



# Enterobacteriales Infection after Intestinal Dominance in Hospitalized Patients

 Krishna Rao,<sup>a</sup>  Anna Seekatz,<sup>a</sup> Christine Bassis,<sup>a</sup> Yuang Sun,<sup>b</sup> Emily Mantlo,<sup>b</sup>  Michael A. Bachman<sup>b</sup>

<sup>a</sup>Department of Internal Medicine, Division of Infectious Diseases, University of Michigan, Ann Arbor, Michigan, USA

<sup>b</sup>Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA

**ABSTRACT** The *Enterobacteriales* order of Gram-negative bacteria includes the common nosocomial pathogens *Klebsiella pneumoniae*, *Escherichia coli*, *Serratia marcescens*, and *Enterobacter* species. Intestinal domination by some colonizing bacterial taxa is associated with subsequent infection, but 16S rRNA gene sequencing is too costly and slow to be used in a clinical setting. The objectives of this study were to develop a PCR-based assay that can measure *Enterobacteriales* density, validate it against 16S rRNA gene sequencing, and measure the association between *Enterobacteriales* dominance and subsequent infection. Two quantitative PCR (qPCR) assays that were developed to quantify the absolute and relative abundance of *Enterobacteriales* had good correlation with 16S rRNA sequence analysis ( $P < 0.0001$ ). Using both PCR assays and 16S sequencing, a matched case-control study was performed comparing rectal swabs from hospitalized patients who later developed bloodstream, urinary tract, or respiratory *Enterobacteriales* infections ( $n = 95$ ) to swabs from patients who remained uninfected ( $n = 189$ ). *Enterobacteriales* abundance measured by sequencing was high in both cases and controls (means, 31.1% and 27.5%, respectively;  $P = 0.322$ ). We observed an increased risk of infection that depended on both the absolute and relative abundance of *Enterobacteriales* as measured by qPCR assay A ( $P = 0.012$ ). After adjustment for albumin levels, central venous catheter presence, and use of cephalosporins at the time of swab collection, this association still approached significance ( $P = 0.061$ ). These results demonstrate that using qPCR to measure intestinal colonization dominance is feasible, indicate that hospitalized patients have high levels of *Enterobacteriales* colonization, and suggest that both relative and absolute abundance may be associated with subsequent infection.

**IMPORTANCE** Increasing antibiotic resistance has resulted in infections that are life-threatening and difficult to treat. Interventions that prevent these infections, particularly without using antibiotics, could save lives. Intestinal colonization by pathogens, including vancomycin-resistant *Enterococcus* and carbapenem-resistant *Enterobacteriaceae* (part of the order *Enterobacteriales*) is associated with subsequent infection, and increased colonization density is associated with increased infection risk. Therefore, colonization offers a window of opportunity for infection prevention if (i) there are rapid and inexpensive assays to detect colonization, (ii) there are safe and effective interventions, and (iii) the risk of infection outweighs the risk of the treatment. Fecal transplants are proof of principle that manipulating the microbiome can reduce such colonization and prevent infections. This study demonstrates the feasibility of implementing rapid and inexpensive assays to quantify colonization and measures the strength of association between *Enterobacteriales* dominance and subsequent infection. The approach described here could be a valuable tool in the prevention of antibiotic-resistant infections.


**KEYWORDS** *Enterobacteriales*, infection prevention, microbiome, prognostic indicators

**Citation** Rao K, Seekatz A, Bassis C, Sun Y, Mantlo E, Bachman MA. 2020. *Enterobacteriales* infection after intestinal dominance in hospitalized patients. *mSphere* 5:e00450-20. <https://doi.org/10.1128/mSphere.00450-20>.

**Editor** Mariana Castanheira, JMI Laboratories

**Copyright** © 2020 Rao et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Michael A. Bachman, [mikebach@med.umich.edu](mailto:mikebach@med.umich.edu).

 Its not all relative! Absolute and relative abundance of Enterobacteriales on rectal swabs interact to associate with subsequent infections, and can be measured by qPCR. @Bachman\_Lab

**Received** 15 May 2020

**Accepted** 2 July 2020

**Published** 22 July 2020

Colonization with *Enterobacterales* species, an order encompassing the *Enterobacteriaceae* family, including *Escherichia coli*, *Klebsiella*, and *Enterobacter*, and other families of Gram-negative pathogens (1), has been associated with subsequent infection, which is both a challenge and opportunity for infection prevention. As an example, for *Klebsiella*, 5.2% of colonized patients became infected compared to 1.3% of noncolonized patients (2). This association appears stronger in antibiotic-resistant *Enterobacteriaceae*—based on a review of 10 studies, carbapenem-resistant *Enterobacteriaceae* (CRE) infections occurred in 16.5% of colonized, hospitalized patients (3). Increasing colonization density may increase the risk of subsequent infection. In patients undergoing stem cell transplantation, a relative abundance of *Proteobacteria* above 30% was associated with subsequent Gram-negative bacteremia (hazard ratio, 5.2) (4). In long-term acute care patients, a relative abundance of CRE *Klebsiella pneumoniae* above 22% was associated with a relative risk of 4.2 for subsequent infection (5). This association of colonization with subsequent infection provides a window of opportunity for intervention with emerging therapies, including decolonization by antibiotics or manipulation of the microbiome. In fact, in 142 patients across 23 reports who underwent fecal transplant for decolonization of multidrug-resistant organisms (MDROs), 78% achieved microbiological clearance (6), and case reports suggest interruption of MDRO infections (7).

Measuring colonization density could be an effective tool for infection prevention if it could be deployed in clinical laboratories. To date, measures of colonization density have required massively parallel 16S rRNA gene sequencing that is expensive and slow. Quantitative PCR (qPCR) assays could be a rapid and less expensive approach to detect and quantify intestinal colonization by pathogens of concern in our hospitals. Since qPCR has become a standard tool in clinical microbiology, deploying these assays would not require additional expertise or equipment. To test the feasibility of qPCR measurement of colonization density and the association between *Enterobacterales* dominance and subsequent infection, we developed two qPCR assays, compared their performance to 16S sequencing, and assessed their association with subsequent *Enterobacterales* infection in a case-control study.

