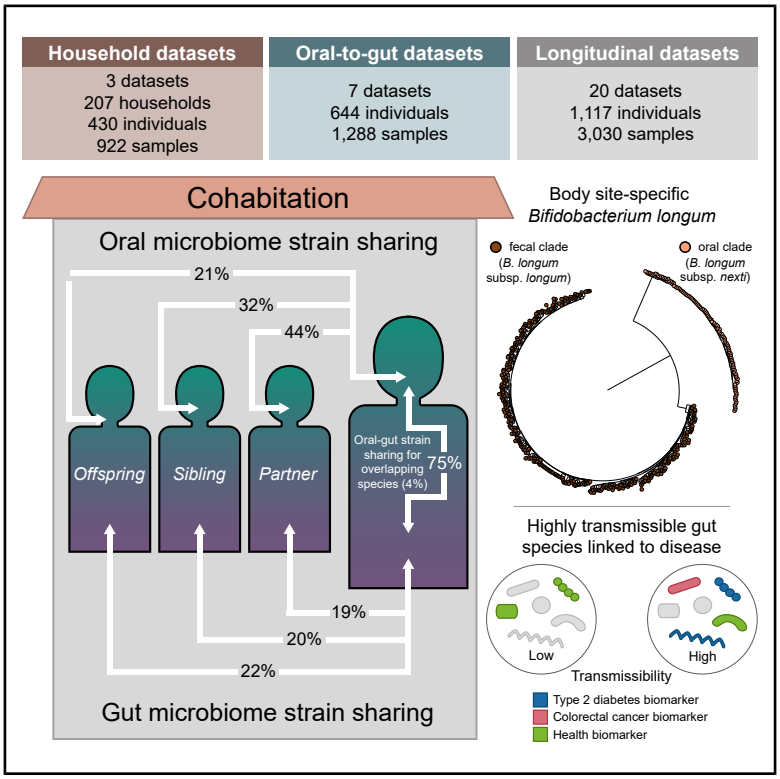
Cell Press Blue

Article

Strain transmission links human microbiomes along the oral-gut axis and across cohabiting individuals

Graphical abstract



Highlights

- Cohabitants share more oral and gut microbial strains than non-cohabitants
- Romantic partners exhibit the highest rates of oral microbiome transmission
- Highly transmissible gut species are associated with poorer cardiometabolic health
- A recently described *Bifidobacterium longum* subspecies colonizes the oral cavity

Authors

Vitor Heidrich, Gloria Fackelmann, Liviana Ricci, ..., Paola Mattarelli, Gianluca Ianiro, Nicola Segata

Correspondence

nicola.segata@unitn.it

In brief

Heidrich and colleagues analyzed paired oral and fecal metagenomes to map the inter-connectedness of human microbiomes across body sites and cohabitants. The work reveals high inter-personal transmissibility of the oral microbiome between partners, links highly transmissible gut species to cardiometabolic disease, and identifies body-site-specific strains among overlapping oral-gut species.

Heidrich et al., 2026, Cell Press Blue *1*, 100034
June 15, 2026
https://doi.org/10.1016/j.cpblue.2026.100034

CellPress

Cell Press Blue

CellPress
OPEN ACCESS

Article

Strain transmission links human microbiomes along the oral-gut axis and across cohabiting individuals

Vitor Heidrich,[1] Gloria Fackelmann,[1] Liviana Ricci,[1] Roan Spadazzi,[1] Gabriel Baldanzi,[1] Michal Punčochář,[1] Giulia Catassi,[2] Paolo Marchi,[1] Monica Modesto,[4] Gianmarco Piccinno,[1] Serena Porcari,[2,5,6] Debora Rondinella,[2,5,6] Francesco Asnicar,[1] Mireia Valles-Colomer,[3] Paola Mattarelli,[4] Gianluca Ianiro,[2,5,6] and Nicola Segata[1,7,8,9,*]

[1]Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy
[2]Department of Translational Medicine and Surgery, Università Cattolica del Sacro Cuore, Rome, Italy
[3]Department MELIS, Pompeu Fabra University, Barcelona, Spain
[4]Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy
[5]Department of Medical and Surgical Sciences, UOC Gastroenterologia, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy
[6]Department of Medical and Surgical Sciences, UOC CEMAD, UOC CEMAD Centro Malattie dell'Apparato Digerente, Medicina Interna e Gastroenterologia, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy
[7]IEO, Istituto Europeo di Oncologia IRCCS, Milan, Italy
[8]Department of Twins Research and Genetic Epidemiology, King's College London, London, UK
[9]Lead contact
*Correspondence: nicola.segata@unitn.it
https://doi.org/10.1016/j.cpblue.2026.100034

SUMMARY

Interpersonal strain transmission shapes the human microbiome, yet a comparative understanding of the transmission dynamics across body sites is lacking. We analyzed 1,644 paired oral and fecal metagenomes to investigate microbiome transmission among healthy cohabitants and intra-individual oral-gut overlap. Cohabitants shared significantly more oral and gut strains than non-cohabitants. Romantic partners exhibited the highest oral strain-sharing rates, exceeding their gut strain sharing. Higher oral transmissibility was associated with increased longitudinal strain replacement, while the most transmissible gut species were linked to poorer cardiometabolic health. Within individuals, 74.5% of cases of species detected in both sites involved the same strains, primarily related to abundant oral species such as *Streptococcus salivarius*, suggesting saliva-mediated transmission. Conversely, *Bifidobacterium longum* strains never overlapped between sites, with the recently proposed *B. longum* subsp. *nexti* uniquely colonizing the oral cavity. These findings extend our understanding of microbiome spread and its potential consequences for human health.

INTRODUCTION

The gut and the oral cavity host the two most diverse microbiomes in the human body.[1] Their significance to humans, including their impact on metabolism, immunity, and various diseases, has been thoroughly demonstrated,[2–4] with an increasing potential of using the microbiome as a clinical therapeutic target.[5,6] However, by focusing on the taxonomic and functional composition of single microbiomes, metagenomic investigations have predominantly treated our microbial communities as isolated entities. This approach overlooks emerging evidence that microbiomes are constantly (re)shaped in their constitutive building blocks—the microbial strains—by microbial exchange between humans throughout life.[7]

After the natural inoculation of maternal microbes during and shortly after birth, microbiome transmission during interactions with any proximate individual is expected to be a key driver of microbiome assembly and dynamics.[7,8] In this regard, cohabitation, as a pervasive and continuous form of human interaction, has already received initial research focus. Current research has found that humans living in the same household share significantly more strains than non-cohabiting individuals across distinct populations and lifestyles (i.e., westernized versus non-westernized),[9–11] with length of cohabitation positively correlated with strain-sharing rates (SSRs).[9,10] However, despite their largely different environmental conditions, anatomy, and microbial compositions,[1,12] little is known about the specific principles governing microbiome transmission in each body site, with a comparative understanding of oral versus gut microbiome transmission patterns still lacking. Furthermore, while oral microbes are constantly swallowed, their intra-individual transmission to the gut is not yet fully understood, as most reports (with a few exceptions[13–17]) have lacked the strain-level resolution necessary to accurately identify these events.[2]

In this study, by leveraging paired oral and fecal metagenomes of the same healthy individuals, we evaluated the microbiome transmission patterns and dynamics across oral and gut microbiomes of individuals with close familial ties who share living spaces. Methodologically, this was enabled by our strain-resolved computational approach that relies on the uniqueness

of microbiome strains across unrelated individuals[7,9] to infer the copresence of strains as evidence of microbiome transmission. Following a multi-cohort characterization of oral and gut microbiome transmission in the household, we compare strain transmission and replacement rates between oral and gut microbiomes across different familial relationship types (e.g., siblings or partners) and provide a strain-level landscape of oral-to-gut transmission.

RESULTS

To analyze and compare the extent of oral versus gut microbiome strain sharing among cohabiting individuals, we leveraged one publicly available metagenomic dataset (BritoIL_2019,[10] *n* individuals = 253, *n* samples = 377) and two new metagenomically sequenced cohorts (CM_MTB, *n* individuals = 74, *n* samples = 335; CM_MTC, *n* individuals = 103, *n* samples = 176; see STAR Methods). Across all these studies, oral and gut metagenomic samples were available, with a total of 207 total households considered (Figure 1A).

In addition to inter-individual microbiome strain sharing, we also evaluated intra-individual microbial strain overlap between oral and gut microbiomes. For this task, we further considered four publicly available metagenomic datasets with matched oral and fecal samples from a total of 378 healthy individuals (HMP_2012,[1] *n* individuals = 86, *n* samples = 172; FerrettiP_2018,[18] *n* individuals = 20, *n* samples = 40; NagataN_2022,[19] *n* individuals = 235, *n* samples = 470; and KartalE_2022,[20] *n* individuals = 37, *n* samples = 74; Figure 1A).

The identification of strain-sharing events is a task requiring accurate profiling of strain-specific genetic features as well as robust means to define strain identity. Here, we adopted a validated pipeline[9] that we expanded with additional longitudinal datasets to fine-tune the strain definition. In brief, the pipeline builds StrainPhlAn 4[21] phylogenetic trees for all species-level genome bins (SGBs; a genome-based definition of microbial species) that are present in metagenomes, then defines SGB-specific same-strain thresholds, and finally evaluates whether the phylogenetic distance between two strains in the same SGB phylogeny is below the same-strain threshold. The operational definition of same-strain thresholds is done by computing distributions of inter-individual and intra-individual (using longitudinal samples) phylogenetic distances between conspecific strains, with the assumption that in temporally close (<6 months apart) intra-personal longitudinal samples the same strain is observed in a large fraction of the cases. To this end, we further incorporated >3,000 longitudinal metagenomes (from 20 publicly available metagenomic datasets spanning >1,000 individuals; see STAR Methods) into the phylogenetic trees (Figure 1A). Finally, a same-strain threshold is defined as the phylogenetic distance that best separates the two distributions (STAR Methods), resulting in a list of strain-sharing events between all pairs of samples evaluated in our study. Considering together the household (*n* = 3) and oral-to-gut (*n* = 4) datasets, this enabled evaluation of strain-sharing patterns for 986 SGBs across 1,644 samples from 808 individuals (Figure 1A; Table S1).

