Relative enrichment of oral bacteria in feces indicates loss of intestinal commensals with implications for host health

Chen Liao[1,8], Thierry Rolling[2,3,4,8], Ana Djukovic[1], Hongbin Liu[5], Lei Dai[5], Bing Zhai[2,3,5], Jonathan U. Peled[6,7], Marcel R.M. van den Brink[6,7], Tobias M. Hohl[2,3,7,*], and Joao B. Xavier[1,*]

[1]Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA
[2]Infectious Disease Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[3]Immunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[4]Division of Infectious Diseases, First Department of Medicine, University Medical Center, Hamburg-Eppendorf, Hamburg 20251, Germany
[5]CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[6]Adult Bone Marrow Transplantation Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[7]Weill Cornell Medical College, New York, NY 10065, USA
[8]These authors contributed equally to this work
*Correspondence: xavierj@mskcc.org (J.B.X.), hohlt@mskcc.org (T.M.H.)

High relative abundance of oral bacteria in the gut is considered abnormal and has been linked to intestinal disorders. However, this reported association has two alternate explanations: ectopic oral bacteria may increase in absolute numbers in the intestine and drive disease (the *driver* hypothesis) or, instead, intestinal commensals may decrease in absolute numbers and the relative rise in oral bacteria is simply an indicator of this loss (the *marker* hypothesis). Here we use experiments in mice and clinical data from patients to provide evidence in favor of the *marker* hypothesis. We demonstrate that depletion of intestinal bacteria by antibiotics increases the relative abundance of oral bacteria in the host feces without increasing their absolute abundance, and the relative enrichment can indicate host health parameters such as stool consistency, pathogen overgrowth, and immunity. Our results reframe the conventional interpretation of microbiome sequencing data with implications for intestinal disorders and microbiome therapies.

Healthy individuals harbor distinct bacterial communities at different body sites, reflecting distinct ecologic niches and conditions[1, 2]. A number of health disorders and associated treatments cause the translocation of bacteria between distinct body sites, leading to a mixture of previously separated microbiome communities[3]. For example, it is well established that bacteria inhabiting the oral cavity can translocate to the lower gastrointestinal tract[4, 5, 6]. Compared to the hematogenous route (oral-blood axis)[7], the enteral route (oral-gut axis) is a natural pathway[8, 9]. On average, humans swallow ~10^{11} oral bacteria cells per day[10], yet oral bacteria remain rare in the healthy intestine under homeostatic conditions. Gastric acids, alkaline bile salts, and antimicrobial peptides kill many ectopic oral bacteria, and survivors must overcome colonization resistance mediated by the intestinal microbiome[9]. As a result, compositional sequencing data typically quantify <2% of fecal 16S rDNA reads from healthy volunteers as oral bacteria[11, 12].

Antibiotics, diet, aging, and intestinal inflammation can disrupt the oral-intestinal barrier and cause a relative enrichment of oral bacteria in the feces[8]. Such enrichment has been observed in patients with a variety of diseases, including Crohn's disease (CD)[13], ulcerative colitis (UC)[14], irritable bowel syndrome[15], colorectal cancer[16] and liver cirrhosis[17]. Recently, the enrichment signature has been found associated with unfavorable clinical response to anti-PD-1 (programmed cell death protein-1) therapy in a melanoma cohort[18]. However, nearly all those associations were obtained from compositional sequencing data, which reports relative abundance that by itself can neither determine if there was an expansion of oral bacteria in absolute numbers nor reveal the true causes of human disorders.

Two alternate explanations exist for this finding (Fig. 1). In the *driver* hypothesis, the high relative abundance of oral bacteria reflects a high absolute abundance. The relative enrichment of oral bacteria reflects an active expansion of the oral bacterial population in the intestine. When the oral population reaches a high density, the *driver* hypothesis predicts that this high density causes an associated disorder. In the *marker* hypothesis, the relative enrichment of oral bacteria is driven by loss of intestinal commensals, and it is this bacterial loss that is associated with the disease state. Determining which hypothesis is true has important implications for our understanding of human health and the pathology of disorders linked to inflammatory and infectious pathology.

Here, we compare the *driver* and *marker* hypotheses by analyzing the dynamics of oral bacteria in feces of both mice and humans. By quantifying the absolute abundance of oral bacteria in feces after antibiotic treatment, we show that their absolute numbers do not increase when their relative numbers do, due to depleted intestinal bacterial populations. Our findings provide strong evidence that the *marker*, not the *driver*, hypothesis is true. The validated *marker* hypothesis further indicates that the link between the relative enrichment of oral bacteria in feces and intestinal disorders is mediated by the loss of intestinal commensals and their associated functions. Recognizing that the increased relative abundance of oral bacteria in the intestine does not reflect an ectopic bloom can critically inform the interpretation of microbiome compositional data and future microbiome-based interventions to treat intestinal disorders.

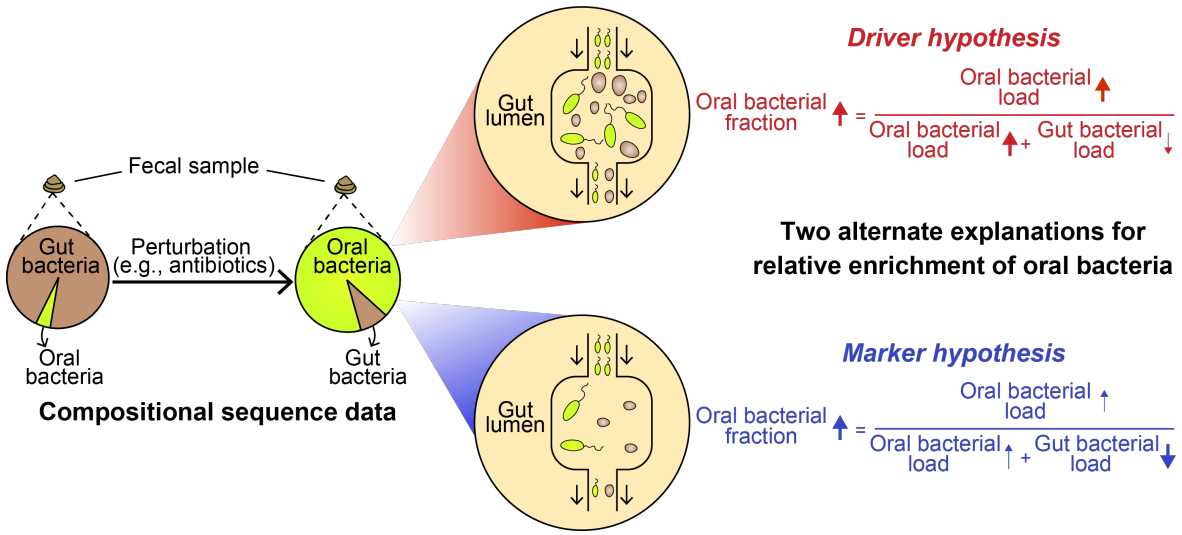


Fig. 1 | The *driver* and *marker* hypotheses provide alternative explanations for the reported association between oral bacterial enrichment in the feces and intestinal disorders. The relative abundance of oral bacteria in the feces can increase in patients with intestinal disorders or after treatment with certain antibiotics. This enrichment has two possible mechanisms: The relative enrichment may be due to increased absolute abundance of oral bacteria in the intestine and this increased population density drives the disorders (the *driver* hypothesis, in red). Alternatively, it may be caused by decreased absolute numbers of intestinal bacterial commensals whose depletion drives the disorders (the *marker* hypothesis, in blue).

Results

Antibiotics raise relative abundance of oral bacteria in mouse feces without raising absolute abundances

The two hypotheses in Fig. 1 give rise to different predictions about the relationships between relative and absolute abundance of oral bacterial populations in the intestine. To distinguish these two possibilities, we treated 3 individually housed C57BL/6J female mice for seven days with an antibiotic cocktail of ampicillin, vancomycin, and neomycin that is known to deplete murine intestinal bacteria[19] (Fig. 2a). We collected paired oral (day 0) and fecal (day 0 and 8) samples from the mice and measured their quantitative profiles via 16S sequencing and 16S qPCR. As expected, the baseline composition of the oral and fecal bacterial communities on day 0 was distinct, reflecting their different habitats (Fig. 2b). Surprisingly, the post-antibiotic fecal microbiomes were more similar (P=0.03; paired t-test of PC 1 distance) to pre-treatment oral microbiomes than to pre-treatment fecal microbiomes (Fig. 2c). 10%-20% of ASVs (amplicon sequence variants) detected in the post-treatment feces (the majority from 6 members; Extended Data Fig. 1) were exclusively found in the corresponding pre-treatment oral samples (Fig. 2d). By defining ASVs typical of the oral cavity (mean relative abundance >0.01% across pre-treatment oral samples and ≤ 0.01% across pre-treatment fecal samples), we showed that antibiotic treatment enriched the total relative abundance of these oral-typical ASVs in the feces from an average of 0.03% at baseline to an average of 27.2% (Fig. 2e). However, the relative enrichment of oral bacteria was not due to an increase in their absolute abundance, which in fact decreased by 72% compared to pre-treatment samples (Fig. 2f). Therefore, the relative enrichment was driven by a much greater decrease (>5,000 fold) in the absolute abundances of intestinal commensals after antibiotic treatment.

The result of this mouse experiment supports the *marker* hypothesis. If this observation were universally true for other mice, we should be able to determine the antibiotic-induced intestinal bacterial loss by quantifying the relative abundance of oral bacteria in mouse feces. However, paired oral and intestinal (or fecal) samples in mice are very rare[20], which prevent us from applying the same paired-sample-analysis to quantify the oral fraction. Alternatively, we compiled a reference set of 149 non-redundant, full-length 16S rRNA genes from culturable bacteria that reside in the oral cavity of mice but do not reside in the murine intestinal tract (Extended Data Fig. 2, Supplementary Table 1, see Methods). This reference-based approach allows us to identify ASVs of oral origin by exact sequence match and to estimate their total fractions in any given mouse fecal samples.

To validate this reference-based approach we analyzed published data from mice treated with an antibiotic cocktail of ampicillin, vancomycin, neomycin, and metronidazole for 5 days[21] (Fig. 2g). Consistent with our mouse experiment, antibiotics enriched three oral ASVs of *Lactobacillus*, *Proteus* and *Escherichia-Shigella* in the feces (Fig. 2h). In agreement with the *marker* hypothesis, their absolute abundances did not increase but decreased by 24.7-fold on average (Fig. 2i, left). The absolute abundance of intestinal bacteria decreased even more, by 189.2-fold (Fig. 2i, right), which caused the observed relative enrichment of oral bacteria. Interestingly, the fraction of oral bacteria in feces showed a negative linear correlation (Pearson's r = -0.52, P = 3.5e-17) with the total bacterial loads on the $\log_{10}$-$\log_{10}$ scale (Fig. 2j, Extended Data Fig. 3), and the anti-correlation remained significant by using the control group data alone (Fig. 2j, data in green, Pearson's r = -0.25, P = 5.9e-3). Taken together, the fraction of oral bacteria in feces can be used to quantitatively determine the level of intestinal commensal bacterial loss in antibiotic-treated subjects and to assess natural fluctuations in their loads even in untreated subjects.