## RESULTS

**In silico analysis of primer and probe binding.** The TaqMan assays were designed to use an *Enterobacterales* probe and a broad-range bacterial probe targeting different regions of the same bacterial 16S rRNA gene amplicon (8, 9) (see Fig. S1A in the supplemental material). To do so, we identified a conserved region of the bacterial 16S rRNA gene amplicon in 35 clinically relevant *Enterobacterales* and designed two probes that accounted for a GG/AC polymorphism in this region. In combination, the two *Enterobacterales* probes exactly matched 274/319 *Enterobacterales* sequences using the Ribosomal Database Project's Probe Match (85.9% [Table 1]). Outside the *Enterobacterales* order, the MGB\_ENT\_GG\_Probe matched only 2,295/1,449,392 (0.2%) non-*Enterobacterales* and the MGB\_ENT\_AC\_Probe matched only 22,455/1,449,392 (1.5%) non-*Enterobacterales*, even when expanding the search to also include nontype strains and uncultured bacteria. Most non-*Enterobacterales* matches were to *Gammaproteobacteria* (24,707/24,750 [99.8%]) but were not likely to be found in human gut samples (e.g., *Vibrionales*, *Oceanospirillales*). However, the MGB\_ENT\_AC\_Probe did match many *Pasteurellales*, including *Haemophilus*, which can be found in human gut samples, although not normally at high levels. MGB\_Uniprobe\_p1, a shorter version of the UniProbe used by Yang et al. (8) had an exact match to 11,627/12,227 (95.1%) 16S sequences in Ribosomal Database Project (RDP) release 11.5 categorized as type strains, isolates, size  $\geq$  1,200, and good quality. The Forward primer A and Reverse primer A (8) did not match many *Bacteroidetes* sequences in the RDP database (62.2% and 36.3%, respectively, Table 1). Therefore, we designed a second assay using the same *Enterobacterales* probes but with primers matching a broader range of bacteria, including >90% *Bacteroidetes* (Forward primer B 1,194/1,315 [90.8%], Reverse primer B, 1,256/1,315 [95.5%]) (Table 1 and Fig. S1B). The second assay also used a different universal

**TABLE 1** Primers and probes used in this study

Oligonucleotide primer or probe <sup>a</sup>	Final Rx concn (nM) <sup>b</sup>	Sequence (5' to 3')	Assay	No. of exact sequence matches/total no. of sequences (%) <sup>c</sup>				
				Enterobacteriales	Proteobacteria (non-Enterobacteriales)	Firmicutes	Bacteroidetes	
MGB_ENT_GG_Probe (FAM)	250	TACCTGGTCTTGACATCCA	Both	62/319 (19.4)	59/4,273 (1.4)	0/2,438 (0)	0/1,315 (0)	
MGB_ENT_AC_Probe (FAM)	250	TTACCTACTCTTGACATCCA	Both	212/319 (66.4)	353/4,273 (8.3)	0/2,438 (0)	0/1,315 (0)	
MGB_Uniprobe_p1 (VIC)	250	ACGAGCTGACGACARC	A	314/319 (98.4)	4,120/4,273 (96.4)	2,357/2,438 (96.7)	1,254/1,315 (95.4)	
Uniprobe_B (VIC)	250	ATTGACGGGRRCCGCAAG	B	315/319 (98.7)	3,762/4,273 (88.0)	2,327/2,438 (95.4)	1,034/1,315 (78.6)	
Forward primer A (P891F)	900	TGGAGCATGTGGTTAATTCGA	A	312/319 (97.8)	3,214/4,273 (75.2)	16,83/2,438 (69.0)	818/1,315 (62.2)	
Reverse primer A (p1022R)	900	TGCGGGACTTAACCCAACA	A	316/319 (99.1)	3,192/4,273 (75.3)	2,259/2,438 (92.7)	478/1,315 (36.3)	
Forward primer B (F785)	900	GGATTAGATACCCTGGTAGTCC	B	316/319 (99.1)	4,138/4,273 (96.8)	2,214/2,438 (90.8)	1,194/1,315 (90.8)	
Reverse primer B (R1097 - P2)	900	GAGCTGACGACARCCATGC	B	314/319 (98.4)	3,980/4,273 (93.1)	2,329/2,438 (95.5)	1,256/1,315 (95.5)	

<sup>a</sup>FAM, 6-carboxyfluorescein.<sup>b</sup>Rx, reaction.<sup>c</sup>Matching 16S sequences from type strains, isolates, size  $\geq 1,200$ , and good quality in RDP release 11.5.

**TABLE 2** Linearity of qPCR assays

Assay	Serial dilution sample	Enterobacterales probe (Ent)			Universal probe (Uni)		
		Range (log <sub>10</sub> CFU)	Slope	r <sup>2</sup>	Range (log <sub>10</sub> CFU)	Slope	r <sup>2</sup>
A	<i>K. pneumoniae</i>	3–9	–3.441	>0.99	3–9	–3.374	>0.99
	<i>S. marcescens</i>	3–8	–3.46	>0.99	3–8	–3.46	>0.99
	<i>K. pneumoniae</i> , <i>E. coli</i> , <i>S. marcescens</i> , <i>Citrobacter freundii</i> , and <i>Enterobacter cloacae</i>	3–9	–3.371	>0.99	3–9	–3.282	>0.99
B	<i>K. pneumoniae</i>	3–9	–3.647	>0.99	3–9	–3.554	>0.99
	<i>S. marcescens</i>	4–8	–3.594	>0.99	4–8	–3.683	>0.99
	<i>K. pneumoniae</i> , <i>E. coli</i> , <i>S. marcescens</i> , <i>Citrobacter freundii</i> , and <i>Enterobacter cloacae</i>	3–9	–3.529	>0.99	3–9	–3.33	>0.99

probe, Uniprobe B, with an exact match to 10,975/12,736 (86.2%) bacterial sequences categorized as type strains, isolates, size  $\geq 1,200$ , and good quality, including wide coverage of the major groups of human gut bacteria *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*.

