**Intestinal enrichment of oral-typical bacteria in feces reflects depletion of gut commensals**

Chen Liao1,8, Thierry Rolling2,3,4,8, Ana Djukovic1, Hongbin Liu6, Lei Dai6, Bing Zhai2,3,6, Jonathan U. Peled5,7, Marcel R.M. van den Brink5,7, Tobias M. Hohl2,3,7,\*, and Joao B. Xavier1,9,\*

1Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

2Infectious Disease Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

3Immunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

4Division of Infectious Diseases, First Department of Medicine, University Medical Center, Hamburg-Eppendorf, Hamburg 20251, Germany

5Adult Bone Marrow Transplantation Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

6CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

7Weill Cornell Medical College, New York, NY 10065, USA

8These authors contributed equally to this work

9Lead contact

\*Correspondence: xavierj@mskcc.org (J.B.X.), hohlt@mskcc.org (T.M.H.)

**SUMMARY**

The association of gastrointestinal diseases with relative enrichment of oral bacteria in human feces can be explained by two alternate hypotheses: it may indicate an increased population density of ectopic oral bacteria in the gut (the *expansion* hypothesis) or a depleted gut commensal bacterial load (the *marker* hypothesis). Through mouse experiments and human microbiome data analysis, we found that the quantitative abundance of oral-typical bacteria in the intestine remains stable as opposed to the marked reduction in the gut bacterial load following antibiotic treatment, thus supporting the *marker* hypothesis. Our mathematical model indicates that the stable oral bacterial population is maintained by a balance between direct antibiotic inhibition and indirect ecological release due to depleted gut commensals. Recognizing that the relative enrichment of oral bacteria does not reflect an ectopic bloom, but results from a depleted gut bacterial population, critically informs the interpretation of microbiome compositional data and interventions to restore healthy gut microbiomes.

**INTRODUCTION**

Healthy people harbor distinct microbiome communities at different body sites (Costello et al., 2009; The Human Microbiome Project Consortium, 2012). The translocation of oral-associated bacteria to the lower gastrointestinal tract and their intestinal enrichment have been linked to a number of digestive system disorders such as inflammatory bowel disease (IBD) (Read et al., 2021) and colorectal cancer (CRC) (Komiya et al., 2019). Compared to the hematogenous route (oral-blood axis) (Abed et al., 2020), the enteral route (oral-gut axis) is a natural and more studied pathway of oral bacterial dissemination (Jin et al., 2022; Kitamoto et al., 2020). For an average person, ~1011 oral-resident bacterial cells are transported to stomach via saliva, food and liquid daily (Segata et al., 2012) but they rarely colonize the healthy gut. Gastric acids, alkaline biles and antimicrobial peptides kill many of those, and survivors face competition from gut-resident bacteria due to colonization resistance (Jin *et al.*, 2022). This is supported by DNA-based surveys that detect very low abundances of oral-resident bacterial DNAs in the feces of health individuals (typically <2%) (Rashidi et al., 2021; Schmidt et al., 2019).

Broad spectrum antibiotics such as ampicillin can weaken the oral-gut barrier by depleting gut anaerobes and promote colonization of translocated oral bacteria (Atarashi et al., 2017). Other than antibiotics, diets, aging, and gut inflammation may also disrupt gut colonization resistance (Kitamoto *et al.*, 2020). All these factors jointly contribute to the clinical associations between ectopic colonization of oral bacteria and a variety of digestive diseases (e.g., Crohn’s disease (Gevers et al., 2014), ulcerative colitis (Schirmer et al., 2018), CRC (Kostic et al., 2013), liver cirrhosis (Qin et al., 2014)). However, nearly all those association studies were based on relative microbiome profiling and thus inconclusive regarding whether the relative enrichment of oral-typical bacteria reflects their active expansion and bloom in the gut. The technical limitation has hampered our ability to associate quantitative microbiome features with disease and led to artificial interpretations from relative microbiome analyses (Vandeputte et al., 2017).

By definition, the relative enrichment of oral bacteria following perturbations can be explained by two alternate hypotheses (Fig. 1). The *expansion* hypothesis proposes that the active expansion of translocated oral bacteria in the gut outweighs the depletion of gut-resident bacteria in determining the relative enrichment. By contrast, the marker hypothesis states the opposite, i.e., the relative enrichment is an artifact caused by loss of gut commensals. The two hypotheses have distinct implications for human health and pathologies: While the expansion hypothesis indicates that the ingested oral bacteria can directly drive or exacerbate digestive diseases, the marker hypothesis indicates indirect effects of microbiome on disease onset and progression due to the loss of gut bacteria. In the following, we distinguished the two hypotheses by quantifying the load of oral-typical bacteria in the gut microbiome of mice and humans.

Diagram

Description automatically generated

**Fig. 1| Expansion versus marker hypotheses to explain the relative enrichment of oral bacteria in human microbiome.** The density of oral-typical bacteria colonizing the human intestine is synergistically determined by a basal transmission influx from the oral cavity, biomass loss due to natural death and fecal excretion, and ecological interactions between bacteria of oral and gut origins. Perturbations such as antibiotics may substantially increase the relative abundance of oral-typical bacteria in the human gut (fecal) microbiome by two alternate mechanisms. The expansion hypothesis proposes that the relative enrichment measures its absolute abundance (i.e., active expansion of the oral populations), while the marker hypothesis views it as an artifact caused by depletion of gut bacterial commensals.

**RESULTS**

**Paired oral and fecal samples from antibiotic-treated mice support marker hypothesis**

To determine whether relative enrichment of oral-derived bacteria reflects their active population expansion, we treated three mice with antibiotic cocktail (ampicillin-vancomycin-neomycin) at day 0 for a week to disrupt colonization resistance by gut commensals (Fig. 2A). We profiled the microbiome compositions at the ASV (amplicon sequence variant) level of oral cavity (oral\_d0) and fecal (fecal\_d0) samples on day 0 prior to antibiotics as well as post-antibiotic fecal samples (fecal\_d8) on day 8 (Fig. 2B). We found that the fecal microbiomes on day 8 shared more identical ASVs (Hamming distance) with the oral samples than the fecal samples on day 0 and were also compositionally similar (Bray-Curtis distance) in 2 out of the 3 mice (Fig. 2C). Notably, 10%-20% ASVs (the majority from 6 members; Fig. S1) present in the post-antibiotic feces were only detected in the pre-antibiotic oral microbiome (Fig. 2D), providing strong evidence of oral-gut transmission of bacteria that contain those ASVs. By identifying ASVs typically colonizing the oral cavity of the three mice (see STAR Methods), we found that the total fractions of the oral-typical ASVs increased substantially from nearly 0% to 30% (Fig. 2E), while their quantitative loads reduced by half a log unit (Fig. 2F). The contrast between relative and absolute abundances of oral bacterial populations supports marker hypothesis.

**Additional evidence of marker hypothesis from the literature**

Due to the lack of large-scale microbiome dataset with paired oral and gut samples in mice (Greenberg et al., 2022), we sought to detect translocation and classify oral-typical ASVs from gut microbiome profiles alone. To this end, we compiled a reference collection of full-length 16S rRNA genes that represent bacteria typically inhabiting mouse oral cavity (Fig. S2A, see STAR Methods). The resulted collection has 149 sequences (Table S1) from diverse genera with most sequences from *Staphylococcus* and *Streptococcus* (Fig. S1B). Given a new ASV, we can classify whether it is oral-typical or -atypical by exactly matching the ASV to all sequences in the reference collection. We applied this approach to a public dataset of paired oral (swab) and gut (distal intestine) samples from 11 healthy pregnant mice (Theis et al., 2020). On average, 4.5% ASVs in the gut microbiome were found to be oral-typical. In 10 out of the 11 mice, >90% of the total fractions of the oral-typical ASVs in the gut were contributed by those present in the oral cavity of the same mice (Fig. S1C).

Using the approach, we first reanalyzed a public study (Reese et al., 2018) that treated mice with an antibiotic cocktail (Ampicillin-Vancomycin-Neomycin-Metronidazole) for 5 days (Fig. 2G). The inferred oral populations, mainly consisting of three ASVs from *Lactobacillus*, *Proteus* and *Escherichia-Shigella*, had substantially higher relative abundances after two days’ exposure (Fig. 2H). However, we did not observe active expansions of the translocated oral populations compared to the control group by tracking the longitudinal changes of the mean oral bacterial loads (Fig. 2I, left). By contrast, the antibiotic cocktail reduced the averaged gut-resident bacterial density by more than 2 orders of magnitudes (Fig. 2I, right). The differential responses of the oral vs. gut populations clearly support the *marker* hypothesis: depletion of gut commensals drove the relative enrichment of oral-typical bacteria in feces. As a result, the oral bacterial fractions are negatively correlated (Pearson’s r = -0.52, P = 3.5e-17) with the total bacterial loads on the log scale (Fig. 2J). The negative association is mostly linear, but several samples with low bacterial loads have notable deviations from the linear fit (Fig. S3). We speculated that the dominant ASVs in these outliers were orally derived but misclassified. Should this be true, the abrupt falls of the oral bacterial fractions (Fig. 2H) and loads (Fig. 2I, left) after day 2 can be explained by the underestimated bacterial translocation.

The *marker* hypothesis was supported by another public dataset from bone-marrow-transplanted mice that received different mono-antibiotic prophylaxis (Staffas et al., 2018). Ampicillin but not Streptomycin, Aztreonam, or Vancomycin reduced gut bacterial loads substantially (Fig. S4A). The oral bacterial loads were stable throughout the experiments for all four antibiotics including ampicillin (Fig. S4B). Similarly, oral bacterial fractions and total bacterial loads exhibited a log-log linear relationship (Pearson’s r = -0.75, P = 5.5e-11) (Fig. S4C).