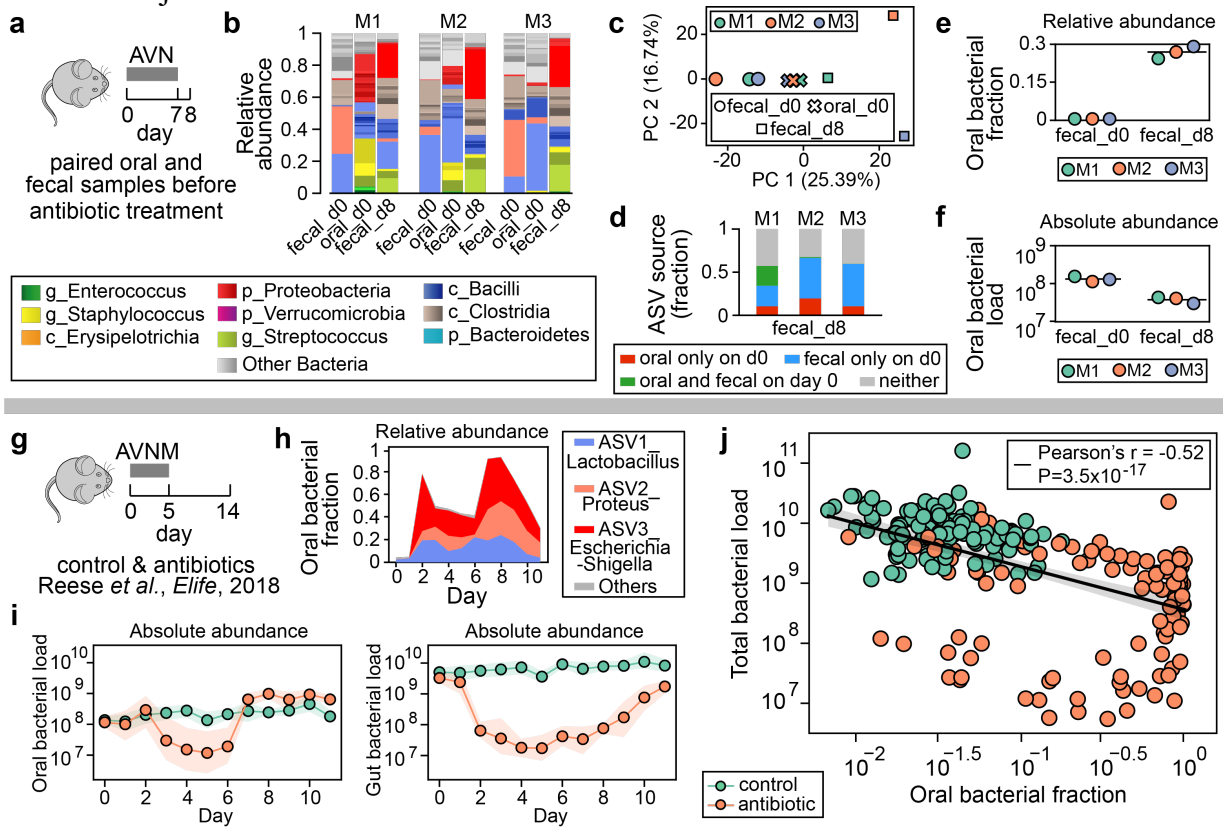


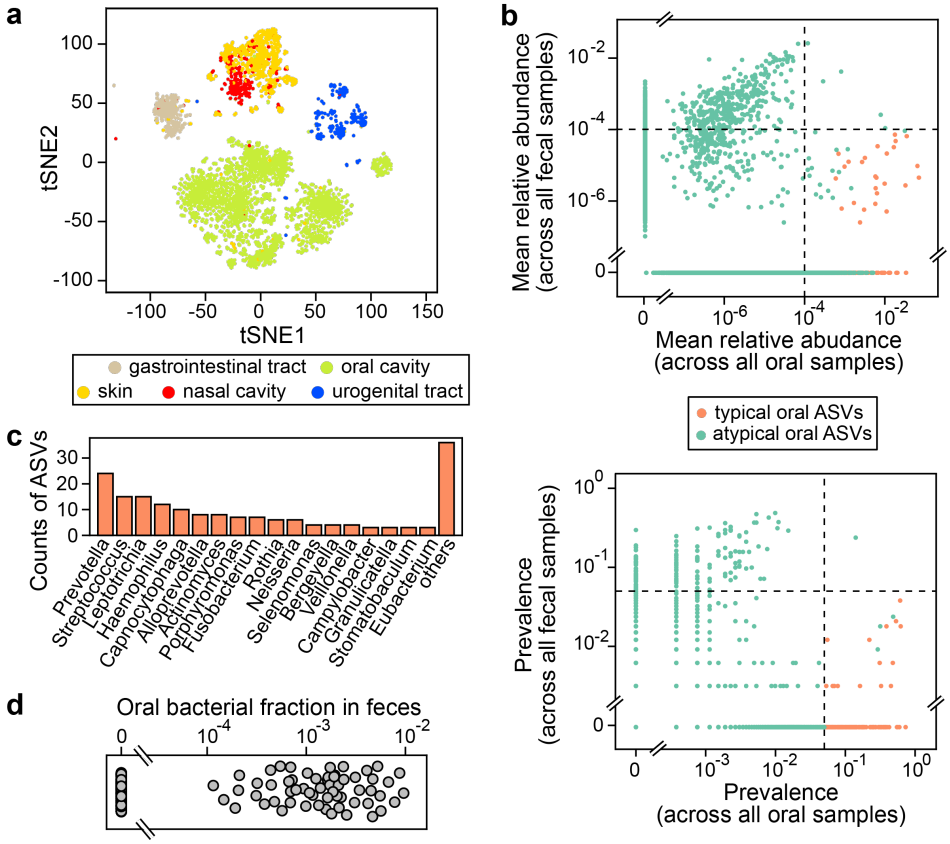
Fig. 2 | Mice treated with antibiotics support the *marker* hypothesis. a, Three mice (labeled as M1, M2, M3) were treated with an antibiotic cocktail of ampicillin, vancomycin, and neomycin (AVN) for a week. The pre-treatment feces (fecal_d0) and oral swabs (oral_d0) were collected on day 0 and the post-treatment feces (fecal_d8) were collected on day 8. **b**, Microbiome compositions of all feces and oral swabs show different oral and fecal microbiomes prior to antibiotics, but an enrichment of oral bacteria was observed in the feces post antibiotics. **c**, Principle component (PC) analysis of all microbiome samples. **d**, Tissue of origin for observed ASVs in the post-treatment fecal samples. **e**,**f**, Relative (e) and absolute (f) abundances of oral bacterial populations in the feces. Horizontal bars: mean. **g-j**, Analysis of a previously published dataset[21] leads to the same conclusion. **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**, The relationship between total fecal bacterial loads and oral bacterial fractions indicates that the fractions can be used to estimate the absolute amounts of bacteria in the fecal microbiome of both antibiotic-treated (orange) and untreated (green) subjects. Line: best linear fit; shading: 95% CI. Unit of bacterial load: 16S copies per gram of feces or per 1 oral swab.

A quantitative approach to estimate the fraction of oral bacteria in human feces

To quantify the oral bacterial fraction in human feces, we followed a similar strategy as used in mice to address the limitation of missing paired oral samples in human intestinal microbiome studies, especially those with quantitative data. To build a reference set of bacterial 16S rRNA gene sequences that typically colonize the human oral cavity (see Methods), we made use of the Human Microbiome Project (HMP) that sequenced paired oral (multiple subsites) and intestinal (stool) microbiomes from 237 healthy volunteers[1]. Distinct bacterial communities inhabit the two body sites (Fig. 3a), which provides the basis to define oral-typical ASVs based on their abundance and prevalence in healthy people. We defined ASVs of oral origin if their mean relative abundance per oral sample was greater than 0.01% and if their mean prevalence was higher than 5%. We excluded ASVs if their relative abundance or their prevalence met either cutoff in fecal samples (Fig. 3b). This filtering step yielded a stringent set of 178 oral ASVs (Supplementary Table 2) that belong to diverse genera such as *Prevotella* and *Streptococcus* (Fig. 3c). Henceforth, we can use the stringent set to infer the oral ASVs by exact sequence match and to estimate their total fraction in feces by summing the abundances of individual taxa.

We found that 219 out of 280 HMP fecal samples contained no oral ASVs. The mean relative abundance of oral ASVs was 0.05%±0.14% (Fig. 3d), which agrees with previous reports that oral bacteria are rare in the feces of healthy people[11, 12]. By contrast, identifying oral taxa via sequence overlaps in the absence of filtering leads to an overestimation of 15.63% ± 22.07% oral ASVs in HMP fecal samples. To further validate our approach in patients, we analyzed a public study that has sequenced paired oral (saliva) and intestinal (stool) samples[3] from CD and UC patients as well as their healthy controls. We first showed that our approach is robust against variations of the filtering cutoffs used to generate the reference set of oral ASVs (Extended Data Fig. 4a). The oral proportions in the feces of CD and UC patients were averagely 4 times as high as the proportions in the feces of HC (Extended Data Fig. 4b), confirming the notion that IBD patients are enriched with oral bacteria in their intestine[22]. Similar to healthy people, our approach to detect oral-intestinal transmission is conservative and less prone to false-

183 positive predictions (i.e., inferred oral ASVs absent from the oral cavity of the same person)
184 when applied to the IBD cohort (Extended Data Fig. 4c,d).
185



186
187 **Fig. 3 | Data from healthy individuals reveals bacterial DNAs typically found in the human**
188 **oral cavity but not the intestine. a**, An overview of the 4,354 microbiome samples from the
189 Human Microbiome Project (HMP). T-distributed stochastic neighbor embedding (tSNE)[23]
190 collapses all samples (dots) into distinct clusters based on their compositions. **b**, The mean
191 relative abundance and prevalence of ASVs (dots) among all HMP oral cavity samples (x axis)
192 and the fecal samples (y axis). 178 ASVs were identified as oral-typical and highlighted in
193 orange. **c**, Genus-level distribution of the 178 oral-typical ASVs. **d**, Distribution of the total
194 fractions of oral-typical ASVs in the feces of HMP subjects (n=280). Each circle represents a
195 fecal sample. We detected no oral ASVs in 219 samples.
196
197 **The intestinal domination of oral bacteria in bone marrow transplant recipients**
198 Using the quantification method developed above, we determined the relative abundance of oral
199 bacteria in a large-scale microbiome dataset from adult cancer patients who received allogeneic
200 hematopoietic cell transplantation (allo-HCT)[24, 25]. Allo-HCT is a potentially curative therapy for
201 a variety of hematologic disorders including leukemias and lymphomas[26]. Allo-HCT disrupts the
202 intestinal microbiome composition of the recipients, likely through the administration of
203 conditioning chemotherapy and the use of antibiotics to prevent and treat bacterial infections.
204 During allo-HCT, *Enterococcus* and *Streptococcus* are the two most abundant genera that
205 frequently dominate (>30% relative abundance of a single ASV[27]) the intestinal bacterial
206 community (Fig. 4a, top)[28]. Interestingly, among 10,433 fecal samples from 1,276 allo-HCT
207 patients, 901 (8.6%) were dominated by at least one oral ASV. *Streptococcus*, *Actinomyces*, and
208 *Abiotrophia* were the leading genera of these dominant oral ASVs (Fig. 4a, bottom). The total

fractions of oral ASVs detected in the feces of allo-HCT recipients was not correlated with sequencing depth (Extended Data Fig. 5).

