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

Chen Liao1,8, Thierry Rolling2,3,4,8, 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 competitive 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**

The healthy persons harbor distinct microbiome communities at different body sites (Costello, Lauber 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, Curtis et al. 2021) and colorectal cancer (CRC) (Komiya, Shimomura et al. 2019). Compared to the hematogenous route (oral-blood axis) (Abed, Maalouf et al. 2020), the enteral route (oral-gut axis) is a natural and more studied pathway of oral bacterial dissemination (Kitamoto, Nagao-Kitamoto et al. 2020, Jin, Wetzel et al. 2022). For an average person, ~1011 oral-resident bacterial cells are transported to stomach via saliva, food and liquid daily (Segata, Haake 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, Wetzel 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%) (Schmidt, Hayward et al. 2019, Rashidi, Ebadi et al. 2021).

Broad spectrum antibiotics such as ampicillin can weaken the oral-gut barrier by depleting gut anaerobes and promote colonization of translocated oral bacteria (Atarashi, Suda et al. 2017). Other than antibiotics, diets, aging, and gut inflammation may also disrupt gut colonization resistance (Kitamoto, Nagao-Kitamoto et al. 2020). All these factors jointly contribute to the clinical associations between ectopic colonization of oral pathobionts and a variety of digestive diseases (e.g., Crohn’s disease (Gevers, Kugathasan et al. 2014), ulcerative colitis (Schirmer, Denson et al. 2018), CRC (Kostic, Chun et al. 2013), liver cirrhosis (Qin, Yang 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 (e.g., trade-off between *Bacteroides* and *Prevotella*) (Vandeputte, Kathagen et al. 2017).

By definition, the relative enrichment of oral bacteria following perturbations can be explained by two alternative 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. Though largely ignored, bacterial load is itself an important variable in microbiome-disease associations and influences host physiology, metabolism, and immunity (Zarrinpar, Chaix et al. 2018, Contijoch, Britton et al. 2019, Tang, Chen et al. 2019). In the following, we will distinguish 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 alternative 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 reduction in the gut bacterial density.

**RESULTS**

**Classification of oral bacteria in the gut microbiome of mice and humans**

Accurate detection of bacterial translocation events requires identification of the same strain-level features (e.g., single nucleotide polymorphisms) in samples collected from multiple body sites (Schmidt, Hayward et al. 2019). However, paired oral and gut microbiome samples are usually unavailable in previous studies (Jin, Wetzel et al. 2022). Moreover, most of the quantitative human gut microbiome studies to date (Vandeputte, Kathagen et al. 2017, Contijoch, Britton et al. 2019, Vieira-Silva, Sabino et al. 2019, Jian, Luukkonen et al. 2020, Schluter, Peled et al. 2020, Rao, Coyte et al. 2021) have focused on profiling the gut microbiome using 16S amplicon sequencing. To reuse these published datasets, we developed reference 16S rRNA sequence databases of oral-typical bacteria for both mice and humans. In this context, the major indicators of bacterial translocation are amplicon sequence variants (ASVs) that are typical for the oral cavity but atypical for the gut.

Since the paired oral-gut samples in mice are extremely rare (Greenberg, Romero et al. 2022), the Mouse Oral Microbiome Database (MOMD) (Joseph, Aduse-Opoku et al. 2021) was leveraged as the starting point for gathering oral-typical sequences (Fig. S1a). A similar approach using the Human Oral Microbiome Database (HOMD) has been previously employed to classify oral bacteria (Coker, Dai et al. 2018, Hu, Amir et al. 2022). We then filtered MOMD by removing sequences closely matching to those in two mouse gut bacterial collections (Lagkouvardos, Pukall et al. 2016, Liu, Zhou et al. 2020) and add oral-typical sequences absent in the MOMD from a small but rare set (n=11) of paired oral-gut microbiome samples (Theis, Romero et al. 2020); see STAR Methods. The resulted filtered-MOMD database has 149 sequences from diverse genera consisting of mainly *Staphylococcus* and *Streptococcus* (Fig. S1b). We show that the filtered-MOMD, but not the original MOMD, provides a conservative estimate of the total fraction of oral ASVs in the distal intestine (Fig. S1c): the estimated oral bacterial fractions remain nearly the same for 10 out of 11 mice after excluding those absent from the paired oral cavity samples (Fig. S1c).