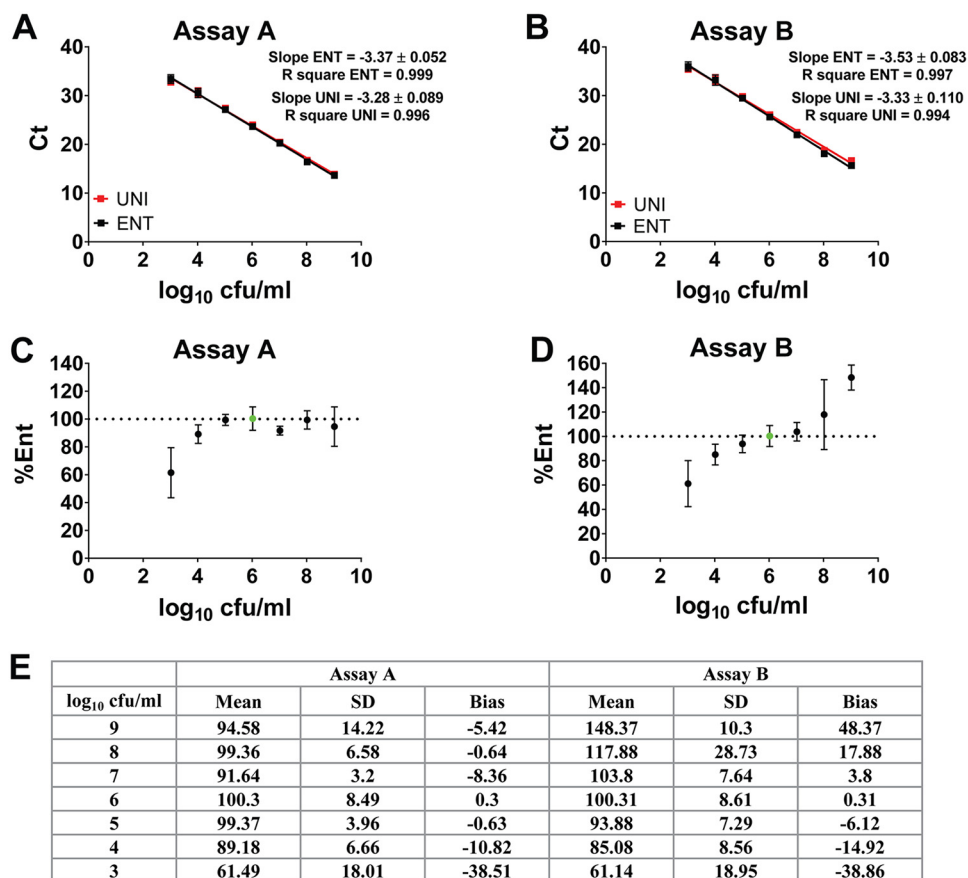
**Assay optimization.** Despite differences in the 16S rRNA gene sequences between *Enterobacterales* and non-*Enterobacterales*, some cross-reactivity was observed in binding of the *Enterobacterales* probe. Although the amplification curves were not exponential, final fluorescence was strong enough to cross the fluorescence threshold used to quantify the PCR result. To increase specificity, a number of PCR additives were tried empirically. Addition of dimethyl sulfoxide (DMSO) (5% [vol/vol]) was able to decrease or eliminate reactivity with three non-*Enterobacterales*: *Pseudomonas aeruginosa*, *Escherichia coli*, and *Lachnospiraceae* (Fig. S2). This additive was used for all subsequent testing.

PCR using universal primers can be confounded by contamination of the reaction mixture from the laboratory or the reagents themselves (10). Indeed, we observed reactivity in no-template control reactions. To eliminate this contaminating DNA, which might alter the quantification of total bacterial DNA, we added a proprietary derivative of propidium monoazide, PMAxx (Biotium, Fremont, CA; 6  $\mu$ M) to the reaction mixture. PMA is an intercalating agent that binds readily to double-stranded DNA (dsDNA). Photolysis (460-nm-wavelength exposure) produces a nitrene that can then covalently bind to DNA, forming cross-links that prevent use as a PCR template (10–12). Addition of PMAxx eliminated reactivity seen in the no-template control while allowing robust amplification of a *K. pneumoniae* 16S template (Fig. S3).

**Linearity, precision, and reportable range.** To compare the linearity and reportable ranges of assay A and assay B, serial dilutions of various templates were performed, and linear regression was performed based on input DNA and cycle threshold ( $C_T$ ) values (Table 2). Both assays were linear, although the slopes were closer to –3.3 for assay A, which would represent perfect PCR efficiency, and closer to each other for the Ent probe and the universal probe. Since the Ent probe is a mixture containing a GG dinucleotide (binds to *K. pneumoniae* among others) and AC (binds to *Serratia marcescens* among others), we tested linearity and reportable range in serial dilutions of *S. marcescens* genomic DNA (Table 2). Again, the slope for assay A was closer to –3.3 than assay B.

Rectal swabs are likely to contain mixtures of *Enterobacterales* with both the AC and GG polymorphisms in the probe target sequence. To assess linearity and reportable range from a mixture of *Enterobacterales*, overnight cultures of *K. pneumoniae*, *E. coli*, *S. marcescens*, *Citrobacter freundii*, and *Enterobacter cloacae* were mixed in equal volumes, serial dilutions were made, and six replicates were tested by assays A and B. As seen for individual bacteria, assay A had a slope close to –3.3 for both Ent and Uniprobe, whereas for assay B, the Ent probe slope (–3.5) diverged from the Uniprobe slope (–3.3; Fig. 1A and B; Table 2).

To calculate percent *Enterobacterales*, the delta-delta- $C_T$  ( $\Delta\Delta C_T$ ) method was used, calculating the ratio of Ent and Uniprobe  $C_T$  values and using the log<sub>10</sub> 6 CFU/ml

