**Relative enrichment of oral 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 digestive diseases with relative enrichment of oral bacteria in human feces can be explained by two alternate hypotheses: it may indicate increased population density of ectopic oral bacteria in the gut (the *expansion* hypothesis) or depleted gut commensal bacteria (the *marker* hypothesis). Using mouse experiments and human microbiome data analysis, we found that antibiotics cause relative enrichment of orally derived bacteria in the intestine but their population do not expand in size, thus supporting the *marker* hypothesis. By computer simulations, we propose that the *marker* hypothesis is achieved by balanced effects of direct antibiotic inhibition and indirect ecological release on the oral populations. 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**

Different body sites of healthy people harbor distinct microbial communities (Costello et al., 2009; The Human Microbiome Project Consortium, 2012) but their microbiome compositions may become similar in people with disease (Imai et al., 2021). The collective 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-derived bacterial cells are transported 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 must overcome colonization resistance of the gut-resident commensal bacteria to populate (Jin *et al.*, 2022). This is supported by DNA-based surveys that detect very low abundances of oral bacterial DNAs in feces of health individuals (typically <2%) (Rashidi et al., 2021; Schmidt et al., 2019).

Multiple factors including antibiotics, diets, aging, and gut inflammation can disrupt oral-gut barrier (Kitamoto *et al.*, 2020) and jointly contribute to the clinical associations between ectopic colonization of oral bacteria and a variety of digestive diseases (e.g., Crohn’s disease (CD) (Gevers et al., 2014), ulcerative colitis (UC) (Schirmer et al., 2018), CRC (Kostic et al., 2013), liver cirrhosis (Qin et al., 2014)). However, nearly all those association studies measured oral bacterial colonization by their relative abundances and are thus inconclusive regarding whether the relative enrichment reflects their active expansion in the gut. By mathematical definition, the increased relative abundance of oral bacteria can be explained by two alternate hypotheses (Fig. 1). In the *expansion* hypothesis, the relative abundance informs absolute abundance, and the relative enrichment reflects active expansion of oral bacterial population in the gut. By contrast, the *marker* hypothesis proposes that the relative enrichment is driven by loss of gut commensals. The two hypotheses have distinct implications for human health and pathologies: While the *expansion* hypothesis implies that the ingested oral bacteria can directly drive or exacerbate digestive diseases, the *marker* hypothesis indicates indirect effects of gut microbiome on disease due to loss of gut beneficial anaerobes and their associated functions. In the following, we distinguished the two hypotheses by bioinformatically detecting translocated oral bacteria and quantifying their absolute abundances 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 quantitative abundance of oral 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 bacteria in the human gut (fecal) microbiome by two alternate mechanisms. The relative enrichment reflects increased absolute abundance (i.e., active expansion of the oral bacterial populations) in the *expansion* hypothesis (red) and reflects depletion of gut bacterial commensals in the *marker* hypothesis (blue).

**RESULTS**

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

We treated 3 mice with antibiotic cocktails of ampicillin, vancomycin, and neomycin since day 0 for a week (Fig. 2A). The microbiome compositions of oral cavity and fecal samples on day 0 prior to the treatment as well as fecal samples on day 8 were profiled at the ASV (amplicon sequence variant) level (Fig. 2B). We found that the post-treatment fecal samples shared more identical ASVs (Hamming distance) with the pre-treatment oral samples than the pre-treatment fecal samples in all 3 mice and were also compositionally similar in 2 mice by taking relative abundances into accounts (Bray-Curtis distance) (Fig. 2C). Notably, 10%-20% ASVs (the majority from 6 members; Fig. S1) found in the post-treatment feces were only detected in the pre-treatment oral microbiomes but not in the pre-treatment fecal microbiomes (Fig. 2D), providing strong evidence of oral-gut transmission of bacteria that contain those ASVs. By identifying ASVs typical of the oral cavity of the 3 mice (mean relative abundance 1e-4 across pre-treatment oral samples and 1e-4 across pre-treatment fecal samples), we found that the antibiotic cocktails increased the total fractions (i.e., relative abundance) of oral-typical ASVs from nearly 0% to 30% on average (Fig. 2E), while their mean loads (i.e., absolute abundance) decreased by half a log density (Fig. 2F). The contrast between the relative and absolute abundances of oral bacterial populations supports the *marker* hypothesis.

**Additional evidence of the *marker* hypothesis from mouse experiments in the literature**

To our best knowledge, antibiotic-treated mouse experiments in the literature lack paired oral and gut samples. This limitation prevents us from applying the same paired-sample-based definition of oral-typical bacteria used above to quantify translocated oral bacteria in feces. To overcome the limitation, we compiled a reference set of 149 non-redundant, full-length 16S rRNA genes of oral-typical bacteria under the operational definition that they are culturable from the oral cavity but not from the gut (Fig. S2, Table S1, see STAR Methods). The reference set allows us to infer whether a 16S ASV in feces is oral-typical in the absence of paired oral samples, by exactly matching the ASV to all sequences in the set. Having established the inference approach, we reanalyzed a public study (Reese et al., 2018) that treated mice with antibiotic cocktails of ampicillin, vancomycin, neomycin, and metronidazole for 5 days (Fig. 2G). The inferred oral bacterial populations, mainly consisting of three ASVs of *Lactobacillus*, *Proteus* and *Escherichia-Shigella*, had substantially higher relative abundances after two days’ exposure (Fig. 2H). However, their absolute abundances did not increase but reduced ~10 folds on average before recovery (Fig. 2I, left). Therefore, the relative enrichment of oral bacteria follows the *marker* hypothesis and is fully driven by reduced gut bacterial load (Fig. 2I, right).