Graphical user interface

Description automatically generated

**Fig. 2| Oral-gut bacterial translocation in antibiotic-treated mice.** (**A-D**) Reanalysis of a public study (Reese *et al.*, 2018). (**A**) Experimental design. Mice were treated with either water (the control group; n=11) or an antibiotic cocktail of ampicillin-vancomycin-neomycin-metronidazole (AVNM) (the antibiotic group; n=10) for five days. (**B**) Compositional dynamics of oral bacterial ASVs during and after antibiotic cocktail treatment. (**C**) Total loads of ASVs classified as of oral (left) and gut (right) origin averaged across mice in the control and antibiotic groups. Lines and dots: mean; shading: 95% confidence interval (CI). (**D**) Linear relationship between total bacterial loads and oral bacterial fractions in the log-log space. Line: linear regression; shading: 95% CI. (**E-I**) New mouse experiments in this study.

**Oral-typical bacterial DNAs in human feces**

To identify 16S rRNA genes for oral-typical bacteria colonizing humans, we leveraged the Human Microbiome Project (HMP) that sequenced paired oral-gut microbiome samples from 237 healthy volunteers (The Human Microbiome Project Consortium, 2012). Distinct bacterial communities were found inhabiting the two body sites (Fig. 3A), which provides the pragmatic basis for distinguishing oral-typical from oral-atypical ASVs. We adopted a similar criteria used before (Thomas et al., 2019) for ASV filtering: the mean relative abundance and prevalence of oral-typical ASVs must be greater than 0.01% and 5% respectively among all oral samples, and no greater than the same cutoffs among all fecal samples (Fig. 2B,C). The filtering step left 178 ASVs (Table S2) that belong to a variety of genera including *Prevotella*, *Streptococcus*, *Leptotrichia*, and *Haemophilus* (Fig. 2D). Using those ASVs as references, we found that 219 out of 280 HMP stool samples have undetectable oral-typical ASVs. Among the remaining 61 samples, the identified oral-typical ASVs in 53 samples were all present in at least one paired oral cavity sample of the same persons (Fig. 2E). The average total fractions of oral-typical bacteria in the HMP stool samples is as low as 0.05%, which are consistent with previous literature findings (Rashidi *et al.*, 2021; Schmidt *et al.*, 2019).

We next show that the 178 oral-typical ASVs identified from healthy people in the HMP study can be generalized to patients who might have different oral sequence signatures. To this end, we reanalyzed paired oral-gut microbiome samples from patients with inflammatory bowel disease (IBD) and their own healthy controls. The mean total fractions of oral-typical ASVs in the gut of the non-IBD healthy controls remains low at 1.1%. As expected (Read *et al.*, 2021), the total oral bacterial fractions in the patients with Crohn’s disease (CD) and ulcerative colitis (UC) increased ~4 folds on average up to 4.2% and 4.3% respectively. Among 81 out of 101 samples, the proportions of oral-typical ASVs in the feces that were also present in the paired oral samples were no less than 85%. Using the same IBD cohort, we demonstrated that the estimated total fractions of oral-typical bacteria in the feces are robust against the cutoff thresholds used to filter HMP (Fig. S5). These computational tests suggest that the oral-typical ASVs we compiled from the HMP study can robustly estimate the total fractions of oral-typical bacteria in other healthy or patient cohorts.

Diagram

Description automatically generated

**Figure 3. Identification of bacterial DNAs typically found in the human oral cavity but not the intestine.** (**A**) An overview of the 4,354 microbiome samples from The Human Microbiome Project (HMP). T-distributed stochastic neighbor embedding (tSNE) (Van der Maaten and Hinton, 2008) collapses all samples (dots) into distinct clusters based on their compositions. (**B, C**) The average relative abundance (B) and prevalence (C) of ASVs (dots) among all HMP oral cavity samples (x axis) and the stool samples (y axis). 178 ASVs were identified as oral-typical and highlighted in orange (see the main text or STAR methods for the filtering criteria). (**D**) Genus-level distribution of the 178 oral-typical ASVs. (**E, F**) The total fractions of oral-typical ASVs in the human fecal samples (circles) and their proportions present in the paired oral samples. (E) Healthy individuals in HMP. (F) Patients with inflammatory bowel disease and their own healthy controls (n=101) (F). HC: healthy control; CD: Crohn’s disease; UC: ulcerative colitis. Data sources: panel A-E (The Human Microbiome Project Consortium, 2012); panel F (Imai et al., 2021).

***Streptococcus*-domination states of allo-HCT recipients indicate oral-gut translocation**

To validate the marker hypothesis, we leveraged a pre-compiled large-scale microbiome dataset with quantitative profiles (16S amplicon sequencing and quantitative PCR) withfor hospitalized patients receiving allogeneic hematopoietic cell transplantation (allo-HCT) in Memorial Sloan Kettering Cancer Center (MSKCC). We chose allo-HCT recipients because the intensive antibiotic exposure during HCT disrupted the gut microbiome colonization resistance.

disrupted the gut microbiome compositions, where *Enterococcus* and *Streptococcus* are the two most abundant genera that frequently dominate (>30% relative abundance of a single ASV; (Taur et al., 2012)) the gut microbiome (Fig. 4A, top row) (Peled et al., 2020). Among the 10,433 samples from 1,276 patients, 1,389 fecal microbiome samples were comprised of >30% oral-typical ASVs that mostly belong to the genera *Streptococcus*, *Actinomyces*, and *Abiotrophia* (Fig. 4A, middle row, Fig. S7). We confirmed that the total fractions of oral-typical ASVs detected in the feces are not correlated with sequencing depths (Fig. S7). The total bacterial loads of a subset of 3,126 samples measured by quantitative PCR showed that the fecal microbiome samples highly enriched with oral-typical ASVs have lower bacterial biomass (Fig. 4A, bottom row).

One limitation of this cohort is the lack of paired oral microbiome samples. To find supporting evidence of bacterial translocation, we used co-occurrence analysis to identify oral-typical ASVs that were present in the feces more than expected by chance. The co-occurring groups of ASVs, if found, indicate collective bacterial transmission from the oral cavity to the gut. Indeed, we found 71 such oral-ASV groups of size from 2 to 5 and 55 groups contain *Streptococcus* ASVs (Table S5). To identify the species of the *Streptococcus* ASVs, we extracted shotgun metagenomes from 9 samples which have any *Streptococcus* ASV domination from the same MSKCC allo-HCT cohort. We were able to recover XX high-quality metagenome-assembled genomes (MAGs) and XX were annotated as *S. thermophilus*, in addition to . To assess the viability of the *Streptococcus* MAGs, we computed the ratio of metagenomic reads mapped to oriC to terminus (i.e., peak-to-trough ratio) to estimating the proportion of cells with active DNA replication forks reads ([Gao and Li, 2018](https://elifesciences.org/articles/45931#bib3); [Korem et al., 2015](https://elifesciences.org/articles/45931" \l "bib4)). We found that the peak-to-trough ratio is 1.5 (Table S6), suggesting a possible mixture of 50% 2-forks and 50% single-fork cells. This suggests that the oral-derived *Streptococcus spp.* may have colonized the gut and were actively multiplying at the time of sampling.

**Validation of maker hypothesis in patients receiving allo-HCT**

The allo-HCT recipients become immunocompromised after conditioning regimen and thus require antibiotics to prophylactically minimize the risk of developing infections before immune system reconstitution and/or to treat infections when they develop (Shono and van den Brink, 2018). The allo-HCT procedure in MSKCC started antibiotic prophylaxis within a week prior to transplantation. The timing of antibacterial drug administration (Fig. 4B, top panel) corresponded to the declined gut bacterial relative abundance (middle panel) and density (bottom panel). The total loads of oral bacteria fluctuated but remained stable (bottom panel), explaining their increased relative abundances (middle panel). As a result, the proportion of bacterial ASVs from oral habitats had a negative correlation with the total bacterial load (Pearson’s r = -0.27, P = 2.2e-54) (Fig. 4C). We further examined the bloodstream infections caused by the most common oral genera (Streptococcus, Actinomyces, Villanella, Hemophilus, Abotiophia) in these patients. In contrast to *Enterococcus* and *Candida spp.* whose intestinal dominations are associated with their bloodstream infections, we only found 5 infection cases (all *Streptococcus spp.*) in 5 patients during transplantation, and none were preceded by intestinal *Streptococcus* domination (Table S7). The few incidences of bloodstream infections further suggest against active expansion of oral bacteria in the intestine.

The complex antibiotic use patterns in allo-HCT recipients makes it a challenge to uncover the specific antibiotics associated ectopic gut colonization by oral bacteria. We conducted a time-varying survival analysis—an approach for analyzing time-to-event data—to assess which antibiotic exposures in the past improve the prediction of future occurrence of the intestinal domination by oral-typical ASVs above null expectation. We identified that piperacillin/tazobactam (P=5.7e-7) and orally administered vancomycin (P=0.049) significantly increase the risk of oral bacterial domination while quinolones (P=6.1e-4) has a negative impact. Intravenous vancomycin showed no association, likely because it fails to reach the GI tract. Piperacillin/tazobactam is a combination of beta-lactam and beta-lactamase inhibitor that can reach the gastrointestinal tract and have a broad spectrum against many gut anaerobes (Morjaria et al., 2019). This association was not only apparent in adults receiving allo-HCT but also seen in an independent cohort of pediatric patients (Bekker et al., 2019). In that cohort, children who have received total gut decontamination (oral piperacillin/tazobactam and amphotericin B) have much higher relative abundance of oral-typical bacteria in feces than those who have received selective gut decontamination (oral polymyxin/neomycin and amphotericin B) (Fig. S8). Since the high oral bacterial fraction in feces indicates gut bacterial depletion, piperacillin/tazobactam is very disruptive to gut microbiome and should be prescribed with caution.