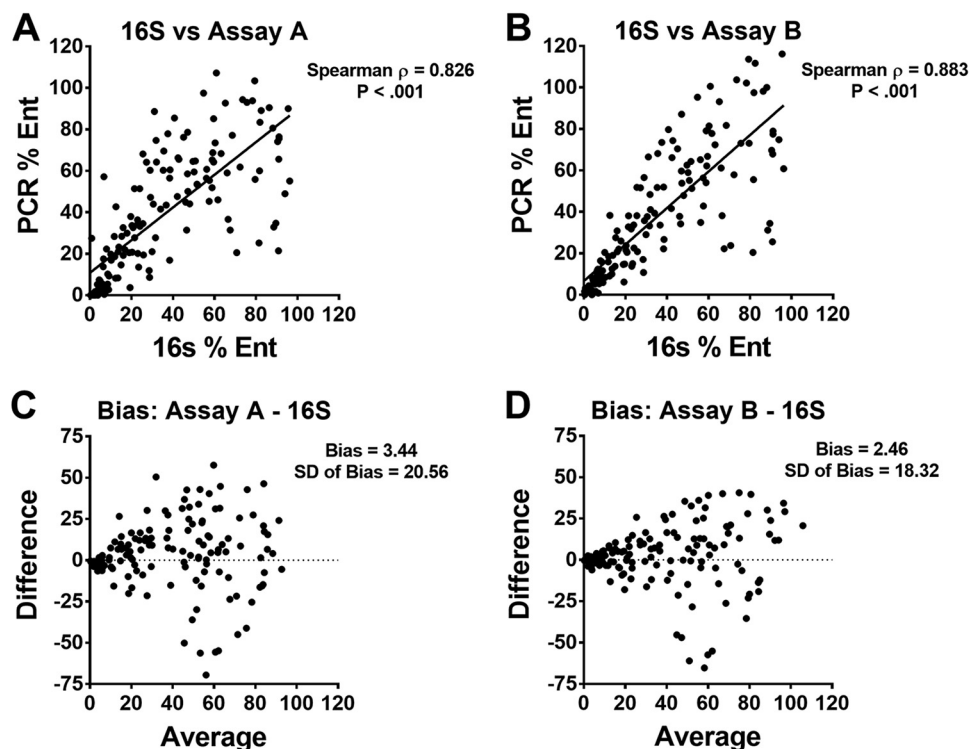


**FIG 1** Linearity and reportable range of assays A and B on a mixture of five *Enterobacteriales* species (*K. pneumoniae*, *E. coli*, *S. marcescens*, *Citrobacter freundii*, and *Enterobacter cloacae*). (A and B) Extracted genomic DNA mixtures in a serial dilution were amplified by assay A (A) and assay B (B) with six technical replicates. (C and D) Percentages of *Enterobacteriales* based on assay A (C) and assay B (D) based on the ddC<sub>7</sub> method relative to the log<sub>10</sub> 6 CFU/ml value (green). (E) Mean, standard deviation (SD), and bias of percent *Enterobacteriales* calculated from assay A (C) and assay B (D).

samples as the value representing 100% *Enterobacteriales*. Assay A did not show significant bias with changing concentrations and showed good precision (standard deviation [SD] < 10) between log<sub>10</sub> 4 to 8 CFU/ml (Fig. 1C and E). In assay B, there was proportional bias in the calculated percent *Enterobacteriales* with overestimation at higher total CFU and underestimation at lower CFU, likely attributable to the differences in PCR efficiency slopes (Fig. 1D; Table 2). The precision was <10 between log<sub>10</sub> 4 to 7, corresponding with the lowest bias (Fig. 1D and E).

The bacterial population on the rectal swabs will be a mixture of *Enterobacteriales* and non-*Enterobacteriales* species, and the total fecal content may vary from swab to swab. To determine the linearity and reportable range based on total bacterial CFU, the five *Enterobacteriales* species were mixed with two non-*Enterobacteriales*, *Bacteroides ovatus* and *Eubacterium rectale*. For the qPCR experiments, the percent *Enterobacteriales* was calculated with the mean dC<sub>7</sub> from positive-control *K. pneumoniae* and *S. marcescens* genomic DNA (gDNA) defined as 100%. All assays were internally consistent in percent *Enterobacteriales* calculated down to a total bacterial CFU of log<sub>10</sub> 5 CFU/ml (Fig. S4 [the dashed line represents the lower limit of reportable range]). Taken together, the linearity and precision data above indicate that the reportable range of qPCR assays A and B are a total bacterial concentration of log<sub>10</sub> 5 to 9 CFU/ml of sample. This corresponded to an average Uniprobe C<sub>7</sub> of 14.5 to 30.1 for assay A and 16.6 to 32.6 for assay B.

**Accuracy.** To assess the accuracy of each PCR assay compared to 16S rRNA gene sequencing (V4 region), we extracted DNA and performed sequencing and PCR



**FIG 2** Accuracy of qPCR assays compared to 16S rRNA gene sequencing from 150 rectal swabs. (A) Percentages of *Enterobacteriales* measured by 16S sequencing compared to assay A (A) and assay B (B). Bland-Altman plot comparing 16S sequencing to assay A (C) and assay B (D).

assays A and B on patient rectal swabs. To test for detection of samples with known *Enterobacteriales*, we included 50 rectal swabs that were culture positive for *K. pneumoniae*. Additionally, we included 100 samples that were culture negative for *K. pneumoniae*. The percent *Enterobacteriales* of the 150 samples by 16S sequencing ranged from 0.332% to 96.2%. For qPCR analysis, *K. pneumoniae* genomic DNA was used as the calibrator to define the  $dC_T$  that corresponds to 100% *Enterobacteriales*. Spearman correlation was  $\rho = 0.826$  ( $P < 0.01$ ) for assay A and  $\rho = 0.883$  ( $P < 0.01$ ) for assay B (Fig. 2A and B). The bias was +3.4 for assay A and +2.5 for assay B (Fig. 2C and D). Despite the predicted differences in inclusivity of the primers in each assay, the correlation between qPCR assays is excellent ( $r = 0.9486$ , 95% CI of 0.9291 to 0.9629, and  $P < 0.0001$ ). Overall, these data indicate that both qPCR assays are highly correlated with 16S sequencing results in calculating percent *Enterobacteriales* from rectal swab samples.

**Initial and unadjusted analyses using 16S data.** The data set from 284 patient samples consisted of 95 cases and 189 matched controls (there was one case that could be matched to only a single control; see Data Set S1 in the supplemental material). Infections were caused by *E. coli* (33), *Klebsiella* (25), *Enterobacter* (13), *Serratia* (9), *Proteus* (6) and *Citrobacter* (2). These infections included 21 bloodstream, 29 urinary tract, and 45 respiratory infections. We had 155 males (54.6%), and mean age was 57.9 years (SD  $\pm 16.3$ ). Based on 16S rRNA gene sequencing (232 samples), the mean/median relative abundance of *Enterobacteriales* in cases versus controls were 31.1%/22.9% and 27.5%/19.6%, with considerable overlap between the two groups. The unadjusted analysis did not show that the difference in means here was statistically significant ( $P = 0.322$ ), and this was true even for other variable constructions such as dichotomization, which argued against a threshold effect in the overall cohort. Only eight patients in our cohort met criteria for neutropenia (absolute neutrophil count of  $< 500$  cells/ $\mu$ l), and this was not associated with infection risk ( $P > 0.99$ ).



**TABLE 3** Multivariable model of infection using 16S data

Variable	OR	95% confidence interval	P value
Central line at baseline	2.27	1.20, 4.30	0.012
Baseline albumin (g/dl)	0.50	0.29, 0.85	0.010
Pressors at baseline	5.47	1.29, 23.22	0.021
Cephalosporins at baseline	3.59	1.26, 10.25	0.017
%Ent	1.01	1.00, 1.02	0.229

Turning to the subgroup analyses for the respiratory and urinary infections revealed a similar lack of statistically detectable association ( $P = 0.896$  and  $0.888$ , respectively). However, bloodstream infections did show an association when dichotomized at the 3rd quartile of 60%—having a percent *Enterobacteriales* by 16S rRNA gene sequence analysis [%Ent] of >60% confers a 10.1-fold increased odds of bloodstream infection (odds ratio [OR], 10.1 [1.17, 85.59],  $P = 0.036$ ).