A consequence of the *marker* hypothesis is that the total bacterial load of the gut microbiome is low at high relative abundance of oral bacteria. Using the mouse data from Reese *et al.*, we showed that the oral bacterial fractions in feces are negatively linearly correlated with the total bacterial loads on the log-log scale (Fig. 2J). The linear relationship is moderate to strong (Pearson’s r = -0.52, P = 3.5e-17) and remained significant by using the control group data alone (Pearson’s r = -0.25, P = 5.9e-3). We noted that 11% samples collected between day 1-6 (Fig. S3A) with low bacterial loads deviated from the trend line of the association and, according to the *marker* hypothesis, should have very high proportions of oral bacteria. We speculated that the dominant ASVs of these outliers (Fig. S3B) were orally derived but undetected by our inference approach. The omit of uncultured bacteria in the reference set is unlikely the major cause of the potential inference failure, because the proportions of uncultured bacteria based on taxonomic annotation are low to intermediate in these samples (Fig. S3C).

Other than Reese *et al.*, we have obtained similar findings in another public dataset from bone-marrow-transplanted mice that received mono-antibiotic prophylaxis (Staffas et al., 2018). Ampicillin but not streptomycin, aztreonam, or vancomycin substantially reduced gut bacterial loads (Fig. S4A). Consistent with the *marker* hypothesis, the oral bacterial loads remained stable without active expansion throughout the experiments for all four antibiotics including ampicillin (Fig. S4B). The stable oral bacteria loads led to a strong log-log linear relationship between the oral bacterial fractions and total bacterial loads (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-F**)Mouse experiment in this study. (A) Experimental design. Mice (n=3, labeled as M1, M2, M3) were treated with antibiotic cocktails of ampicillin, vancomycin, and neomycin (AVN) for a week. The pre-treatment fecal (fecal\_d0) and oral (oral\_d0) samples were collected on day 0 and the post-treatment fecal samples (fecal\_d8) were collected on day 8. (**B**) Microbiome compositions of all fecal and oral samples**.** (**C**) Compositional dissimilarity between post-treatment fecal samples and pre-treatment fecal or oral samples. (**D**) Sources of ASVs in the post-treatment fecal samples. (**E,F**) Relative (E) and absolute (F) abundances of oral bacterial populations in the feces. (**G-J**) Reanalysis of a public study (Reese *et al.*, 2018). (**G**) Experimental design. Mice were treated with either water (the control group; n=11) or antibiotic cocktails of ampicillin, vancomycin, neomycin, and metronidazole (AVNM) (the antibiotic group; n=10) for 5 days. (**H**) Compositional dynamics of oral bacterial ASVs in feces. (**I**) Absolute abundances of oral (left) and gut (right) bacteria in feces averaged across mice in the control or treatment group. Lines and dots: mean; shading: 95% confidence interval (CI). (**J**) Linear relationship between total bacterial loads and oral bacterial fractions. Line: best linear fit; shading: 95% CI. Unit of bacterial load: 16S copies per gram of feces.

**Detection of oral bacterial DNAs in human feces without paired oral samples**

The existing human gut microbiome studies with quantitative profiles mostly lack sequencing of paired oral samples. To reuse the quantitative data, we followed a similar strategy as used in mice to build a reference set of bacterial 16S rRNA gene sequences that typically colonize the human oral cavity (see STAR Methods). Rather than being limited to cultured bacteria, we leveraged the Human Microbiome Project (HMP) that have sequenced paired oral (multiple subsites)-gut (stool) microbiomes from 237 healthy volunteers (The Human Microbiome Project Consortium, 2012). Distinct bacterial communities were found inhabiting the two body sites (Fig. 3A), which provides a pragmatic basis to define oral-typical ASVs based on their abundance and prevalence. Here we used the following filtering criteria for oral-typical ASVs: their mean relative abundance and prevalence must be greater than 0.01% and 5% respectively among all oral cavity samples, and no greater than the same cutoffs among all fecal samples (Fig. 3B,C). The rationale of filtering is that the more typical ASV of the oral cavity, the less likely it colonizes the gut due to niche specificity, and the more likely its presence in feces indicates oral-gut translocation.

The filtering step left a reference set of 178 unique ASVs sequenced for 16S rRNA at the V4-V5 region (Table S2) and the leading genera are *Prevotella* and *Streptococcus* (Fig. 3D). Using the reference set, we found that 219 out of 280 HMP fecal samples do not contain oral ASVs, and their mean relative abundance is as low as 0.05% (Fig. 3E). This is consistent with previous findings that oral bacteria are very minor members of healthy gut microbiome (Rashidi *et al.*, 2021; Schmidt *et al.*, 2019). To show that filtering is key to 16S rRNA-based inference of oral bacteria, we took an alternative approach by counting every single ASV in feces as of oral origin if it is found in any of paired oral cavity subsite samples. This naïve approach led to an average of 247 oral ASVs per HMP fecal sample and their mean relative abundance rises >300-fold to 15.6%. The severe overestimation suggests that the majority of shared ASVs are not indicators of oral-gut transmission but coincidences of closely related bacteria occupying both niches.

To explore whether the selected oral ASVs from HMP healthy subjects can be applied to identify oral bacteria in the gut of patients, we reanalyzed a public study that sequenced paired oral (saliva)-gut (stool) samples from IBD (CD and UC) patients and their healthy controls (HC). We first showed that the estimated oral bacterial fractions in feces are generally robust against variations of the filtering cutoffs used to generate the reference set of oral bacterial sequences (Fig. S5). Using the default reference set, we estimated that the total fractions of oral ASVs in the feces of IBD patients were averagely 4 times (CD: 4.2%, UC: 4.3%) as high as the mean fraction in the feces of HC (1.1%) (Fig. 3F), confirming the notion that IBD patients are enriched with oral bacteria in their gut. Among 99 fecal samples (HC:41, CD:16, UC:42) that contained at least one oral ASV, the proportion of oral ASVs that were also found in paired saliva samples is >70% in 87 samples (HC:36, CD:15, UC:36; Fig. 3G). Similarly, >70% of the total fractions of oral ASVs in 90 out of 99 fecal samples was contributed by those found in the paired saliva samples (HC:38, CD:15, UC:37; Fig. 3H). Both computational validations indicate that our inference approach is conservative and less prone to false-positive predictions (i.e., inferred oral ASVs absent from the oral cavity).