Graphical user interface

Description automatically generated

**Figure 4. Relative enrichment of oral-typical bacteria in the feces of allo-HCT recipients.** (**A**)The total fractions of oral-typical bacteria and total bacterial loads of fecal microbiome samples with diverse microbiome compositions. (**B**) Dynamics of relative and absolute abundances of the oral-typical and gut-resident populations s. (**C**) Negative association between oral bacterial fraction and total bacterial load. Red line: best linear fit; shading and bars: 95% confidence interval (CI). (**D**) Hazard ratio (vertical line) and 95% CIs (bar width) of antibiotics that induce relative enrichment of oral-typical bacteria >30%. \*\*\*\*: P<0.0001; \*: P<0.05; \*\*\*: P<0.001. xx test.

**Ecological mechanisms of the marker hypothesis**

To understand why the oral bacterial populations in the gut stabilize following antibiotic perturbation, we developed a mathematical model that simulates oral and gut bacterial growth in the gut lumen. Our model considers two competing forces that regulate the loads of oral bacteria in the intestine (Fig. 5A): antibiotics inhibit their growth but relieve them from the competitive inhibitions from the gut commensals (i.e., ecological release). The kinetic equations are shown below (see Supplementary Text for details)

Eq. (1)

Eq. (2)

Here and are the oral and gut bacterial loads respectively, is the oral-gut transmission rate due to saliva, food and liquid intake, is the rate constant of bacterial loss, and are the host carrying capacities, and and are the maximum growth rates. The physiological values of these parameters were either taken from the literature ( and ) or estimated by theoretical modeling ( and ) or fitting human data ( and ); see Table S1 for their values and Supplementary Text for the details of parameter estimation. Based on the parameter values, the gut bacteria strongly inhibit the oral populations but are only weakly affected

oral bacterial fraction ()

Although we have not observed active expansions of oral bacterial populations in the mouse and human intestine, our model predicts that the oral bacterial loads increase when antibiotics have stronger inhibitory effects on the gut bacteria compared to the oral bacteria (Fig. SX). To explore what relative enrichment of oral bacteria represent at broader parameter space, we fixed and varied from 0 to 1. During the full parameter range, we have reproduced relative enrichment of oral bacteria but their totrla bacterial loads can decrease, remain stable, or increase up to 2 orders of magnitudes. However, the change of oral bacterial loads, despite sometimes increase, decrease several orders of magnitude are compara ble smaller magitnudet than that of the oral bacterial lodad reduction. Our previous asusmpotion that marker hypothesis is no active expansion of gut bacteria. In theory, active but it can still To distinguish between marker and expansion hypotheses under complex scenarios, we compared the deviation of oral bacterial fraction at any time point from values computed using pre-treatment oral bacterial loads and pre-treatment gut bacterial loads. If the deviation from the fixed oral bacteria is smaller than that from the fixed gut bacteria, it is closer to a marker hypothesis as the change in gut bcacterial explains more than that of the oral bacteria. Narly lal the simulations support makrer hypothesis in the extended definition. The underlying reason is that the total oral bacteria load is limited by its carrying capacity (Supplementary Text): we estimated it is ~20 fold less than the gut bacterial carrying capacity . This difference in carrying capacity suggests that the gut-resident population adapt better to the nutrient resources available in the gut environment.

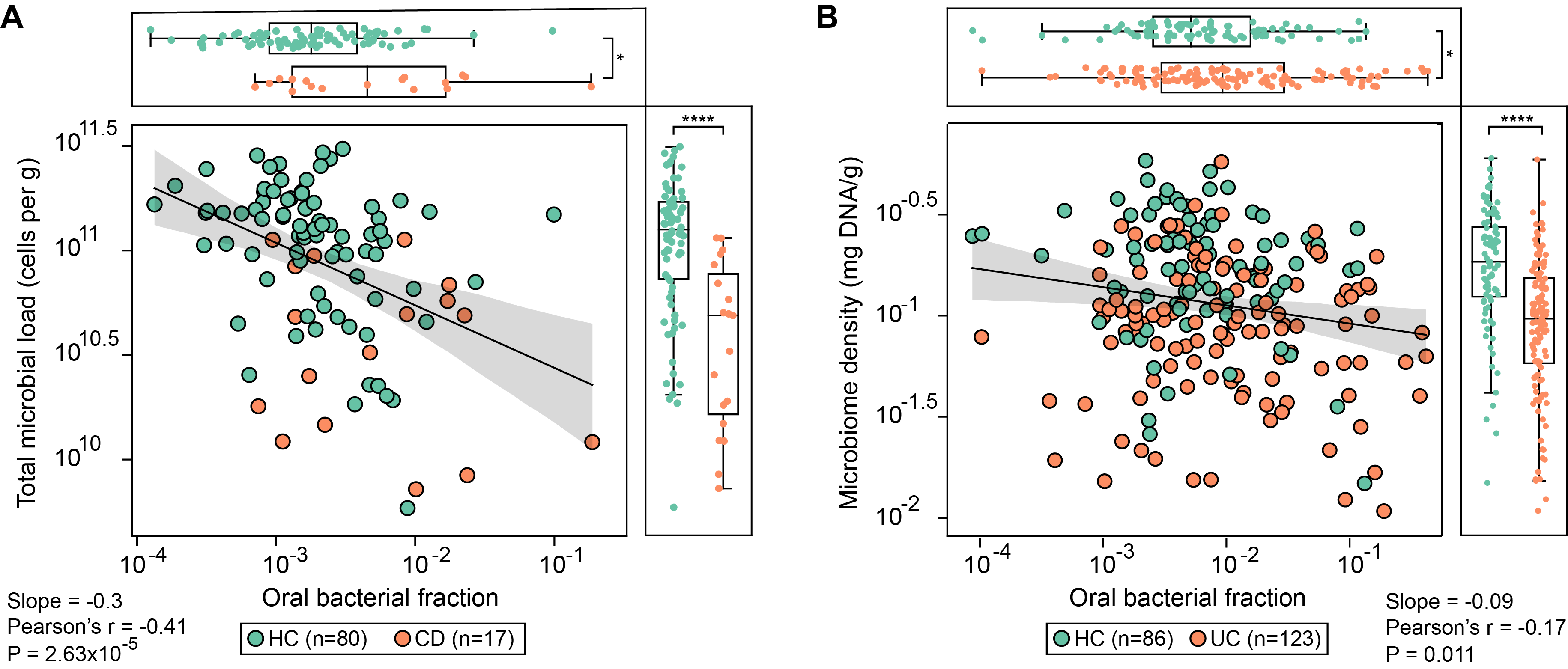
Logo

Description automatically generated

**Fig. 2| Mathematical modeling supports the marker hypothesis. a.** The oral bacterial density is determined by two two competing forces.,density of oral bacteria in the gut lumen is synergistically determined by a constant biomass inflow from the oral cavity, biomass loss due to natural death and fecal excretion, and ecological interactions between bacteria of oral and gut origins. Perturbations such as antibiotics may increase the relative abundance of oral bacteria by two mechanisms. The driver hypothesis assumes that the relative enrichment is mainly driven by the active expansion of the oral bacteria due to compromised colonization resistance by the gut commensals. By contrast, the marker hypothesis assumes that the oral bacterial do not expand in population size and their relative enrichment is largely resulted from the reduction in the gut bacterial density.

**Generalization of the marker hypothesis in IBD patients**

Relative enrichment of oral-associated bacterial species in the gut has long recognized as a characteristic of the gut microbiome in patients with IBD, including both Crohn’s disease (CD) and ulcerative colitis (UC). Using our method to quantify the total fractions of bacteria of the oral origin in feces, we confirmed that this association remains significant for all oral-typical bacteria as a whole (Fig. S10). Other than the enrichment of oral bacteria, it was revealed that both patients with CD and UC have low microbiome density compared to the healthy controls (Contijoch et al., 2019; Vandeputte *et al.*, 2017). The marker hypothesis, if it can be generalized to IBD patients, would predict a negative association between the two signatures. By reanalyzing two public datasets (Contijoch *et al.*, 2019; Vandeputte *et al.*, 2017), we found that the oral bacterial fractions and total bacterial loads in the feces of patients with CD (Fig. 6A) and UC (Fig. 6B) were indeed negatively correlated in the log-log space. Since the two signatures are significantly different between the patients and their healthy controls in both studies, the marker hypothesis integrates the two signatures that each distinguishes IBD patients from the healthy individuals into a single mechanism. For the UC cohort, the association became insignificant (P=0.32) after excluding patients with prior antibiotics or an unclear antibiotic history, due to the loss of significant differences of oral bacterial fractions between the control and UC patients (P=0.15). However, relative enrichment of oral-typical bacteria was found in IBD patients without previous antibiotic exposure history. The heterogeneous results indicate a possibility that the relative enrichment of oral bacteria in IBD patients may be synergistically caused by antibiotics in IBD management and gut inflammation.



**Figure 4. Negative association between oral bacterial fractions and total bacterial loads in the feces of patients with Crohn’s disease (A) and Ulcerative Colitis (B).** Each circle represents a fecal sample. Lines: best linear fits; shading: 95% confidence intervals. Data sources: panel A (Vandeputte *et al.*, 2017); panel B (Contijoch *et al.*, 2019). Abbreviations: HC (healthy control); CD (Crohn’s disease); UC (ulcerative colitis). \*\*\*\*P<0.0001; \*P<0.05; Welch's t-test.