To test further whether the identified genera were of oral origin, we used co-occurrence analysis (see Methods) to identify oral ASVs that were simultaneously present in the feces more frequently than expected by chance. The co-occurring oral ASVs, if such existed, might indicate collective bacterial transmission from the oral cavity to the intestine. We found 71 such oral-ASV groups that included 2 to 5 ASVs each. *Streptococcus* accounted for 55 of these 71 ASV groups (Supplementary Table 3). To identify the species in the *Streptococcus* ASV groups, we generated shotgun metagenomes from 19 samples which contained 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 (MAGs) of *Streptococcus spp.*, among which 4 contains ASV_8. All four MAGs were annotated as *S. thermophilus*—a probiotic species of lactic acid bacteria[29]. To infer the viability of these MAGs, we computed the ratio of metagenomic reads near the replication origin to those near the replication terminus (i.e., peak-to-trough (PTR) ratio)[30]. This ratio estimates the simultaneous replication events averaged across the entire population. We found that the averaged PTR ratio was 1.35 ± 0.014 (Supplementary Table 4), consistent with a mixture of 35% two-replication-fork and 65% single-replication-fork bacterial cells. The PTR-based growth estimate is consistent with the idea that the *Streptococcus* species that encodes ASV_8 was growing slowly in the intestine.

Antibiotics, particularly piperacillin/tazobactam, deplete intestinal bacterial commensals and drives relative enrichment of oral bacteria in feces

Allo-HCT recipients commenced antibiotic prophylaxis on median six days prior to transplantation. The timing of antibacterial antibiotic exposures (Fig. 4b, top) corresponded well with the decline in relative (Fig. 4b, middle) and absolute (Fig. 4b, bottom) intestinal bacterial abundances. Since the mean oral bacterial loads fluctuated around a stable average, the increased relative abundance of oral ASVs was driven by the decline in intestinal bacterial load. This finding supports the *marker* hypothesis and generates a negative correlation (Pearson's $r = -0.27$, $P = 2.2e\text{-}54$) between the oral bacterial fraction and the total bacterial loads in feces (Fig. 4c). The inverse correlation of the oral bacterial fraction and the total bacterial load stands in contrast to pathogens such as *Enterococcus*[31] and *Candida spp.*[32, 33], in which an increase in the relative abundance reflects an increase in the absolute quantity of these organisms.

We next conducted a time-varying survival analysis to quantify the effect-size of the different antibiotics that are associated with enriched oral ASVs in feces. Piperacillin/tazobactam (TZP, $P=5.7e\text{-}7$) exposure strongly increased the risk of intestinal domination by oral-derived bacteria (Fig. 4d). We confirmed this association in an independent cohort of pediatric allo-HCT recipients[34]. Children who have received oral TZP for gut decontamination had much higher relative abundance of oral bacteria in feces than those who received oral polymyxin/neomycin (Extended Data Fig. 6). Other than TZP, oral vancomycin ($P=0.049$) but not intravenous ($P=0.58$) vancomycin was associated with intestinal domination of oral bacteria, likely because intravenous vancomycin is not excreted into the intestine. We also saw a negative association between quinolones and the fraction of oral bacteria in feces ($P=6.1e\text{-}4$).

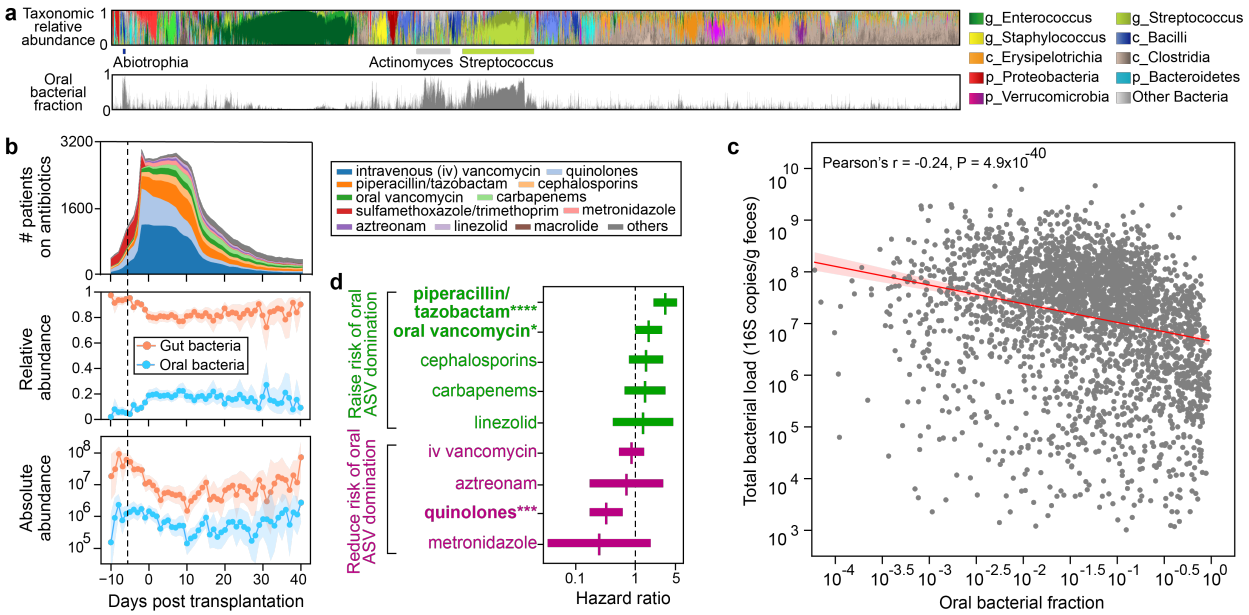


Fig. 4 | Allo-HCT injures the intestinal microbiome and increases the relative, but not the absolute, abundances of oral bacteria in feces. a, Fecal microbiome compositions of all 10,433 fecal samples. The taxonomic composition (top) and the estimated fractions of oral bacteria in the feces (bottom) were aligned for each sample. **b**, Antibiotic exposures (top) and population dynamics of the relative (middle) and absolute (bottom) abundances (16S copies per gram feces) of oral (light blue line) and intestinal (orange line) bacteria in the feces. Line and dots: mean; shading: 95% confidence interval (CI). The vertical dotted line indicates the median day -6 when antibiotic prophylaxis started. **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. Red line: best linear fit; shading: 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.

The relative abundance of oral bacteria in the feces is associated with host health

We next sought to determine whether the relative abundance of oral bacteria is correlated with the health status of allo-HCT recipients. The stool consistency, as assessed by technicians at the time of aliquoting fecal samples, was looser in patient samples with higher relative abundance of oral bacteria than in those with a low abundance of oral bacteria (Extended Data Fig. 7a). This association may be partially attributed to the loss of intestinal bacteria: Previous work[35] showed that water content, which strongly correlated with stool consistency, is negatively associated with microbiome density. We further found that the intestinal domination by any oral ASV over 30% of relative abundance reduced the risk of intestinal domination (relative abundance >30%) by *Enterococcus* but increased the risk of *Candida* domination (relative abundance >90%) (Supplementary Table 5), both of which are predisposing factors of bloodstream infections[27, 33]. Consistently, the oral bacterial fractions are significantly higher in samples with positive than negative fungal culture (Extended Data Fig. 7b). It is likely that the administration of antibiotics successfully eliminated bacterial pathogens including *Enterococcus spp.* when intestinal bacterial commensals were depleted but opened the niche for fungal pathobionts to expand.

Our previous work investigated interactions between specific intestinal bacteria and white blood cell dynamics (i.e., rate of white blood cell count changes)[36]. We revealed that oral bacteria in feces, as a new metric of intestinal microbiome damage, is also associated with white

blood cell dynamics. When we accounted for immunomodulatory drugs (e.g., granulocyte colony-stimulating factor) as potential covariates, the oral bacterial fraction remained positively associated with the day-to-day change in the total white blood cell, neutrophil, and monocyte counts (Supplementary Table 6). We further reanalyzed a public dataset from bone-marrow-transplanted mice that received antibiotics with a single drug[37] (Extended Data Fig. 8a). In that study, antibiotics reduced intestinal bacterial loads (Extended Data Fig. 8b) without altering oral bacterial loads (Extended Data Fig. 8c), causing a strong linear relationship (Pearson's r = -0.81, P = 2.9e-11) between oral bacterial fractions and total bacterial loads in the $\log_{10}$-$\log_{10}$ space (Extended Data Fig. 8d). Most importantly, the number of bone marrow cells at day 14 showed a negative linear relationship (Pearson's r = -0.90, P = 5.9e-5) with $\log_{10}$-transformed oral bacterial fractions in feces (Extended Data Fig. 8e). Taken together, these data from human and mice indicated that the relative abundance of oral bacteria in the feces, which indicates the total bacterial load, can inform host immune system status.

The *marker* hypothesis unifies competing microbiome biomarkers of IBD

Intestinal microbiome dysbiosis is a hallmark of intestinal disorders, but there are competing ways to determine dysbiosis. For example, clinical studies of IBD have revealed its associations with relative enrichment of oral-associated bacterial species[22] or low absolute values of microbial loads[35, 38]. Since these associations were reported by separate studies, they raise a critical question: are the two IBD biomarkers independent or related?

According to the *marker* hypothesis, the two quantities should be dependent and negatively correlated with each other. Indeed, the fraction of oral bacteria 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 feature was 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 the feces of IBD patients who have not taken antibiotics weeks prior to study[3, 39] (Extended Data Fig. 9), antibiotics may amplify the intestinal microbiome dysbiosis in IBD patients[13] and strengthen the negative relationships between the two biomarkers. In conclusion, both increased oral bacterial fraction and reduced total number of microbes in fecal samples indicate a loss of intestinal commensal bacteria. This bacterial loss, rather than the reported association of oral bacterial enrichment in fecal samples, likely represents a fundamental risk factor of IBD.