To classify oral bacteria in human gut microbiomes, we leveraged the Human Microbiome Project (HMP) dataset that provides large-scale surveys of paired gut and oral microbiome samples from XXX healthy volunteers (The Human Microbiome Project Consortium 2012). Distinct bacterial communities were found inhabiting the human gut and oral cavity (Fig. 2a), suggesting that the majority of bacterial ASVs do not colonize the two habitats equally abundant. Indeed, the ASVs highly abundant in one body site are generally not representative of the other site (Fig. 2b). To be quantitative, we adopted a similar strategy used before (Thomas, Manghi et al. 2019) and defined “oral-typical bacterial ASVs” as those whose mean relative abundance > 10-4 and prevalence > 0.05 across all oral samples, and whose mean relative abundance and prevalence must not exceed the same cutoff thresholds across all fecal samples (Fig. 2b,c). Altering the cutoff thresholds changes the number of oral-typical ASVs by at most 2%. We compared the filtered HMP to HOMD and showed that the former is more conservative and predicts oral bacterial fraction that are all present in the paired oral cavities in the HMP dataset. The resulted filtered-HMP database has 178 sequences from diverse genera consisting of mainly *Staphylococcus* and *Streptococcus* (Fig. S1b). We further validated the conservancy of our approach using IBD cohort with paired oral-gut samples using the HMP dataset (Fig. 2d) and an independent IBD cohort (Fig. 2e). Notably, the filtered-HMP database covers the V4-V5 region, as opposed to the filtered-MOMD which contains full length 16S rRNAs.

**Total load of oral bacterial populations remains stable in mouse feces**

Having established the quantitative methods, we sought to test the expansion vs. the marker hypothesis in mice. We first reanalyzed antibiotic-treated microbiome dysbiosis from a public study. In the study, mice were treated with an antibiotic cocktail AVNM (Ampicillin-Vancomycin-Neomycin-Metronidazole) for 5 days and their fecal samples were collected daily for 2 weeks. We found that the fraction of oral populations in the intestine increases at day 2 after antibiotics, showing relative enrichment of oral bacteria. The enrichment was driven by three major ASVs belong to the genera, Escherihci coi, Lactobacillus. More importantly, the mean oral bacterial loads increase slightly at day 2 but remains generally stable over the course while that of the gut bacterial loads decreases substantially. This suggests the marker hypothesis. When the data were plotted on the log-log space, we found a remarkable linear relationship with slope -0.70. This corresponds to the regime of antibiotics that inhibit stronger on gut bacteria than oral bacteria. The outliers in the plot (see red box in ) are not similar in sequences than any of the filtered MMOMD database and they mostly come from day 2-6. That explains the dip and fluctuations in Fig. 2B and these further suggest the low coverage of oral bacteria. A similar slope was found in a allogenetic stem cell tranpslatnion model that also received antibiotic treatment but one antibiotic (not cocktail) per mouse experiment.

The reanalysis of Dr. Reese’s and Anna;s experiments were limited by the lack of paired-oral samples and we thus cannot prove that the predicted oral sequences indeed come from the oral cavity. We therefore conducted a specific experiment that we treated mice with AVN for 7 days and collected fecal sample at day 0 and 8, and oral sample at day 0. The microbiome composition at day 8 are more similar to the oral sample at day 0, than the fecal sample at day 0 in tersm fo the shared ASVs (quantified by hamming distance). This finding is conserved on 2/3 mice by taking relative abundance into accounts (quantified by bray-curtis distance). More importantly, we found about 10% of ASVs only come from the oral cavity (not present in the d0 feces at all) and they came from ASVs of strpetococus, stapyhylcso, ananeor and peptococcus. Despite the few darta point,s we computed the slope of -1.23 and confirmed this value using oral-typical ASVs from the paired end samples (not using HOMD). This suggests that AVN in this experiment has stronger killing effects than AVNM in the Reese’s experiment. This might be attributed to longer antibiotic treatment or stronger doses used, in addition to other potential factors such as antibiotic brand.