**DISCUSSION**

Except for a few quantitative studies (Contijoch *et al.*, 2019; Jian et al., 2020; Rao et al., 2021; Schluter et al., 2020; Vandeputte *et al.*, 2017; Vieira-Silva et al., 2019), most human gut microbiome research to date have focused on amplicon-based profiling of microbiome composition. Due to the limitation, gut microbiome dysbiosis has been mostly characterized by loss of biodiversity from the compositional data. Gut microbiome density, another fundamental parameter that impacts host-microbiome interactions (Contijoch *et al.*, 2019; Tang et al., 2019; Zarrinpar et al., 2018), has been much less studied. For example, the total microbial load in the gut regulates the proportion of mucosal RORγt+ Treg cells in mice (Britton et al., 2020). Notably, changes in gut microbiome density do not strictly reflect changes in the gut microbiome biodiversity (Contijoch *et al.*, 2019), suggesting that the two metrics describe different aspects of dysbiosis.

Both cell-based (e.g., flow cytometry) and molecular-based (e.g., qPCR, DNA mass) methods have been developed to measure absolute microbial abundances (Galazzo et al., 2020). In the fecal samples of healthy volunteers, the correlations of microbial loads between flow cytometry and qPCR are moderate (Pearson’s r close to -0.5) (Galazzo *et al.*, 2020; Vandeputte *et al.*, 2017). Although these methods all have their own limitations, we argued that qPCR is conceptually superior for our study due to its compatibility with next generation sequencing (Jian et al., 2021). To compute absolute abundance of any bacterial ASV, we should theoretically multiply its relative abundance by the total bacterial loads in the prepared samples before sequencing, rather than loads in the original samples (see Supplementary Text for mathematical explanataion). The technical biases introduced in the extraction, purification, and amplification steps of 16S amplicon sequencing are shared with qPCR, but not flow cytometry. Additionally, 16S qPCR measures bacterial loads (not microbial loads) and is compatible with 16S amplicon sequencing. Therefore, we used qPCR-based quantification to validate the marker hypothesis in the mouse and human experiments (Figs. 2-4).

Quantitative microbiome profiling bypasses compositionality effects and can thus calibrate the associations learned from relative microbiome profiling. One notable example is the negative association between *Bacteroides* and *Prevotella*, which is an artifact of relative microbiome analysis and disappears by taking absolute cell counts into accounts (Vandeputte *et al.*, 2017). Resonating with this finding, our study revealed another artifact of relative abundance-based association: the relative enrichment of oral-derived bacteria in the gut microbiome does not reflect the active expansion of the translocated oral populations, but simply indicates a depleted gut microbiome state. Therefore, the various bacterial species (e.g., *Streptococcus spp.*, *Veillonella parvula*, *Fusobacterium nucleatum*) found enriched in the intestine of IBD patients (Read *et al.*, 2021) may not be biologically different; they could be all biomarkers of declined gut bacterial density (Duvallet et al., 2017). Despite low biomass of orally derived bacterial populations in the gut, they may have functional impacts on human health. For example, the orally translocated *Klebsiella spp.* causes Th1 cell expansion in the mouse gut (Atarashi *et al.*, 2017).

**ACKNOWLEDGMENTS**

C.L. is supported by National Institutes of Health (NIH) grant nos. U01 AI124275 (J.B.X.), R01 AI137269 (J.B.X.) and U54 CA209975 (J.B.X.). T.R. is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant no. RO-5328/1-2 (T.R.), NIH grant nos. R01 AI093808 (T.M.H.), R21 AI105617 (T.M.H.) and R21 AI156157 (T.M.H.). We thank Dr. Zhenjiang Xu for discussion so foral bacterial fraction estimation and thank Dr. Anna Staffas for providien the medadata of the bone-marrow-transplanted mouse.

**AUTHOR CONTRIBUTIONS**

Conceptualization, C.L., T.R. and J.B.X; Public microbiome dataset analysis, C.L; Methodology, C.L.; Investigation, C.L. and T.R; Writing – Original Draft, C.L., T.R.; Writing – Review and Editing, T.M.H., J.B.X., J.U.P., B.Z., and M.R.M.v.d.B; Supervision, J.B.X. and T.M.H.

**DECLARATION OF INTERESTS**

J.U.P. reports research funding, intellectual property fees and travel reimbursement from Seres Therapeutics and consulting fees from DaVolterra, CSL Behring and from Maat Pharma. He has filed intellectual property applications related to the microbiome (reference nos. 62/843,849, 62/977,908 and 15/756,845). M.R.M.v.d.B. has received research support from Seres Therapeutics; he has consulted, received honorarium from or participated in advisory boards for Seres Therapeutics, WindMIL Therapeutics, Rheos, Frazier Healthcare Partners, Nektar Therapeutics, Notch Therapeutics, Forty Seven Inc., Priothera, Ceramedix, Lygenesis, Pluto Immunotherapeutics, Magenta Therapeutics, Merck & Co., Inc. and DKMS Medical Council (Board); and he has IP Licensing with Seres Therapeutics, Juno Therapeutics and stock options from Seres and Notch Therapeutics. T.M.H. has participated in a scientific advisory board for Boehringer-Ingelheim Inc.

**STAR METHODS**

**KEY RESOURCES TABLE**

|  |  |  |
| --- | --- | --- |
| **REAGENT or RESOURCE** | **RESOURCE** | **IDENTIFIER** |
| **Deposited Data** | | |
| Human Microbiome Project | Human Microbiome Project Consortium, 2012 | SRA study PRJNA48489 |
| MSK cohort | Liao et al., 2021; Yan et al., 2021 | SRA studies PRJNA394877, PRJNA607574, PRJNA606262, PRJNA548153, PRJNA545312 |
| PRISM, LifeLines DEEP and NLIDB cohorts | Franzosa et al., 2019 | SRA study [PRJNA400072](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA400072/) |
| UMH cohort | Schubert et al., 2014 | mothur.org/CDI\_MicrobiomeModeling/ |
| RISK cohort | Gevers et al., 2014 | SRA study SRP040765 |
| Belgium PSC/IBD cohort | Vieira-Silva et al., 2019 | European Genome-phenome Archive  EGAS00001003600 |
| PROTECT cohort | Schirmer et al., 2018 | SRA study PRJNA436359 |
| Mount Sinai cohort | Contijoch et al., 2019 | SRA study PRJNA413199 |
| LUMC cohort | Bekker et al., 2019 | SRA study PRJEB28845 |
| UPenn cohort | Lewis et al., 2015 | SRA study SRP057027 |
| **Databases** | | |
| SILVA v138 | Quast et al., 2013 | www.arb-silva.de |
| MicrobiomeHD | Duvallet et al., 2017 | [doi.org/10.5281/zenodo.840333](https://doi.org/10.5281/zenodo.840333) |
| **Softwares** | | |
| Python v3.7.9 | Python | python.org |
| Scikit learn v0.24.0 | Pedregosa et al., 2011 | scikit-learn.org |
| Non-negative matrix factorization | Lee and Seung, 1999 | scikit-learn.org |
| t-distributed stochastic neighbor embedding | van der Maaten and Hinton, 2008 | scikit-learn.org |
| Random Forest classifier | Ho, 1995 | scikit-learn.org |
| Lifelines v0.25.6 | Davidson-Pilon, 2019 | github.com/CamDavidsonPilon/lifelines |
| SourceTracker2 | Knights et al., 2011 | github.com/biota/sourcetracker2 |
| iRep | Brown et al., 2016 | github.com/christophertbrown/iRep |
| Cutadapt v3.4 | Martin, 2011 | github.com/marcelm/cutadapt |
| Naive Bayesian classifier | Wang et al., 2007 | rdp.cme.msu.edu |
| DADA2 | Callahan et al., 2016 | benjjneb.github.io/data2 |
| Bhatt lab workflow | Siranosian et al., 2021 | github.com/bhattlab/bhattlab\_workflows |
| MEGAHIT | Li et al., 2015 | github.com/voutcn/megahit |
| MetaBAT 2 | Kang et al., 2019 | bitbucket.org/berkeleylab/metabat/src/master/ |
| CONCOCT | Alneberg et al., 2014 | github.com/BinPro/CONCOCT |
| DAS Tool | Sieber et al., 2018 | github.com/cmks/DAS\_Tool |
| PATRIC | Davis et al., 2020 | www.patricbrc.org |
| Kraken2 | Wood et al., 2019 | github.com/DerrickWood/kraken2 |

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Joao B. Xavier (xavierj@mskcc.org)

**Materials Availability**

This study did not generate new unique reagents.

**Data and code availability**

* All raw sequences of the microbiome datasets analyzed in this study are public available with accession numbers listed in the key resource table.
* All processed data supporting the findings of this study are available within the article and its supplementary materials.
* Customized Python scripts for all analyses included in this study are available on Github (https://github.com/liaochen1988/Source\_codes\_for\_oral\_microbiome\_contamination).

METHOD DETAILS

**Reference 16S rRNA gene sequences of oral-typical bacteria**

To detect oral-derived bacteria in the human gut, we reused the HMP data and defined oral-typical 16S rRNA gene ASVs based on the two criteria: (1) the mean relative abundance of each oral-typical ASV across all oral cavity samples is greater than 1e-4 and its relative abundance averaged across all gut samples is no bigger than the same threshold; (2) the prevalence of each oral-typical ASV across all oral cavity samples is greater than 0.05 and its prevalence averaged across all gut samples is smaller than or equal to the same cutoff value. Prevalence of an ASV is computed as the proportion of samples that contain the ASV at a relative abundance above 1e-3. Similar thresholds of relative abundance and prevalence were used in a previous study to identify oral-typical species from metagenomic profiles (Thomas *et al.*, 2019). The threshold of occurrence (or presence) of an ASV was also adopted from the literature (Machado et al., 2021). Importantly, we have shown that the estimation of total fractions of oral-typical bacteria is robust against variations of these cutoff thresholds (Fig. 5).

