–1632.
9. Drewes JL, Corona A, Sanchez U, et al. Transmission and clearance of potential procarcinogenic bacteria during fecal microbiota transplantation for recurrent *Clostridioides difficile*. JCI Insight 2019;4:e130848.
10. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, et al. Mutational signature in colorectal cancer caused by genotoxic *pks*+ *E. coli*. Nature 2020;580:269–273.
11. Dubinsky V, Dotan I, Gophna U. Carriage of colibactin-producing bacteria and colorectal cancer risk. Trends Microbiol 2020;28:P874–P876.
12. Olier M, Marcq I, Salvador-Cartier C, et al. Genotoxicity of *Escherichia coli* Nissle 1917 strain cannot be dissociated from its probiotic activity. Gut Microbes 2012;3:501–509.
13. Nougayrède J-P, Homburg S, Taieb F, et al. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. Science 2006;313:848–851.
14. Li Z-R, Li J, Cai W, et al. Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. Nat Chem 2019;11:880–889.
15. Dziubańska-Kusibab PJ, Berger H, Battistini F, et al. Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. Nat Med 2020;26:1063–1069.
16. Terveer EM, Beurden YH van, Goorhuis A, et al. How to: establish and run a stool bank. Clin Microbiol Infect 2017;23:924–930.
17. Schmitz D, Zwagemaker F, Nooij S, et al. DennisSchmitz/Jovian: Release 1.0.01. Zenodo, 2020. https://doi.org/10.5281/zenodo.3929618. Accessed July 14, 2020.
18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–2120.
19. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–359.
20. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–2079.
21. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–842.
22. Nurk S, Meleshko D, Korobeynikov A, et al. metaSPAdes: a new versatile metagenomic assembler. Genome Res 2017;27:824–834.
23. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.
24. Rubino F, Creevey CJ. MGkit: Metagenomic Framework For The Study Of Microbial Communities. https://figshare.com/articles/poster/MGkit_Metagenomic_Framework_For_The_Study_Of_Microbial_Communities/1269288/1; 2014. Accessed November 10, 2020; 2014.

25. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio] 2013. http://arxiv.org/abs/1303.3997. Accessed November 10, 2020.
26. Nooij S. *jovian-screener*. Zenodo. https://doi.org/10.5281/zenodo.4288679; 2020. Accessed November 10, 2020.
27. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2019. https://www.R-project.org/. Accessed November 10, 2020.
28. Lee C, Kim J, Shin SG, et al. Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. J Biotechnol 2006;123:273–280.
29. Chern EC, Siefring S, Paar J, et al. Comparison of quantitative PCR assays for *Escherichia coli* targeting ribosomal RNA and single copy genes. Lett Appl Microbiol 2011;52:298–306.
30. Wirth T, Falush D, Lan R, et al. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 2006;60:1136–1151.
31. Adiri RS, Gophna U, Ron EZ. Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. FEMS Microbiol Lett 2003;222:199–203.
32. Pritchard L, Glover RH, Humphris S, et al. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 2015;8:12–24.
33. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 2010;11:595.
34. 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;23:229–240.e5.
35. He Z, Gharaibeh RZ, Newsome RC, et al. *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin. Gut 2019;68:289–300.
36. Putze J, Hennequin C, Nougayrède J-P, et al. Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. Infect Immun 2009;77:4696–4703.
37. Fukugaiti MH, Ignacio A, Fernandes MR, et al. High occurrence of *Fusobacterium nucleatum* and *Clostridium difficile* in the intestinal microbiota of colorectal carcinoma patients. Braz J Microbiol 2015;46:1135–1140.

Author names in bold designate shared co-first authorship.

Received February 24, 2021. Accepted June 3, 2021.

Correspondence
Address correspondence to: Sam Nooij, MSc, Leiden University Medical Center, Department of Medical Microbiology, PO Box 9600, Post Zone E4-P, 2300 RC Leiden, the Netherlands. e-mail: s.nooij@lumc.nl.

Acknowledgments
The authors thank Dr Szymon M. Kiełbasa and the other members of the interdepartmental next-generation sequencing discussion group at the Leiden University Medical Center for their advice on the analysis and statistics, and the Experimental Bacteriology group for helpful comments during work discussions. Also, the authors thank Dr Bastian V.H. Hornung for assistance with handling the data set.