A picture containing text

Description automatically generated

**Figure 3. 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 mean relative abundance (B) and prevalence (C) of ASVs (dots) among all HMP oral cavity samples (x axis) and the fecal samples (y axis). 178 ASVs were identified as oral-typical and highlighted in orange. (**D**) Genus-level distribution of the 178 oral-typical ASVs. (**E, F**) The total fractions of oral-typical ASVs in the feces of (E) HMP subjects (n=280) and (F) patients with inflammatory bowel disease as well as their own healthy controls (n=101). Each circle represents a fecal sample. HC: healthy control (n=43); CD: Crohn’s disease (n=16); UC: ulcerative colitis (n=42). (**G**) Proportions of fecal ASVs inferred as of oral origin that can also be found in paired saliva samples. (**H**) Total fractions of inferred oral ASVs that were also found in the paired saliva samples, divided by total fractions of all inferred oral ASVs in the feces. Data sources: panels A-E (The Human Microbiome Project Consortium, 2012); panels F-H (Imai *et al.*, 2021).

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

We applied the inference approach established above to a large-scale microbiome dataset with quantitative profiles (16S amplicon sequencing and qPCR) (Liao et al., 2021; Yan et al., 2022). The dataset was obtained from adult patients receiving allogeneic hematopoietic cell transplantation (allo-HCT)—the only curative therapy for a variety of hematologic disorders including leukemias, lymphomas, and myeloma (Shono and van den Brink, 2018). Along with other related procedures (conditioning regime, antibiotic exposure, etc.), allo-HCT disrupted the gut microbiome compositions of the recipients, 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 intestinal bacterial communities (Fig. 4A, top) (Peled et al., 2020). Among the 10,433 samples from 1,276 patients, 901 were dominated by at least one oral ASV and the leading genera of these ASVs are *Streptococcus*, *Actinomyces*, and *Abiotrophia* (Fig. 4A, middle). The total bacterial loads of a subset of 3,126 samples showed that the fecal samples highly enriched with oral ASVs have lower bacterial biomass (Fig. 4A, bottom). We confirmed that the total fractions of oral ASVs detected in the feces of allo-HCT recipients are not correlated with sequencing depths (Fig. S6).

Due to the lack of paired oral samples, we are unclear whether the inferred oral ASVs originated from the oral cavity of the same patients. To attenuate the limitation, we used co-occurrence analysis (see STAR Methods) to identify combinations of oral ASVs that were simultaneously present in the feces more than expected by chance. The co-occurring oral ASVs, if existed, would indicate collective bacterial transmission from the oral cavity to the gut. We found 71 such oral-ASV groups of size from 2 to 5, among which 55 contain *Streptococcus* ASVs (Table S3). To identify the species of the *Streptococcus* ASVs, we extracted shotgun metagenomes from 19 samples which have at least 10% (measured by 16S amplicon sequencing) ASV\_8 (the most abundant *Streptococcus* ASV inferred as of oral origin) from the same allo-HCT cohort. We were able to recover 22 high-quality metagenome-assembled genomes of *Streptococcus spp.*, among which 15 were annotated as *S. thermophilus*—a probiotic species of lactic acid bacteria. To assess the viability of the *Streptococcus* genomes, we computed the ratio of metagenomic reads near the replication origin to the replication terminus (i.e., peak-to-trough (PTR) ratio) to estimate the simultaneous replication events averaged across the population ([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 averaged PTR ratio is 1.50 0.16 (Table S4), suggesting a possible mixture of 50% 2-fork and 50% single-fork cells. The PTR-based growth estimates indicated that these *Streptococcus* genomes generally had slow to intermediate growth at the time of sampling.

**Validation of the *marker* hypothesis in the gut microbiome of allo-HCT recipients**

The allo-HCT recipients 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). For the cohort we used, the allo-HCT procedure usually started antibiotic prophylaxis within a week prior to transplantation. The timing of antibacterial antibiotic administration (Fig. 4B, top) corresponded well to the declined gut bacterial relative (Fig. 4B, middle) and absolute (Fig. 4B, bottom) abundances. Since the mean oral bacterial loads fluctuated around a stable average, the increased relative abundance of oral ASVs was mainly driven by the declined gut bacterial load. This supports the *marker* hypothesis and generates a negative correlation (Pearson’s r = -0.27, P = 2.2e-54) between the oral bacterial proportions and the total bacterial loads in feces (Fig. 4C). The inverse correlation of oral bacteria is in direct contrast to *Enterococcus* (Stein-Thoeringer et al., 2019) and *Candida spp.* (Rolling et al., 2021; Zhai et al., 2020) whose relative abundances in the gut informs their absolute quantities.