Due to the lack of large-scale paired oral-gut microbiome samples in mouse, we used the full-length 16S rRNA genes from the Mouse Oral Microbiome Database (MOMD) as the starting point. Similar approaches were implemented to identify oral bacteria in the human gut by profiling sequences against oral bacteria in the Human Oral Microbiome Database (HOMD) (Coker et al., 2018; Hu et al., 2022). However, the original MOMD contains sequences from bacteria that colonize both mouse oral cavity and gut. We thus filtered MOMD by excluding sequences that have exact matches in the Mouse Gut Microbial Biobank (Liu et al., 2020) and the Mouse Intestinal Bacterial Collection (Lagkouvardos et al., 2016). Both sources contain cultured gut microbes that also have full length 16S rRNA genes. To further expand the database, we profiled the sequences in a paired mouse oral-gut microbiome dataset and identified oral-typical ASVs using the same criteria mentioned above. For each oral-typical ASV that cannot be matched to any sequences in MOMD, we searched against the NCBI NT database for full-length 16S rRNA sequences that were isolated from mouse and contain the oral-typical ASV as a partial sequence.

The 16S rRNA sequences in the Mouse Oral Microbiome Database (MOMD) (Joseph et al., 2021) were used as the starting point for gathering oral-typical sequences (Fig. S1A). We filtered MOMD by removing gut-colonizing sequences closely matching to those in two large-scale mouse gut bacterial collections (Lagkouvardos *et al.*, 2016; Liu *et al.*, 2020) and added oral-typical sequences absent in the MOMD from a small but rare set (n=11) of paired oral-gut samples from pregnant mice (Theis *et al.*, 2020);.

The oral bacterial loads are computed by multiplying total bacterial loads with total fractions of oral-typical bacterial ASVs. The gut bacterial loads are thetotl bacterailloads struacted from oral bacterial loads. They represent the toal density of oral-atyocal ASVs, which may include ASVs that translocated from body sites other than oral cavity. Our study ignored sthis possibility and consider oral-atypical as gut-resident bacteria.

**Ecological modeling**

Our model (Eq. (1)-(2) in the main text) describes several key ecological processes associated with the oral-derived and gut-resident bacterial populations: transmission of oral bacteria to the gut lumen (zero-order kinetics), excretion of both populations in feces (first-order kinetics), population growth, and intra-/inter-population competition for resources (Lotka-Volterra-type (Stein et al., 2013)). The host effects such as gut inflammation are not explicitly modeled. We focus on antibiotic-induced perturbations and assume that antibiotics inhibit growth of both types of bacteria equally. If the averaged inhibitory effects between oral and gut bacteria are proportional and differ by a constant, the coefficients representing the unequal inhibitory effects (Carr et al., 2020) can be absorbed into the growth rates ( and ). Since oral cavity would be only briefly exposed to antibiotics during swallowing (Carr et al., 2020), we assume that ϵ does not affect . Antibiotics may also change dilution rate (e.g., inducing diarrhea) but this effect is indirect and generally mild (Polage et al., 2012). The details of model derivation and analytical solutions are available in **Supplemental Information**. Unless otherwise noted, the model parameters are assigned to their default values: , , , , , .

**Quantifying oral bacteria fraction in microbiome samples**

A similar approach using the Human Oral Microbiome Database (HOMD) has been previously employed to classify oral bacteria (Coker *et al.*, 2018; Hu *et al.*, 2022).

We model the bacterial community in any microbiome sample as a linear combination of five source communities, corresponding to the typical communities of the five bacterial habitats revealed by the HMP dataset in the tSNE plot. Our approach is based on two assumptions: (1) bacteria translocate among different body sites (O’Boyle et al., 1998) within individuals and (2) each human body site is a distinct habitat that selects for a particular microbiome composition, such that the microbiome composition at a specific body site in one person can be approximated by the typical composition shared among the entire population. The second assumption is supported by findings that microbiome compositions vary more between body sites within the same individuals than between the same body site but in different individuals (Chu et al., 2017; Human Microbiome Project Consortium, 2012; Segata et al., 2012).

We simultaneously obtained the typical compositions of these habitats () and their mixing proportions () in HMP samples by decomposing the HMP relative abundance table () using Non-negative Matrix Factorization (NMF). The following objective function was minimized by NMF under the constraints that both and are non-negative: , where is the element-wise L1 norm, is the Frobenius norm, and is the Kullback-Leibler distance. Both (a constant that multiplies the regularization term) and (a regularization mixing parameter) are set to zero because the decomposition error (i.e.,) is optimized without regularization (see **Fig. S1A**). We normalized such that each row sum equals to 1 (i.e., the relative abundance for each community sums up to 1) and is updated accordingly to keep their product () the same. Given the microbiome relative abundances of new test samples (), the mixing proportions of these samples () can be similarly solved by minimizing where is obtained in the previous step and fixed here. Unless otherwise specified, our approach sums up the fractions of oral bacteria from both oral cavity habitat 1 and 2 to compute “oral bacteria in feces”.

We choose to collapse bacterial taxa at the lowest classified level up to genus for two reasons: (1) the interindividual variability in gut microbiome is lower in higher taxonomic ranks (Eckburg et al., 2005; Human Microbiome Project Consortium, 2012) and (2) higher taxonomic rank reduces batch effects when comparing results across multiple studies (DeSantis et al., 2006). Since the test samples may be classified using any taxonomic database, its relative abundance table () needs to reconstructed to match the taxonomic labels assigned to HMP sequences () using the GreenGenes taxonomy (DeSantis et al., 2006). When a taxon is not a GreenGenes taxonomy, it will be mapped to the lowest classified taxonomy in GreenGenes along the lineage of the taxon until the rank kingdom. Non-bacterial sequences are removed from all relative abundance tables.

**Alternative oral bacteria quantification approaches**

We adopted two well-established microbial source tracking methods to validate our approach. A Random Forest classifier (1,000 trees with maximum tree depth 5) was trained on all HMP samples to predict sample collection body sites from bacterial composition. The prediction is probabilistic with probabilities calculated as the number of trees voting for each body site divided by the total number of trees in the forest. For new test samples, the proportion from each body site is therefore an average of the relative abundance of all bacterial taxa weighted by the probabilities that they come from the body site.

SourceTracker 2 (Knights et al., 2011) uses a Bayesian approach to estimate the mixing proportions of all source communities into a sink community. For any fecal sample as a sink, we used all other samples from different body sites of the same individuals as sources. Rarefaction is performed at 100 sequences/sample for both sources and sinks. The total fraction of oral bacteria in feces is the sum of mean proportions from all oral cavity samples (from different subsites).

**Identifying co-occurring ASVs**

The algorithm for computing co-occurrence is described in detail elsewhere (Machado et al., 2021). Briefly, it begins with ASV pairs and iteratively identifies co-occurring ASVs of larger combination sizes. In the next iteration, larger combinations are created by extending all co-occurring groups identified in the current iteration with one new ASV. The presence/absence of an ASV is determined by a relative abundance cutoff of 0.001. A combination of ASVs must satisfy the following criteria to be considered as co-occurring: (1) they must co-occur in at least 200 patients and 500 samples; (2) they must co-occur at least twice more than expected by chance, which is estimated by assuming each ASV is observed independently and its number of observation can be modeled by a binomial distribution; (3) the (false discovery rate) FDR-corrected *P*-values for the independent observation hypothesis must be less than 0.05. An oral group is a co-occurring group of oral ASVs. In this analysis, ASVs assigned to the following genera are considered to have an oral origin: *Actinomyces*, *Leptotrichia*, *Campylobacter*, *Fusobacterium*, *Neisseria*, *Corynebacterium*, *Rothia*, *Treponema*, *Veillonella*, *Prevotella*, *Streptococcus*, *Capnocytophaga*, *Haemophilus*. This list is a conservative enumeration based on the overlaps between principal bacterial genera found in the healthy oral cavity (Deo and Deshmukh, 2019) and the major signature taxa (relative abundance 1%) of the oral cavity habitat 1 and 2 of the HMP samples.

**16S rRNA gene amplicon sequencing data analysis**

Generally, we prefer to adopt the metadata, OTU/ASV abundances and taxonomy used in the original studies if provided. This helps us contrast our results with their findings by minimizing the technical differences in processing their 16S sequences. Sample metadata, OTU table (including taxonomy) and sequences of the HMP cohort were downloaded from the HMP website (https://www.hmpdacc.org/HMQCP/all/). Sample, subject and clinical metadata, taxonomy and counts of ASVs, 16S/18S quantitative PCR data, and binarized fungal CFU counts of the MSK cohort (Liao et al., 2021; Yan et al., 2021) were downloaded from Figshare (https://figshare.com/collections/Compilation\_of\_longitudinal\_microbiome\_data\_and\_hospitalome\_from\_hematopoietic\_cell\_transplantation\_patients/5271128). Subject and clinical metadata, and microbial species relative abundance profiles of the PRISM, LifeLines DEEP and NLIBD cohorts were downloaded from the supplementary material of Franzosa et al., 2019. Subject and clinical metadata, OTU taxonomy and counts of the UMH (Schubert et al., 2014) and RISK (Gevers et al., 2014) cohorts were downloaded from the MicrobiomeHD database (<https://zenodo.org/record/569601#.YTqb8y9h1TY>). Subject metadata and genus-level quantitative microbiome profiling matrix of the Belgium PSC/IBD cohort were downloaded from the supplementary material of Vieira-Silva et al., 2019. Sample, subject and clinical metadata as well as the OTU table of the PROTECT cohort were downloaded from the supplementary material of the associated publication (Schirmer et al., 2018). Sample, subject, clinical metadata as well as other microbiology data of the Mount Sinai cohort (Contijoch et al., 2019), the LUMC cohort (Bekker et al., 2019), and the UPenn cohort (Lewis et al., 2015) were downloaded from the supplementary material of the associated studies.