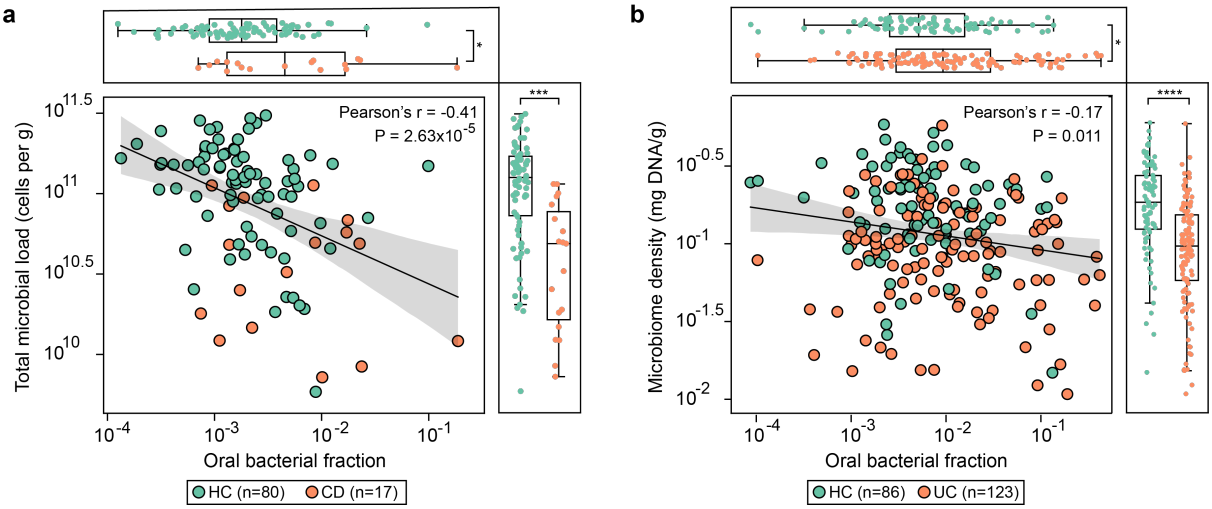


Fig. 5 | The *marker* hypothesis unifies previous work showing associations between oral bacterial enrichment or low microbial loads in the intestine of patients with inflammatory bowel disease. a, Crohn's disease (CD)[35]. **b**, Ulcerative colitis (UC)[38]. Each circle represents a fecal sample and samples that contained no oral bacteria were omitted in the plots. Lines: best linear fits; shading: 95% confidence intervals; HC: healthy control. ****P<0.0001; ***P<0.001; *P<0.05; Welch's t-test.

Discussion

Our study combines mouse experiments, human cohort studies, and quantitative microbiome analyses to provide strong evidence that, when a high relative abundance of oral bacteria is observed in fecal samples, this does not indicate a proportionally high density of oral bacteria in the intestine. Instead, the high relative abundance of oral bacteria represents a marker of the loss of intestinal commensal bacteria. These findings challenge the view that the oral cavity is a reservoir of opportunistic pathogens, whose translocation to and ectopic colonization of the intestine drives intestinal inflammation and pathologic immune responses[4, 6, 40]. Consistent with the *marker* hypothesis, oral bacterial populations in the intestine did not expand after antibiotic treatment, although their relative abundance increased substantially. As in a previous study[29], we recovered *Streptococcus thermophilus* from fecal metagenomic samples of adult allo-HCT recipients. *S. thermophilus* is generally not considered as a pathogen and is a commercially available probiotic that has been reported to attenuate inflammation[41]. Our findings do not formally exclude the possibility that translocated oral bacteria may exert pathologic effects in the intestine, even when present at low absolute levels. More experimental evidence is needed to broadly understand the pathogenicity of translocated oral bacteria at the strain level.

Amplicon-based sequencing produces compositional data, which requires careful interpretation. Quantitative microbiome profiling bypasses compositionality effects and can thus calibrate the associations learned from relative microbiome profiling[35, 42, 43]. 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[35]. To date, there is a limited number of quantitative microbiome studies in the literature[35, 36, 38, 42, 44, 45]. Our study adds to the limited literature by revealing another discrepancy between relative and absolute abundances: the relative enrichment of oral-derived bacteria in the intestine does not reflect their active population expansion, but simply indicates a microbiome state of depleted intestinal bacteria. Therefore, various oral bacterial species (e.g., *Streptococcus spp.*, *Veillonella*

353 *parvula*, *Fusobacterium nucleatum*) found enriched in the intestine of IBD patients[22] may all
354 represent biomarkers of a reduced intestinal bacterial load.
355 Our study discovered that exposure to TZP and quinolones are associated with higher and
356 lower risks of intestinal dominations of oral bacteria, respectively (Fig. 3d). According to the
357 *marker* hypothesis, the positive association of TZP reflects the strong anaerobe-killing capacity
358 of this β-lactam/β-lactam inhibitor drug combination[46]. Consistently, TZP used to treat bacterial
359 infections is known as a double-edged sword that significantly increases the GVHD (graft-
360 versus-host disease)-related mortality following allo-HCT[47]. Quinolones are widely used as
361 prophylaxis to reduce the incidences of gram-negative bacterial infections (e.g., *Pseudomonas*
362 and *Enterobacteriaceae*) in patients with neutropenia[48]. One possible explanation for its negative
363 association is that most quinolones[49] are more effective against aerobic and facultative anaerobic
364 bacteria (e.g., oral flora) than against anaerobic bacteria. Another potential explanation might be
365 that allo-HCT patients are universally started on quinolone prophylaxis and then they are
366 switched to other agents such as TZP if and when fevers or bloodstream infections occur. As
367 such, fecal samples exposed only to quinolones are enriched for samples collected prior to acute
368 febrile episodes, or from patients with more benign treatment courses who never got a fever and
369 remained on quinolone prophylaxis.
370 Research over the past two decades has accumulated unequivocal evidence that the
371 intestinal microbiome contributes to human health[50, 51]. Compared to biodiversity, the absolute
372 load of intestinal bacteria has received much less attention as a different indicator of microbiome
373 dysfunction[38]. Losing intestinal commensals can fundamentally alter host-microbiome
374 interactions and host health[38, 52, 53]. Consistent with our findings, the total microbial load in the
375 intestine was shown to be positively correlated with the abundance of host immune cells, such as
376 the proportion of mucosal RORγt+ Treg cells in mice[54]. For host disorders such as IBD, the loss
377 of intestinal bacteria may contribute to the pathogenesis via its regulations of host immunity.
378 Since the total fraction of oral bacteria in feces reflects the total bacterial load in the intestine, it
379 provides a convenient way to assess the size of intestinal bacterial population and even host
380 status. The robust relationship between the two quantities will lead to new opportunities for
381 diagnosis, prognosis, and treatment of microbiome-related human diseases. For example, the
382 finding of a high relative abundance of oral-typical bacteria such as *Streptococcus salivarius* and
383 *Dialister invisus* in the recipient microbiome, as an indicator of low bacterial load, could
384 facilitate engraftment in studies of therapeutic fecal microbiota transplantation[55].

Acknowledgements
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.) and R21 AI156157 (T.M.H.).

Author contributions
Conceptualization, C.L., T.R., T.M.H., 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.

Competing 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 serves on an Advisory board of and holds equity in Postbiotics Plus Research. 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. T.R. is currently an employee of BioNTech SE. Memorial Sloan Kettering Cancer Center (MSK) has financial interests relative to Seres Therapeutics.

Data availability
All raw sequences of the microbiome datasets analyzed in this study are publicly available. All processed data supporting the findings of this study are available within the article and its supplementary materials. Customized codes for the main text figures are available on Github (https://github.com/liaochen1988/marker_hypothesis).

Methods
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[56] 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[57]. 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[56]. 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. The raw sequences have been uploaded to Sequence Read Archive (SRA) with accession number PRJNA873058.

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.

Identification of oral bacteria in feces. 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 intestinal 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[58]. 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[59].

Our approach of constructing the mouse reference set is schematically shown in Extended Data Fig. 2. Due to a lack of large-scale microbiome datasets with paired oral and intestinal samples, we used the Mouse Oral Microbiome Database (MOMD)[60] as the starting point for gathering oral-typical sequences. MOMD has 164 full length rRNA sequences from cultured bacterial isolates. To detect if any MOMD sequence is also typically found in the intestine, we

480 matched them to a combined collection of 344 cultured intestinal 16S rRNA sequences from the
481 Mouse Intestinal Microbial Biobank (mGMB)[61] and the Mouse Intestinal Bacterial Collection
482 (miBC)[62] using reciprocal blast search (-perc_identity 99 -qcov_hsp_perc 100 -ungapped). We
483 excluded 37 MOMD sequences that either have exact matches or belong to any taxa that contain
484 at least 3 matched sequences.
485 Meanwhile, we profiled the paired oral (swab) and intestinal (distal colon) samples from
486 a small set (n=11) of pregnant mice in a public study[20] and identified 17 oral-typical ASVs using
487 definitions (1) and (2) for humans above. Prevalence-based filter was not implemented here due
488 to too few samples. Our results showed that only 5 of 17 ASVs were covered by MOMD,
489 suggesting that MOMD is incomplete. By searching the nucleotide database in NCBI (National
490 Center for Biotechnology Information), we obtained 29 full length mouse 16S rRNA sequences
491 that contain 6 of the 12 missed oral-typical ASVs. The 29 NCBI sequences were added to
492 MOMD. Finally, 7 redundant sequences were removed.
493
494 **16S amplicon sequencing.** The >10,000 microbiome samples from the adult allo-HCT
495 recipients were previously analyzed by an in-house processing pipeline and the ASV profiles
496 were available in a recent compilation study[24]. Briefly, reads were trimmed to the first 180 bp or
497 the first point with a quality score Q<2, and removed if they contained ambiguous nucleotides
498 (N) or if two or more errors were expected based on the quality of the trimmed reads. ASVs were
499 identified using DADA2 (Divisive Amplicon Denoising Algorithm)[63] and classified by IDTaxa[64]
500 and the SILVA v138 database[65].
501 The demultiplexed and primer-trimmed HMP 16S sequences were downloaded from the
502 Qitta repository[66] and processed by QIIME (Quantitative Insights Into Microbial Ecology) 2[67] .
503 DADA2 was used to denoise data and generate an ASV per sample count table, using the QIIME
504 denoise-pyro plugin[67]. Parameter --p-trunc-len 395 was used to remove low quality tails.
505 Taxonomy classification of the representative ASV sequences was performed using the QIIME
506 plugin "feature-classifier"[68] and the SILVA v138 database[65]. The classification took three steps.
507 We first extracted the V3-V5 region of the SILVA reference sequences using the extract-reads
508 method. Then we created a classifier by using the fit-classifier-naïve-bayes method with
509 extracted reads and the SILVA reference taxonomy. Finally, we ran the classifier on the ASV
510 sequences using the classify-sklearn method to get their taxonomy.
511 All other microbiome datasets used in this study were similarly processed using QIIME
512 2[67]. When needed, primers were removed from demultiplexed short reads using the QIIME
513 cutadapt plugin[69] with parameters "--p-error-rate 0.1" and "--p-overlap 3". The trimmed reads
514 were denoised using the QIIME dada2 plugin with truncation lengths determined by per-base
515 quality scores to generate feature tables at the ASV level. Taxonomic classification was
516 performed using the QIIME plugin "feature-classifier classify-sklearn"[70] against the SILVA 138
517 database [65] at a cutoff of 80% and inferred to the lowest possible taxonomic level.
518 For each sample, we only kept bacterial ASVs and removed those whose taxonomy
519 contains keywords "Chloroplast" or "Mitochondria". We further discarded samples whose total
520 sequencing depths are smaller than 1,000 reads.
521
522 **Shotgun metagenomic sequencing.** We adapted a recently published pipeline[29] to assemble the
523 contigs from short reads using MEGAHIT[71] and then bin the contigs into Metagenome-
524 assembled genomes (MAGs) using two different methods: Metabat2[72] and CONCOCT[73]. The
525 results were then aggregated to produce an optimized, non-redundant set of MAGs by DAS
526 Tool[74]. The high-quality MAGs (≥75% complete, ≤175 fragments/Mbp sequence, and ≤2%