Given the complex antibiotic use pattern, we conducted a time-varying survival analysis to identify which antibiotic exposures in the past predict the future occurrence of oral ASV domination in the gut above null expectation. Piperacillin/tazobactam (TZP, P=5.7e-7)—a combination of beta-lactam and beta-lactamase inhibitor—strongly increased the risk of domination by oral-derived bacteria (Fig. 4D). In an independent cohort of pediatric allo-HCT recipients (Bekker et al., 2019), we confirmed the association by observing that children who have received oral TZP had much higher relative abundance of oral bacteria in feces than those who have received oral polymyxin/neomycin (Fig. S7). According to the *marker* hypothesis, the positive association of TZP reflects the strong anaerobe-killing capacity of this drug (Morjaria et al., 2019). Other than TZP, orally (P=0.049) but not intravenously (P=0.58) administered vancomycin is associated with intestinal domination of oral bacteria, likely because the latter fails to reach the gut. Interestingly, quinolones (P=6.1e-4) reduced the risk and thus preserved the total bacterial loads.

Graphical user interface

Description automatically generated

**Figure 4. Relative enrichment of oral bacteria in the feces of allo-HCT recipients.** (**A**)Overview of the gut microbiome states of all 10,433 fecal samples. The taxonomic composition (top), total fractions of oral bacteria (middle), and total bacterial loads (bottom) were aligned for each sample. (**B**) Antibiotic administration (top) and population dynamics of the relative (middle) and absolute (bottom) abundances (16S copies per gram feces) of oral and gut bacteria in the feces. Line and dots: mean; shading: 95% confidence interval (CI). (**C**) Negative association between proportions of oral bacteria in feces and total bacterial loads across 3,055 samples (dots) with loads no less than 1,000 16S copies per gram feces. Bar height: mean; red line: best linear fit; shading and bars: 95% CI. (**D**) Association between antibiotic exposure and intestinal domination of oral ASVs. Vertical line: hazard ratio; bar width: 95% CI. \*\*\*\*: P<0.0001; \*\*\*: P<0.001, \*: P<0.05. Chi-square test.

**Ecological mechanisms underlying the *marker* hypothesis**

To elucidate the mechanisms of the *marker* hypothesis, we developed a mathematical model by considering two competing forces that regulate the oral bacterial load in the intestine (Fig. 5A): antibiotics inhibit their growth but relieve them from the competitive inhibitions from the gut commensals (i.e., ecological release effect). The kinetic equations of the model are shown below

Eq. (1)

Eq. (2)

Here and are the oral and gut bacterial loads respectively, is the rate of oral-gut transmission, is the rate constant of bacterial loss, and are the carrying capacities, and and are the maximum growth rates. The physiological values of these parameters (Table S5) were taken from the literature ( and ), predicted by theoretical modeling ( and ), or estimated by fitting human data ( and ). The antibiotic susceptibility of the two populations was described by and (the higher , the less susceptible) as free parameters. See Supplementary Text and Table S6 for basic model assumptions, analytical solutions, and parameter estimations.

The *marker* hypothesis states that antibiotics caused relative enrichment, but not absolute expansion, of oral bacterial population in feces. By simulating the model (Eq. 1-2) under constant antibiotic exposure, we aimed to identify the parameter ranges of and that correspond to the regime of the *marker* hypothesis. We found that the oral bacterial fractions (relative abundance) and loads (absolute abundance) can increase, remain nearly unchanged, or decrease, depending on the values of and (Fig. S8). However, the regime of the *marker* hypothesis (i.e., increased relative abundance but reduced absolute abundance) is limited to a narrow region of the parameter space (Fig. 5B, blue shading), where the two populations need to be both susceptible and have similar antibiotic susceptibility (Fig. 5B). In this scenario, the antibiotic-mediated growth inhibition and ecological release are generally balanced for the oral bacterial population. Though not observed in the data analysis of this study, our simulations suggested the condition when antibiotics increased both relative and absolute abundances of oral bacterial populations in the gut (Fig. 5B, green). This condition requires that gut bacteria are more susceptible to antibiotics (), which can be met when the oral populations consist of multi-drug resistant bacteria.