For samples whose relative microbiome profiles are not provided (the Mount Sinai cohort and LUMC cohort), their 16S reads were downloaded and analyzed using an in-house processing pipeline (Liao et al., 2021). Briefly, ASVs were identified using the Divisive Amplicon Denoising Algorithm (DADA2) pipeline including filtering and trimming of the reads (Callahan et al., 2016). Reads were trimmed to the first 180 bp or the first point with a quality score Q<2 (Martin, 2011). Reads were removed if they contained ambiguous nucleotides (N) or if two or more errors were expected based on the quality of the trimmed read. We assigned taxonomy to ASVs using the naive Bayesian classifier (Wang et al., 2007) and the SILVA 138 database (Quast et al., 2013).

**Shotgun metagenomic sequencing data analysis**

The Shotgun data in Lewis *et al.* 2015 was analyzed using the taxonomic classification services provided by the Pathosystems Resource Integration Center (PATRIC (Davis et al., 2020)). PATRIC can take the SRA accession number of each sample as an input without downloading or transferring the sequencing files, and output the taxonomic classification of metagenomic DNA sequences included in the sample. The default algorithm is Kraken2 (Wood et al., 2019), which uses exact k-mer matches to achieve high accuracy and fast classification speeds.

We used iRep (Brown et al., 2016) to determine the replication rates of *Streptococcus spp.* that dominate stool samples in the MSK cohort. We first identified 33 Shotgun samples with at least 10% ASV\_8 or ASV\_16 based on their relative abundances in the 16S samples. We adapted a recently published pipeline (Siranosian et al., 2021) to assemble the contigs from short reads using MEGAHIT (Li et al., 2015) and then bin the contigs into Metagenome-assembled genomes (MAGs) using two different methods: Metabat2 (Kang et al., 2019) and CONCOCT (Alneberg et al., 2014). The results were then aggregated to produce an optimized, non-redundant set of MAGs by DAS Tool (Sieber et al., 2018). The high-quality MAGs (75% complete, 175 fragments/Mbp sequence, and 2% contamination) classified as *Streptococcus spp.* by Kraken2 (Wood et al., 2019) were analyzed by iRep. The iRep value of a MAG represents the average number of replication events over different subpopulations of the MAG weighted by their relative abundances.

**Statistical analysis**

Pairwise Kruskal-Wallis test was used exclusively for multiple comparisons. The raw *P*-values were adjusted for multiple testing using Benjamini-Hochberg correction to control the FDR and associations with post-correction *P*<0.05 were considered significant. When different stratified fecal sample groups are compared for their oral bacteria fractions, we replaced values less than 0.034 with 0.034 in the test. This is because comparing oral bacteria fractions below this threshold is meaningless for detecting significant differences in the total bacterial density between the groups (see **Fig. 3D** in the main text). The Student’s t-test was used to assess the significance of slopes of linear fits in the scatter plots between bacterial relative abundance and total bacterial or microbiome density.

We used the Cox’s time-varying proportional hazard model to regress fecal dominations of oral bacteria (relative abundance30%) as a microbial endpoint of interest against antibiotic administration (aminoglycosides, aztreonam, piperacillin/tazobactam, linezolid, sulfamethoxazole/trimethoprim, cephalosporins, oral vancomycin, intravenous vancomycin, amoxicillin clavulanate, quinolones, metronidazole, doxycycline, clindamycin, macrolide derivatives, carbapenems, tigecycline, other antibiotics) as multivariable predictors. Although the enrichment of oral species in gut microbiome has been linked to liver cirrhosis (Dubinkina et al., 2017; Qin et al., 2014), liver dysfunction such as altered bile acid secretion is less likely a causal factor for allo-HCT patients since the collateral damage on liver caused by the pretransplant regimen is not as severe and patients with severe liver disease would not undergo transplantation. Next, we again examined domination of oral bacteria in fecal samples, this time as a univariable predictor of *Enterococcus* infection using the same time-varying Cox’s hazard model. Penalty was not added for all Cox hazard calculations. The oral domination states in missing samples were imputed as the same as their nearest preceding samples within 4 days and assumed false if their nearest preceding samples are 5 or more days earlier. For survival analysis of mortality, we looked at the most common likely origin of fecal samples collected between day 0 and engraftment, as the exposure. Outcome analysis started at engraftment. To compare overall survival Kaplan-Meier curves were plotted and between-group *P*-values were calculated using Wald test.

All statistical analyses were performed using Python v3.7.9. Information on the statistical tests and simulations can be found in the figure legends and in the corresponding Method Details.

REFERENCES

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nat. Methods *11*, 1144–1146.

Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motoo, I., Narushima, S., Kiguchi, Y., Yasuma, K., Watanabe, E., Tanoue, T., et al. (2017). Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. Science *358*, 359–365.

Bekker, V., Zwittink, R.D., Knetsch, C.W., Sanders, I.M.J.G., Berghuis, D., Heidt, P.J., Vossen, J.M.J.J., de Vos, W.M., Belzer, C., Bredius, R.G.M., et al. (2019). Dynamics of the Gut Microbiota in Children Receiving Selective or Total Gut Decontamination Treatment during Hematopoietic Stem Cell Transplantation. Biol. Blood Marrow Transplant. *25*, 1164–1171.

Brown, C.T., Olm, M.R., Thomas, B.C., and Banfield, J.F. (2016). Measurement of bacterial replication rates in microbial communities. Nat. Biotechnol. *34*, 1256–1263.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583.

Carr, V.R., Witherden, E.A., Lee, S., Shoaie, S., Mullany, P., Proctor, G.B., Gomez-Cabrero, D., and Moyes, D.L. (2020). Abundance and diversity of resistomes differ between healthy human oral cavities and gut. Nat. Commun. *11*, 693.

Chu, D.M., Ma, J., Prince, A.L., Antony, K.M., Seferovic, M.D., and Aagaard, K.M. (2017). Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. Nat. Med. *23*, 314–326.

Contijoch, E.J., Britton, G.J., Yang, C., Mogno, I., Li, Z., Ng, R., Llewellyn, S.R., Hira, S., Johnson, C., Rabinowitz, K.M., et al. (2019). Gut microbiota density influences host physiology and is shaped by host and microbial factors. Elife *8*.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. Science *326*, 1694–1697.

Davidson-Pilon, C. (2019). lifelines: survival analysis in Python. J. Open Source Softw. *4*, 1317.

Davis, J.J., Wattam, A.R., Aziz, R.K., Brettin, T., Butler, R., Butler, R.M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E.M., et al. (2020). The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res. *48*, D606–D612.

Deo, P.N., and Deshmukh, R. (2019). Oral microbiome: Unveiling the fundamentals. J. Oral Maxillofac. Pathol. *23*, 122–128.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. *72*, 5069–5072.

Dubinkina, V.B., Tyakht, A.V., Odintsova, V.Y., Yarygin, K.S., Kovarsky, B.A., Pavlenko, A.V., Ischenko, D.S., Popenko, A.S., Alexeev, D.G., Taraskina, A.Y., et al. (2017). Links of gut microbiota composition with alcohol dependence syndrome and alcoholic liver disease. Microbiome *5*, 141.

Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat. Commun. *8*, 1784.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the human intestinal microbial flora. Science *308*, 1635–1638.

Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H.J., Reinker, S., Vatanen, T., Hall, A.B., Mallick, H., McIver, L.J., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat. Microbiol. *4*, 293–305.

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The treatment-naive microbiome in new-onset Crohn’s disease. Cell Host Microbe *15*, 382–392.

Goyal, A., Dubinkina, V., and Maslov, S. (2018). Multiple stable states in microbial communities explained by the stable marriage problem. ISME J. *12*, 2823–2834.

Ho, T.K. (1995). Random decision forests. Proceedings of 3rd International Conference On.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359.

Kitamoto, S., Nagao-Kitamoto, H., Hein, R., Schmidt, T.M., and Kamada, N. (2020a). The Bacterial Connection between the Oral Cavity and the Gut Diseases. J. Dent. Res. *99*, 1021–1029.

Kitamoto, S., Nagao-Kitamoto, H., Jiao, Y., Gillilland, M.G., Hayashi, A., Imai, J., Sugihara, K., Miyoshi, M., Brazil, J.C., Kuffa, P., et al. (2020b). The Intermucosal Connection between the Mouth and Gut in Commensal Pathobiont-Driven Colitis. Cell *182*, 447–462.e14.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., and Kelley, S.T. (2011). Bayesian community-wide culture-independent microbial source tracking. Nat. Methods *8*, 761–763.

Lahti, L., Salojärvi, J., Salonen, A., Scheffer, M., and de Vos, W.M. (2014). Tipping elements in the human intestinal ecosystem. Nat. Commun. *5*, 4344.

Lamont, R.J., Koo, H., and Hajishengallis, G. (2018). The oral microbiota: dynamic communities and host interactions. Nat. Rev. Microbiol. *16*, 745–759.

Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788–791.

Lewis, J.D., Chen, E.Z., Baldassano, R.N., Otley, A.R., Griffiths, A.M., Lee, D., Bittinger, K., Bailey, A., Friedman, E.S., Hoffmann, C., et al. (2015). Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric crohn’s disease. Cell Host Microbe *18*, 489–500.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676.

Liao, C., Taylor, B.P., Ceccarani, C., Fontana, E., Amoretti, L.A., Wright, R.J., Gomes, A.L.C., Peled, J.U., Taur, Y., Perales, M.-A., et al. (2021). Compilation of longitudinal microbiota data and hospitalome from hematopoietic cell transplantation patients. Sci. Data *8*, 71.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research *9*, 2579–2605.

Machado, D., Maistrenko, O.M., Andrejev, S., Kim, Y., Bork, P., Patil, K.R., and Patil, K.R. (2021). Polarization of microbial communities between competitive and cooperative metabolism. Nat. Ecol. Evol. *5*, 195–203.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10.

Morjaria, S., Schluter, J., Taylor, B.P., Littmann, E.R., Carter, R.A., Fontana, E., Peled, J.U., van den Brink, M.R.M., Xavier, J.B., and Taur, Y. (2019). Antibiotic-Induced Shifts in Fecal Microbiota Density and Composition during Hematopoietic Stem Cell Transplantation. Infect. Immun. *87*.

O’Boyle, C.J., MacFie, J., Mitchell, C.J., Johnstone, D., Sagar, P.M., and Sedman, P.C. (1998). Microbiology of bacterial translocation in humans. Gut *42*, 29–35.

Olsen, I., and Yamazaki, K. (2019). Can oral bacteria affect the microbiome of the gut? J. Oral Microbiol. *11*, 1586422.

Pedregosa, F., Varoquaux, G., and Gramfort, A. (2011). Scikit-learn: Machine learning in Python. Of Machine Learning.

Peled, J.U., Gomes, A.L.C., Devlin, S.M., Littmann, E.R., Taur, Y., Sung, A.D., Weber, D., Hashimoto, D., Slingerland, A.E., Slingerland, J.B., et al. (2020). Microbiota as Predictor of Mortality in Allogeneic Hematopoietic-Cell Transplantation. N. Engl. J. Med. *382*, 822–834.

Polage, C.R., Solnick, J.V., and Cohen, S.H. (2012). Nosocomial diarrhea: evaluation and treatment of causes other than Clostridium difficile. Clin. Infect. Dis. *55*, 982–989.

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., Guo, J., Le Chatelier, E., Yao, J., Wu, L., et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. Nature *513*, 59–64.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590-6.

Ramirez, J., Guarner, F., Bustos Fernandez, L., Maruy, A., Sdepanian, V.L., and Cohen, H. (2020). Antibiotics as major disruptors of gut microbiota. Front. Cell Infect. Microbiol. *10*, 572912.

Rao, C., Coyte, K.Z., Bainter, W., Geha, R.S., Martin, C.R., and Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. Nature *591*, 633–638.

Rashidi, A., Ebadi, M., Weisdorf, D.J., Costalonga, M., and Staley, C. (2021). No evidence for colonization of oral bacteria in the distal gut in healthy adults. Proc. Natl. Acad. Sci. USA *118*.

Read, E., Curtis, M.A., and Neves, J.F. (2021). The role of oral bacteria in inflammatory bowel disease. Nat. Rev. Gastroenterol. Hepatol. *18*, 731–742.

Rolling, T., Zhai, B., Gjonbalaj, M., Tosini, N., Yasuma-Mitobe, K., Fontana, E., Amoretti, L.A., Wright, R.J., Ponce, D.M., Perales, M.A., et al. (2021). Haematopoietic cell transplantation outcomes are linked to intestinal mycobiota dynamics and an expansion of Candida parapsilosis complex species. Nat. Microbiol.

Schirmer, M., Denson, L., Vlamakis, H., Franzosa, E.A., Thomas, S., Gotman, N.M., Rufo, P., Baker, S.S., Sauer, C., Markowitz, J., et al. (2018). Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. Cell Host Microbe *24*, 600–610.e4.

Schmidt, T.S., Hayward, M.R., Coelho, L.P., Li, S.S., Costea, P.I., Voigt, A.Y., Wirbel, J., Maistrenko, O.M., Alves, R.J., Bergsten, E., et al. (2019). Extensive transmission of microbes along the gastrointestinal tract. Elife *8*.

Schubert, A.M., Rogers, M.A.M., Ring, C., Mogle, J., Petrosino, J.P., Young, V.B., Aronoff, D.M., and Schloss, P.D. (2014). Microbiome data distinguish patients with Clostridium difficile infection and non-C. difficile-associated diarrhea from healthy controls. MBio *5*, e01021-14.

Segata, N., Haake, S.K., Mannon, P., Lemon, K.P., Waldron, L., Gevers, D., Huttenhower, C., and Izard, J. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biol. *13*, R42.

Shenhav, L., Thompson, M., Joseph, T.A., Briscoe, L., Furman, O., Bogumil, D., Mizrahi, I., Pe’er, I., and Halperin, E. (2019). FEAST: fast expectation-maximization for microbial source tracking. Nat. Methods *16*, 627–632.

Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat. Microbiol. *3*, 836–843.

Siranosian, B.A., Brooks, E., Andermann, T., Rezvani, A.R., Banaei, N., Tang, H., and Bhatt, A.S. (2021). Rare transmission of commensal and pathogenic bacteria in the gut microbiome of hospitalized adults. BioRxiv.

Stein, R.R., Bucci, V., Toussaint, N.C., Buffie, C.G., Rätsch, G., Pamer, E.G., Sander, C., and Xavier, J.B. (2013). Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput. Biol. *9*, e1003388.

Tamburini, F.B., Andermann, T.M., Tkachenko, E., Senchyna, F., Banaei, N., and Bhatt, A.S. (2018). Precision identification of diverse bloodstream pathogens in the gut microbiome. Nat. Med. *24*, 1809–1814.

Tang, T.W.H., Chen, H.-C., Chen, C.-Y., Yen, C.Y.T., Lin, C.-J., Prajnamitra, R.P., Chen, L.-L., Ruan, S.-C., Lin, J.-H., Lin, P.-J., et al. (2019). Loss of gut microbiota alters immune system composition and cripples postinfarction cardiac repair. Circulation *139*, 647–659.

Taur, Y., Xavier, J.B., Lipuma, L., Ubeda, C., Goldberg, J., Gobourne, A., Lee, Y.J., Dubin, K.A., Socci, N.D., Viale, A., et al. (2012). Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clin. Infect. Dis. *55*, 905–914.

Tett, A., Pasolli, E., Masetti, G., Ercolini, D., and Segata, N. (2021). Prevotella diversity, niches and interactions with the human host. Nat. Rev. Microbiol. *19*, 585–599.

Vandeputte, D., Kathagen, G., D’hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. Nature *551*, 507–511.

Vieira-Silva, S., Sabino, J., Valles-Colomer, M., Falony, G., Kathagen, G., Caenepeel, C., Cleynen, I., van der Merwe, S., Vermeire, S., and Raes, J. (2019). Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. Nat. Microbiol. *4*, 1826–1831.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. *73*, 5261–5267.

Weiss, G.A., and Hennet, T. (2017). Mechanisms and consequences of intestinal dysbiosis. Cell Mol. Life Sci. *74*, 2959–2977.

Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257.

Yan, J., Liao, C., Taylor, B.P., Fontana, E., Amoretti, L.A., Wright, R.J., Dai, A., Waters, N., Peled, J.U., Taur, Y., et al. (2021). A compilation of fecal microbiome shotgun metagenomics from hospitalized patients undergoing hematopoietic cell transplantation. BioRxiv.

Zhai, B., Ola, M., Rolling, T., Tosini, N.L., Joshowitz, S., Littmann, E.R., Amoretti, L.A., Fontana, E., Wright, R.J., Miranda, E., et al. (2020). High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis. Nat. Med. *26*, 59–64.

Abed, J., Maalouf, N., Manson, A.L., Earl, A.M., Parhi, L., Emgård, J.E., Klutstein, M., Tayeb, S., Almogy, G., and Atlan, K.A. (2020). Colon cancer-associated Fusobacterium nucleatum may originate from the oral cavity and reach colon tumors via the circulatory system. Frontiers in cellular and infection microbiology *10*, 400.

Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motoo, I., Narushima, S., Kiguchi, Y., Yasuma, K., Watanabe, E., and Tanoue, T. (2017). Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. Science *358*, 359-365.

Bekker, V., Zwittink, R.D., Knetsch, C.W., Sanders, I.M., Berghuis, D., Heidt, P.J., Vossen, J.M., de Vos, W.M., Belzer, C., and Bredius, R.G. (2019). Dynamics of the gut microbiota in children receiving selective or total gut decontamination treatment during hematopoietic stem cell transplantation. Biology of Blood and Marrow Transplantation *25*, 1164-1171.

Britton, G.J., Contijoch, E.J., Spindler, M.P., Aggarwala, V., Dogan, B., Bongers, G., San Mateo, L., Baltus, A., Das, A., and Gevers, D. (2020). Defined microbiota transplant restores Th17/RORγt+ regulatory T cell balance in mice colonized with inflammatory bowel disease microbiotas. Proceedings of the National Academy of Sciences *117*, 21536-21545.

Coker, O.O., Dai, Z., Nie, Y., Zhao, G., Cao, L., Nakatsu, G., Wu, W.K., Wong, S.H., Chen, Z., and Sung, J.J. (2018). Mucosal microbiome dysbiosis in gastric carcinogenesis. Gut *67*, 1024-1032.