CRediT Authorship Contributions
Sam Nooij, MSc (Conceptualization: Equal; Data curation: Lead; Formal analysis: Lead; Methodology: Lead; Software: Lead; Validation: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead).

Quinten Ducarmon, MSc (Conceptualization: Supporting; Investigation: Supporting; Methodology: Supporting; Software: Supporting; Validation: Supporting; Visualization: Supporting; Writing – original draft: Supporting; Writing – review & editing: Equal).

Jeroen Laros, PhD (Methodology: Equal; Supervision: Supporting; Visualization: Equal; Writing – review & editing: Supporting).

Romy Zwittink, PhD (Supervision: Equal; Visualization: Supporting; Writing – review & editing: Supporting).

Jason Norman, PhD (Funding acquisition: Equal; Methodology: Supporting; Resources: Lead; Validation: Supporting; Writing – review & editing: Supporting).

Wiep Klaas Smits, PhD (Supervision: Supporting; Visualization: Supporting; Writing – review & editing: Supporting).

Hein Verspaget, PhD (Data curation: Supporting; Methodology: Supporting; Validation: Supporting; Writing – review & editing: Supporting).

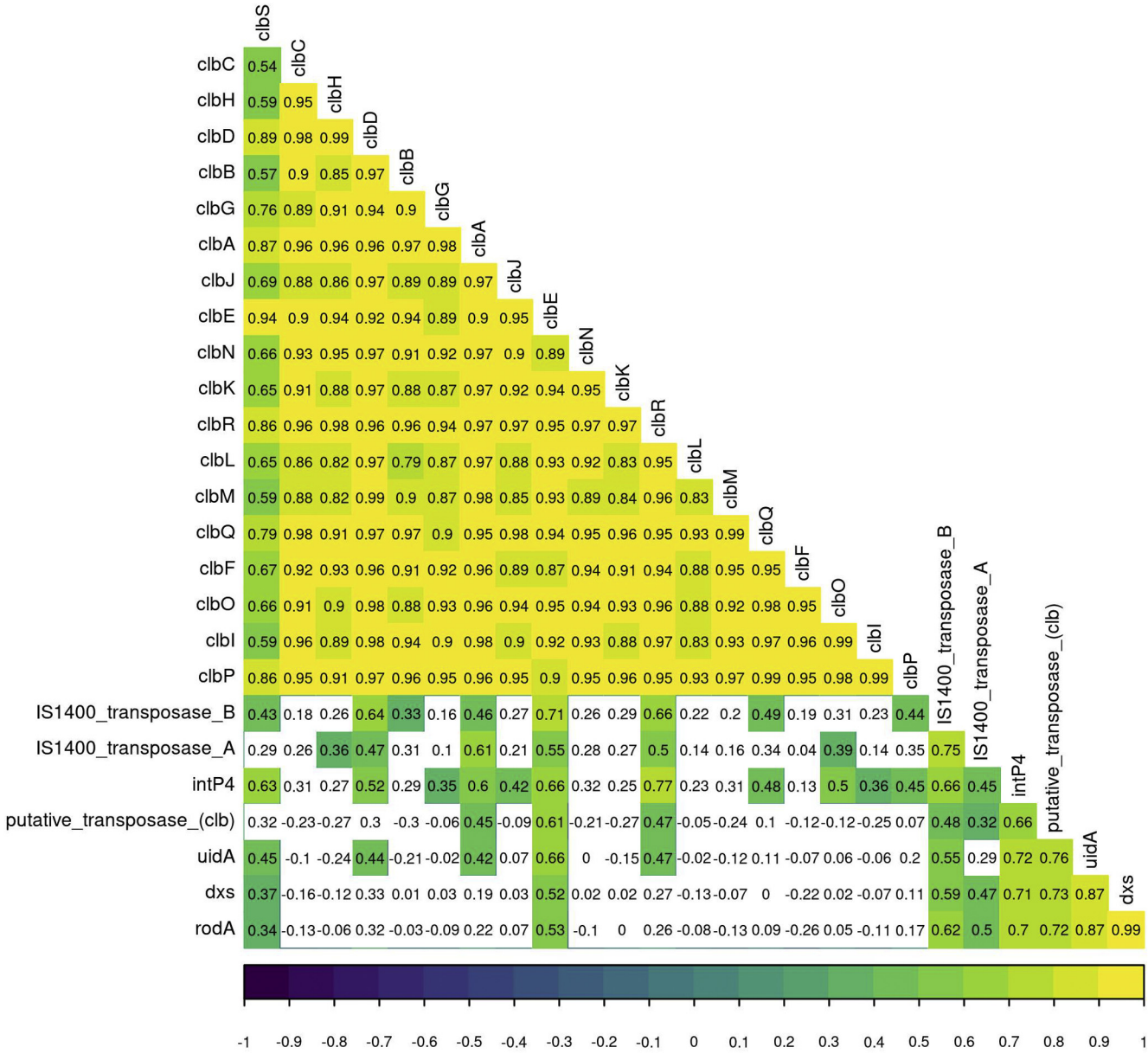
Josbert Keller, MD, PhD (Funding acquisition: Equal; Validation: Supporting; Writing – review & editing: Supporting).