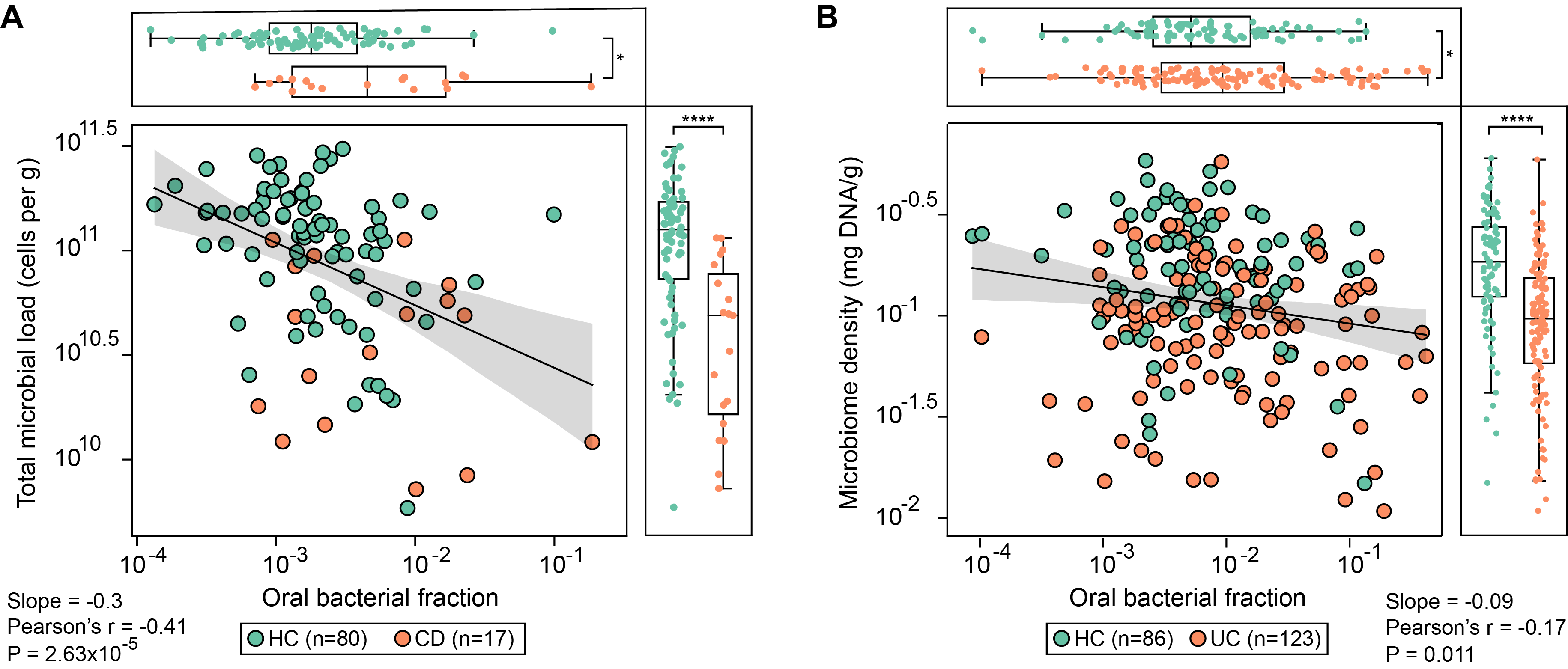
A picture containing chart

Description automatically generated

**Fig. 5| Mathematical modeling of oral bacterial colonization in the gut.** (**A**) Microbiome ecology of oral bacterial growth in the gut. The oral bacterial load (absolute abundance) is regulated by two mechanisms in opposite directions: antibiotics directly inhibit oral bacterial growth but indirectly promote their growth by releasing their inhibitions from the gut bacteria. and quantify the inhibitory levels of antibiotics on the oral and gut bacterial populations respectively. (**B**) Distinct response patterns of oral bacteria to antibiotics. The patterns were distinguished by qualitative changes ( , increase; , decrease; , no change) in the time-averaged relative abundances (rel. abun.) and absolute abundances (abs. abun.).

**The *marker* hypothesis unifies biomarkers of IBD**

Both the relative enrichment of oral-associated bacterial species (Read *et al.*, 2021) and low microbial loads (Contijoch et al., 2019; Vandeputte et al., 2017) in the gut have been identified as signatures of the gut microbiome in patients with IBD (both UC and CD). Reported by separate studies, the two signatures created a puzzle whether they are independent or related. In this study, we propose that the *marker* hypothesis can unify the two biomarkers by predicting a negative association between the two signatures. Indeed, the oral bacterial fractions and total bacterial loads in the feces of patients with CD (Pearson’s r = -0.41, P = 2.63e-5, Fig. 6A) and UC (Pearson’s r = -0.17, P = 0.011; Fig. 6B) were negatively correlated, and each signature is also significantly different between the patients and their healthy controls. 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 patients and controls. Since oral bacteria can be significantly enriched in IBD patients who have not taken antibiotics weeks prior to study (Fig. S9), antibiotics can amplify the gut microbiome dysbiosis in IBD patients (Gevers *et al.*, 2014) and strengthen the negative relationships between the two biomarkers.



**Figure 4. Negative association between oral bacterial fractions and total bacterial loads in the feces of patients with inflammatory bowel disease.** (**A**)Crohn’s disease (CD)(Vandeputte *et al.*, 2017)and (**B**)Ulcerative colitis (UC) (Contijoch *et al.*, 2019).Each circle represents a fecal sample. Lines: best linear fits; shading: 95% confidence intervals; HC: healthy control. \*\*\*\*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 microbial load, which has received much less attention, is another fundamental parameter that impacts host-microbiome interactions (Contijoch *et al.*, 2019; Tang et al., 2019; Zarrinpar et al., 2018). 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 microbial load do not strictly reflect changes in the biodiversity (Contijoch *et al.*, 2019), suggesting that the two metrics describe different aspects of dysbiosis.

Both cell- (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). Although these methods all have their own limitations, qPCR is conceptually superior for quantification of oral bacteria in feces due to its compatibility with 16S rRNA sequencing (Jian et al., 2021). The compatibility is found at three levels. First, both 16S sequencing and qPCR do not distinguish live from dead (including free DNAs) cells, but their combination measures absolute abundances of live, not total (live and dead), oral bacteria (see Supplementary Text for justification). Second, the technical biases introduced in the extraction, purification, and amplification steps of 16S sequencing are shared with qPCR, but not other approaches. Third, 16S qPCR measures bacterial loads (not microbial loads) and is again compatible with 16S sequencing. Therefore, we exclusively used 16S qPCR-based quantification unless the data type is unavailable (e.g., Fig. 6).

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 outcome 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 discrepancy between relative and absolute abundances: the relative enrichment of oral-derived bacteria in the gut does not reflect their active population expansion, 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 all be biomarkers of declined gut bacterial load (Duvallet et al., 2017). Despite low absolute amount of orally derived bacteria 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). The functional characterization (e.g., metabolomics) of the depleted gut microbiome state with high relative abundance of oral bacteria certainly warrants further study in the future.