**Theoretical population ecology model**

The human gut lumen is an ecosystem with rhythmic food intake and fecal excretion (Goyal et al., 2018). Assuming the transient ecological dynamics induced by these cyclic events have short time scales, we approximated its long-term dynamics and quasi-steady state by a continuous, coarse-grained model of oral-derived (]) and gut-resident (]) bacteria (**Fig. 1A**)

Eq. (1)

Eq. (2)

Here is the oral-gut transmission rate due to saliva, food, and liquid intake, is the fecal excretion rate constant, and are the host carrying capacities, and and are the maximum growth rates as a proxy of population fitness. Our model focuses on antibiotic perturbations—the major disruptor of gut microbiome (Ramirez et al., 2020)—and represents the impacts of antibiotics on the fitness of both types of bacteria.

The model makes two predictions on the fraction of oral bacteria in feces () at steady state. First, its value changes abruptly from low (nearly absent) to high (very abundant) when the perturbation exceeds a threshold (**Fig. 1B**, red line). The ultrasensitive response to the perturbation strength suggests that a unimodal distribution of the perturbation effects can generate a bimodal distribution in the relative oral bacteria abundance (**Fig. 1C**). The mathematical result indicates that humans with perturbed gut bacteria may constitute two subpopulations with very low or very high relative abundance of oral bacteria in their feces. We find evidence for this model prediction in the fecal microbiota of healthy Western adults, which show bimodal distributions of *Prevotella* species commonly found in the oral cavity (Lahti et al., 2014; Tett et al., 2021). We further derived that is inversely proportional to the total bacterial density () in the gut lumen regardless of antibiotic-induced perturbations

Eq. (3)

This implies a negative linear relationship in the log-log space between the fraction of oral bacteria and the total bacterial density in the gut, irrespective of the perturbations (**Fig. 1D**).

**Quantifying oral bacteria in feces**

Quantifying the fraction of oral bacteria in feces is a microbial source tracking problem (Knights et al., 2011; Shenhav et al., 2019). Paired oral and gut microbiome samples are generally unavailable for most literature studies. To circumvent this limitation, we developed a new computational approach that mathematically models any bacterial population—including those in the fecal and oral samples—as a mixture of stereotypical bacterial communities inhabiting different body sites of a typical average person. We identified five habitats from Human Microbiome Project (HMP) (Human Microbiome Project Consortium, 2012): All 4,587 HMP samples collected from five major body sites (GI tract, oral cavity, nasal cavity, skin, and urogenital tract) of 235 healthy individuals clustered into five habitats of similar compositions that match roughly, though not exactly, their collection sites (**Fig. 2A**). The nasal cavity and skin clustered together, and the oral cavity formed two clusters, each corresponding to different oral cavity subsites (Lamont et al., 2018; Segata et al., 2012).

We extracted the typical composition of each habitat from the HMP samples using non-negative matrix factorization (Lee and Seung, 1999) (**Fig. 2B**, upper panel; **Fig. S1A,B**). This revealed that each habitat had one or a few dominant bacterial genera (**Fig. 2C**). For instance, *Streptococcus* (54%) and *Prevotella* (20%) were the most abundant genera in the oral cavity habitat 1 (e.g., saliva) and 2 (supra-/sub-gingival plaques) respectively. The unmixing step also quantified the proportions of oral bacteria in HMP samples. We compared our approach to two alternative methods based on Random Forest classifier (RFC) (Ho, 1995) and SourceTracker 2 (Knights et al., 2011), both showing Pearson’s correlation coefficients > 0.6 (**Fig. S1C**). The predicted fractions confirmed the general notion that oral bacteria are rarely abundant in the feces of healthy people (Costello et al., 2009; Human Microbiome Project Consortium, 2012) (**Fig. 2D**). Although the majority of HMP samples are composed of bacteria typically inhabiting their collection body sites, 37 (11.5%) fecal samples have 10 % oral genera (**Fig. 2D**, violet box) such as *Veillonella* (**Fig. 2E**, upper panel) and *Prevotella* (**Fig. 2E**, lower panel). This finding supports the purpose of unmixing: even healthy people can have significant proportions of oral bacteria in their feces.