Elisabeth Terveer, MD (Conceptualization: Supporting; Data curation: Equal; Funding acquisition: Equal; Investigation: Supporting; Project administration: Equal; Supervision: Equal; Writing – review & editing: Supporting).

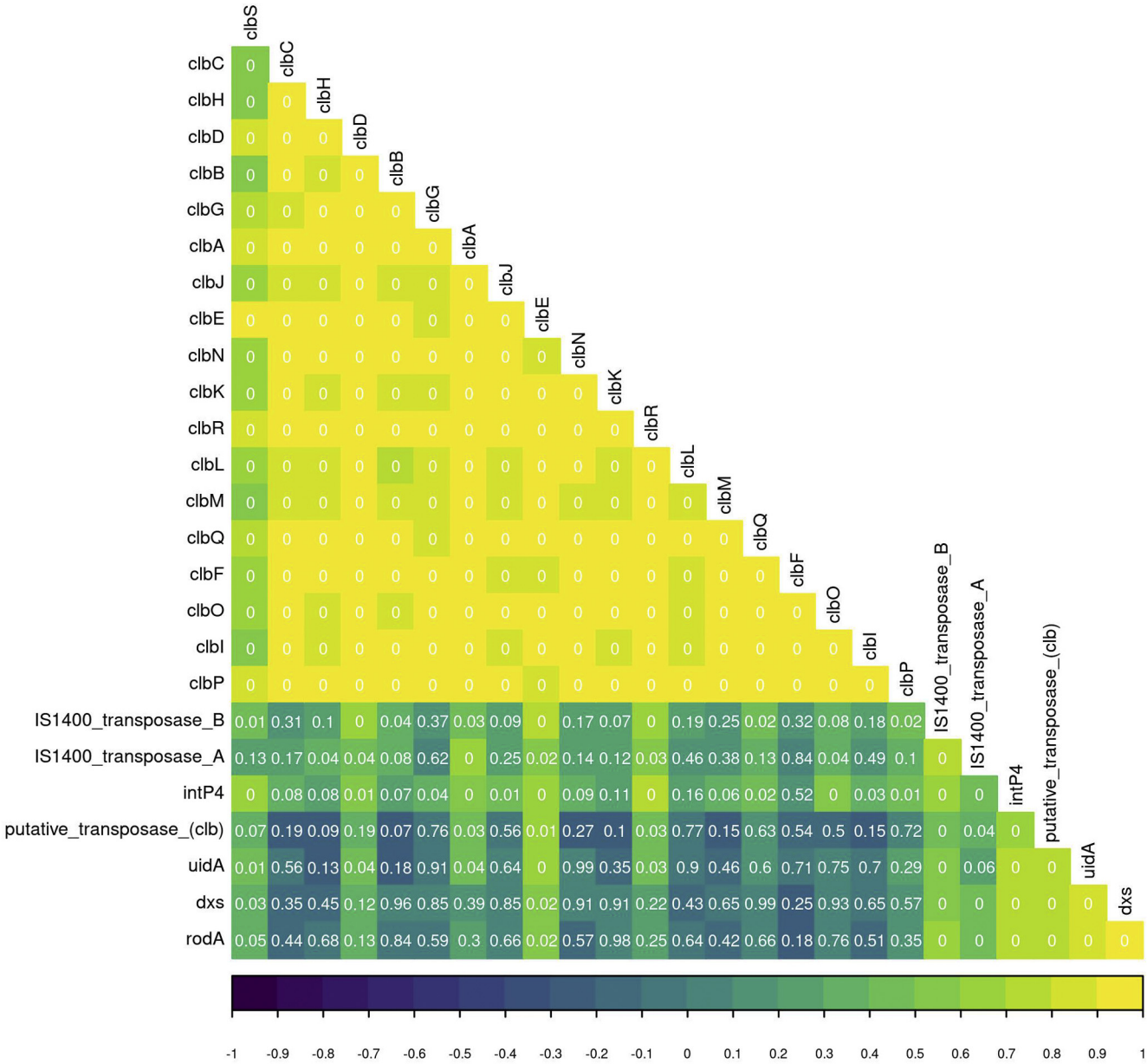
Ed Kuijper, MD, PhD (Conceptualization: Lead; Funding acquisition: Lead; Supervision: Lead; Validation: Supporting; Visualization: Supporting; Writing – review & editing: Supporting).