Confirmed extensive gut and oral microbiome strain sharing between cohabiting individuals

For each of the three household datasets considered, we validated and substantiated the previously reported increased sharing of both gut and oral microbiome strains among individuals living in the same household[9–11] (Figures 1B and 1C; Table S2). The BritoIL_2019 dataset presented the sole partial exception, with no significant increase in same-household oral microbiome sharing with respect to different households in the same dataset. This is likely a consequence of the very low number of intra-household oral microbiome SSRs available for this dataset (*n* = 6), since this metric is defined only for pairs having ≥10 SGBs typed at the strain level in common (STAR Methods). Indeed, increased intra-household oral microbiome strain sharing can be detected also in BritoIL_2019 when comparing the numbers of shared strains (irrespective of the number of typed SGBs in common) instead of SSRs (Mann-Whitney U test, *n* = 15,766, *U* = 1,454,982, $p = 0.017$). Pooling all datasets together, we found that cohabiting individuals have a median 19.0% gut microbiome SSR (versus 6.0% for individuals living in different households; Mann-Whitney U test, *n* = 17,776, *U* = 490,405, $p < 1e{-}20$; Figure 1B) and a median 25.8% oral microbiome SSR (versus 0%; Mann-Whitney U test, *n* = 3,919, *U* = 15,758, $p < 1e{-}20$; Figure 1C). This result also holds for both the gut and the oral microbiome when factoring in the distinct relationship types (e.g., siblings or partners) present in the household (Figures S1A and S1B; Table S3).

In line with previous observations[9–11] and consistent with the expected strain transmission dynamics,[7] individuals living in different households but from the same dataset/cohort often had a non-zero SSR (83.8% of gut pairs and 44.5% of oral pairs). This means that sharing at least one strain was not uncommon, which might be attributed to them being part of the same population (Vanua Levu, Fiji, in BritoIL_2019; Trento, Italy, in CM_MTB; and Lazio, Italy, in CM_MTC) and therefore with increased potential coacquisition from shared spaces and/or transmission during untraced interactions. This hypothesis is supported by the observation that this modality of strain sharing was significantly higher than that observed between pairs of individuals from different datasets (Figure 1A; Table S4), which more rarely shared strains (25.2% of gut pairs and 22.8% of oral pairs in different datasets; 8.7% of gut pairs and 17.7% of oral pairs in different studies and countries), translating into a median 0% gut and oral microbiome SSR among unrelated individuals from different populations.

In terms of number of strains shared (and still considering only pairs of individuals having ≥10 SGBs typed at the strain level in common), individuals from different datasets shared an average 0.3 oral and 0.5 fecal strains, individuals from different households in the same dataset shared an average 0.8 oral and 2.4 fecal strains, and individuals from the same household shared an average 8.8 oral and 10.6 fecal strains (Figure S1C). These patterns were broadly corroborated by the alternative strain-sharing inference approach SameStr[22] (STAR Methods; Figure S1D), with our StrainPhlAn 4-based methodology comparing slightly favorably with SameStr for the identification of strain-sharing events among cohabitant and sympatric individuals (Figures S1E and S1F).

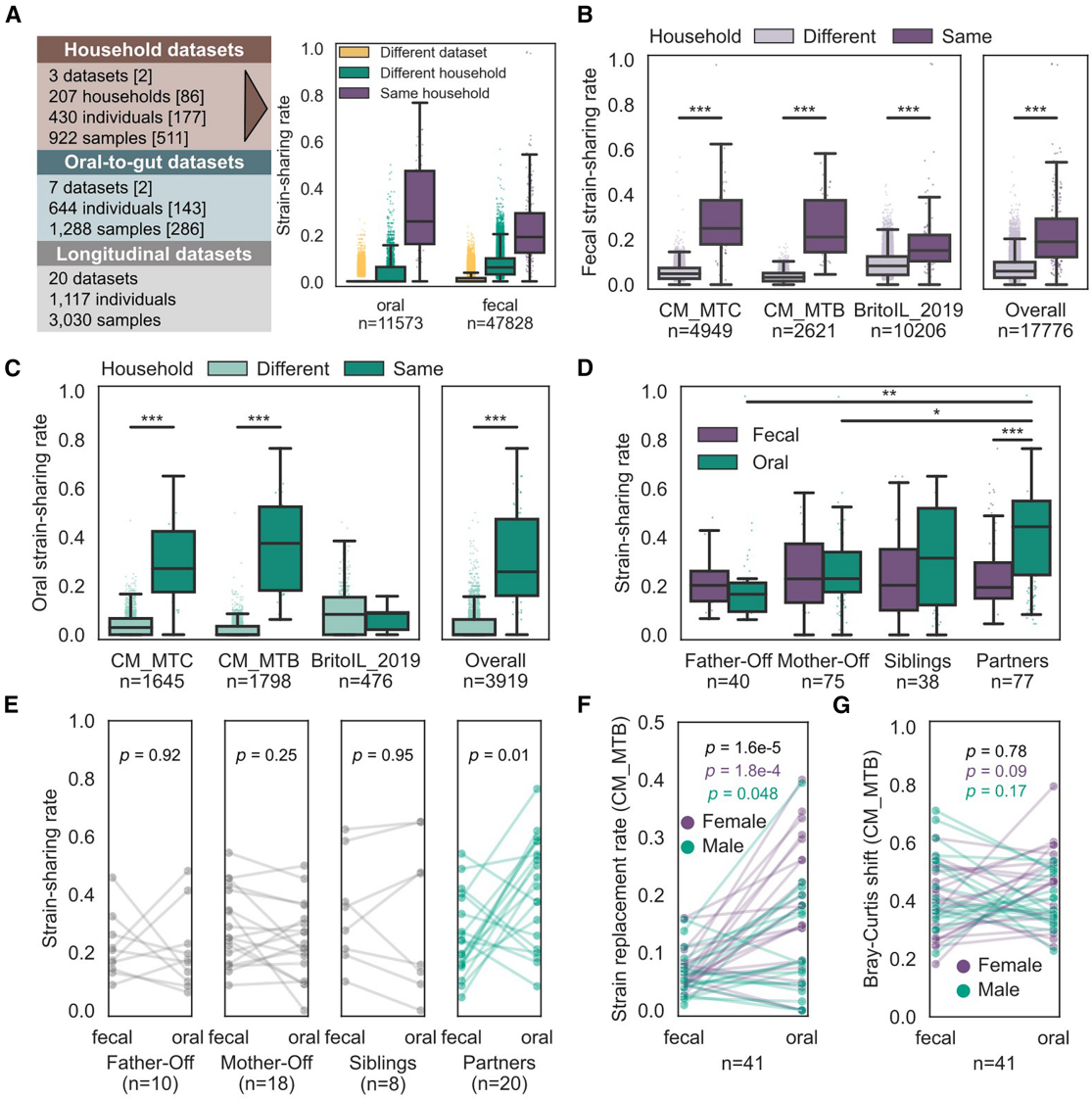


Figure 1. Gut and oral microbiome transmission among cohabitants and longitudinal strain replacement

(A) The different types of datasets (household, oral-to-gut, and longitudinal) leveraged in this study, with indication of total number of households, individuals, and samples. Numbers between squared brackets refer to the data newly added by this study. As an overview of the patterns to be explored in more detail in (B)–(G), strain-sharing rates (SSRs) for same versus different households versus different datasets are represented in boxplots.

(B) Gut microbiome SSRs for same versus different households in each dataset or in all combined (overall). In BritoIL_2019, SSRs between different households in the same village are discarded.

(C) Oral microbiome SSRs for same versus different households in each dataset or in all combined (overall). In BritoIL_2019, SSRs between different households in the same village are discarded.

(D) Gut versus oral microbiome SSRs for different relationship types within households. In (B)–(D), the Mann-Whitney U test was used, with statistical significance reported as $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. The *n* below the *x*-tick labels indicates the number of observations.

(E) Gut versus oral microbiome SSRs for different relationship types within households, with pairing across individuals for which gut and oral microbiome SSRs were both available. The green color in the partners subplot emphasizes the only significant result. In (D) and (E): Father-Off, father-offspring; Mother-Off, mother-offspring.

(F) Gut versus oral microbiome strain replacement rate (STAR Methods) among females (purple), males (green), or overall (black) in the CM_MTB dataset.

(G) Gut versus oral microbiome Bray-Curtis dissimilarity between consecutive time points among females (purple), males (green), or overall (black) in the CM_MTB dataset. In (E)–(G), the Wilcoxon signed-rank test was used.

n indicates the number of individuals. See also Figure S1 and Tables S1, S2, S3, S4, and S5.

Oral strains are much more shared and less longitudinally retained than gut strains

As previously hypothesized that partners can have higher oral microbiome SSRs than other pairs of cohabiting individuals,[9,23] we set out to test whether particular relationship types were associated with differential SSRs. While the portions of gut microbiome sharing between cohabiting partners, siblings, mother-offspring, and father-offspring were comparable (no significant SSR difference between them; Figure 1D; Table S5), partners showed significantly higher oral microbiome SSRs than other relationship types (Figure 1D; Table S5). Interestingly, partners were also the only relationship type with differential (increased) oral versus gut microbiome strain sharing (median 44.4% oral versus 19.5% gut SSR; Mann-Whitney U test, n = 77, U = 261, p = 3.4e−4; Figure 1D), which we confirmed to be true among partners in a paired analysis (including only pairs of subjects for which both oral and gut microbiome SSRs were available; Wilcoxon signed-rank test, n = 20 couples, T = 38, p = 0.011; Figure 1E). Importantly, the estimated strain-sharing false-positive rates (FPRs) (STAR Methods) for oral versus gut SGBs (n = 986, with SGBs classified as oral or gut according to the body site in which they were more frequently typed) were comparable (3.02% median strain-sharing FPR for oral versus 3.03% for gut SGBs; Mann-Whitney U test, n = 986, U = 73,233, p = 0.25; Figure S1G), confirming that finding greater oral microbiome transmission as compared with gut is most likely not a technical artifact. Together, these results not only confirm that cohabitation is a strong driver of strain sharing across all relationship types but also show that among romantic partners there is significantly more extensive oral than gut microbiome strain sharing.