**Validation of model predictions**

The typical composition of the five habitats allowed us to estimate the proportion of oral bacteria in any fecal microbiome sample—without the need of paired oral sample—by using a simple constrained regression (**Fig. 2B**, lower panel). We applied this method to a large-scale dataset containing >10,000 fecal microbiome samples from >1,000 patients (Liao et al., 2021). All patients in this dataset (MSK cohort) had hematological diseases including leukemias, lymphomas, and myeloma and underwent allo-HCT at Memorial Sloan-Kettering Cancer Center. We validated the predicted oral bacterial fractions using RFC (**Fig. S2A**).

In most allo-HCT patients, the initial conditioning regimens (e.g., chemo- or radio-therapy) and antibacterial prophylaxis caused marked loss of bacterial diversity and compositional changes that are not observed in healthy individuals (**Fig. S2B**). In this cohort, only 55% of the fecal samples contained gut-resident bacteria as their largest fraction. The remaining 45% of samples were dominated by ectopic bacteria, including 18% of samples that were highly abundant in oral bacterial genera such as *Streptococcus* and *Actinomyces* (**Fig. 3A**). By co-occurrence analysis, we found 32 distinct combinations of oral bacterial ASVs (amplicon sequence variants) that co-occurred in feces more than expected by chance (**Table S1**), supporting that the fecal enrichment of oral bacteria in these allo-HCT patients is driven by their collective transmission from oral to gut.

The relative abundance of oral bacteria in these fecal samples was bimodally distributed (**Fig. 3B**), which is consistent with our model prediction (**Fig. 1C**). The distribution was heavily skewed toward 0, but 383 samples from 182 unique patients were composed of 95% oral bacteria. Even single bacterial ASVs (**Fig. S2C**) could dominate these samples by ≥30%, with dominations of *Streptococcus* (ASV8 and ASV16) among the most frequent domination states. Shotgun-sequenced samples showed that the two ASVs contained a variety of *Streptococcus* species, mostly *S. thermophilus*. Some of these *Streptococci* could be alive and replicating in the fecal sample, as evidenced by peak-to-trough ratio (i.e., the ratio of reads mapped to oriC to terminus) calculated by iRep (Brown et al., 2016) analysis (**Table S2**).

To validate the proposed quantitative relationship (**Fig. 1D**), we measured the total bacterial density in the feces (16S rRNA copies per gram of feces) by qPCR (Liao et al., 2021). The timing of antibacterial prophylaxis (**Fig. 3C**, lower panel) corresponded to the period in which gut-resident bacteria declined in both relative (upper panel) and absolute (middle panel) abundances. Consistent with our model prediction, the mean absolute abundance of oral bacteria (both habitat 1 and 2) remained constant despite antibiotic exposure, explaining their increased relative abundances (**Fig. 3C**). As a result, the proportion of bacteria from oral habitats 1 and 2 had the strongest negative correlations with the total bacterial density (Slope=-0.34 and -0.39, *P*<0.001) among all five habitats (**Fig. S3A-E**). When combined, the two oral populations showed an even stronger association (Slope=-0.49, *P*<0.001; **Fig. 3D**, bottom row). We determined 3.4% as the threshold percentage of oral bacteria in feces above which the linear fitting is nearly optimal (middle row) and the best-fit slope matches the predicted value -1 (top row). This threshold delineates the dynamic range of our model: the negative association becomes less apparent or even obscured by noise at oral bacteria fraction below this threshold.