**Initial and unadjusted analyses using PCR assays A and B.** To independently assess the performance of both PCR assays on the novel task of predicting *Enterobacteriales* infection from rectal swabs, we evaluated both relative abundance calculations and cycle thresholds ( $C_T$ ) for the numerator (Ent) and denominator (uniprobe) components. For assays A and B, we observed a similar lack of significant differentiation between cases and controls by relative abundance (median/mean for cases versus controls 32.2%/35% versus 30.7%/33.8% for assay A [ $P = 0.8$ ,  $n = 235$ ] and 24.5%/35% versus 24.9%/33.5% for assay B [ $P = 0.794$ ,  $n = 222$ ]). However, for assay A, there was an interaction between the numerator  $C_T$  and percent *Enterobacteriales*, and when accounting for this, the association with infection became statistically significant ( $P = 0.012$  for interaction as a whole). That is, the absolute and relative abundance of *Enterobacteriales* interacted with each other additively to increase the risk of infection. As  $C_T$  values are the number of PCR cycles required to amplify the target to a detectable level, a lower  $C_T$  indicates a higher abundance of target in the sample. For a specific example, at an *Enterobacteriales* relative abundance of 35%, for every five-cycle decrease in  $C_T$  (i.e., 32-fold increase in absolute *Enterobacteriales* abundance), there is an increased odds of infection (OR, 3.36 [1.88, 5.98];  $P = 0.012$ ). The denominator  $C_T$  for assay A associated with infection—for every 5-unit decrease in  $C_T$  (i.e., 32-fold increased total bacterial abundance), we see a 2.8-fold increased odds of invasive infection (OR, 2.8 [1.53, 5.2];  $P < 0.001$ ).

For assay B, percent *Enterobacteriales* did not associate with infection ( $P = 0.794$ ). However, the numerator  $C_T$  did show a significant association, but not as strong as for assay A (OR, 1.35 [1.09, 1.67];  $P = 0.007$ ). The denominator  $C_T$  value did associate with infection, but it was also not as strong as for assay A (OR, 1.9 [1.1–3.2];  $P = 0.012$ ). Thus, only PCR assay A was considered for multivariable modeling, presented below.

**Adjusted analyses using 16S rRNA sequence data.** Stepwise regression constructed an adjusted multivariable model that included %Ent calculated from 16S sequence data (Table 3). This failed to show an association between %Ent and infection ( $P = 0.229$ ) but identified clinical variables, notably cephalosporin and pressor use, as putative predictors of infection. We also attempted to construct an adjusted model for the bloodstream infection subgroup; however, this was much more limited as a result of having only 21 cases. The only clinical variable retained in this model was the presence of a central line at baseline. After this adjustment, for those with >60% *Enterobacteriales* dominance in the gut, there is a 7.28-fold increase in the odds of bloodstream infection of borderline significance (OR, 7.28 [0.81, 65.36];  $P = 0.076$ ).

**Adjusted analyses using PCR assay A.** For adjusted analysis of PCR assay A, the interaction with numerator  $C_T$  was carried forward. Denominator  $C_T$  values were considered but did not make it into the final model and when forced in (either alone or with various interactions), were not significant and did not improve model fit (data not shown); thus, the denominator  $C_T$  was not included in adjusted analyses. However, percent *Enterobacteriales*, being a relative abundance metric, does implicitly carry some

**TABLE 4** Multivariable model of infection using qPCR assay A

Variable	$\beta$ coefficient <sup>a</sup>	Standard error	OR [CI] <sup>b</sup>	P value
Central line at baseline	0.909	0.396	2.48 [1.14, 5.4]	0.022
Baseline albumin (g/dl)	-0.829	0.324	0.43 [0.23, 0.82]	0.011
Cephalosporins at baseline	1.89	1.26	6.64 [1.85, 23.8]	0.017
%Ent (per 10% increase)	0.377	0.262		0.061
$C_T$ (per 5-cycle decrease)	0.38	0.172		
%Ent $\times$ $C_T$ interaction	0.133	0.074		

<sup>a</sup>Taken from the logit model. These can be interpreted as increased risk (positive coefficient) or decreased risk (negative coefficient).

<sup>b</sup>Values provided only for independent covariates (i.e., no interaction present). CI, confidence interval.

denominator information into the final model. Save for use of pressors at baseline, the same clinical variables as in the 16S analysis made it into the final model (Table 4). We observed an increased risk of infection that depended on both the absolute and relative abundance of *Enterobacteriales* as assessed by PCR assay A.

Given the interaction term, the percent *Enterobacteriales* and  $C_T$  cannot be interpreted independently, so for these variables  $\beta$  coefficients and standard errors are presented in the table in lieu of odds ratios and confidence intervals, and the *P* value is presented for both terms and the interaction together. The  $\beta$  coefficients can be interpreted as increased risk (positive coefficient) or decreased risk (negative coefficient), but for the interaction terms, the coefficients must be summed and interpreted at a specific level. As an example of how to interpret the interaction results, after adjustment for clinical confounders and at the average numerator  $C_T$  of 25, every 10% increase in percent *Enterobacteriales* increases the odds of infection 2.8-fold (OR, 2.83 [1.47, 5.44]; *P* = 0.061). Conversely, at a relative abundance of 35% *Enterobacteriales*, every five-cycle decrease in  $C_T$  (i.e., 32-fold increase in absolute *Enterobacteriales* abundance), there is an increased odds of infection (Table 4; OR, 2.32 [1.21, 4.48], *P* = 0.061).

## DISCUSSION

The aims of this study were to develop a qPCR assay to measure *Enterobacteriales* abundance on rectal swabs, validate this PCR assay against 16S rRNA sequence analysis as the gold standard, and measure the association between *Enterobacteriales* abundance and subsequent infection. Towards these aims, we developed and validated two real-time PCR assays that measure *Enterobacteriales* as a proportion of total bacterial 16S rRNA-encoding DNA in a sample. These assays are somewhat complementary, as assay A had better analytical performance in terms of precision, linearity, and reportable range, but assay B was inclusive of *Bacteroides* species. Compared to the percent *Enterobacteriales* calculated from 16S sequencing as the gold standard, both assays correlated well with minimal bias in patient rectal swab samples. This was true both in specimens that were culture positive for *Enterobacteriales* (*K. pneumoniae*) and in the specimens used in the case-control cohort. Although there were challenges in terms of eliminating contaminating DNA from reagents, minimizing probe cross-reactivity, and optimizing inclusivity, these results indicate that relative abundance of bacterial taxa in microbiome samples can be measured accurately by qPCR.