As higher oral versus gut microbiome transmission in partners suggests that strain exchange is linked to more direct accessibility to the external environment, we hypothesized that within-individual strain replacement occurs more prominently in the oral cavity than in the gut. To evaluate that, we leveraged the longitudinal design of CM_MTB, which includes 66 healthy adults (37 females and 29 males) sampled thrice over a period of ~3.5 months (STAR Methods), to compare the strain replacement rate (defined as 1 minus the longitudinal SSR) between gut and oral microbiomes (STAR Methods). Following subject pairing, we found that oral microbiome strains are in general significantly more longitudinally replaced than gut strains (14.7% median oral versus 5.8% gut strain replacement rate; Wilcoxon signed-rank test, n = 41, T = 117, p = 1.6e−5; Figure 1F), which is also observed when analyzing separately females (18.2% versus 6.0% gut strain replacement rate; Wilcoxon signed-rank test, n = 23, T = 27, p = 1.8e−4; Figure 1F) and males (8.2% versus 5.3%; Wilcoxon signed-rank test, n = 18, T = 40, p = 0.048; Figure 1F) or after removing individuals (n = 6) undergoing antibiotic treatment between time points (Figure S1G). Because our strain-sharing inference approach estimates false-negative rates (FNRs) based on the assumption that most strains are longitudinally retained in individuals from external longitudinal datasets (STAR Methods), the observed slightly higher FNR for oral versus gut SGBs is further indirect confirmation of the more intense strain replacement taking place in the oral cavity as compared with the gut (3.57% median strain-sharing FNR for oral versus 0% for gut SGBs; Mann-Whitney U test, n = 986, U = 54,958, p = 0.014; Figure S1H). At the SGB level, however, this effect cannot be observed (0.41 median oral versus 0.39 gut Bray-Curtis compositional shift; Wilcoxon signed-rank test, n = 41, T = 408, p = 0.78; Figures 1G and S1I), illustrating that healthy oral and gut microbiomes can have equally stable species-level configurations but with more intense (dominant) strain replacement dynamics in the oral cavity.

Gut microbiome transmissibility correlates with cardiometabolic health associations

We then estimated the transmissibility score (TS; STAR Methods) of each SGB both generally within household settings (i.e., regardless of relationship types) and more specifically across specific relationship types (e.g., between partners). Out of the 197 gut SGBs for which a household TS was computed, 27 were classified as having high transmissibility (gut TS > $^1/_3$, i.e., for every three pairs of cohabiting individuals sharing an SGB, on average, more than one pair shares the same strain, and significantly higher TS in the household than in the population; Figure 2A; Table S6). These highly transmitted SGBs included the archaeal *Methanobrevibacter smithii* (SGB714; household TS = 0.42), the GABA-consumer *Evtepia gabavorous* (SGB15120; household TS = 0.36),[24] and the GABA-producer *Bifidobacterium adolescentis* (SGB17244; household TS = 0.35).[25] This list of highly transmissible SGBs in the household reasonably overlaps with previous estimates of household gut microbiome transmissibility.[9] Despite transmissibility values not being available in this past study for 6 out of the 27 high-gut-TS SGBs here reported (due to a combination of different populations being assessed and an older version of the SGB database being used there), we found that 11 out of the remaining 21 SGBs were also among the top-transmitted SGBs (TS > $^1/_3$) in this past study (Figure 2A).

In the effort to assess associations between characteristics of gut SGBs and their transmissibility, we inferred their phenotypes with Traitar and compared household TSs between SGBs where each phenotype was either present or absent (Table S7; STAR Methods). We found that household transmissibility was positively associated with being Gram negative, as previously reported and hypothesized to be linked to resistance to sanitizers.[9,27] Transmissibility also correlated significantly with positivity in the Voges-Proskauer test, which indicates the production of neutral pH metabolic end-products such as 2,3-butanediol that stabilize environmental pH, promoting long-term survival by preventing niche acidification,[28] and with the ability to decarboxylate ornithine to putrescine, a process that enhances bacterial fitness through acid stress resistance[29] and may therefore promote survival through the host gastric barrier. *Sutterella wadsworthensis* (SGB9283), the eighth most transmissible SGB among cohabitants in this study (household TS = 0.55), was the only gut species showing all three features significantly associated with transmissibility.

We also evaluated ZOE MB health ranks.[26] We observed that gut SGBs with a high household TS are usually pathobionts associated with poorer cardiometabolic health (Figure 2A). The only notable exception was a health-associated (ZOE MB health rank = 0.13) unknown Oscilospiraceae (SGB15106), also the only

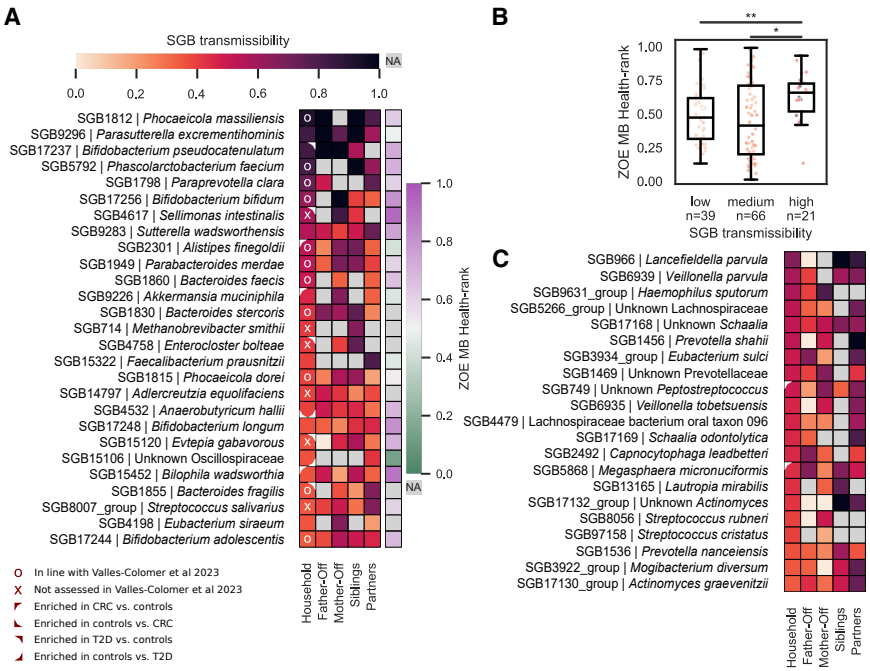


Figure 2. Gut and oral microbiome species transmissibility

(A) Transmissibility scores of gut microbiome SGBs with high transmissibility (STAR Methods) generally in the household and across particular relationship types. Central symbols indicate whether the SGB was previously identified as highly transmissible ("o") or if this was not assessed due to absence of the SGB in the previously used MetaPhlAn 4 database ("x"). Corner symbols indicate whether the SGB is enriched in CRC versus controls or enriched in T2D versus controls. When available, ZOE MB health rank of each SGB is also shown.[26]

(B) ZOE MB health ranks of gut microbiome SGBs with low versus medium versus high household transmissibility (STAR Methods). The Mann-Whitney U test was used with statistical significance represented as $*p < 0.05$ and $**p < 0.01$. The *n* below the *x*-tick labels indicates the number of SGBs.

(C) Transmissibility scores of oral microbiome SGBs with high transmissibility (STAR Methods) generally in the household and across particular relationship types. In (A) and (C), transmissibility across particular relationship types is not available (NA) for cases with <3 SGB-sharing pairs (STAR Methods). In (A) and (C): Father-Off, father-offspring; Mother-Off, mother-offspring.

See also Figure S2 and Tables S6, S7, S8, and S9.

to-be-characterized SGB among these highly transmissible gut microbiome members. In addition, when expanding the association analysis to all the 126 SGBs for which a ZOE MB health rank and a gut household TS was available, we found that SGBs with a high TS (TS > $^1/_3$) have significantly higher ranks (i.e., are associated with poorer cardiometabolic health) than SGBs with low TS (TS < ⅙, regardless of their population transmissibility; Mann-Whitney U, $n = 60$, $U = 233$, $p = 0.006$; Figure 2B) and than SGBs with medium TS (⅙ ≤ TS ≤ $^1/_3$, regardless of their population transmissibility; Mann-Whitney U, $n = 87$, $U = 453$, $p = 0.018$; Figure 2B), which is also consistent—albeit noisier—across particular relationship types (Figure S2A).

We further evaluated the relationship between gut microbiome transmissibility and health by associating SGB household TS with biomarker SGBs from disease association meta-analyses. We first leveraged results from a recently published meta-analysis on the gut microbiome of colorectal cancer (CRC) in comparison with healthy matched controls,[30] analyzing 47 significant ($q < 0.1$) biomarker SGBs for which household TSs were available. When associating enrichment direction (i.e., enriched in CRC or in controls) with household TS categories (i.e., low, medium, or high), we found no significant association between transmissibility and CRC/controls enrichment (Fisher-Freeman-Halton test, $p = 0.45$). Noteworthily, three high household transmissibility SGBs were significant CRC-enriched biomarkers, including *Akkermansia muciniphila* (SGB9226), while *Anaerobutyricum hallii* (SGB4532) was significantly enriched among controls (Figure 2A).