**Associations with stool consistency, antibiotic exposure and allo-HCT patient outcomes**

Stool consistency was looser in allo-HCT patients with higher relative abundance of oral bacteria (**Fig. 3E**). We further identified two antibacterial antibiotics that were significantly associated with an increased fraction of oral bacteria in fecal samples above 30% (Taur et al., 2012): oral vancomycin and piperacillin/tazobactam (**Fig. 3F**). Intravenous vancomycin showed no association, likely because it fails to reach the GI tract. Piperacillin/tazobactam, a combination of beta-lactam and beta-lactamase inhibitor that can reach the GI tract and have a broad spectrum against many gut bacteria (Morjaria et al., 2019), had the strongest association. 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) (**Fig. S4**).

Regarding hematopoietic cell transplant outcomes, the allo-HCT patients with high oral bacteria during the pre-engraftment period had lower overall survival compared to those whose microbiome was mainly composed of gut-resident bacteria (**Fig. 3G**). Although the mechanism underlying this association is unclear, existing evidence suggest against bloodstream infections (BSIs) by ectopic oral bacteria: Oral-typical bacteria only caused 5 infection cases (all *Streptococcus spp.*) in 5 patients during transplantation and none were preceded by intestinal *Streptococcus* domination (**Table S3**). The loss of gut bacteria may actually preclude microbiota dominations by true pathogens—such as *Enterococcus*—and reduce the risk for the development of BSI caused by gut-to-blood transmission (Tamburini et al., 2018) (**Fig. 3H**). Unlike oral bacteria domination states, *Enterococcus* dominations represent true blooms since a high relative abundance positively correlates with a high total bacterial density (**Fig. S3F**).

Finally, the fecal samples with higher proportion of oral bacteria tended to have higher fungal densities (**Fig. 3I**, **S3G**), suggesting that fungi could expand to fill the niche left open by the loss of gut bacteria. The gut colonization and overgrowth of fungal pathogens such as *Candida parapsilosis* typically occurs when gut bacteria are depleted, consistent with the finding that gut-resident bacteria can protect against fungal blooms and infection (Rolling et al., 2021; Zhai et al., 2020).

**Generalization to other patient cohorts**

Supporting the *marker* hypothesis, another key characteristic of IBD besides the relative enrichment of oral bacteria is the reduced density of the gut bacterial population. This observation suggests that the apparent rise in relative abundance of oral bacteria in human feces may simply indicate a loss of the gut bacterial density.

We next validated the quantitative relationship between oral bacteria in feces and gut bacterial density for three GI diseases (Crohn’s disease (CD), Ulcerative colitis (UC), and *Clostridioides difficile* infection (CDI)) using the Mount Sinai cohort (Contijoch et al., 2019) and the LUMC cohort (Bekker et al., 2019) which had measurements of microbiome density. Previous studies reported, separately, relative expansion of oral bacteria in the gut microbiome and lower microbiome density in both CD and UC patients (Contijoch et al., 2019; Read et al., 2021; Vandeputte et al., 2017). Our study links the two signatures by showing a negative interdependence between each other (slope=-0.26, *P*<0.001 for the Mount Sinai cohort; slope=-0.25, *P*=0.04 for the LUMC cohort) using data with oral bacteria 3.4% (**Fig. 4A**,**B**). Oral bacteria from habitat 1 but not habitat 2 significantly contributed to the correlation (**Fig. S5**). Notably, the negative association remained significant (slope=-0.22, *P*=0.018) after excluding patients with prior antibiotics or an unclear antibiotic history in the Mount Sinai cohort (**Fig. S6A**), indicating the relationship is independent of antibiotic exposure.