Conflicts of interest
The Netherlands Donor Feces Bank received an unrestricted grant from Vedanta Biosciences, Inc (Cambridge, MA) from 2017 to 2020. Metagenomic sequencing was funded by Vedanta Biosciences, Inc. The authors received no financial support for the research, authorship, and/or publication. Jason M. Norman is an employee of Vedanta Biosciences. The other authors declare no competing interests.

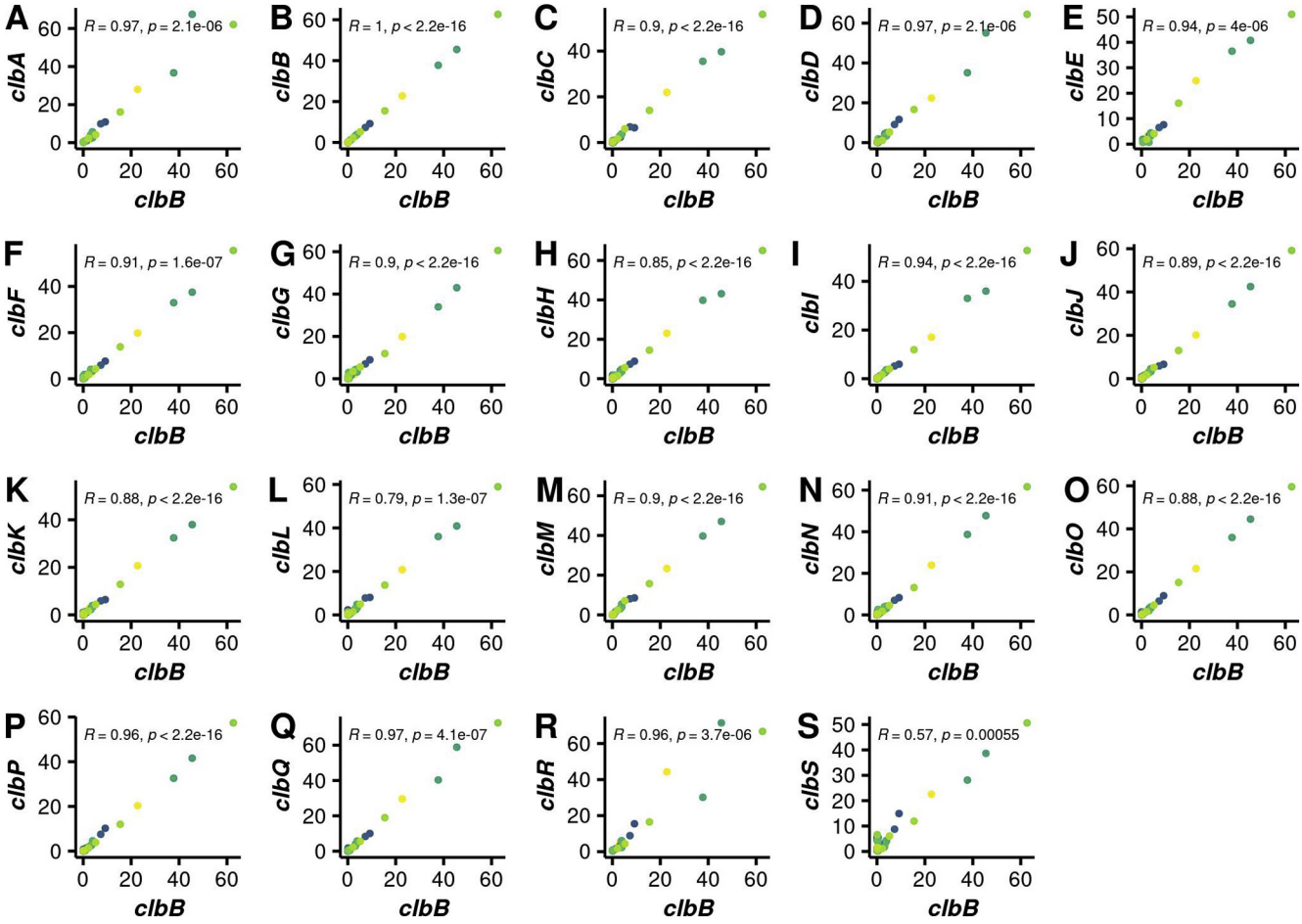
Funding
This work was funded by an unrestricted grant by Vedanta Biosciences, Inc.