Next, to evaluate transmissibility links with a cardiometabolic disease, we performed a similar meta-analysis in the context of type 2 diabetes (T2D; STAR Methods). Among 99 significant ($q < 0.1$) biomarkers, T2D-enriched SGBs ($n = 15$) were significantly overrepresented in the high transmissibility category (5 out of 7 significant SGBs; 71%) compared with the medium (8 out of 62 significant SGBs; 13%) and low (2 out of 30 significant SGBs; 7%) transmissibility categories (Fisher-Freeman-Halton test, $p = 3.0e{-}4$). T2D-enriched SGBs included *Sellimonas intestinalis* (SGB4617), the most strongly associated with poor cardiometabolic health among highly transmissible SGBs (ZOE MB health rank = 0.93); conversely, the unknown Oscillospiraceae (SGB15106) previously highlighted for its strong association with good cardiometabolic health was found to be controls enriched. Despite a negative association with cardiometabolic health markers (ZOE MB health rank = 0.69), *Anaerobutyricum hallii* (SGB4532) was—as observed in the CRC meta-analysis—significantly enriched among controls, suggesting that the implications of high transmissibility can be multi-faceted and context dependent, although causal links remain to be determined.

Opportunistic pathogens and colorectal cancer markers among highly transmissible oral species

We also computed the household transmissibility for 51 oral SGBs, 20 of which were classified as having high transmissibility (Figure 2C; Table S8). Interestingly, a significantly greater proportion of SGBs are highly transmissible in the oral microbiome as compared with the gut microbiome (highly transmitted SGBs in the oral microbiome 41.2% versus 13.7% in the gut microbiome; Pearson's chi-squared test, $n = 248$, $\chi^2 = 17.9$, $p = 2.4e{-}05$), in line with the higher transmissibility of the oral microbiome previously reported (Figures 1D and 1E). The only SGB with household TS computed for both oral and gut microbiomes (as this depends on both environment-dwelling strains being

reconstructed in a sufficiently high number of individuals; STAR Methods) was the oral commensal *Streptococcus salivarius* (SGB8007_group), which, despite a higher prevalence in the oral cavity (97.8% oral versus 80.7% gut prevalence), shows higher transmissibility of gut strains (oral TS = 0.20 versus gut TS = 0.35), although not significantly (Pearson's chi-squared test, $n = 105$, $\chi^2 = 2.57$, $p = 0.11$).

Although no microbial phenotypes were significantly linked to oral microbiome transmissibility (Table S9), and despite cardiometabolic health ranks not being available for oral species, we noticed that several opportunistic pathogens were among the most transmissible oral SGBs, including *Haemophilus sputorum* (SGB9631_group; household TS = 0.57), *Schaalia odontolytica* (SGB17169; household TS = 0.45), *Streptococcus cristatus* (SGB97158; household TS = 0.40), and *Actinomyces gravenitzii* (SGB17130_group; household TS = 0.35). Consistent with the enrichment of several oral species in the gut microbiome of CRC patients,[30] the list of highly transmissible oral SGBs also included two CRC-enriched gut biomarkers (*Megasphaera micronuciformis* [SGB5868] and an unknown *Peptostreptococcus* [SGB749]; Figure 2C).

Oral-gut microbiome overlap is mostly due to oral-to-gut transmission

Next, we considered matched oral and fecal samples to evaluate the within-individual overlap between oral and gut microbiomes. To this end, in addition to the individuals with matched gut and oral microbiomes from the three datasets previously described, we evaluated healthy individuals with matched fecal and oral samples from four additional studies[1,18–20] (Figure 1A), resulting in a total of 644 oral-gut pairs with enough sequencing depth (STAR Methods).

Analyzing MetaPhlAn 4 taxonomic profiles, we identified a total of 814 SGBs in oral samples and 2,659 SGBs in fecal samples present in at least two individuals. Only 4% ($n = 127$) of the total SGBs were detected at prevalence >1% (i.e., >6 individuals) in both body sites, and for those we noted higher prevalence in the oral cavity ($n = 103$, 81.1%; Figure 3A). Considering matched oral-gut pairs, we found that a median of 7 SGBs (range: 0–82) could be detected in both body sites, while a median of 140 SGBs (13–365) were detected only in the oral cavity (or 93.96% of the median oral cavity richness) and a median of 195 SGBs (34–765) were detected only in feces (or 95.12% of the median gut richness). By factoring in the abundance of those SGBs in each body site to better grasp their relative importance in each body site, we found that the cumulative relative abundance of overlapping SGBs was much higher in the oral cavity (median oral 17.8% versus 0.4% median gut; Wilcoxon signed-rank test, $n = 644$ individuals, $T = 12{,}586$, $p < 1e{-}20$; Figure 3B), with only 6.1% ($n = 39$) of subjects showing higher abundance of overlapping SGBs in feces.

We also found that SGBs detected in both body sites had a higher relative abundance rank (RAR) in the oral cavity than oral-only SGBs (median RAR in the oral cavity when detected in both body sites 26 versus 83 when oral only; Mann-Whitney U, $n = 101{,}272$, $U = 1{,}479{,}521{,}166$, $p < 1e{-}20$; Figure 3C) but a lower RAR in feces than gut-only SGBs (median RAR in feces when detected in both body sites 145 versus 122 when gut only; Mann-Whitney U, $n = 151{,}324$, $U = 493{,}105{,}919$, $p < 1e{-}20$; Figure 3D). This indicates that SGBs detected in both environments are among the most abundant in the oral cavity and among the least abundant in feces. Together, these results are highly suggestive of saliva-mediated oral-to-gut transmission being a major factor responsible for the observed overlap, but strain-level resolution is needed to confirm this hypothesis.

Oral species in the gut comprise mainly oral-derived strains but also gut-specific ones

We then used our strain-sharing inference framework to evaluate the instances in which SGB sharing between the oral cavity and the gut is the consequence of strain sharing. We managed to reconstruct the genetic signature via StrainPhlAn for strains dwelling in both body sites in 4.3% of the 6,161 single cases in which an oral-gut SGB overlap was detected. Indeed, at the individual level, the number of strain-level-reconstructed overlapping SGBs ranged between 0 and 7, despite this range being 0–82 overlapping SGBs based on SGB-level data, reflecting the different coverage requirements for strain- and species-level profiling. Among the 267 cases of conspecific strains being reconstructed in both body sites, in 74.5% ($n = 199$) the same strain was present, with the number of detected oral strains translocated to the gut per individual ranging between 0 and 6 (average 1.22), likely an underestimation linked to limited coverage.

To better evaluate this, we analyzed how the number of oral-typical SGBs (see STAR Methods) detected in the gut (based on taxonomic profiles) varied in relation to the number of strains found in common between oral cavity and feces. We found them to be strongly positively linked (*post hoc* Dunn, $p \leq 0.013$; Figures 3E; Table S10), with an average of seven oral-typical SGBs detected in feces for every new oral strain detected, suggesting that the greater overlap observed at SGB level would likely be confirmed at strain level if sufficient coverage were reached for lower-abundance SGBs in feces.

Similar to what was observed at the SGB level (Figure 3C), oral strains also detected in the gut were more typically among the most abundant in the oral cavity (mode = 1st, count = 27) than oral strains for which a distinct conspecific strain was detected in the gut (median RAR for same strain 5 versus 7 for different strain; Mann-Whitney U, $n = 267$, $U = 8{,}063$, $p = 0.018$; Figure 3F). Those translocated oral strains have a slightly lower RAR in the gut than gut strains conspecific with but distinct from oral strains, although not significantly (median RAR for same strain 24 versus 23 for different strain; Mann-Whitney U, $n = 267$, $U = 6{,}259$, $p = 0.36$; Figure 3G). Considering that microbiome abundance estimates are contingent on the microbiome diversity of the population investigated and influenced by batch effects, we investigated this association further through a meta-analysis ($n = 6$ datasets) ran directly on strain relative abundances (instead of RARs). Specifically, we compared the abundance of a strain in a given body site based on whether the same strain or a different strain of the same SGB was detected in the other body site. Again, we found higher abundance in the oral cavity of strains also detected in feces (as compared with oral strains not detected in feces but for which a distinct conspecific strain was present in the gut) (Hedges' $g = 0.30$,

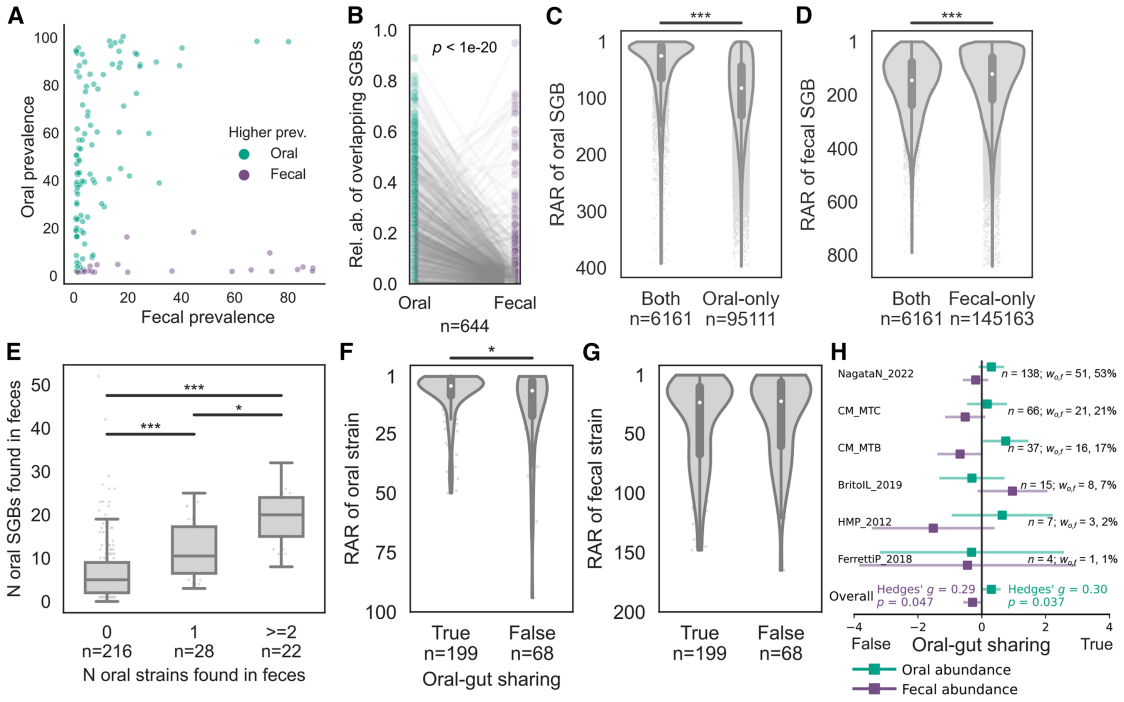


Figure 3. Oral-gut microbiome overlap

(A) Oral versus gut microbiome prevalence for the 127 SGBs present at >1% prevalence in both body sites, with indication of whether the SGB has higher prevalence in the oral cavity (green) or the gut (purple).