Measured by 16S sequencing, the relative abundance of *Enterobacteriales* in rectal swabs had a limited association with subsequent *Enterobacteriales* infection in our cohort of intensive care and hematology/oncology patients. Overall, there was no significant difference in relative abundance between cases and controls as measured by 16S sequencing or qPCR. In adjusted analysis, relative abundance calculated from 16S sequencing was not significantly associated with infection. Many patients had percent *Enterobacteriales* values that met prior definitions of dominance (>30%) (4), suggesting that both cases and controls may have an abnormal microbiome. However, *Enterobacteriales* of >60% was associated with bloodstream infection. With only 21 cases in our data set, this is a preliminary finding as the ability to adjust for clinical confounders was limited. Notably, though, this is consistent with prior findings that intestinal domination



by *Proteobacteria* and certain *Enterobacteriales* species is associated with subsequent Gram-negative bacteremia in allogeneic hematopoietic cell transplant patients (4, 13). The potential association with bloodstream infections outside of transplant patients would be important to explore further, as bacteremia and sepsis are devastating clinical outcomes.

Although these qPCR assays were designed to measure relative abundance, our findings indicate that absolute abundance is important information that can be obtained from qPCR but not from 16S sequencing and that both absolute and relative abundance are best considered together in assessing the risk of infection. In unadjusted analysis, there was an interaction between  $C_T$  values, a measure of absolute abundance, and percent *Enterobacteriales*. Accounting for this interaction revealed a significant association between assay A qPCR results and subsequent infection. The  $C_T$  value of the denominator in assay A was also associated with infection in unadjusted analysis. The design of these assays may be critical, as assay B, which sacrificed some analytical performance for increased inclusivity, was less informative. It is also unclear to what degree this superior performance was contingent upon the infections studied and whether this would change for other infections, which should be explored in future research. Adjustment for clinical variables still supported an association between increasing gut *Enterobacteriales* and subsequent *Enterobacteriales* infection. Our final clinical model for adjustment of qPCR variables included the presence of a central line, cephalosporin antibiotics, and albumin as associated with subsequent infection. In this context, the denominator  $C_T$  value did not improve the performance of the model, but the interacting variables of relative and absolute *Enterobacteriales* abundance remained informative.

Compared to 16S sequencing that provides relative abundance, qPCR can provide at least two additional measurements of the microbiota in patient samples: the absolute abundance of particular taxa and the total microbiota density in the sample. This approach may have broader applications, as reduced gut microbiota density has been associated with inflammatory bowel disease and recurrent *Clostridioides difficile* infection and is reversed by fecal transplantation (14). If the overall gut microbiota density varies between patients, measurement of both relative and absolute abundance of specific taxa may be required to assess associations with human health and disease. qPCR assays that predict infections based on measurements of the microbiota could be translated into practice, since many clinical laboratories are proficient in the validation and implementation of these assays. In summary, these results demonstrate that using qPCR to measure intestinal dominance by specific bacterial taxa is feasible, provides additional information compared to 16S rRNA gene sequence analysis, and suggests that increased intestinal *Enterobacteriales* levels are associated with subsequent infection.

## MATERIALS AND METHODS

**Case definitions.** We collected discarded swabs sent for vancomycin-resistant enterococcus (VRE) screening during a 3-month period from patients in the intensive care units (ICUs) and select wards (hematology, oncology, and hematopoietic stem cell transplant) at Michigan Medicine. We then observed whether patients developed a bloodstream, respiratory, or urinary infection from a member of the *Enterobacteriales* within 90 days after swab collection (including infections that occurred after hospital discharge) and identified them as putative cases. The putative cases were confirmed via manual chart review by the study team to decide whether they met clinical criteria derived from the Infectious Diseases Society of America, National Health Safety Network, and/or the American Thoracic Society guidelines (15–19). Following this, cases were matched 1:2 to controls based on age ( $\pm 5$  years), gender, and number of days ( $\pm 30$ ) between swab collections (to account for any temporal trends). This study was approved by the University of Michigan Institutional Review Board.

**Samples for PCR and 16S rRNA gene sequence analysis.** Rectal swab cultures were collected using the ESwab liquid Amies collection and transport system (Copan Diagnostics, Inc., Murrieta, CA), which elutes the swab sample into 1 ml of liquid Amies medium. The Amies medium samples were retrieved from  $-80^\circ\text{C}$  storage, thawed, and extracted for DNA as described in Text S1 in the supplemental material. 16S rRNA gene sequencing and PCR assays A and B (described below) were performed for each sample from the same extracted DNA. *Klebsiella pneumoniae* and *Serratia marcescens* were cultured overnight in Luria-Bertani broth. The following species and strain were selected for a simple contrived community representative of common gut inhabitants: *Eubacterium rectale*, *Bacteroides ovatus*, *Roseburia intestinalis*, and *Lachnospiraceae* strain DW28. Strains were grown anaerobically in a Coy vinyl anaerobic chamber on

modified or enriched brain heart infusion (BHI) liquid medium until reaching an optical density of 0.5 to 1, at which point individually grown strains were combined into an aliquot consisting of 40% *Lachnospiraceae* strain DW28, 30% *E. rectale*, 15% *R. intestinalis*, and 15% *B. ovatus* (Text S1). This contrived community was combined with different ratios of *K. pneumoniae* for “spiked” community analysis. Individual, combined, and spiked communities were used for DNA extractions.

**DNA extraction.** Samples were added to wells of the bead plate included in the MagAttract PowerMicrobiome DNA/RNA kit (catalog no. 27500-4-EP; Qiagen) (formerly known as the MoBio PowerMag microbiome DNA/RNA kit) designed to be used in conjunction with the automated liquid handling Eppendorf EPmotion 5075 TMX (catalog no. 960020033; Eppendorf). Bacterial cultures were extracted from 200  $\mu\text{l}$  of culture. Rectal swab samples were extracted from 250  $\mu\text{l}$  of liquid Amies transport medium in the E-swab transport system. Subsequent steps to DNA extraction were conducted following the manufacturer’s directions using the Eppendorf EPmotion 5075 TMX. The resulting DNA was aliquoted into two replicate plates of 50  $\mu\text{l}$  each, one stored for archive and the other used to generate the 16S rRNA gene-based sequencing.