We also assessed the generality of the associations of oral bacterial fraction in feces with disease status, antibiotic exposure, and gut fungal growth in patients with the three GI diseases. Re-analyzing multiple case-control cohort studies confirmed known associations between the fecal proportion of oral cavity bacteria and disease status, including CD (*P*=0.029) and UC (*P*=0.037) patients in the PRISM, LifeLines DEEP and NLIBD cohorts (Franzosa et al., 2019) (**Fig. 4C**), CDI patients in the UMH cohort (Schubert et al., 2014) (*P*<0.001; **Fig. 4D**) and pediatric CD patients in the RISK cohort (Gevers et al., 2014) (*P*<0.045; **Fig. S7A**). Notably, these associations were much stronger for bacteria from oral cavity habitat 1 than habitat 2 (**Fig. S8**). More interestingly, all associations for subsets of (pediatric) IBD patients who have not previously received antibiotics became insignificant (**Fig. S6B**), suggesting that the relative enrichment of oral bacteria in feces is not a signature of IBD but, instead, reflects collateral damage of antibiotic treatment. Finally, the percentage of oral bacteria increased with disease severity in pediatric UC patients from the PROTECT cohort (Schirmer et al., 2018) (*P*<0.001; **Fig. 4E**). Although similar associations with disease severity were not found in the pediatric CD patients enrolled in the UPenn cohort (**Fig. S7B**), those patients with enriched oral bacteria had a significantly higher percentage of fungal DNA in stool samples (**Fig. 4F**).

**DISCUSSION**

A general tenet of theoretical ecology holds that population density impacts ecosystem function as much as the population composition. Yet almost all studies of human gut microbiome identify the associations of ectopic oral bacteria in the gut with specific diseases based on bacterial compositional changes (Kitamoto et al., 2020a; Read et al., 2021; Schmidt et al., 2019) and neglect variations in population density (Rao et al., 2021; Vandeputte et al., 2017). For example, >10 different oral bacterial species (e.g., *Streptococcus spp.*, *Veillonella parvula*, *Fusobacterium nucleatum*) were reportedly enriched in the intestine of IBD patients by 12 studies, where 10 drew conclusions based on bacterial relative abundances estimated from 16S amplicon sequencing (Read et al., 2021). Mouse experiments show the possibility that the enriched oral bacteria in feces may be oral pathobionts (Atarashi et al., 2017; Kitamoto et al., 2020b), the driver hypothesis, but this remains to be demonstrated in humans. Here we favor an alternative interpretation, the marker hypothesis: oral bacteria in feces neither correspond to absolute expansions of oral pathogens in the gut nor do they reflect species- or disease-specific biological differences. Rather, they simply indicate a shared decline of gut bacterial density (Duvallet et al., 2017).

**Figure 1.** **Ecological modeling links relative abundance of oral bacteria to total bacterial density in fecal samples.** (**A**)The marker hypothesis and its ecological model. The human gut lumen has a constant inflow of oral bacteria (at rate ) and a first-order outflow of bacteria of both oral and gut origins (at rate constant ). Perturbations such as antibiotics, which impact the fitness of bacteria in the gut, can reduce total bacterial density and relatively increase the abundance of oral bacteria detected in human feces. (**B**) Predicted response of the fecal proportion of oral bacteria to the perturbation level that impacts bacterial fitness. The compositional switch is more sensitive when is lower. (**C**) Bimodal distribution of the relative abundance of oral bacteria in feces results from its ultrasensitive response to a unimodal distribution of the perturbed fitness (). P.d.f.: probability density function. (**D**) Predicted linear relationship in log-log space between relative abundance of oral bacteria and total bacterial density in feces.