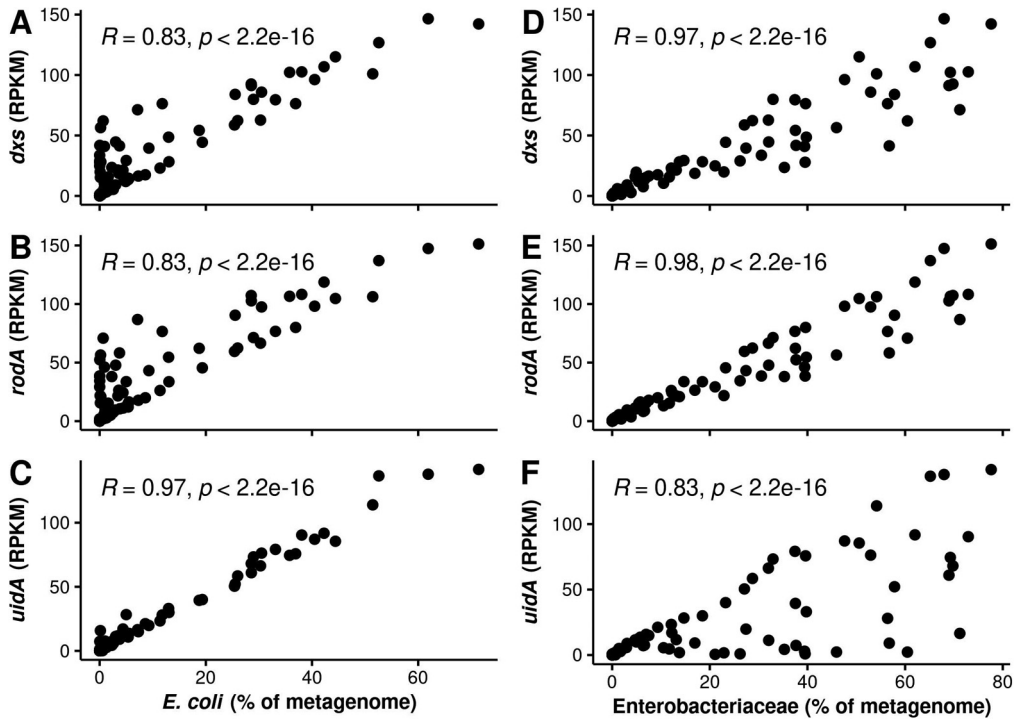
Supplementary Figure 1. Spearman's correlations matrix of RPKM values of *pks* island genes, flanking regions, and selected *E coli* marker genes. Trimmed and human-filtered reads from all donor and patient metagenomes were mapped to the *pks* island (accession ID AM227896), and putative *E coli* single-copy marker genes and values for each gene were normalized by gene length and library size (RPKM). The RPKM values for each of these genes were compared between samples using Spearman's correlations and rounded to 2 decimals. Each cell is labeled with the Spearman's correlation value and is also colored according to the *r* value. The *clb* genes (encoding the genotoxin colibactin) are highly correlated to one another and are less correlated to the other genes.