(B) Cumulative relative abundance in the oral versus the gut microbiome of the 127 overlapping SGBs for 644 individuals with paired samples available. The Wilcoxon signed-rank test was used.

(C) Relative abundance rank of oral microbiome SGBs detectable in both body sites versus only in the oral cavity.

(D) Relative abundance rank of gut microbiome SGBs detectable in both body sites versus only in the gut. In (C) and (D), the Mann-Whitney U test was used with statistical significance represented as ***$p < 0.001$. The *n* below the *x*-tick labels indicates each observation within the 644 individuals with paired samples available.

(E) Number of oral-typical SGBs found in feces in cases in which 0 versus 1 versus ≥2 oral-derived strains are found in feces. The *post hoc* Dunn test for Kruskal-Wallis was used, with statistical significance represented as **$p < 0.01$ and ***$p < 0.001$. The *n* below the *x*-tick labels indicates the number of individuals, restricting the analysis to those not used in the definition of oral-typical SGBs.

(F) Relative abundance rank of oral microbiome strains when the same strain is detected in the gut (true) versus when a different conspecific strain is detected in the gut (false).

(G) Relative abundance rank of fecal microbiome strains when the same strain is detected in the oral cavity (true) versus when a different conspecific strain is detected in the oral cavity (false). In (F) and (G), the Mann-Whitney U test was used, with statistical significance represented as *$p < 0.05$. The *n* below the *x*-tick labels indicates each observation within the 644 individuals with paired samples available.

(H) Meta-analysis of strain abundances across oral-to-gut datasets including 267 cases overall in which conspecific strains were detected in both body sites (STAR Methods). The abundances of the strains in a given body site (green for the oral cavity and purple for the gut) are compared between situations in which the same strain is detected in the other body site versus when a different conspecific strain is detected in the other body site. The number of observations for each dataset and their weight in oral (w_o) and fecal (w_f) abundance meta-analyses are indicated respectively. Error bars indicate the 95% confidence interval of the effect size.

See also Tables S10 and S11.

$p = 0.037$; Figure 3H; Table S11) and a lower abundance in feces of strains also detected in the oral cavity (as compared with gut strains not detected in the oral cavity but for which a distinct conspecific strain was present in the mouth) (Hedges' $g = -0.29$, $p = 0.047$; Figure 3H; Table S11). The latter also means that, among oral-gut shared SGBs, gut-specific strains have higher abundance in stool than oral-gut shared strains.

Overall, our data support that observation of oral commensals in feces is mainly due to oral-to-gut strain transmission rather than independent acquisition of niche-specific strains in each body site, with high-abundance oral strains being the ones reaching the gut at detectable levels. At the same time, distinct conspecific strains can occasionally be found in each body site, with higher abundance in feces of those strains than of oral-to-gut transmitted strains.

Evidence of oral species with strains established in the gut environment

Overall, we could reconstruct conspecific strains in matched oral-fecal samples for 22 distinct SGBs. Reflective of its relatively higher abundance across both body sites, the most frequent was *S. salivarius* (SGB8007_group), reconstructed from both oral and fecal samples of 138 individuals, 68% of which displayed the same *S. salivarius* strain in both body sites (Figure 4A). This percentage—which, akin to the SGB inter-individual TS, can be interpreted as an SGB intra-individual oral-to-gut TS (OGS)—is

meaningful only for SGBs with strains reconstructed in both body sites of a reasonably large number of individuals (≥10) and ranged between 0.68 and 0.93 for the four SGBs for which this score was computed. Surprisingly, only two of them were among the most-household-transmitted SGBs (TS > $^1/_3$; Figures 2A–2C): *Veillonella parvula* (SGB6939) in the oral microbiome (oral TS = 0.59; Figure 2C) and *S. salivarius* (SGB8007_group) in the gut microbiome (gut TS = 0.36; Figure 2A). The highly oral-to-gut-transmitted (OGS = 0.89) *Streptococcus parasanguinis* (SGB8071) had the lowest (out of 51 oral SGBs) computed household TS (oral TS = 0.13), indicating that high odds of oral-to-gut transmission for a given SGB might not be linked with inter-individual SGB transmissibility.

Despite the higher microbiome diversity in the gut compared with the oral cavity, leading to a greater number of low-relative-abundance microbes (Figures S3A and S3B), there were 32/267 cases in which the gut strain of a given SGB had greater relative abundance than the oral counterpart (Figure 4A). Interestingly, this happened significantly more often when different conspecific strains were present in the oral cavity and feces (50.0% different versus 22.1% same strain; Pearson's chi-squared test, $n = 267$, $\chi^2 = 10.1$, $p = 0.001$). Even after filtering out observations related to the two SGBs with higher prevalence in feces (SGB17248 and SGB17234; Figure 4A), this association remained statistically significant (Pearson's chi-squared test, $n = 261$, $\chi^2 = 7.2$, $p = 0.007$). Taken together as a testament to their establishment in the gut environment, gut strains that are conspecific with but different from oral strains in the same individual not only have higher abundance than oral-to-gut-transmitted strains (Figure 3H) but often have higher abundance than their oral counterpart, which holds even when considering only typically oral SGBs.

Identification of an oral-specific *Bifidobacterium longum* subspecies

Among the 22 SGBs involved in the oral-gut overlap (Figure 4A), there were also SGBs for which finding the same strain in both body sites was infrequent, as better evidenced by the case of *B. longum* (SGB17248). This SGB was the only among the 22 with considerably higher (>8×) prevalence in feces (9.2% oral versus 73.8% gut prevalence; Figure 4A) and, despite being among the most-household-transmitted gut SGBs (gut TS = 0.43; Figure 2A), was never present as the same strain in both body sites (out of five cases). Since this observation could be linked to the existence of distinct *B. longum* subspecies inhabiting each body site, we contextualized the *B. longum* strains reconstructed here with metagenomically reconstructed *B. longum* strains from additional publicly available oral and food metagenomic samples and reference genomes for each previously described *B. longum* subspecies (STAR Methods). Those included the foundational subspecies *longum*, *infantis*, and *suis*[31]; the later described subspecies *suillum*[32] and *iuvenis*[33]; as well as the very recently proposed subspecies *nexti*.[34]

The phylogenetic analysis confirmed that the oral and gut strains of *B. longum* fall in clear-cut distinct clades (Figure 4B). Moreover, we found that all gut strains (along with strains of documented food origin[35]) fall within the *B. longum* subsp. *longum* cluster (except one gut strain closer to *B. longum* subsp. *suis*), while oral strains seem to compose their own subspecies, clustering with a genome of the provisionally named *B. longum* subsp. *nexti*.[34] As StrainPhlAn 4 relies on SGB-specific marker genes, we further validated the existence of this oral-related subspecies by assessing the average nucleotide identity between full *B. longum* genomes reconstructed from the same samples or retrieved from the Human Reference Oral Microbiome (HROM) database[36] (STAR Methods). This analysis, which spanned eight oral metagenomic datasets across five countries from three continents, revealed the same subspecies structure observed in the phylogeny, including the clear presence of an oral-specific clade (Figure S4). Together, our data suggest that the recently proposed *B. longum* subsp. *nexti* is an undercharacterized oral-dwelling *B. longum* subspecies.

Inter-individual oral-gut transmission is detectable and usually mediated by intra-individual oral-to-gut transmission

Finally, having explored the transmission landscape of gut and oral microbiomes separately and having described the extent of oral-to-gut strain transmission, we analyzed whether intra-household transmission could also take place across different body sites, which was done by identifying strains present in the oral cavity of a household member and in the feces of another household member. Overall, we found very few examples of this kind of strain sharing ($n = 33$ total events, 1.2% of the intra-household inter-subject strain-sharing events) but significantly more than by chance, as evidenced by the higher proportion of times the same strain was detected in different body sites of co-habiting individuals when they had strain-level reconstructed SGBs in common (i.e., the SGB SSR) versus the same value for non-cohabiting individuals (22.8% same household versus 3.4% different households; Pearson's chi-squared test, $n = 11{,}381$, $\chi^2 = 113.9$, $p < 1e{-}20$; Figure 4C; Table S12).

We then evaluated if the inter-individual oral-gut strain sharing was simply due to oral microbiome transmission followed by oral-to-gut transmission in the recipient or if also direct inter-individual oral-to-gut transmission (i.e., without stable colonization of the oral cavity in the recipient) was possible. To evaluate this hypothesis, we looked for oral microbiome sharing between the same pair of individuals involved in the 33 oral-gut strain-sharing events. Out of the 26 pairs with oral samples available for both individuals, we successfully reconstructed the strain in the oral cavity of both individuals in 23 cases (with the SGB still being detected in the remaining 3 cases but not with enough coverage for reconstruction of the strain). In the majority of cases (18/23), we could detect the oral microbiome strain-sharing event, supporting a model of inter-individual oral-microbiome transmission followed by intra-individual oral-to-gut transmission to explain the oral-gut strain-sharing event between cohabiting individuals (Figure 4D).