**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 fruitful discussions.

**AUTHOR CONTRIBUTIONS**

Conceptualization, C.L., T.R. and J.B.X.; Mouse experiment: A.D.; Microbiome data processing: C.L. and H.L.; Microbiome data analysis, C.L.; Methodology, C.L.; Investigation, C.L. and T.R.; Writing – Original Draft, C.L., T.R., A.D., H.L.; Writing – Review and Editing, T.M.H., J.B.X., J.U.P., B.Z., L.D., 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** | **SOURCE** | **IDENTIFIER** |
| **Biological samples** | | |
| Fecal and oral samples from mouse | This study | NA |
| **Chemicals** | | |
| Ampicillin |  | Cat# |
| Vancomycin |  | Cat# |
| Neomycin |  | Cat# |
| **Deposited Data** | | |
| This study | SRA | PRJNA873058 |
| Reese *et al.* (2018) | SRA | PRJEB26446 |
| Staffas *et al.* (2018) | SRA | PRJEB24887 |
| Theis et al. (2020) | SRA | PRJNA594727 |
| The Human Microbiome Project Consortium (2012) | Qiita | 1928 |
| Imai *et al.* (2021) | SRA | PRJNA684508,  PRJNA684584 |
| Pascal et al. (2017) | SRA | PRJNA422193 |
| Liao *et al.* (2021); Yan *et al.* (2022) | SRA | PRJNA394877, PRJNA607574, PRJNA606262, PRJNA548153, PRJNA545312 |
| Bekker *et al.* (2019) | SRA | PRJEB28845 |
| Vandeputte *et al.* (2017) | SRA | PRJEB21504 |
| Contijoch *et al.* (2019) | SRA | PRJNA413199 |
| **Experimental Models: Organisms/Strains** | | |
| Female C57BL6/J Mice | The Jackson Laboratory | JAX Cat# 000664; RRID: IMSR\_JAX:000664 |
| **Oligonucleotides** | | |
| Primer: 16S\_Forward\_563F 5’-AYTGGGYDTAAAGNG-3’ | Taur et al. (2018) | NA |
| Primer: 16S\_Reverse\_926R 5’-CCGTCAATTYHTTTRAGT-3’ | Taur *et al.* (2018) | NA |
| qPCR: 16S\_Forward\_27F 5’-AGAGTTTGATCMTGGCTCAG-3’ |  | NA |
| qPCR: 16S\_Reverse\_338R 5’TGCTGCCTCCCGTAGGAGT-3’ |  | NA |
| **Databases** | | |
| SILVA v138 | Quast et al. (2012) | www.arb-silva.de |
| Virtual Metabolic Human (VMH) | Magnúsdóttir et al. (2017) | www.vmh.life |
| The Mouse Gut Microbial Bank (mGMB) | Liu et al. (2020) | NA |
| The Mouse Intestinal Bacterial Collection (miBC) | Lagkouvardos et al. (2016) | NA |
| Mouse Oral Microbiome Database (MOMD) | Joseph et al. (2021) | momd.org |
| **Softwares** | | |
| Python v3.7.9 | Python | python.org |
| Scikit learn v0.24.0 | Pedregosa et al. (2011) | scikit-learn.org |
| Lifelines v0.25.6 | Davidson-Pilon (2019) | github.com/CamDavidsonPilon/lifelines |
| iRep v1.10 | Brown et al. (2016) | github.com/christophertbrown/iRep |
| DADA2 v1.20 | Callahan et al. (2016) | benjjneb.github.io/data2 |
| Cutadapt v3.4 | Martin (2011) | github.com/marcelm/cutadapt |
| Bhatt lab workflow | Siranosian et al. (2022) | github.com/bhattlab/bhattlab\_workflows |
| MEGAHIT v1.2.9 | Li et al. (2015) | github.com/voutcn/megahit |
| MetaBAT 2 v2.12.1 | Kang et al. (2019) | bitbucket.org/berkeleylab/metabat/src/master/ |
| CONCOCT v1.1.0 | Alneberg et al. (2014) | github.com/BinPro/CONCOCT |
| DAS Tool v1.1.2 | Sieber et al. (2018) | github.com/cmks/DAS\_Tool |
| Kraken2 v2.1.1 | Wood et al. (2019) | github.com/DerrickWood/kraken2 |
| QIIME2 v2021.8 | Bolyen et al. (2019) | qiime2.org |

**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/marker_hypothesis>).

METHOD DETAILS

**Mouse experiment setup**

We treated C57BL/6J female mice with cocktail of ampicillin (0.5g/l), vancomycin (0.5g/l) and neomycin (1g/l) for one week in drinking water. Antibiotics were changed once during the course of the treatment. Animals were single-housed in autoclaved cages. Autoclaved water supplemented with antibiotics and 5053 irradiated food was provided *ad libitum*. Fecal pellets were collected immediately before and one week after the initiation of antibiotic treatment. Oral swabs were collected as per (Abusleme et al., 2017) before the treatment was started. Briefly, mice were hand-held while sterile swab was introduced into mouth and swiped for at least 30 seconds. After, the swab was put into 150ul of TE, the tip was cut off so that the eppendorf can be closed. Samples were put immediately to dry ice. One negative control swab was taken by pulling out the swab from the pouch and swirling through air for at least 30 sec, after which it was put in TE and on dry ice. Fecal samples and oral swabs were kept at -80ºC until further processing.

**DNA extraction and sequencing**

Fecal DNA was extracted and 16S rRNA gene was amplified using the previously described protocol (Taur *et al.*, 2018). Illumina TruSeq Sample Preparation protocol was used for generating libraries, that were later quantified, normalized, and sequenced using MiSeq Reagent Kit V3. Oral DNA was extracted by using modified DNeasy Blood and Tissue Kit protocol as described in (Abusleme et al., 2017). After extraction 16S rRNA gene was amplified and KAPA LTP Library Preparation Kit was used to generate sequencing libraries that were later quantified, normalized and sequenced using MiSeq Reagent Kit V3.