Contijoch, E.J., Britton, G.J., Yang, C., Mogno, I., Li, Z., Ng, R., Llewellyn, S.R., Hira, S., Johnson, C., and Rabinowitz, K.M. (2019). Gut microbiota density influences host physiology and is shaped by host and microbial factors. Elife *8*.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. science *326*, 1694-1697.

Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nature communications *8*, 1-10.

Galazzo, G., Van Best, N., Benedikter, B.J., Janssen, K., Bervoets, L., Driessen, C., Oomen, M., Lucchesi, M., van Eijck, P.H., and Becker, H.E. (2020). How to count our microbes? The effect of different quantitative microbiome profiling approaches. Frontiers in cellular and infection microbiology *10*, 403.

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., and Yassour, M. (2014). The treatment-naive microbiome in new-onset Crohn’s disease. Cell host & microbe *15*, 382-392.

Greenberg, J.M., Romero, R., Winters, A.D., Galaz, J., Garcia-Flores, V., Arenas-Hernandez, M., Panzer, J., Shaffer, Z., Kracht, D.J., and Gomez-Lopez, N. (2022). Microbiota of the pregnant mouse: characterization of the bacterial communities in the oral cavity, lung, intestine, and vagina through culture and DNA sequencing. bioRxiv.

Hu, Y., Amir, A., Huang, X., Li, Y., Huang, S., Wolfe, E., Weiss, S., Knight, R., and Xu, Z.Z. (2022). Diurnal and eating-associated microbial patterns revealed via high-frequency saliva sampling. Genome Research *32*, 1112-1123.

Imai, J., Ichikawa, H., Kitamoto, S., Golob, J.L., Kaneko, M., Nagata, J., Takahashi, M., Gillilland III, M.G., Tanaka, R., and Nagao-Kitamoto, H. (2021). A potential pathogenic association between periodontal disease and Crohn’s disease. JCI insight *6*.

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., and Korpela, K. (2020). Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PloS one *15*, e0227285.

Jian, C., Salonen, A., and Korpela, K. (2021). Commentary: How to count our microbes? The effect of different quantitative microbiome profiling approaches. Frontiers in Cellular and Infection Microbiology *11*, 627910.

Jin, S., Wetzel, D., and Schirmer, M. (2022). Deciphering mechanisms and implications of bacterial translocation in human health and disease. Current Opinion in Microbiology *67*, 102147.

Joseph, S., Aduse-Opoku, J., Hashim, A., Hanski, E., Streich, R., Knowles, S.C., Pedersen, A.B., Wade, W.G., and Curtis, M.A. (2021). A 16S rRNA gene and draft genome database for the murine oral bacterial community. Msystems *6*, e01222-01220.

Kitamoto, S., Nagao-Kitamoto, H., Hein, R., Schmidt, T., and Kamada, N. (2020). The bacterial connection between the oral cavity and the gut diseases. Journal of dental research *99*, 1021-1029.

Komiya, Y., Shimomura, Y., Higurashi, T., Sugi, Y., Arimoto, J., Umezawa, S., Uchiyama, S., Matsumoto, M., and Nakajima, A. (2019). Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity. Gut *68*, 1335-1337.

Kostic, A.D., Chun, E., Robertson, L., Glickman, J.N., Gallini, C.A., Michaud, M., Clancy, T.E., Chung, D.C., Lochhead, P., and Hold, G.L. (2013). Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell host & microbe *14*, 207-215.

Lagkouvardos, I., Pukall, R., Abt, B., Foesel, B.U., Meier-Kolthoff, J.P., Kumar, N., Bresciani, A., Martínez, I., Just, S., and Ziegler, C. (2016). The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. Nature microbiology *1*, 1-15.

Liu, C., Zhou, N., Du, M.-X., Sun, Y.-T., Wang, K., Wang, Y.-J., Li, D.-H., Yu, H.-Y., Song, Y., and Bai, B.-B. (2020). The Mouse Gut Microbial Biobank expands the coverage of cultured bacteria. Nature communications *11*, 1-12.

Machado, D., Maistrenko, O.M., Andrejev, S., Kim, Y., Bork, P., Patil, K.R., and Patil, K.R. (2021). Polarization of microbial communities between competitive and cooperative metabolism. Nature ecology & evolution *5*, 195-203.

Morjaria, S., Schluter, J., Taylor, B.P., Littmann, E.R., Carter, R.A., Fontana, E., Peled, J.U., van den Brink, M.R., Xavier, J.B., and Taur, Y. (2019). Antibiotic-induced shifts in fecal microbiota density and composition during hematopoietic stem cell transplantation. Infection and immunity *87*, e00206-00219.

Peled, J.U., Gomes, A.L., Devlin, S.M., Littmann, E.R., Taur, Y., Sung, A.D., Weber, D., Hashimoto, D., Slingerland, A.E., and Slingerland, J.B. (2020). Microbiota as predictor of mortality in allogeneic hematopoietic-cell transplantation. New England Journal of Medicine *382*, 822-834.

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., Guo, J., Le Chatelier, E., Yao, J., and Wu, L. (2014). Alterations of the human gut microbiome in liver cirrhosis. Nature *513*, 59-64.

Rao, C., Coyte, K.Z., Bainter, W., Geha, R.S., Martin, C.R., and Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. Nature *591*, 633-638.

Rashidi, A., Ebadi, M., Weisdorf, D.J., Costalonga, M., and Staley, C. (2021). No evidence for colonization of oral bacteria in the distal gut in healthy adults. Proceedings of the National Academy of Sciences *118*, e2114152118.

Read, E., Curtis, M.A., and Neves, J.F. (2021). The role of oral bacteria in inflammatory bowel disease. Nature Reviews Gastroenterology & Hepatology *18*, 731-742.

Reese, A.T., Cho, E.H., Klitzman, B., Nichols, S.P., Wisniewski, N.A., Villa, M.M., Durand, H.K., Jiang, S., Midani, F.S., and Nimmagadda, S.N. (2018). Antibiotic-induced changes in the microbiota disrupt redox dynamics in the gut. Elife *7*, e35987.

Schirmer, M., Denson, L., Vlamakis, H., Franzosa, E.A., Thomas, S., Gotman, N.M., Rufo, P., Baker, S.S., Sauer, C., and Markowitz, J. (2018). Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. Cell host & microbe *24*, 600-610. e604.

Schluter, J., Peled, J.U., Taylor, B.P., Markey, K.A., Smith, M., Taur, Y., Niehus, R., Staffas, A., Dai, A., and Fontana, E. (2020). The gut microbiota is associated with immune cell dynamics in humans. Nature *588*, 303-307.

Schmidt, T.S., Hayward, M.R., Coelho, L.P., Li, S.S., Costea, P.I., Voigt, A.Y., Wirbel, J., Maistrenko, O.M., Alves, R.J., and Bergsten, E. (2019). Extensive transmission of microbes along the gastrointestinal tract. Elife *8*, e42693.

Segata, N., Haake, S.K., Mannon, P., Lemon, K.P., Waldron, L., Gevers, D., Huttenhower, C., and Izard, J. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome biology *13*, 1-18.

Shono, Y., and van den Brink, M.R. (2018). Gut microbiota injury in allogeneic haematopoietic stem cell transplantation. Nature Reviews Cancer *18*, 283-295.

Staffas, A., da Silva, M.B., Slingerland, A.E., Lazrak, A., Bare, C.J., Holman, C.D., Docampo, M.D., Shono, Y., Durham, B., and Pickard, A.J. (2018). Nutritional support from the intestinal microbiota improves hematopoietic reconstitution after bone marrow transplantation in mice. Cell host & microbe *23*, 447-457. e444.

Tang, T.W., Chen, H.-C., Chen, C.-Y., Yen, C.Y., Lin, C.-J., Prajnamitra, R.P., Chen, L.-L., Ruan, S.-C., Lin, J.-H., and Lin, P.-J. (2019). Loss of gut microbiota alters immune system composition and cripples postinfarction cardiac repair. Circulation *139*, 647-659.

Taur, Y., Xavier, J.B., Lipuma, L., Ubeda, C., Goldberg, J., Gobourne, A., Lee, Y.J., Dubin, K.A., Socci, N.D., and Viale, A. (2012). Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clinical infectious diseases *55*, 905-914.

The Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. nature *486*, 207-214.

Theis, K.R., Romero, R., Greenberg, J.M., Winters, A.D., Garcia-Flores, V., Motomura, K., Ahmad, M.M., Galaz, J., Arenas-Hernandez, M., and Gomez-Lopez, N. (2020). No consistent evidence for microbiota in murine placental and fetal tissues. Msphere *5*, e00933-00919.

Thomas, A.M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., and Pozzi, C. (2019). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nature medicine *25*, 667-678.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research *9*.

Vandeputte, D., Kathagen, G., D’hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., and Darzi, Y. (2017). Quantitative microbiome profiling links gut community variation to microbial load. Nature *551*, 507-511.

Vieira-Silva, S., Sabino, J., Valles-Colomer, M., Falony, G., Kathagen, G., Caenepeel, C., Cleynen, I., van der Merwe, S., Vermeire, S., and Raes, J. (2019). Quantitative microbiome profiling disentangles inflammation-and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. Nature microbiology *4*, 1826-1831.

Zarrinpar, A., Chaix, A., Xu, Z.Z., Chang, M.W., Marotz, C.A., Saghatelian, A., Knight, R., and Panda, S. (2018). Antibiotic-induced microbiome depletion alters metabolic homeostasis by affecting gut signaling and colonic metabolism. Nature communications *9*, 1-13.