**Primer design and evaluation.** Candidate primer and probe assays were optimized using Visual OMP software (DNA Software, Ann Arbor, MI), which uses secondary structures and thermodynamics of DNA hybridization to predict the melting temperatures of primers and probes compared to those of unintended dimerization products and calculate the expected efficiency of the PCR. Primer and probe coverage and specificity were evaluated using the Ribosomal Database Project’s Probe Match (RDP database release 11, update 5) (2).

**PCR assay protocol.** To balance amplification efficiency and inclusivity of the universal probe, two assays (A and B) were designed. Real-time PCR was performed using primers (IDT) and probes (Thermo Fisher Scientific) with sequences and concentrations listed in Table 1 in combination with TaqMan Gene Expression Master Mix (Thermo Fisher Scientific), 5% dimethyl sulfoxide (DMSO) (Sigma), 6  $\mu\text{M}$  PMAxx (Biotium) (a proprietary derivative of propidium monoazide), and 10  $\mu\text{l}$  template in a final reaction volume of 50  $\mu\text{l}$ . Prior to template addition, the reaction mixture was incubated for 10 min at room temperature and then treated in a Biotium PMA-Lite LED (light-emitting diode) photolysis device for 10 min. PCR conditions were 50°C for 2 min, 95°C for 10 min, and then 40 cycles with 1 cycle consisting of 95°C for 15 s and 62°C for 1 min on a QuantStudio 3 real-time thermocycler (Thermo Fisher Scientific). As positive controls, genomic DNA from *Klebsiella pneumoniae* was used for the MGB\_ENT\_GG\_Probe, *Serratia marcescens* was used for the MGB\_ENT\_AC\_Probe and *Eubacterium rectale* was used for MGB\_Uniprobe\_p1 and Uniprobe\_B. DNA-free water (Qiagen) was used as a negative control.

**16S rRNA gene sequencing and analysis.** For details of the 16S rRNA gene sequencing and analysis, see Text S1 in the supplemental material.

**Statistical methods. (i) Initial analyses.** Initial descriptive statistics such as measures of central tendency/spread and counts were used to examine closely the primary variables of interest (percent *Enterobacteriales* by 16S rRNA gene sequence analysis [hereafter referred to as %Ent] and the two PCR assays [assay A and assay B]). These data were used to explore different variable constructions such as log transformation, dichotomization, and categorization. This was done to ensure that we were not missing important associations due to nonnormal distributions in the assay results or nonlinear associations with threshold effects. For the PCR assays, we examined the cycle threshold ( $C_T$ ) values for both the numerator and denominator, as this reflected absolute abundance.

**(ii) Unadjusted analyses.** Following initial analyses, we conducted unadjusted analysis of cases versus controls and the relative abundance of *Enterobacteriales* from 16S data and the two *Enterobacteriales* PCR assays via conditional logistic regression. This was repeated for the subgroups of bloodstream, respiratory, urinary tract infections, and extended-spectrum  $\beta$ -lactamase-producing (ESBL+) infections to assess whether the associations differed by type of infection. For the PCR assays, we also assessed the  $C_T$  values, and for the numerators, we looked for interactions with the *Enterobacteriales* relative abundance. We then conducted unadjusted analysis of the clinical variables versus cases and arrived at a set of clinical variables that were significantly associated with infection.

**(iii) Multivariable modeling.** These unadjusted analyses were then used to construct adjusted models that included the assays of interest (along with any significant interactions) as well as the potential clinical confounders that were significant by unadjusted analysis. This was again done by conditional logistic regression. Models were built using stepwise addition. Starting with a null model, the likelihood ratio test with a cutoff of  $P < 0.05$  was used to add variables with the lowest  $P$  value iteratively until no additional candidates existed. If the assay of interest was not selected during this procedure, then it was forced back into the model.

The equation to model the log odds of infection conditional upon stratum  $j$  (a matched set of cases and controls), with independent and interacting variables was specified as follows:

$$\text{logit}(\text{infection}|\text{stratum}_j) = \sum_{i=1}^n \beta_0 + \beta_r X_{ij} + \beta_p X_{pj} + \beta_q X_{qj} + \beta_{pq} X_{pj} X_{qj}$$

where  $X_i$  is each independent variable (i.e., predictor) from 1 to  $n$  variables,  $X_p$  and  $X_q$  are the interacting variables, and the modeled log of the odds of infection (i.e., logit) is conditioned upon stratum  $j$ . If no interaction is present, the  $X_p$  and  $X_q$  variables are omitted. For statistical inference regarding the interaction (i.e., the  $P$  value), an analysis of variance (ANOVA) was performed comparing the models with/without the interaction term,  $X_p \times X_q$ .

**(iv) Missing data.** There were many missing clinical data in our set, especially with infrequently ordered clinical laboratory results such as albumin. There were also rectal swab samples where insufficient DNA was extracted for 16S rRNA gene sequencing and PCR. Thus, prior to multivariable modeling,

we applied random forest multiple imputation implemented in the R package *missForest* to fill in missing data and utilize the whole data set, as this method has shown superiority to others for building biomarker-based predictive models (20, 21).

**Data availability.** All sequence files have been deposited in the Sequence Read Archive under BioProject accession numbers [PRJNA631262](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA631262) and [PRJNA556249](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA556249).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**TEXT S1**, DOCX file, 0.02 MB.

**FIG S1**, PDF file, 0.3 MB.

**FIG S2**, TIF file, 0.7 MB.

**FIG S3**, TIF file, 0.5 MB.

**FIG S4**, TIF file, 0.2 MB.

**DATA SET S1**, XLSX file, 0.1 MB.

## ACKNOWLEDGMENTS

This project was funded by the Centers for Disease Control and Prevention (contract number 200-2016-91966) to M.A.B.