**Quantitative PCR (qPCR) for determining bacterial load**

For assessing the bacterial load in fecal and oral samples, qPCR against standard curve was used to determine 16S rRNA copy number. For this purpose, the PowerUP qPCR Kit was used. Briefly, for each sample, 20μl PCR triplicates were prepared with each containing 2μl of the DNA used as template, 10μl of mix provided by the manufacturer, and 1μl of forward and reverse primers at the final concentration of 0.5μM (F- AGAGTTTGATCMTGGCTCAG; R- TGCTGCCTCCCGTAGGAGT). In order to complete the volume of the reaction, 6μl of water was added. A PCR product of the 16S rRNA gene from *Enterococcus faecium* ATCC 700221 strain was used for obtaining a standard curve by amplifying its 16S rRNA gene and purifying the product. The copy number of the PCR product was determined based on its concentration and 16S rRNA sequence. A standard curve was obtained by using 10-fold dilutions.

Cycling conditions of the qPCR were 50ºC for 2 minutes, 95ºC for 2 minutes, and 40 cycles of 95ºC for 15 seconds, 56ºC for 15 seconds and 72ºC for 60 seconds. By extrapolating results by looking the ones obtained from standard curve samples, the number of 16S rRNA genes was determined for each sample. The final number of 16S rRNA genes per 1g of fecal sample was calculated by multiplying the number of 16S rRNA molecules obtained by qPCR with DNA elution volume after DNA extraction and dividing this number with the weight of the fecal pellet from which DNA extraction was performed.

**Inference of oral-typical bacterial sequences**

We constructed two reference sets of 16S rRNA sequences, separately for humans (V4-V5 region) and mice (full length), to identify bacteria typically colonizing the oral cavity. A 16S rRNA sequence detected in human or mouse gut microbiome can be classified to be oral-typical or oral-atypical by exactly matching the sequence to the corresponding reference set. The 16S rRNA regions typical of the human oral cavity were selected by filtering the microbiome profiles of the Human Microbiome Project at the ASV (V4-V5 region of 16S rRNA gene) level. An ASV is oral-typical if (1) its relative abundance averaged across all oral cavity samples (including samples from all oral cavity subsites, the same below) is greater than 1e-4; and (2) its relative abundance averaged across all fecal samples is no greater than 1e-4; and (3) its prevalence across all oral cavity samples is greater than 0.05; and (4) its prevalence across all fecal samples is no greater than 0.05. Prevalence of an ASV was computed as the proportion of samples that contain the ASV at a relative abundance above 1e-3 (Machado et al., 2021). Similar thresholds of relative abundance (1e-3 at the species level) and prevalence (0.05) were used in a previous study to identify oral-typical species (not ASVs) from metagenomic profiles (Thomas et al., 2019).

Our approach of constructing the mouse reference set is visualized in Fig. S2. Due to a lack of large-scale microbiome datasets with paired oral and gut samples, we used the Mouse Oral Microbiome Database (MOMD) (Joseph *et al.*, 2021) as the starting point for gathering oral-typical sequences. MOMD has 164 full length rRNA sequences from cultured isolates. By reciprocal blast search (-perc\_identity 99 -qcov\_hsp\_perc 100 -ungapped) between MOMD and a combined collection of 344 cultured gut bacterial 16S rRNA sequences from the Mouse Gut Microbial Biobank (mGMB) (Liu *et al.*, 2020) and the Mouse Intestinal Bacterial Collection (miBC) (Lagkouvardos *et al.*, 2016), we found that 38 sequences are likely from common gut bacteria as well. Meanwhile, we profiled the paired oral (swab) and gut (distal colon) samples from a small set (n=11) of pregnant mice in a public study (Theis *et al.*, 2020) and identified 17 oral-typical ASVs by following conditions (1) and (2) defined for humans above. Only 5 of 17 ASVs were found in MOMD; by searching the nucleotide database in NCBI (National Center for Biotechnology Information), we obtained 29 full length mouse 16S rRNA sequences that cover 6 of the 12 missed oral-typical ASVs in full. Given these preliminary analyses, a filtered (and expanded) MOMD to be used for inference was constructed in three steps. First, 37 sequences that belong to either the set of 39 potential common gut bacterial sequences or the taxa that contains at least 3 sequences in the set, but not matched to the 17 oral-typical ASVs, were removed. Next, the 29 sequences matching the 17 oral-typical ASVs from the coverage test were added. Finally, 7 redundant sequences were found and removed if there exist longer sequences that fully contain them as parts.

**16S amplicon sequencing**

The >10,000 microbiome samples from the adult allo-HCT recipients were previously analyzed by an in-house processing pipeline and the ASV profiles were available in a recent compilation study (Liao et al., 2021). Briefly, reads were trimmed to the first 180 bp or the first point with a quality score Q<2, and removed if they contained ambiguous nucleotides (N) or if two or more errors were expected based on the quality of the trimmed reads. ASVs were identified using DADA2 (Divisive Amplicon Denoising Algorithm) (Callahan *et al.*, 2016) and classified by IDTaxa (Murali et al., 2018) and the SILVA v138 database (Quast *et al.*, 2012).

The demultiplexed and primer-trimmed HMP 16S sequences were downloaded from the Qitta repository (Gonzalez et al., 2018) and processed by QIIME (Quantitative Insights Into Microbial Ecology) 2 (Bolyen *et al.*, 2019) . DADA2 was used to denoise data and generate an ASV per sample counts table, using the QIIME denoise-pyro plugin (Bolyen *et al.*, 2019). Parameter --p-trunc-len 395 was used to remove low quality tails. Taxonomy classification of the representative ASV sequences was performed using the QIIME plugin “feature-classifier” (Bokulich et al., 2018) and the SILVA v138 database (Quast *et al.*, 2012). The classification took three steps. We first extracted the V3-V5 region of the SILVA reference sequences using the extract-reads method. Then we created a classifier by using the fit-classifier-naïve-bayes method with extracted reads and the SILVA reference taxonomy. Finally, we ran the classifier on the ASV sequences using the classify-sklearn method to get their taxonomy.