**Figure 2. Bacterial populations collected from different body sites of healthy people reflect five distinct habitats within the human microbiome.** (**A**) t-distributed stochastic neighbor embedding (tSNE) (van der Maaten and Hinton, 2008) of HMP data clusters according to five distinct habitats. (**B**)Schematic diagram of our computational approach for quantifying fraction of oral bacteria in microbiome samples. Our approach has two steps: (1) extracting the typical compositions of the five habitats () by decomposing the HMP dataset () and (2) estimating bacterial proportions originating from the five habitats () for any new microbiome samples () under fixed . (**C**)Relative abundance of the top 5 most abundant genera (signature taxa) in the five habitats. (**D**) Compositions of HMP samples represented by proportions of bacteria originating from the five habitats (represented by black & white color gradient). Each row is a bacterial habitat and each thin column is a sample. The body sites where samples were collected are indicated by colors at the top. (**E**)Two representative stool samples with 10% *Veillonella* (upper panel) or *Prevotella* (lower panel). The distributions of relative abundance of the two genera across HMP samples collected from the GI tract and oral cavity are shown to the right of each panel.

**Figure 3. Oral bacteria in feces of MSK allo-HCT patients are associated with bacterial density, stool consistency, antibiotic use, and transplant outcomes.** (**A**)Relative abundances of bacteria represented by their habitat origins (first row) and taxonomy (second row). Each thin column is a sample and all samples are clustered by their most abundant sources of origin (labeled above each cluster). (**B**) Bimodal distribution of oral bacteria in feces.(**C**) Dynamics of relative or absolute abundances of bacterial populations from the gastrointestinal tract and the two oral cavity habitats. (**D**) Negative relationship between fraction of oral bacteria in feces and total bacterial density. Red and blue lines (associated shading area: 95% confidence interval) in the scatter plot represent the best linear fits using data within the dynamic range (oral bacteria 3.4%, orange shading) and the entire range of data respectively. R2 and the fitted slopes were determined stepwisely by varying the range of data included in the linear regression. (**E**) Fraction of oral bacteria in feces stratified by stool consistency. (**F**)Hazard ratio and 95% confidence intervals of antibacterial antibiotic predictors for gut microbiome domination by oral bacteria (30%). (**G**)Overall survival curves of allo-HCT patients stratified by the dominant bacterial sources in their gut microbiome pre-(bone marrow) engraftment. The two oral cavity habitats are combined. *P*-value: Wald test. (**H**)Hazard ratio and 95% confidence interval of oral bacteria domination (30%) as a predictor for *Enterococcus* bloodstream infection. (**I**)Fraction of oral bacteria in feces stratified by fungal cultivability. cfu: colony-forming unit. In panels (E) and (I), the numbers represent the percentages of data above the thresholds indicated by dashed lines. \*\*\**P*<0.001; Kruskal-Wallis test. ns, not significant.

**Figure 4. Oral bacteria in feces indicates gut microbiome density, disease status and severity, and intestinal fungal growth in patients with gastrointestinal diseases.** (**A**,**B**) Quantitative relationship between proportion of oral bacteria and microbiome density (as a proxy of bacterial density) in the Mount Sinai cohort (A) and the LUMC cohort (B). Red and blue lines and associated shading areas represent best linear fits and 95% confidence intervals using data within the gray shaded area (oral bacteria 3.4%) and the entire range respectively. Abbreviations: healthy control (HC); Crohn’s disease (CD); ulcerative colitis (UC); *Clostridioides difficile* infection (CDI); recurrent *Clostridioides difficile* infection (rCDI); ileal pouch-anal anastomosis (IPAA); fecal microbiome transplantation (FMT); primary sclerosing cholangitis (PSC). (**C**-**F**) Fractions of oral bacteria in feces are associated with disease status (C,D), severity (E) and fungal DNA fraction (F). Each dot represents a sample. Median and interquartile range are indicated by colored circles and bars respectively. The numbers above dashed lines indicate the percentage of samples with 3.4% oral bacteria. Data sources: CD and UC (the PRISM, LifeLinesDEEP and NLIBD cohorts), CDI (the UMH cohort), pediatric UC (the PROTECT cohort), and pediatric CD (the UPenn cohort). \*\*\**P*<0.001; \**P*<0.05; Kruskal-Wallis test.