The views expressed in written conference materials or publications and by speakers and moderators at Department of Health and Human Services-sponsored conferences do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

## REFERENCES

- Adeolu M, Alnajjar S, Naushad S, Gupta RS. 2016. Genome-based phylogeny and taxonomy of the ‘*Enterobacterales*’: proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganeliaceae* fam. nov., and *Budviciaceae* fam. nov. *Int J Syst Evol Microbiol* 66:5575–5599. <https://doi.org/10.1099/ijsem.0.001485>.
- Martin RM, Cao J, Brisse S, Passet V, Wu W, Zhao L, Malani PN, Rao K, Bachman MA. 2016. Molecular epidemiology of colonizing and infecting isolates of *Klebsiella pneumoniae*. *mSphere* 1:e00261-16. <https://doi.org/10.1128/mSphere.00261-16>.
- Tischendorf J, de Avila RA, Safdar N. 2016. Risk of infection following colonization with carbapenem-resistant *Enterobacteriaceae*: a systematic review. *Am J Infect Control* 44:539–543. <https://doi.org/10.1016/j.ajic.2015.12.005>.
- Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne A, Lee YJ, Dubin KA, Succi ND, Viale A, Perales MA, Jenq RR, van den Brink MR, Pamer EG. 2012. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. *Clin Infect Dis* 55:905–914. <https://doi.org/10.1093/cid/cis580>.
- Shimasaki T, Seekatz A, Bassis C, Rhee Y, Yelin RD, Fogg L, Dangana T, Cisneros EC, Weinstein RA, Okamoto K, Lolans K, Schoeny M, Lin MY, Moore NM, Young VB, Hayden MK, Centers for Disease Control and Prevention Epicenters Program. 2019. Increased relative abundance of *Klebsiella pneumoniae* carbapenemase-producing *Klebsiella pneumoniae* within the gut microbiota is associated with risk of bloodstream infection in long-term acute care hospital patients. *Clin Infect Dis* 68: 2053–2059. <https://doi.org/10.1093/cid/ciy796>.
- Gargiullo L, Del Chierico F, D’Argenio P, Putignani L. 2019. Gut microbiota modulation for multidrug-resistant organism decolonization: present and future perspectives. *Front Microbiol* 10:1704. <https://doi.org/10.3389/fmicb.2019.01704>.
- Wang T, Kraft CS, Woodworth MH, Dhare T, Eaton ME. 2018. Fecal microbiota transplant for refractory *Clostridium difficile* infection interrupts 25-year history of recurrent urinary tract infections. *Open Forum Infect Dis* 5:ofy016. <https://doi.org/10.1093/ofid/ofy016>.
- Yang S, Lin S, Kelen GD, Quinn TC, Dick JD, Gaydos CA, Rothman RE. 2002. Quantitative multiprobe PCR assay for simultaneous detection and identification to species level of bacterial pathogens. *J Clin Microbiol* 40:3449–3454. <https://doi.org/10.1128/jcm.40.9.3449-3454.2002>.
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42:D633–D642. <https://doi.org/10.1093/nar/gkt1244>.
- Humphrey B, McLeod N, Turner C, Sutton JM, Dark PM, Warhurst G. 2015. Removal of contaminant DNA by combined UV-EMA treatment allows low copy number detection of clinically relevant bacteria using pan-bacterial real-time PCR. *PLoS One* 10:e0132954. <https://doi.org/10.1371/journal.pone.0132954>.
- Hein I, Schneeweiss W, Stanek C, Wagner M. 2007. Ethidium monoazide and propidium monoazide for elimination of unspecific DNA background in quantitative universal real-time PCR. *J Microbiol Methods* 71:336–339. <https://doi.org/10.1016/j.mimet.2007.09.005>.
- Patel P, Garson JA, Tettmar KI, Ancliff S, McDonald C, Pitt T, Coelho J, Tedder RS. 2012. Development of an ethidium monoazide-enhanced internally controlled universal 16S rDNA real-time polymerase chain reaction assay for detection of bacterial contamination in platelet concentrates. *Transfusion* 52:1423–1432. <https://doi.org/10.1111/j.1537-2995.2011.03484.x>.
- Stoma I, Littmann ER, Peled JU, Giralt S, van den Brink MRM, Pamer EG, Taur Y. 2020. Compositional flux within the intestinal microbiota and risk for bloodstream infection with Gram-negative bacteria. *Clin Infect Dis* <https://doi.org/10.1093/cid/ciaa068>.
- Contijoch EJ, Britton GJ, Yang C, Mogno I, Li Z, Ng R, Llewellyn SR, Hira S, Johnson C, Rabinowitz KM, Barkan R, Dotan I, Hirten RP, Fu S-C, Luo Y, Yang N, Luong T, Labrias PR, Lira S, Peter I, Grinspan A, Clemente JC, Kosoy R, Kim-Schulze S, Qin X, Castillo A, Hurley A, Atreja A, Rogers J, Fasihuddin F, Saliq M, Nolan A, Reyes-Mercedes P, Rodriguez C, Aly S, Santa-Cruz K, Peters L, Suárez-Fariñas M, Huang R, Hao K, Zhu J, Zhang B, Losic B, Irizar H, Song W-M, Di Narzo A, Wang W, Cohen BL, DiMaio C, Greenwald D, Itzkowitz S, et al. 2019. Gut microbiota density influences host physiology and is shaped by host and microbial factors. *Elife* 8:e40553. [CrossRef] <https://doi.org/10.7554/eLife.40553>.
- Centers for Disease Control and Prevention. 2014. National Healthcare Safety Network device-associated module: ventilator-associated events. Centers for Disease Control and Prevention, Atlanta, GA.
- Hooton TM, Bradley SF, Cardenas DD, Colgan R, Geerlings SE, Rice JC, Saint S, Schaeffer AJ, Tambayh PA, Tenke P, Nicolle LE, Infectious Diseases Society of America. 2010. Diagnosis, prevention, and treatment of catheter-associated urinary tract infection in adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases Society of America. *Clin Infect Dis* 50:625–663. <https://doi.org/10.1086/650482>.

17. Centers for Disease Control and Prevention. 2016. National Healthcare Safety Network device-associated module: urinary tract infection (catheter-associated urinary tract infection [CAUTI] and non-catheter-associated urinary tract infection [UTI]) and other urinary system infection [USI] events. Centers for Disease Control and Prevention, Atlanta, GA.
18. American Thoracic Society, Infectious Diseases Society of America. 2005. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. *Am J Respir Crit Care Med* 171:388–416. <https://doi.org/10.1164/rccm.200405-6445T>.
19. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, Dowell SF, File TM, Jr, Musher DM, Niederman MS, Torres A, Whitney CG, Infectious Diseases Society of America, American Thoracic Society. 2007. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis* 44(Suppl 2):S27–S72. <https://doi.org/10.1086/511159>.
20. Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, Marrero J, Zhu J, Higgins PD. 2013. Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open* 3:e002847. <https://doi.org/10.1136/bmjopen-2013-002847>.
21. Stekhoven DJ, Buhlmann P. 2012. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28:112–118. <https://doi.org/10.1093/bioinformatics/btr597>.