527 contamination) classified as *Streptococcus spp.* by Kraken2[75] were analyzed by iRep[30]. The iRep
528 value of a MAG represents the average number of replication events over different
529 subpopulations of the MAGs weighted by their relative abundances.
530
531 **Co-occurrence analysis.** The algorithm for computing co-occurring bacterial communities is
532 described in detail elsewhere[58]. Briefly, it begins with ASV pairs and iteratively identifies co-
533 occurring ASVs of larger combination sizes. The presence/absence of an ASV is determined by a
534 relative abundance cutoff of 0.001. A combination of ASVs must satisfy the following criteria to
535 be considered as co-occurring: (1) they must co-occur in at least 200 patients and 500 samples;
536 (2) they must co-occur at least twice more than expected by chance, which is estimated by
537 assuming each ASV was observed independently and its number of observation can be modeled
538 by a binomial distribution; (3) the (false discovery rate) FDR-corrected *P*-values for the
539 independent observation hypothesis must be less than 0.05.
540
541 **Cox's proportional hazard model.** We used the Cox's time-varying proportional hazard model
542 to regress fecal domination (relative abundance >30%) of any oral bacterial ASV as a microbial
543 endpoint of interest against antibiotic administration as multivariable predictors. The exposure to
544 each antibiotic takes a value of 1 on the day the antibiotic was administered and 0 otherwise.
545 Next, we again examined domination of oral bacteria in fecal samples, this time as a univariable
546 predictor of intestinal domination of *Enterococcus* and *Candida* ASVs using the same time-
547 varying Cox's hazard model. Penalty was not added for all Cox hazard calculations. The time
548 interval of allo-HCT recipients starts from 10 days prior to transplantation and ends by 40 days
549 post-transplantation. Patients with less than 5 samples during the time interval were excluded.
550 Rarely administered antibiotics that have been administered less than 10 times were also
551 excluded. We assumed no intestinal domination of oral ASVs on days when samples were not
552 collected.
553
554 **Linear mixed-effects model.** To identify the effects of oral bacterial fractions in feces on white
555 blood cell dynamics, we followed our previous strategy[36] and developed four linear mixed-effect
556 models for total white blood cells, neutrophils, lymphocytes, and monocytes separately. These
557 models were specified by the Wilkinson notation '$\Delta\log W_i \sim \text{TransplantType} + \text{GCSF}_i + \text{MM}_i +$
558 $\text{cetirizine}_i + \text{OralFraction}_i + W_i + (1|\text{PatientID})$', where W_i is any of the four white blood cell
559 counts measured on day i, TransplantType represents different graft sources (unmodified bone
560 marrow, T-cell depleted, peripheral blood stem cell, umbilical cord) of bone marrow
561 transplantation, GCSF (Granulocyte colony-stimulating factor), MM (Mycophenolate mofetil),
562 and cetirizine are three immunomodulatory drugs and their values represent binary exposure on
563 day i, and PatientID represents a unique identifier of each patient. $\Delta\log W_i$ was computed as the
564 difference between log-transformed white blood cell counts on day i and its next day i+1. The
565 microbiome data and clinical metadata were limited to the periods between neutrophil
566 engraftment and 100 days post engraftment. The expression '(1|PatientID)' indicates that the
567 patient heterogeneity was modeled as random effects. The model was solved by fitlme in Matlab.
568
569 **Absolute microbiome quantification.** The oral and intestinal bacterial loads in the intestine
570 were computed by multiplying total bacterial loads (qPCR, flow cytometry, DNA mass) with
571 relative abundance of oral-typical and oral-atypical ASVs respectively. By interpreting oral-
572 atypical bacteria as intestinal bacterial members, we ignored bacterial transmission from other
573 non-oral body sites to the intestine.

Both cell-counting (e.g., flow cytometry) and molecular-based methods (e.g., qPCR, DNA mass) have been applied to measure absolute microbial abundances[76]. All methods have their own limitations, but qPCR is a convenient approach as it makes use of the same nucleic-acid preparation that is used for sequencing[77]. Technical biases introduced in the extraction, purification, and amplification steps of 16S sequencing are shared with qPCR, but not other approaches. In addition, 16S qPCR measures bacterial loads, not total microbial loads, which is again compatible with 16S data. Therefore, we exclusively used 16S qPCR-based quantification unless the data type was unavailable (e.g., Fig. 5).

Statistical analysis. All statistical analysis, including linear regression, correlation between variables (Pearson's and Spearman's correlation), and hypothesis testing (Kruskal-Wallis test, paired sample t-test, and Welch's t-test) were performed using python. The descriptive statistics was shown as mean±standard deviation.

References

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

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

3. Imai, J. *et al.* A potential pathogenic association between periodontal disease and Crohn's disease. *JCI Insight* **6**, e148543 (2021).

4. Atarashi, K. *et al.* Ectopic colonization of oral bacteria in the intestine drives T_H1 cell induction and inflammation. *Science* **358**, 359-365 (2017).

5. Li, B. *et al.* Oral bacteria colonize and compete with gut microbiota in gnotobiotic mice. *International journal of oral science* **11**, 10 (2019).

6. Kitamoto, S. *et al.* The intermucosal connection between the mouth and gut in commensal pathobiont-driven colitis. *Cell* **182**, 447-462 (2020).

7. Abed, J. *et al.* Colon cancer-associated *Fusobacterium nucleatum* may originate from the oral cavity and reach colon tumors via the circulatory system. *Front. Cell. Infect. Microbiol.* **10**, 400 (2020).

8. Kitamoto, S., Nagao-Kitamoto, H., Hein, R., Schmidt, T. & Kamada, N. The bacterial connection between the oral cavity and the gut diseases. *J. Dent. Res.* **99**, 1021-1029 (2020).

9. Jin, S., Wetzel, D. & Schirmer, M. Deciphering mechanisms and implications of bacterial translocation in human health and disease. *Curr. Opin. Microbiol.* **67**, 102147 (2022).

10. Segata, N. *et al.* Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* **13**, R42 (2012).

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

12. Rashidi, A., Ebadi, M., Weisdorf, D.J., Costalonga, M. & Staley, C. No evidence for colonization of oral bacteria in the distal gut in healthy adults. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2114152118 (2021).

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

14. Schirmer, M. *et al.* Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. *Cell Host Microbe* **24**, 600-610 (2018).

15. Priya, S. *et al.* Identification of shared and disease-specific host gene–microbiome associations across human diseases using multi-omic integration. *Nat. Microbiol.* **7**, 780-795 (2022).

16. Kostic, A.D. *et al. Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207-215 (2013).

17. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59-64 (2014).

18. McCulloch, J.A. *et al.* Intestinal microbiota signatures of clinical response and immune-related adverse events in melanoma patients treated with anti-PD-1. *Nat. Med.* **28**, 545-556 (2022).

19. Reikvam, D.H. *et al.* Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PLoS One* **6**, e17996 (2011).

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

21. Reese, A.T. *et al.* Antibiotic-induced changes in the microbiota disrupt redox dynamics in the gut. *Elife* **7**, e35987 (2018).

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

23. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579-2605 (2008).

24. Liao, C. *et al.* Compilation of longitudinal microbiota data and hospitalome from hematopoietic cell transplantation patients. *Sci. Data* **8**, 71 (2021).

25. Yan, J. *et al.* A compilation of fecal microbiome shotgun metagenomics from hematopoietic cell transplantation patients. *Sci. Data* **9**, 219 (2022).

669
670 26. Shono, Y. & van den Brink, M.R. Gut microbiota injury in allogeneic haematopoietic
671 stem cell transplantation. *Nat. Rev. Cancer* **18**, 283-295 (2018).
672
673 27. Taur, Y. *et al.* Intestinal domination and the risk of bacteremia in patients undergoing
674 allogeneic hematopoietic stem cell transplantation. *Clin. Infect. Dis.* **55**, 905-914 (2012).
675
676 28. Peled, J.U. *et al.* Microbiota as predictor of mortality in allogeneic hematopoietic-cell
677 transplantation. *N. Engl. J. Med.* **382**, 822-834 (2020).
678
679 29. Siranosian, B.A. *et al.* Rare transmission of commensal and pathogenic bacteria in the gut
680 microbiome of hospitalized adults. *Nat. Commun.* **13**, 586 (2022).
681
682 30. Brown, C.T., Olm, M.R., Thomas, B.C. & Banfield, J.F. Measurement of bacterial
683 replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256-1263 (2016).
684
685 31. Stein-Thoeringer, C. *et al.* Lactose drives *Enterococcus* expansion to promote graft-
686 versus-host disease. *Science* **366**, 1143-1149 (2019).
687
688 32. Rolling, T. *et al.* Haematopoietic cell transplantation outcomes are linked to intestinal
689 mycobiota dynamics and an expansion of *Candida parapsilosis* complex species. *Nat.*
690 *Microbiol.* **6**, 1505-1515 (2021).
691
692 33. Zhai, B. *et al.* High-resolution mycobiota analysis reveals dynamic intestinal
693 translocation preceding invasive candidiasis. *Nat. Med.* **26**, 59-64 (2020).
694
695 34. Bekker, V. *et al.* Dynamics of the gut microbiota in children receiving selective or total
696 gut decontamination treatment during hematopoietic stem cell transplantation. *Biol.*
697 *Blood Marrow Transplant.* **25**, 1164-1171 (2019).
698
699 35. Vandeputte, D. *et al.* Quantitative microbiome profiling links gut community variation to
700 microbial load. *Nature* **551**, 507-511 (2017).
701
702 36. Schluter, J. *et al.* The gut microbiota is associated with immune cell dynamics in humans.
703 *Nature* **588**, 303-307 (2020).
704
705 37. Staffas, A. *et al.* Nutritional support from the intestinal microbiota improves
706 hematopoietic reconstitution after bone marrow transplantation in mice. *Cell Host*
707 *Microbe* **23**, 447-457 (2018).
708
709 38. Contijoch, E.J. *et al.* Gut microbiota density influences host physiology and is shaped by
710 host and microbial factors. *Elife* **8**, e40553 (2019).
711
712 39. Pascal, V. *et al.* A microbial signature for Crohn's disease. *Gut* **66**, 813-822 (2017).
713
714 40. Rojas-Tapias, D.F. *et al.* Inflammation-associated nitrate facilitates ectopic colonization
715 of oral bacterium *Veillonella parvula* in the intestine. *Nat. Microbiol.* **7**, 1673-1685
716 (2022).