However, this also means that direct oral-to-gut (or gut-to-oral) strain transmission is detectable, with 5/23 cases of an oral strain being shared with the gut but not with the oral cavity of another individual, which contained a distinct conspecific strain (Figure 4D). Specifically, we found two couples in CM_MTB in which a strain of *Veillonella atypica* (SGB6936) in

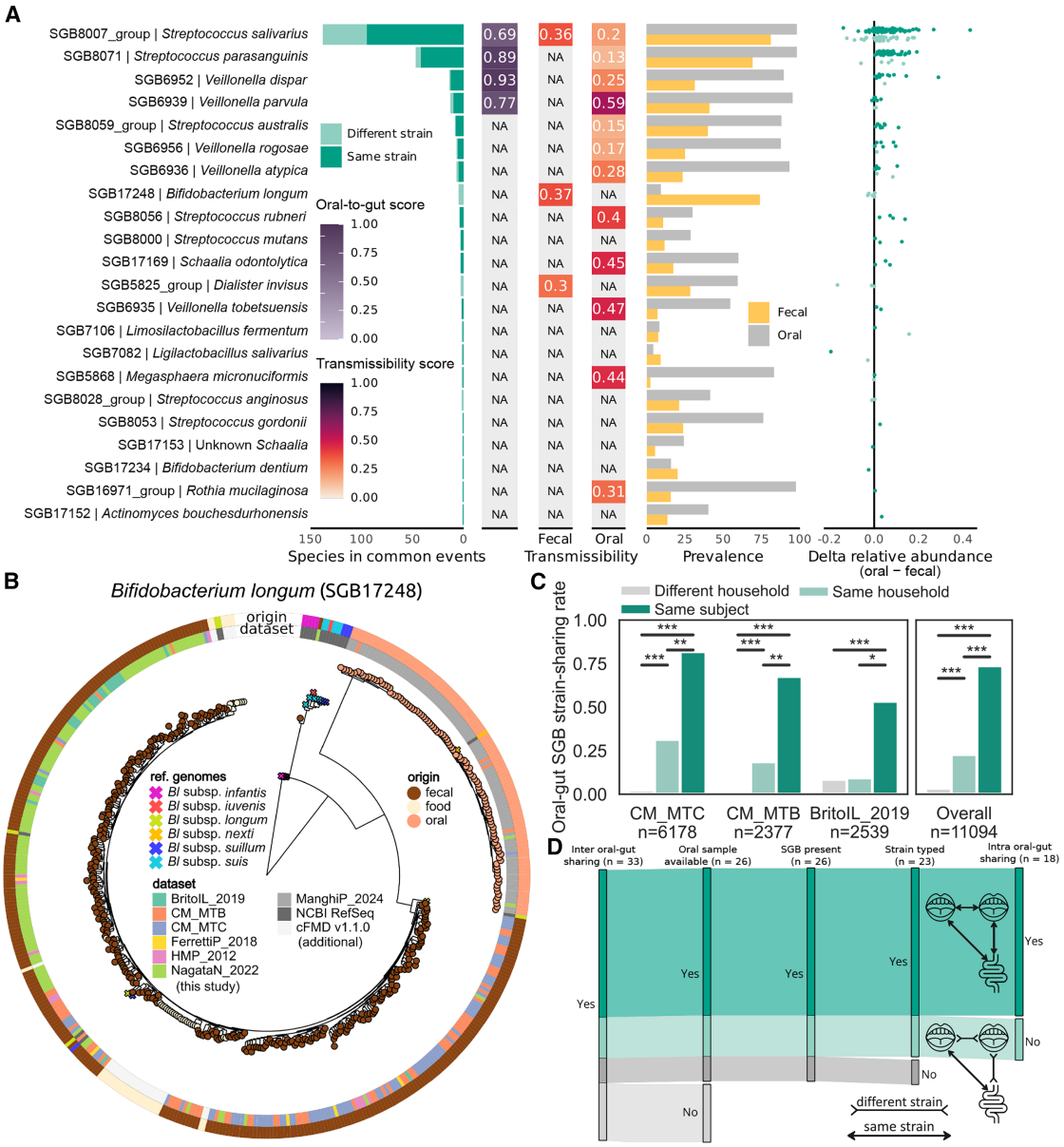


Figure 4. Strain transmission across body sites

(A) List of SGBs typed at strain level in both body sites of the same individual, with the indication of how many times the same or a different strain was found (also summarized as the SGB oral-to-gut transmissibility score). For context, the SGB household transmissibility and the prevalence in each body site are shown. Finally, the delta in the relative abundance between body sites (oral *minus* fecal) is shown for cases in which the same strain versus different strains were found in each body site. For both oral-to-gut and household transmissibility scores, values are not available (NA) for cases with <10 SGB-sharing pairs (STAR Methods).

(B) *Bifidobacterium longum* (SGB1724) StrainPhlAn 4 phylogenetic tree. This tree includes strains reconstructed from metagenomes analyzed in this study and from metagenomes from additional datasets used specifically to enrich the tree (ManghiP_2024 and cFMD v.1.1.0), as well as strains from isolate genomes representing all described *B. longum* subspecies.

(C) Proportion of same strain cases for all the cases in which the same SGB was typed at strain level in both body sites of different individuals from different households versus different individuals from the same household versus the same individual. The Pearson's chi-squared test was used, with statistical significance represented as $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

(D) Schematic representation of inter-individual inter-body site household transmission and its mediation by oral-to-gut transmission. Out of the 33 cases of oral-gut sharing across different individuals, the oral sample for the individual having the strain in the gut was available in 26 cases, all of which with the SGB involved in the strain-sharing event present in taxonomic profiles. Out of these 26 cases, in 23 the SGB could be typed at strain level (in the oral cavity of the other individual). Out of these 23 cases, in 18 the same strain involved in the across-individual oral-gut sharing was detected in the oral cavity of the other individual (pointing to inter-individual oral-gut transmission mediated by intra-individual oral-to-gut transmission), with the remaining 5 individuals having a different oral strain.

See also Figures S3 and S4 and Table S12.

one case and of *Veillonella dispar* (SGB6952) in the other was found in the oral cavity of a participant and in the feces of their partner, who had a distinct conspecific strain in the oral cavity. We also identified an *S. salivarius* (SGB8007_group) strain in CM_MTC being shared between the oral cavity of an 18-year-old female (or her 50-year-old father) and the feces (but not the oral cavity, which had a distinct conspecific strain) of her 16-year-old brother. Also in CM_MTC, the same was observed between 7-year-old female twins, which in this case, besides oral-gut sharing, also showed gut-gut sharing of the same *Veillonella parvula* (SGB6939) strain, despite a distinct conspecific strain being present in the oral cavity of one of the twins. As all the SGBs involved are typically oral, a plausible model to explain those events is one in which an oral strain was transferred to the oral cavity of another individual, did not colonize this environment (due to the previous presence of a distinct conspecific strain occupying the same niche), but then engrafted in the gut of the recipient.

DISCUSSION

In this study, we investigated transmission patterns of oral and gut microbiomes in the context of individuals sharing households. As observed previously with independent oral and gut microbiome datasets,[9] we confirmed the elevated oral and gut microbiome SSRs among cohabiting individuals. Our unique study design with oral and gut samples from the same cohabitants allowed us to reveal additionally a higher transmissibility of the oral versus the gut microbiome among romantic partners, who also showed higher oral microbiome transmissibility compared with other relationship types. As previously proposed,[9,23] these observations point to saliva exchange due to intimate kissing habits as a key factor in shaping the healthy oral microbiome in addition to their notorious role as a mode of transmission of pathogens.[37] Therefore, as kissing is an ancient and universal human habit,[38] its extensively debated biological impact throughout human history, such as through mate-pairing mediated by chemical cues or as a facilitator of sexual arousal,[39] additionally includes a likely profound role in influencing the development and evolution of the oral microbiome.

The oral microbiome also showed more intense strain replacement dynamics compared with the gut microbiome, despite similar species-level stability. Comparable species-level stability can be explained by the maintenance of niches in both body sites over time, as volunteers in the analyzed studies did not generally undergo major lifestyle alterations between assessed time points, such as dietary shift or antibiotic use. The higher strain replacement dynamics (as well as the higher transmissibility of the oral microbiome) might be explained by the exposure of the oral cavity to the environment and the corresponding higher aerotolerance of oral microbiome members compared with gut microbes enabling a more continued inflow of competitive strains. This frequent competition for the available niches is likely in contrast to what is observed in the gut, where outcompeting resident strains requires the daunting task of surviving aerobic conditions and reaching the anaerobic intestinal environment alive.[7]

Reaching the gut and colonizing a niche is arguably an even bigger challenge for gut microbiome species associated with favorable host cardiometabolic markers, as we demonstrated that they are usually not among the most highly person-to-person transmitted. This list of highly transmitted species in households, which is consistent with those we previously reported,[9] was rather enriched by gut microbiome members associated with T2D and more generally negatively associated with cardiometabolic health.[26] This observation raises the question of whether, akin to pathogens, microbiome species correlated with a given disease have particular mechanisms for effective spread or receive less colonization resistance. That being the case, microbiome modulation strategies aiming to promote colonization by beneficial species will have to account for their potential inherent disadvantages in colonization-resulting transmissibility.

Our unique household datasets with paired oral and fecal metagenomes were further augmented by additional cohorts with this characteristic and allowed us to provide a strain-level landscape of the oral-gut microbial overlap during health. We found little overlap between paired oral and gut microbiomes (4% of species in common), but we noticed that, among the few cases of commonly present species, the same strain was present in both body sites in three out of four cases, remarkably similar to a previous estimate.[13] Importantly, analysis of species prevalences and strain abundances in each body site confirmed that detection of the same microbiome species in the oral cavity and in the gut can be in most cases explained by saliva-mediated translocation as opposed to other intra-individual transmission modes (e.g., fecal-oral route or bloodstream translocation). This observation is especially relevant as the study of oral-gut microbiome connections gains traction,[2] as recently exemplified by the finding of several oral species among the top gut microbiome markers of CRC.[30] Indeed, although our study focuses on healthy subjects aiming to provide baseline estimates for that phenomenon, higher levels of true or perceived (due to lower absolute abundance of gut microbes[40]) oral-to-gut microbiome transfer are to be expected during disease,[13,15,41] likely due to compromised gastric (linked to medication use) and ecological (linked to oral/gut microbiome dysbiosis) barriers.