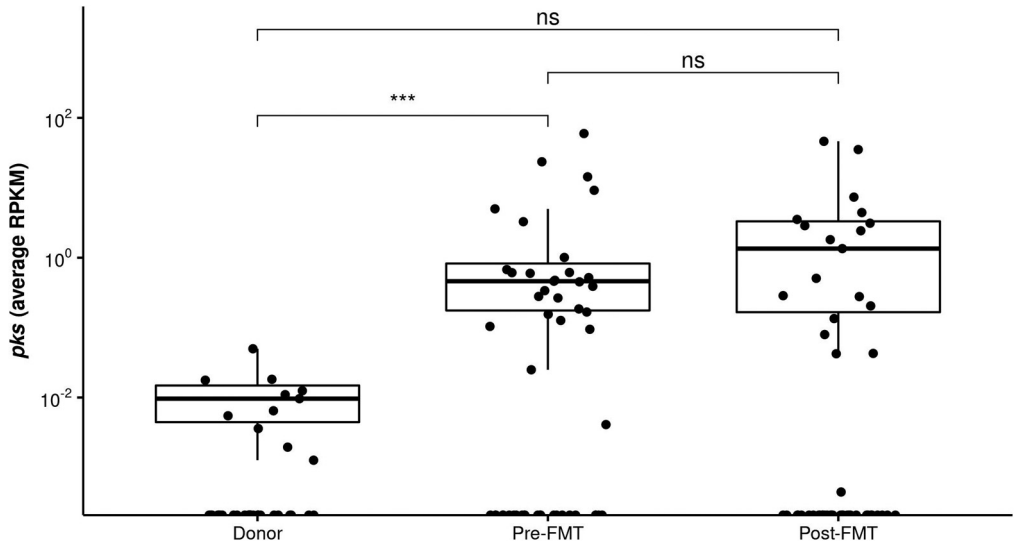
Supplementary Figure 2. *P* values of Spearman's correlations between RPKM values of *pks* island genes, flanking regions, and selected *E coli* marker genes calculated using methods identical to Supplementary Figure 1. Now the values in the cells represent *P* values rounded to 2 decimals. *P* values between *clb* genes are all $< .01$.



Supplementary Figure 3. (*A–S*) RPKM values of each *clb* gene in each sample plotted against *clbB*, with Spearman's correlations. Correlations between *clb* genes are mostly around $r = .9$ and highly significant ($P < .001$), only the correlation between *clbB* and *clbS* is less strong ($r = .57$).



Supplementary Figure 4. Spearman's correlations between 3 single-copy marker genes and the relative abundance of *E coli* and Enterobacteriaceae per sample. Abundances are based on de novo assembled scaffolds, quantified by mapping reads back to the scaffolds and normalized to percentages of all reads in the metagenome. Enterobacteriaceae abundances include *E coli*. (*A*) The correlation between *uidA* and *E coli* is extremely high ($r = .97$). The correlation between *E coli* and (*B*) *dxs* or (*C*) *rodA* is high ($r = .86$), but lower than with *uidA*. *E coli* is absent in a number of metagenomes with relatively high levels of *dxs* or *rodA*. (*D*) The correlation between *uidA* and Enterobacteriaceae is high ($r = .84$), but *uidA* is sometimes absent in metagenomes with $>$ 10% Enterobacteriaceae. Enterobacteriaceae are more strongly correlated with the abundance of (*E*) *dxs* ($r = .96$) and (*F*) *rodA* ($r = .97$).