41. Han, F. *et al. Streptococcus thermophilus* attenuates inflammation in septic mice mediated by gut microbiota. *Front. Microbiol.* **11**, 598010 (2020).

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

43. Vandeputte, D. *et al.* Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nat. Commun.* **12**, 6740 (2021).

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

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

46. Morjaria, S. *et al.* Antibiotic-induced shifts in fecal microbiota density and composition during hematopoietic stem cell transplantation. *Infect. Immun.* **87**, e00206-00219 (2019).

47. Shono, Y. *et al.* Increased GVHD-related mortality with broad-spectrum antibiotic use after allogeneic hematopoietic stem cell transplantation in human patients and mice. *Sci. Transl. Med.* **8**, 339ra371-339ra371 (2016).

48. Cometta, A., Calandra, T., Bille, J. & Glauser, M.P. *Escherichia coli* resistant to fluoroquinolones in patients with cancer and neutropenia. *N. Engl. J. Med.* **330**, 1240-1241 (1994).

49. Appelbaum, P.C. Quinolone activity against anaerobes. *Drugs* **58**, 60-64 (1999).

50. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55-71 (2021).

51. Clemente, J.C., Ursell, L.K., Parfrey, L.W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **148**, 1258-1270 (2012).

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

53. Zarrinpar, A. *et al.* Antibiotic-induced microbiome depletion alters metabolic homeostasis by affecting gut signaling and colonic metabolism. *Nat. Commun.* **9**, 2872 (2018).

54. Britton, G.J. *et al.* Defined microbiota transplant restores Th17/RORγt+ regulatory T cell balance in mice colonized with inflammatory bowel disease microbiotas. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21536-21545 (2020).

55. Schmidt, T.S. *et al.* Drivers and determinants of strain dynamics following faecal microbiota transplantation. *Nat. Med.* **28**, 1902-1912 (2022).

56. Abusleme, L. *et al.* Oral microbiome characterization in murine models. *Bio-protoc.* **7**, e2655-e2655 (2017).

57. Taur, Y. *et al.* Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. *Sci. Transl. Med.* **10**, eaap9489 (2018).

58. Machado, D. *et al.* Polarization of microbial communities between competitive and cooperative metabolism. *Nat. Ecol. Evol.* **5**, 195-203 (2021).

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

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

61. Liu, C. *et al.* The Mouse Gut Microbial Biobank expands the coverage of cultured bacteria. *Nat. Commun.* **11**, 79 (2020).

62. Lagkouvardos, I. *et al.* The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat. Microbiol.* **1**, 16131 (2016).

63. Callahan, B.J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581-583 (2016).

64. Murali, A., Bhargava, A. & Wright, E.S. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome* **6**, 140 (2018).

65. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-D596 (2012).

66. Gonzalez, A. *et al.* Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods* **15**, 796-798 (2018).

67. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852-857 (2019).

68. Bokulich, N.A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).

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

70. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825-2830 (2011).

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

72. Kang, D.D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

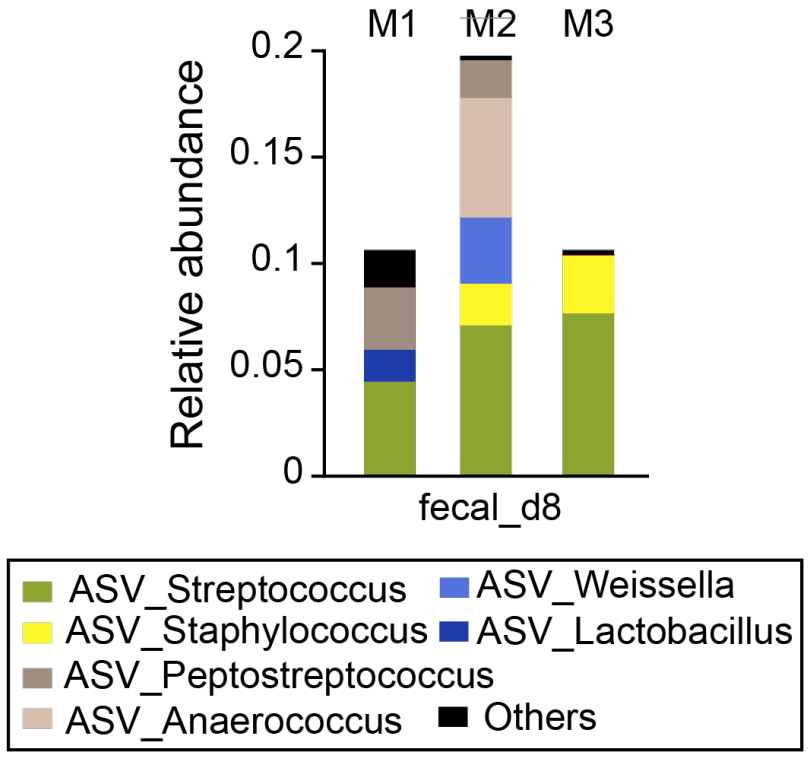
73. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144-1146 (2014).

74. Sieber, C.M. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836-843 (2018).

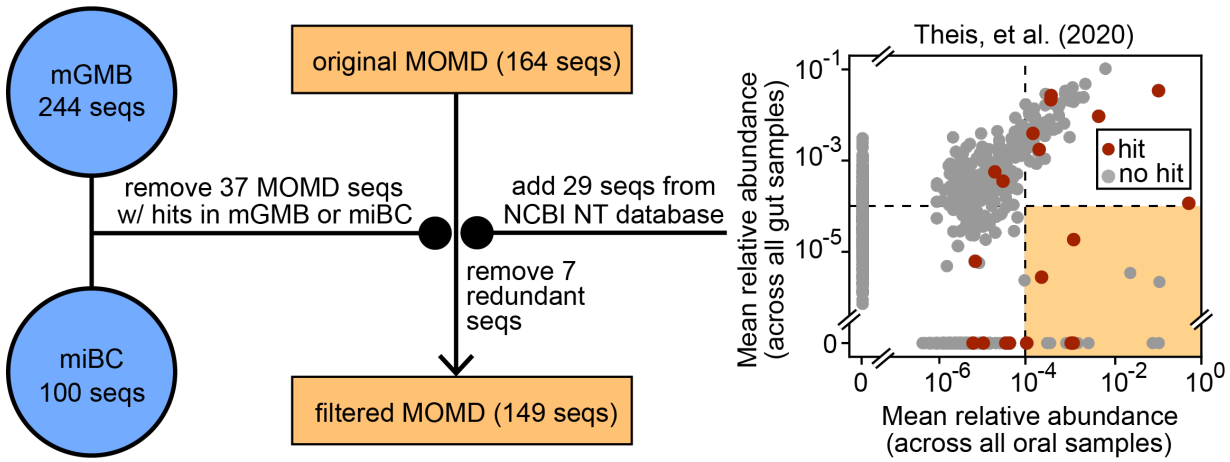
75. Wood, D.E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).

76. Galazzo, G. *et al.* How to count our microbes? The effect of different quantitative microbiome profiling approaches. *Front. Cell. Infect. Microbiol.* **10**, 403 (2020).

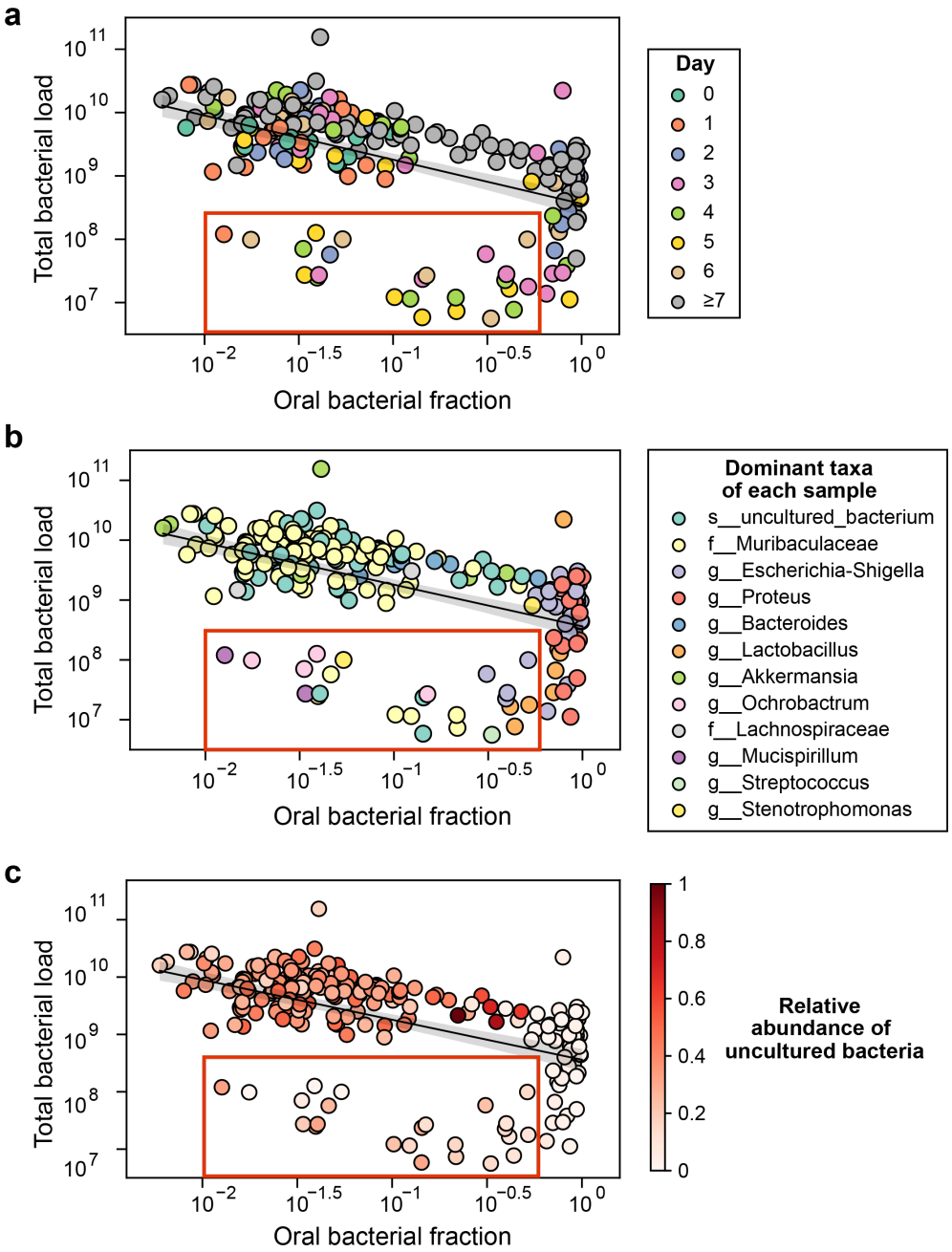
77. Jian, C., Salonen, A. & Korpela, K. Commentary: How to count our microbes? The effect of different quantitative microbiome profiling approaches. *Front. Cell. Infect. Microbiol.* **11**, 627910 (2021).