A notable exception to this trend involved *B. longum*, a species detected in both body sites but never as the same strain. Surprisingly, we found that to be the case due to the presence of different *B. longum* subspecies in each body site. While *B. longum* subsp. *longum* was found, as expected, in the gut, the recently proposed *B. longum* subsp. *nexti*[34] (also described as *B. longum* subsp. *X/leicesterensis*[42]), although first isolated from infant stool, seems to have particular tropism for the oral cavity (in line with previously described enrichment in non-gut environments[42]), making it the first of its kind among human-dwelling *B. longum* subspecies. Given its low prevalence (9.2%) and very low abundance (median 0.03% relative abundance when present) in the oral microbiome, this subspecies has likely remained elusive in previous surveys, requiring high sequencing depth as in the ManghiP_2024[43] dataset for its detection. The case of *B. longum* subsp. *nexti*, alongside the identification of gut-specific strains of oral-typical species, underscores that strain-sharing analysis extends beyond mapping microbiome transmission. Sharing patterns immediately highlight the potential existence of functional genetic adaptations

constraining or facilitating microbial spread. Characterizing the drivers of this niche specificity, such as the functional characteristics allowing *B. longum* subsp. *nexti* to establish in the oral cavity, should be a goal of future investigations. These efforts will likely benefit from the integration of metagenomic surveys and cultivation-based functional analyses,[44] an approach that has proven successful in defining the genetic basis for the oral-to-CRC transition of *Fusobacterium nucleatum* clades.[45]

Beyond oral-to-gut strain transmission, we also found for the first time evidence of oral-gut strain sharing between cohabiting individuals. This can usually (in at least 78% of the cases) be explained by inter-individual oral microbiome transmission followed by intra-individual oral-to-gut transmission, but we also found indications that a transmitted strain can colonize a different body site of another individual without colonizing at detectable levels its original niche in the recipient. Such passage through the oral cavity without detectable colonization, with subsequent engraftment and expansion in the gut, is an illustration of the mechanism that is likely behind the transmission of gut microbiome strains.

In summary, our strain-level investigation of microbiome transmission in households confirms previous observations and provides further insights into the intra- and inter-individual connectedness of human microbiomes. We have described the high extent to which cohabitation, especially when close interpersonal contact is present, can influence the composition and dynamics of the gut and the oral microbiome. In line with the hypothesis that human microbiome transmission can have direct impacts on host health, we have also revealed a link between microbiome transmissibility and cardiometabolic risk. Furthermore, we have investigated oral-gut strain-sharing with unprecedented resolution, confirming the predominant role of saliva in mediating oral-to-gut transmission and revealing the oral tropism of *B. longum* subsp. *nexti.* Overall, our findings delineate the power of high-resolution strain-sharing analysis in establishing a framework for understanding how human interactions and microbial features come together to dictate microbiome transmission and its consequences for human health.

Limitations of the study

Although strain sharing and strain transmission are two directly linked phenomena, it is generally not possible to disentangle strain sharing between two individuals from direct transmission between them without very densely longitudinally sampled microbiomes, let alone to infer transmission directionality. In addition, similar to most strain-level metagenomic investigations, our approach can identify only the most abundant strain of each microbiome species in a given metagenomic sample, as previously detailed.[46] Although most human microbiome species are expected to be present as single-strain populations,[47] we might have missed cases of subdominant strain transmission, meaning that SSRs among cohabiting individuals could be even higher. Finally, although we excluded infants and young toddlers (age <2 years) from our analyses to focus on horizontal rather than vertical transmission (defined as mother-to-infant transmission happening during birth and soon thereafter[7]), many of the remaining same household pairs in our study were of the type mother-offspring (25.7%, although only 10.5% of the total involved mothers and children aged <12 years), cases in which it is impossible to tell early-life acquisition from the mother followed by long-term retention (expected to be extensive[9]) apart from household transmission later in life. Still, we reinforce that the majority (78%) of our same-household participant pairs do not involve children, even after excluding partners (61%), indicating that the differential transmission patterns observed among partners versus other relationship types is unlikely driven by age structure.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Nicola Segata (nicola.segata@unitn.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Shotgun metagenomic data of newly sequenced datasets (CM_MTB: PRJNA1140720; CM_MTC: PRJNA1308818) and public datasets (BengtssonPalmeJ_2015: PRJEB7369; BritoIL_2019: PRJNA217052; CosteaPI_2017: PRJEB17632; FerrettiP_2018: PRJNA352475; GhensiP_2020: PRJNA547717; HallAB_2017: PRJNA385949; HansenLBS_2018: PRJNA491335; HMP_2012: PRJNA48479; HMP_2019_ibdmdb: PRJNA398089; HMP_2019_t2d: PRJNA497499; KartalE_2022: PRJEB38625 and PRJEB42013; KosticAD_2015: PRJNA231909; LeChatelierE_2013: PRJEB4336; LouisS_2016: PRJNA290729; MehtaRS_2018: PRJNA354235; NagataN_2022: PRJNA832909; NielsenHB_2014: PRJNA32811; PehrssonE_2016: PRJNA300541; RaymondF_2016: PRJEB8094; VatanenT_2016: PRJNA290380; VincentC_2016: PRJNA297252; YassourM_2016: PRJEB1690; and YassourM_2018: PRJNA475246) are available at the NCBI-SRA. *B. longum* MAGs and reference genomes analyzed in this study either are available at Zenodo (https://doi.org/10.5281/zenodo.18155149) or can be retrieved from the HROM database (Table S13).
- All the software used in this study is available as open-source code in Zenodo (https://doi.org/10.5281/zenodo.19682711) and in the 4.1.0 release of the maintained MetaPhlAn 4 repository (https://github.com/biobakery/MetaPhlAn), including StrainPhlAn 4 and the script for strain transmission inference. Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC-CoG microTOUCH-101045015) and by European Union – NextGenerationEU – PNRR POC-2023-12377196 (CUP E63C24000650003) to N.S. and by the PNRR 2023 (project PNRR–POC–2023–12377319) of the Italian Ministry of Health, by the Next Gen Clinician Scientist 2024 of the AIRC (project 30203), by the Fondo Italiano per la Scienza of the Italian Ministry of Research (project FIS00001711), and by the European Research Council (project ERC-StG MicroRestore-101221279) to G.I. The staff of the Fondazione Policlinico Gemelli IRCCS thanks the Fondazione Roma for the invaluable support for their scientific research and is supported by the Ricerca Corrente 2024 of the Italian Ministry of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank B. Servaes for designing the graphical abstract.

AUTHOR CONTRIBUTIONS

N.S., G.I., and V.H. designed the study. L.R. managed cohort recruitment and sample acquisition from CM_MTB. G.C., D.R., and S.P. managed cohort

recruitment and sample acquisition from CM_MTC. V.H. performed the computational analyses with support from G.F., R.S., G.B., M.P., G.P., F.A., and P. Marchi. V.H. and N.S. interpreted the data, with support from M.V.-C., M.M., and P. Mattarelli. V.H. and N.S. wrote the manuscript. All authors revised the manuscript and approved its final version before submission.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Description of newly sequenced cohorts
  - Survey of public metagenomic datasets
- METHOD DETAILS
  - Sample collection, DNA extraction, and metagenomic sequencing
  - Metagenomic sequences quality control
  - Species- and strain-level microbiome profiling
  - Strain-sharing assessment
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Strain-sharing rate, strain replacement rate, and SGB transmissibility
  - Evaluation of strain-sharing with SameStr
  - Definition of oral-typical SGBs
  - *Bifidobacterium longum* (SGB17248) analysis
  - Type 2 diabetes meta-analysis
  - Traitar SGB profiling and association with household transmissibility
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cpblue.2026.100034.

Received: September 26, 2025
Revised: February 27, 2026
Accepted: May 11, 2026
Published: June 15, 2026