Supplementary Figure 5. Comparison of *pks* levels (as average RPKM) in all donor and patient samples, including *pks*$^-$ samples. To compare *pks* between donors and patients, we quantified *pks* as the average number of reads mapped to 19 *clb* genes, normalized to RPKM. A pairwise Wilcoxon's rank sum test indicates that *pks* levels differ between donor samples and patients with rCDI pre-FMT ($P < .001$), but are not different between donor and post-FMT ($P = .078$) and between patients pre- and post-FMT ($P = .179$). Of 38 donor samples included, 27 had 0 *pks*, and of 49 patient samples, 22 pre-FMT and 30 post-FMT had 0 *pks* (indicated as *half dots* on the *x*-axis). The *horizontal line* in the middle of each *box* indicates the median; the *top and bottom borders* of the box mark the 75th and 25th percentiles, respectively, and the *vertical lines* mark minimum and maximum of all the data. ***$P < .001$; ns, not significant.

Donor	Pre-FMT	Post-FMT	
0.0000	0.0000	0.0000	P1
0.0000	0.0000	0.0000	P2
0.0000	0.0000	0.0000	P5
0.0000	0.0000	0.0000	P6
0.0000	0.0000	0.0000	P8
0.0000	0.0000	0.0000	P19
0.0000	0.0000	0.0000	P22
0.0000	0.0000	0.0000	P28
0.0000	0.0000	0.0000	P29
0.0000	0.0000	0.0000	P33
0.0000	0.0000	0.0000	P40
0.0000	0.0000	0.0000	P41
0.0000	0.0000	0.0423	P10
0.0000	0.0000	0.1348	P46
0.0000	0.0000	0.2046	P49
0.0000	0.0000	0.5064	P44
0.0000	0.0000	1.3479	P26
0.0000	0.0041	0.0000	P25
0.0000	0.0249	0.0000	P3
0.0000	0.0945	0.0000	P23
0.0000	0.1040	0.0000	P39
0.0000	0.1550	0.0000	P48
0.0000	0.1850	0.0004	P42
0.0000	0.2659	0.0000	P20
0.0000	0.2792	0.0000	P21
0.0000	0.3899	3.1095	P7
0.0000	0.4520	0.0000	P27
0.0000	0.4614	0.0000	P24
0.0000	0.4745	0.0000	P36
0.0000	0.6108	0.0000	P11
0.0000	1.0079	0.0000	P38
0.0000	4.9871	1.8087	P47
0.0000	14.3814	0.0000	P45
0.0000	23.5230	2.4323	P9
0.0000	59.7469	0.2779	P43
0.0013	0.0000	0.0000	P14
0.0019	0.0000	0.0000	P17
0.0019	3.2755	7.3547	P18
0.0036	0.3384	35.2527	P32
0.0055	0.0000	0.0000	P16
0.0055	9.1926	0.0428	P15
0.0065	0.6145	0.0000	P4
0.0096	0.5979	2.8743	P37
0.0110	0.5179	4.4302	P34
0.0125	0.0000	0.0000	P30
0.0125	0.6769	46.2759	P31
0.0177	0.1267	0.0796	P12
0.0182	0.0000	0.2863	P13
0.0498	0.1670	3.5270	P35

Supplementary Table 1. Sequence Types of Assembled *Escherichia coli* Scaffolds As Determined by In Silico Multilocus Sequence Typing

Patient	Sequence typing	
	Pre-FMT	Post-FMT
P7	540	…
P8	1611	…
P10	38	…
P11	297	…
P12	399	…
P15	69	…
P21	131	131
P23	69	…
P26	…	405
P27	10	…
P31	393	…
P32	58	…
P36	155	155
P37	405	…
P39	131	…
P43	73	…
P44	657	…
P45	…	131
P48	2453	…
P49	…	453

Supplementary Figure 6. *pks* levels (as average RPKM) for each donor-patient combination. For each patient, listed on the right, the pre-FMT and post-FMT *pks* values are listed, as well as the *pks* level of the donor sample with which the patient was treated. (Except for P2 and P5, for which the donor samples collected 7 and 19 days later are included.)