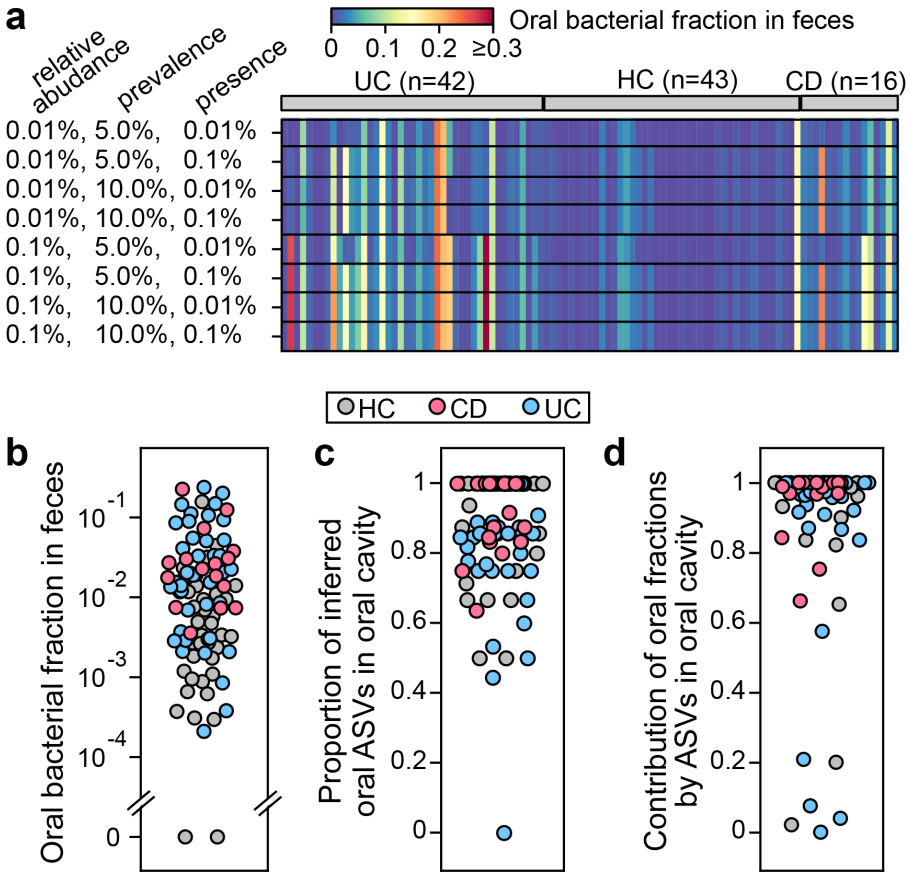
Extended Data Fig. 1 | Taxonomy of ASVs in the post-treatment fecal samples. Shown are the ASVs exclusively found in the pre-treatment oral samples (i.e., not in the pre-treatment fecal samples) of the same mice. M1, M2, and M3 represent three mice. fecal_d8: post-treatment fecal sample on day 8.



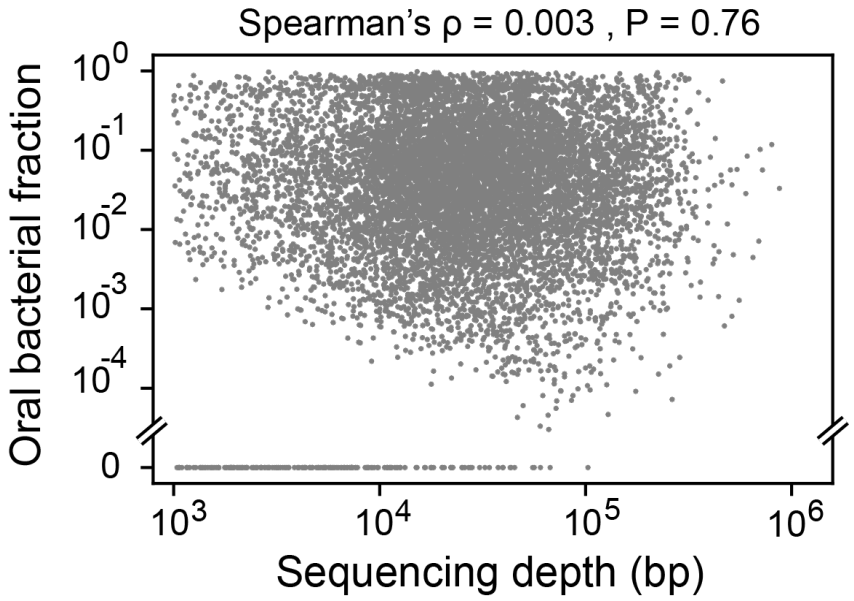
Extended Data Fig. 2 | The schematic of our approach for constructing a reference set of oral bacterial 16S rRNA sequences in mouse. Briefly, we started with 164 full length 16S rRNA sequences from cultivable oral bacteria in MOMD (Mouse Oral Microbiome Database[60]). We removed sequences from cultivable intestinal bacteria in mouse (those collected in mGMB[61] and miBC[62]) and added a few new sequences missed by MOMD but representative of the oral cavity of 11 mice in a public study[20]. In the scatter plot, hits refer to ASVs in the 11 mice matching at least one sequence in the original MOMD. The dashed lines (equal to 10^{-4}) represent the relative abundance cutoffs to identify oral-typical ASVs (those in the shading area). Here we did not use prevalence-based filter due to the small sample size. mGMB: the mouse Gut Microbial Biobank; miBC: the mouse intestinal Bacterial Collection; Seqs: sequences. See Methods of the main text for details.



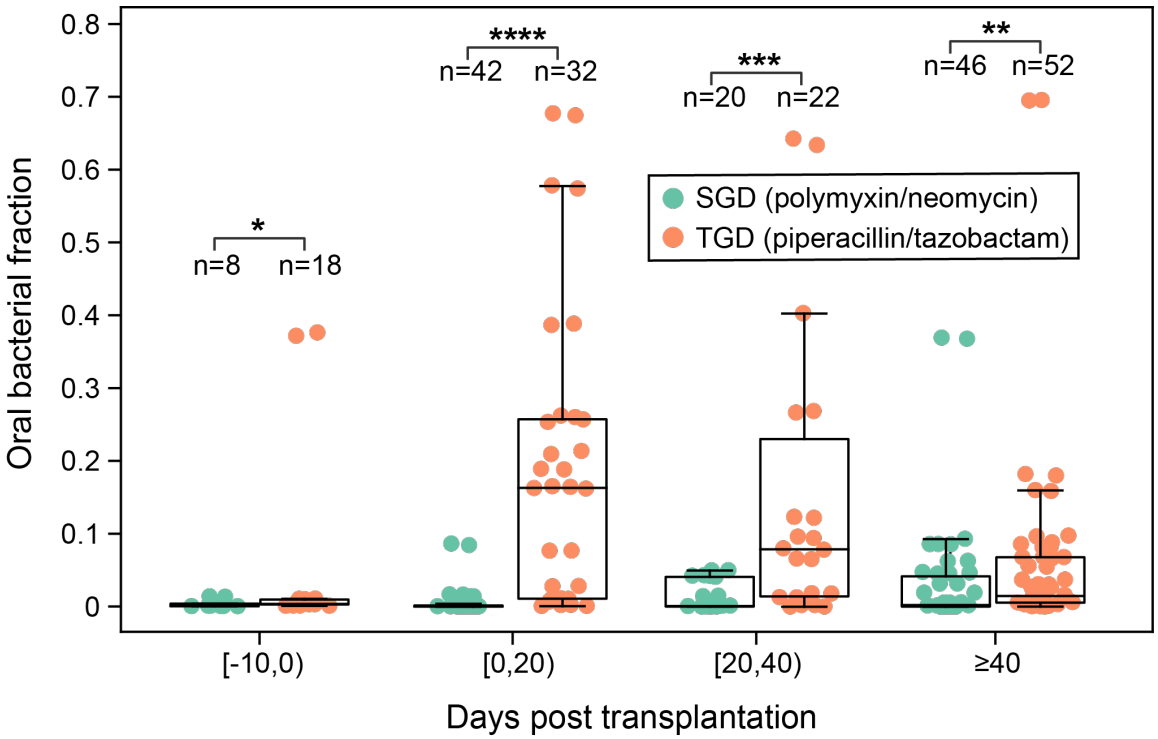
Extended Data Fig. 3 | Fig. 2j of the main text colored by sample collection day (a), dominant bacterial taxa (b), and relative abundance of uncultured bacteria (c). The red box outlined 11% samples collected between day 1-6. These samples have both low fractions of oral bacteria in feces ($<10^{-0.25}$) and low total bacterial loads ($<10^{8.5}$), and deviated from the trend line of the association (black line: best linear fit; gray shading: 95% confidence interval). According to the *marker* hypothesis, they should be detected with very high proportions of oral bacteria. We speculated that the most abundant ASVs of these outliers 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. Unit of total bacterial load: 16S copies per gram feces.