**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.).

**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

**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**

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., et al. (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., et al. (2017). "Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation." Science **358**(6361): 359-365.

Coker, O. O., et al. (2018). "Mucosal microbiome dysbiosis in gastric carcinogenesis." Gut **67**(6): 1024-1032.

Contijoch, E. J., et al. (2019). "Gut microbiota density influences host physiology and is shaped by host and microbial factors." Elife **8**.

Costello, E. K., et al. (2009). "Bacterial community variation in human body habitats across space and time." Science **326**(5960): 1694-1697.

Gevers, D., et al. (2014). "The treatment-naive microbiome in new-onset Crohn’s disease." Cell host & microbe **15**(3): 382-392.

Greenberg, J. M., et al. (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., et al. (2022). "Diurnal and eating-associated microbial patterns revealed via high-frequency saliva sampling." Genome Research **32**(6): 1112-1123.

Jian, C., et al. (2020). "Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling." PloS one **15**(1): e0227285.

Jin, S., et al. (2022). "Deciphering mechanisms and implications of bacterial translocation in human health and disease." Current Opinion in Microbiology **67**: 102147.

Joseph, S., et al. (2021). "A 16S rRNA gene and draft genome database for the murine oral bacterial community." Msystems **6**(1): e01222-01220.

Kitamoto, S., et al. (2020). "The bacterial connection between the oral cavity and the gut diseases." Journal of dental research **99**(9): 1021-1029.

Komiya, Y., et al. (2019). "Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity." Gut **68**(7): 1335-1337.

Kostic, A. D., et al. (2013). "Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment." Cell host & microbe **14**(2): 207-215.

Lagkouvardos, I., et al. (2016). "The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota." Nature microbiology **1**(10): 1-15.

Liu, C., et al. (2020). "The Mouse Gut Microbial Biobank expands the coverage of cultured bacteria." Nature communications **11**(1): 1-12.

Qin, N., et al. (2014). "Alterations of the human gut microbiome in liver cirrhosis." nature **513**(7516): 59-64.

Rao, C., et al. (2021). "Multi-kingdom ecological drivers of microbiota assembly in preterm infants." nature **591**(7851): 633-638.

Rashidi, A., et al. (2021). "No evidence for colonization of oral bacteria in the distal gut in healthy adults." Proceedings of the National Academy of Sciences **118**(42): e2114152118.

Read, E., et al. (2021). "The role of oral bacteria in inflammatory bowel disease." Nature Reviews Gastroenterology & Hepatology **18**(10): 731-742.

Schirmer, M., 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**(4): 600-610. e604.

Schluter, J., et al. (2020). "The gut microbiota is associated with immune cell dynamics in humans." nature **588**(7837): 303-307.

Schmidt, T. S., et al. (2019). "Extensive transmission of microbes along the gastrointestinal tract." Elife **8**: e42693.

Segata, N., et al. (2012). "Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples." Genome biology **13**(6): 1-18.

Tang, T. W., et al. (2019). "Loss of gut microbiota alters immune system composition and cripples postinfarction cardiac repair." Circulation **139**(5): 647-659.

The Human Microbiome Project Consortium (2012). "Structure, function and diversity of the healthy human microbiome." nature **486**(7402): 207-214.

Theis, K. R., et al. (2020). "No consistent evidence for microbiota in murine placental and fetal tissues." Msphere **5**(1): e00933-00919.

Thomas, A. M., et al. (2019). "Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation." Nature medicine **25**(4): 667-678.

Vandeputte, D., et al. (2017). "Quantitative microbiome profiling links gut community variation to microbial load." nature **551**(7681): 507-511.

Vieira-Silva, S., et al. (2019). "Quantitative microbiome profiling disentangles inflammation-and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses." Nature microbiology **4**(11): 1826-1831.

Zarrinpar, A., et al. (2018). "Antibiotic-induced microbiome depletion alters metabolic homeostasis by affecting gut signaling and colonic metabolism." Nature communications **9**(1): 1-13.