All other microbiome datasets used in this study were similarly processed using QIIME 2 (Bolyen *et al.*, 2019). When needed, primers were removed from demultiplexed short reads using the QIIME cutadapt plugin (Martin, 2011) with parameters “--p-error-rate 0.1” and “--p-overlap 3”. The trimmed reads were denoised using the QIIME dada2 plugin with truncation lengths determined by per-base quality scores to generate feature tables at the ASV level. Taxonomic classification was performed using the QIIME plugin “feature-classifier classify-sklearn” (Pedregosa *et al.*, 2011) against the SILVA 138 database (Quast *et al.*, 2012) at a cutoff of 80% and inferred to the lowest possible taxonomic level.

For each sample, we only kept bacterial ASVs and removed those whose taxonomy contains keywords “Chloroplast” or “Mitochondria”. We further discarded samples whose total sequencing depths are smaller than 1,000 reads.

**Shotgun metagenomic sequencing**

We adapted a recently published pipeline (Siranosian *et al.*, 2022) 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 (Brown *et al.*, 2016). The iRep value of a MAG represents the average number of replication events over different subpopulations of the MAG weighted by their relative abundances.

**Co-occurrence analysis**

The algorithm for computing co-occurring bacterial communities 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. 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.

**Cox’s proportional hazard model**

We used the Cox’s time-varying proportional hazard model to regress fecal domination (relative abundance > 30%) of any oral bacterial ASV as a microbial endpoint of interest against antibiotic administration as multivariable predictors. The exposure to each antibiotic takes a value of 1 on the day the antibiotic was administered and 0 otherwise. The time interval of allo-HCT recipients starts from 10 days prior to transplantation and ends by 40 days post-transplantation. Patients with less than 5 samples during the time interval were excluded. Rarely administered antibiotics that have been administered less than 10 times were also excluded.

**Quantification and statistical analysis**

The oral and gut bacterial loads in the intestine were computed by multiplying total bacterial loads (qPCR, flow cytometry, DNA mass) with relative abundance of oral-typical and oral-atypical ASVs respectively. By interpreting oral-atypical bacteria as gut bacterial members, we ignored bacterial transmission from other non-oral body sites to the gut. All statistical analysis, including linear regression, correlation between variables (Pearson’s and Spearman’s correlation), quantitation of microbiome similarity (Hamming and Bray-Curtis distances), and hypothesis testing (Kruskal-Wallis test and Welch’s t-test) were performed using python.

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.

Abusleme, L., Hong, B.-Y., Hoare, A., Konkel, J.E., Diaz, P.I., and Moutsopoulos, N.M. (2017). Oral microbiome characterization in murine models. Bio-protocol *7*, e2655-e2655.

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. Nature methods *11*, 1144-1146.

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.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., and Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2’s q2-feature-classifier plugin. Microbiome *6*, 1-17.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., and Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology *37*, 852-857.

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.

Brown, C.T., Olm, M.R., Thomas, B.C., and Banfield, J.F. (2016). Measurement of bacterial replication rates in microbial communities. Nature biotechnology *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. Nature methods *13*, 581-583.

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.

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

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.

Gonzalez, A., Navas-Molina, J.A., Kosciolek, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., DeReus, J., Janssen, S., Swafford, A.D., and Orchanian, S.B. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. Nature methods *15*, 796-798.

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.

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., 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.

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., Peled, J.U., Taur, Y., and Perales, M.-A. (2021). Compilation of longitudinal microbiota data and hospitalome from hematopoietic cell transplantation patients. Scientific data *8*, 1-12.

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.

Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D.A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., and Wilmes, P. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nature biotechnology *35*, 81-89.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal *17*, 10-12.

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.

Murali, A., Bhargava, A., and Wright, E.S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome *6*, 1-14.

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., Martinez, X., Varela, E., Sarrabayrouse, G., and Machiels, K. (2017). A microbial signature for Crohn's disease. Gut *66*, 813-822.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research *12*, 2825-2830.

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.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research *41*, D590-D596.

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.

Rolling, T., Zhai, B., Gjonbalaj, M., Tosini, N., Yasuma-Mitobe, K., Fontana, E., Amoretti, L.A., Wright, R.J., Ponce, D.M., and Perales, M.A. (2021). Haematopoietic cell transplantation outcomes are linked to intestinal mycobiota dynamics and an expansion of Candida parapsilosis complex species. Nature microbiology *6*, 1505-1515.

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.

Sieber, C.M., 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. Nature microbiology *3*, 836-843.

Siranosian, B.A., Brooks, E.F., Andermann, T., Rezvani, A.R., Banaei, N., Tang, H., and Bhatt, A.S. (2022). Rare transmission of commensal and pathogenic bacteria in the gut microbiome of hospitalized adults. Nature communications *13*, 1-17.

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.

Stein-Thoeringer, C., Nichols, K., Lazrak, A., Docampo, M., Slingerland, A., Slingerland, J., Clurman, A., Armijo, G., Gomes, A., and Shono, Y. (2019). Lactose drives Enterococcus expansion to promote graft-versus-host disease. Science *366*, 1143-1149.

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., Coyte, K., Schluter, J., Robilotti, E., Figueroa, C., Gjonbalaj, M., Littmann, E.R., Ling, L., Miller, L., and Gyaltshen, Y. (2018). Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. Science translational medicine *10*, eaap9489.

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.

Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome biology *20*, 1-13.

Yan, J., Liao, C., Taylor, B.P., Fontana, E., Amoretti, L.A., Wright, R.J., Littmann, E.R., Dai, A., Waters, N., and Peled, J.U. (2022). A compilation of fecal microbiome shotgun metagenomics from hematopoietic cell transplantation patients. Scientific Data *9*, 1-11.

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.

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