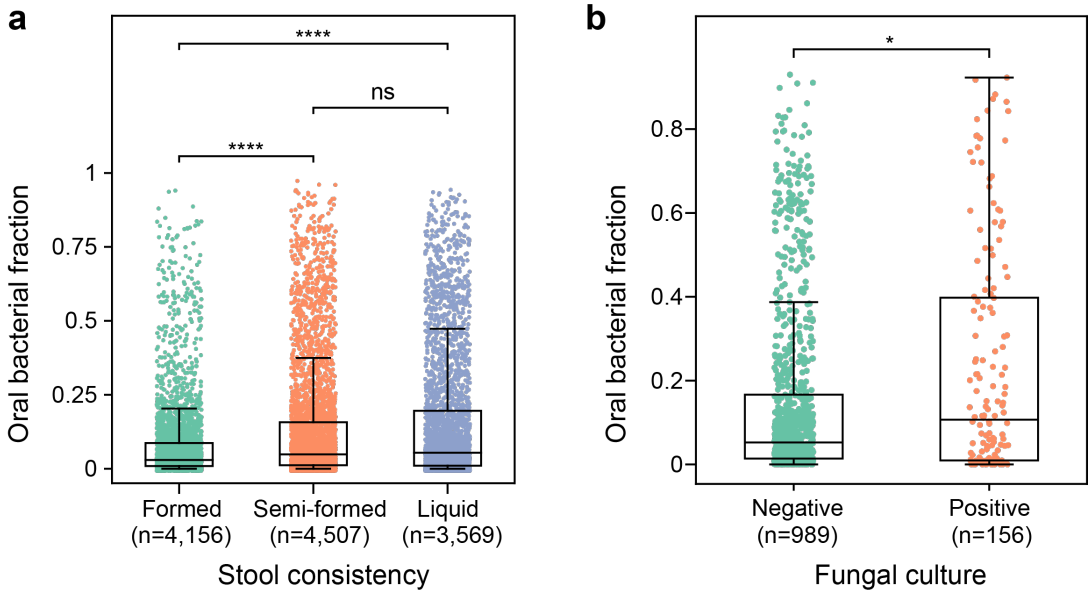
Extended Data Fig. 4 | Validations of our approach to identify oral bacteria in human feces using a cohort of patients with inflammatory bowel disease and their healthy controls[3]. a, Effects of filtering cutoffs to construct the reference 16S rRNA sequences of oral-typical bacteria. We varied the thresholds of mean relative abundance (0.01%, 0.1%), prevalence (5%, 10%), and the definition of ASV presence (0.01%, 0.1%) in the computation of prevalence. The reference oral-typical ASVs generated by each combination of the thresholds was used to compute the total fractions of oral bacteria in the feces of all subjects. We saw that the estimated oral proportions formed vertical colored bands in the heatmap, suggesting that they are largely insensitive to the variations in filtering thresholds. CD: Crohn's disease; UC: ulcerative colitis; HC: healthy controls. **b**, A strip plot showing the distributions of oral bacterial fraction in the feces. **c**, Proportions of fecal ASVs inferred as of oral origin that can also be found in paired saliva samples. 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). **d**, Total fractions of inferred oral ASVs contributed by those found in the paired saliva samples (i.e., total fraction of inferred oral ASVs that were also present in the paired oral samples divided by total fraction of all inferred oral ASVs). We showed that >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).



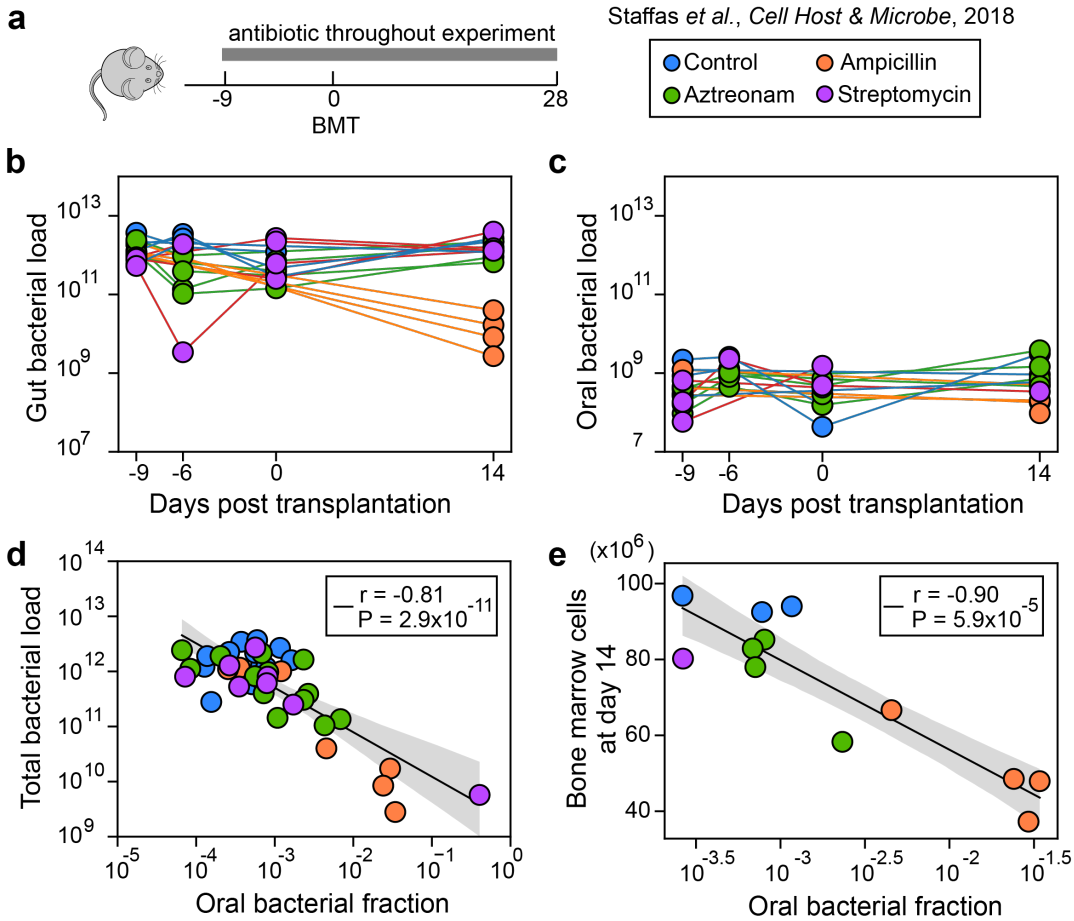
Extended Data Fig. 5 | Total fractions of oral bacteria in feces are not associated with sequencing depths. Each dot represents a fecal sample from adult allo-HCT (allogeneic hematopoietic cell transplantation) recipients hospitalized at Memorial Sloan Kettering Cancer Center.



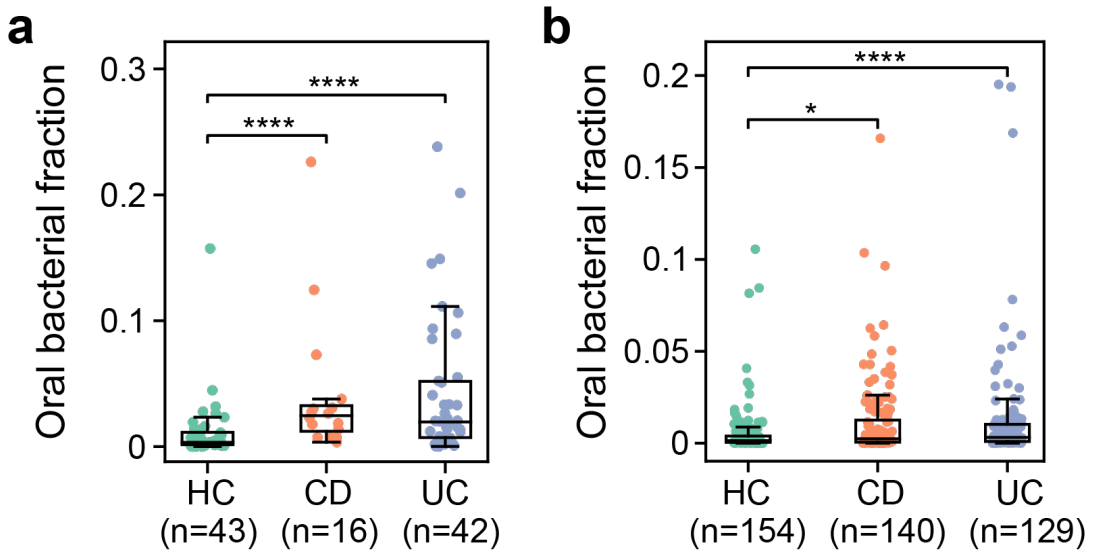
Extended Data Fig. 6 | Piperacillin/tazobactam (TZP) caused significant enrichment of oral bacteria in the feces of 19 pediatric allo-HCT (allogeneic hematopoietic cell transplantation) recipients (1-17 year old, 10.1 year old on average)[34]. Samples were grouped into four different stages of transplantation. TZP was used in total gut decontamination (TGD) as the only antibacterial antibiotic, while polymycin/neomycin was used in selective gut decontamination (SGD). The administration of SGD and TGD started 10 days before transplantation until engraftment or 21 days after transplantation, whichever occurred later. The comparison is visualized using box plot overlaid on top of a strip plot (each dot represents a fecal sample), with central line representing the median, box limits representing the first and third quartiles, and whiskers extending to the smallest and largest values or at most to 1.5× the interquartile range, whichever is smaller. ****, P<0.0001; ***, P<0.001, **, P<0.01; *, P<0.05. Kruskal-Wallis test.



Extended Data Fig. 7 | Distribution of oral bacterial fractions in feces stratified by stool consistency (a) and fungal cultivability (b). Each dot represents a fecal sample from adult allo-HCT (allogeneic hematopoietic cell transplantation) recipients hospitalized at Memorial Sloan Kettering Cancer Center. The comparison is visualized using box plot overlaid on top of a strip plot (each dot represents a fecal sample), with central line representing the median, box limits representing the first and third quartiles, and whiskers extending to the smallest and largest values or at most to 1.5× the interquartile range, whichever is smaller. ****, P<0.0001; ns, not significant. Kruskal-Wallis test.



Extended Data Fig. 8 | Correlations between oral bacteria in feces and host immunity in bone marrow transplanted mice[37]. a, Experiment design. Mice were given a split 1100cGy radiation dose and administered 5×10^6 bone marrow cells via tail vein injection at day 0. Ampicillin, aztreonam, and streptomycin were started 9 days before bone marrow transplantation (BMT) and administered throughout the experiment. Except for ampicillin (4 mice), all other antibiotic treatments (including the control group) were repeated in 5 mice. **b**,**c**, Dynamics of gut (b) and oral (c) bacterial loads during BMT. Each trajectory represents a mouse. **d**, Linear relationship between oral bacterial fractions and total bacterial loads in the $\log_{10}$-$\log_{10}$ space. Black line: best linear fit; shading: 95% confidence interval (CI). **e**, Negative linear relationship between log-transformed oral bacterial fractions in feces and bone marrow cells at day 14. Black line: best linear fit; shading: 95% CI. Samples that do not contain oral ASVs were omitted in panels d and e. Unit of bacterial loads: 16S copies per gram feces.



Extended Data Fig. 9 | Oral bacteria showed significantly higher enrichment in the feces of patients with Crohn's disease (CD) and ulcerative colitis (UC) compared to their healthy controls (HC). a, Imai et al.[3]. Participants who took any antibiotics within the last 3 months were excluded. **b**, Pascal et al.[39]. Participants who were treated by antibiotics within the previous 4 weeks were excluded. The comparison is visualized using box plot overlaid on top of a strip plot (each dot represents one sample), with central line representing the median, box limits representing the first and third quartiles, and whiskers extending to the smallest and largest values or at most to 1.5× the interquartile range, whichever is smaller. ****, $P<0.0001$; *, $P<0.05$. Kruskal-Wallis test.