REFERENCES

1. Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214. https://doi.org/10.1038/nature11234.
2. Kunath, B.J., De Rudder, C., Laczny, C.C., Letellier, E., and Wilmes, P. (2024). The oral-gut microbiome axis in health and disease. Nat. Rev. Microbiol. *22*, 791–805. https://doi.org/10.1038/s41579-024-01075-5.
3. Gilbert, J.A., Quinn, R.A., Debelius, J., Xu, Z.Z., Morton, J., Garg, N., Jansson, J.K., Dorrestein, P.C., and Knight, R. (2016). Microbiome-wide association studies link dynamic microbial consortia to disease. Nature *535*, 94–103. https://doi.org/10.1038/nature18850.
4. Baker, J.L., Mark Welch, J.L., Kauffman, K.M., McLean, J.S., and He, X. (2024). The oral microbiome: diversity, biogeography and human health. Nat. Rev. Microbiol. *22*, 89–104. https://doi.org/10.1038/s41579-023-00963-6.
5. Gilbert, J.A., Azad, M.B., Bäckhed, F., Blaser, M.J., Byndloss, M., Chiu, C.Y., Chu, H., Dugas, L.R., Elinav, E., Gibbons, S.M., et al. (2025). Clinical translation of microbiome research. Nat. Med. *31*, 1099–1113. https://doi.org/10.1038/s41591-025-03615-9.
6. Porcari, S., Ng, S.C., Zitvogel, L., Sokol, H., Weersma, R.K., Elinav, E., Gasbarrini, A., Cammarota, G., Tilg, H., and Ianiro, G. (2025). The microbiome for clinicians. Cell *188*, 2836–2844. https://doi.org/10.1016/j.cell.2025.04.016.
7. Heidrich, V., Valles-Colomer, M., and Segata, N. (2025). Human microbiome acquisition and transmission. Nat. Rev. Microbiol. *23*, 568–584. https://doi.org/10.1038/s41579-025-01166-x.
8. Browne, H.P., Neville, B.A., Forster, S.C., and Lawley, T.D. (2017). Transmission of the gut microbiota: spreading of health. Nat. Rev. Microbiol. *15*, 531–543. https://doi.org/10.1038/nrmicro.2017.50.
9. Valles-Colomer, M., Blanco-Míguez, A., Manghi, P., Asnicar, F., Dubois, L., Golzato, D., Armanini, F., Cumbo, F., Huang, K.D., Manara, S., et al. (2023). The person-to-person transmission landscape of the gut and oral microbiomes. Nature *614*, 125–135. https://doi.org/10.1038/s41586-022-05620-1.
10. Brito, I.L., Gurry, T., Zhao, S., Huang, K., Young, S.K., Shea, T.P., Naisilisili, W., Jenkins, A.P., Jupiter, S.D., Gevers, D., and Alm, E.J. (2019). Transmission of human-associated microbiota along family and social networks. Nat. Microbiol. *4*, 964–971. https://doi.org/10.1038/s41564-019-0409-6.
11. Beghini, F., Pullman, J., Alexander, M., Shridhar, S.V., Prinster, D., Singh, A., Matute Juárez, R., Airoldi, E.M., Brito, I.L., and Christakis, N.A. (2025). Gut microbiome strain-sharing within isolated village social networks. Nature *637*, 167–175. https://doi.org/10.1038/s41586-024-08222-1.
12. 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. https://doi.org/10.1186/gb-2012-13-6-r42.
13. 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*, e42693. https://doi.org/10.7554/eLife.42693.
14. Rashidi, A., Gem, H., McLean, J.S., Kerns, K., Dean, D.R., Dey, N., and Minot, S. (2024). Multi-cohort shotgun metagenomic analysis of oral and gut microbiota overlap in healthy adults. Sci. Data *11*, 75. https://doi.org/10.1038/s41597-024-02916-x.
15. Jin, S., Cenier, A., Wetzel, D., Arefaine, B., Moreno-Gonzalez, M., Stamouli, M., Mohamad, M., Lupatsii, M., Ríos, E., Lee, S., et al. (2026). Microbial collagenase activity is linked to oral-gut translocation in advanced chronic liver disease. Nat. Microbiol. *11*, 211–227. https://doi.org/10.1038/s41564-025-02223-0.
16. Kunath, B.J., Hickl, O., Queirós, P., Martin-Gallausiaux, C., Lebrun, L.A., Halder, R., Laczny, C.C., Schmidt, T.S.B., Hayward, M.R., Becher, D., et al. (2022). Alterations of oral microbiota and impact on the gut microbiome in type 1 diabetes mellitus revealed by integrated multi-omic analyses. Microbiome *10*, 243. https://doi.org/10.1186/s40168-022-01435-4.
17. Yersin, S., Gody, J.-C., Mazel, F., Djimbele, E., Nigateloum, S.N., Gondje, B.P., Vondo, S.S., Kaleb Jephté Estimé, K., Raub, A., Teo, Y., et al. (2026). Strain-level translocation and enrichment dynamics of oral bacteria in the lower gastrointestinal tract of stunted children. Gut Microbes *18*, 2653550. https://doi.org/10.1080/19490976.2026.2653550.
18. Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D.T., Manara, S., Zolfo, M., et al. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. Cell Host Microbe *24*, 133–145.e5. https://doi.org/10.1016/j.chom.2018.06.005.
19. Nagata, N., Nishijima, S., Kojima, Y., Hisada, Y., Imbe, K., Miyoshi-Akiyama, T., Suda, W., Kimura, M., Aoki, R., Sekine, K., et al. (2022). Metagenomic identification of microbial signatures predicting pancreatic cancer from a multinational study. Gastroenterology *163*, 222–238. https://doi.org/10.1053/j.gastro.2022.03.054.
20. Kartal, E., Schmidt, T.S.B., Molina-Montes, E., Rodríguez-Perales, S., Wirbel, J., Maistrenko, O.M., Akanni, W.A., Alashkar Alhamwe, B., Alves, R.J., Carrato, A., et al. (2022). A faecal microbiota signature with high specificity

for pancreatic cancer. Gut *71*, 1359–1372. https://doi.org/10.1136/gutjnl-2021-324755.

21. Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L.J., Thompson, K.N., Zolfo, M., Manghi, P., Dubois, L., Huang, K.D., Thomas, A.M., et al. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat. Biotechnol. *41*, 1633–1644. https://doi.org/10.1038/s41587-023-01688-w.
22. Podlesny, D., Arze, C., Dörner, E., Verma, S., Dutta, S., Walter, J., and Fricke, W.F. (2022). Metagenomic strain detection with SameStr: identification of a persisting core gut microbiota transferable by fecal transplantation. Microbiome *10*, 53. https://doi.org/10.1186/s40168-022-01251-w.
23. Kort, R., Caspers, M., van de Graaf, A., van Egmond, W., Keijser, B., and Roeselers, G. (2014). Shaping the oral microbiota through intimate kissing. Microbiome *2*, 41. https://doi.org/10.1186/2049-2618-2-41.
24. Strandwitz, P., Kim, K.H., Terekhova, D., Liu, J.K., Sharma, A., Levering, J., McDonald, D., Dietrich, D., Ramadhar, T.R., Lekbua, A., et al. (2019). GABA-modulating bacteria of the human gut microbiota. Nat. Microbiol. *4*, 396–403. https://doi.org/10.1038/s41564-018-0307-3.
25. Duranti, S., Ruiz, L., Lugli, G.A., Tames, H., Milani, C., Mancabelli, L., Mancino, W., Longhi, G., Carnevali, L., Sgoifo, A., et al. (2020). Bifidobacterium adolescentis as a key member of the human gut microbiota in the production of GABA. Sci. Rep. *10*, 14112. https://doi.org/10.1038/s41598-020-70986-z.
26. Asnicar, F., Manghi, P., Fackelmann, G., Baldanzi, G., Bakker, E., Ricci, L., Piccinno, G., Piperni, E., Mladenovic, K., Amati, F., et al. (2026). Gut microorganisms associated with health, nutrition and dietary interventions. Nature *650*, 450–458. https://doi.org/10.1038/s41586-025-09854-7.
27. Mahnert, A., Moissl-Eichinger, C., Zojer, M., Bogumil, D., Mizrahi, I., Rattei, T., Martinez, J.L., and Berg, G. (2019). Man-made microbial resistances in built environments. Nat. Commun. *10*, 968. https://doi.org/10.1038/s41467-019-08864-0.
28. Xiao, Z., and Xu, P. (2007). Acetoin metabolism in bacteria. Crit. Rev. Microbiol. *33*, 127–140. https://doi.org/10.1080/10408410701364604.
29. Zhao, B., and Houry, W.A. (2010). Acid stress response in enteropathogenic gammaproteobacteria: an aptitude for survival. Biochem. Cell. Biol. *88*, 301–314. https://doi.org/10.1139/o09-182.
30. Piccinno, G., Thompson, K.N., Manghi, P., Ghazi, A.R., Thomas, A.M., Blanco-Míguez, A., Asnicar, F., Mladenovic, K., Pinto, F., Armanini, F., et al. (2025). Pooled analysis of 3,741 stool metagenomes from 18 cohorts for cross-stage and strain-level reproducible microbial biomarkers of colorectal cancer. Nat. Med. *31*, 2416–2429. https://doi.org/10.1038/s41591-025-03693-9.
31. Mattarelli, P., Bonaparte, C., Pot, B., and Biavati, B. (2008). Proposal to reclassify the three biotypes of Bifidobacterium longum as three subspecies: Bifidobacterium longum subsp. longum subsp. nov., Bifidobacterium longum subsp. infantis comb. nov. and Bifidobacterium longum subsp. suis comb. nov. Int. J. Syst. Evol. Microbiol. *58*, 767–772. https://doi.org/10.1099/ijs.0.65319-0.
32. Yanokura, E., Oki, K., Makino, H., Modesto, M., Pot, B., Mattarelli, P., Biavati, B., and Watanabe, K. (2015). Subspeciation of Bifidobacterium longum by multilocus approaches and amplified fragment length polymorphism: Description of B. longum subsp. suillum subsp. nov., isolated from the faeces of piglets. Syst. Appl. Microbiol. *38*, 305–314. https://doi.org/10.1016/j.syapm.2015.05.001.
33. Modesto, M., Ngom-Bru, C., Scarafile, D., Bruttin, A., Pruvost, S., Sarker, S.A., Ahmed, T., Sakwinska, O., Mattarelli, P., and Duboux, S. (2023). Bifidobacterium longum subsp. iuvenis subsp. nov., a novel subspecies isolated from the faeces of weaning infants. Int. J. Syst. Evol. Microbiol. *73*. https://doi.org/10.1099/ijsem.0.006013.
34. Ennis, D., Andreu-Sanchez, S., Levin, Y., Goldberger, O., Mallon, C.A., Jayaprakash, Y.K., Shao, Y., Zhao, S., Sinha, T., Fu, J., et al. (2025). *Bifidobacterium longum* subsp. *nexti* subsp. nov., a novel subspecies isolated from infant stool. Preprint at bioRxiv *2025*. https://doi.org/10.1101/2025.06.23.661074.
35. Carlino, N., Blanco-Míguez, A., Punčochář, M., Mengoni, C., Pinto, F., Tatti, A., Manghi, P., Armanini, F., Avagliano, M., Barcenilla, C., et al. (2024). Unexplored microbial diversity from 2,500 food metagenomes and links with the human microbiome. Cell *187*, 5775–5795.
36. Cha, J.H., Kim, N., Ma, J., Lee, S., Koh, G., Yang, S., Beck, S., Byeon, I., Lee, B., and Lee, I. (2025). A high-quality genomic catalog of the human oral microbiome broadens its phylogeny and clinical insights. Cell Host Microbe *33*, 1977–1994.e8. https://doi.org/10.1016/j.chom.2025.10.001.
37. Slots, J., and Slots, H. (2011). Bacterial and viral pathogens in saliva: disease relationship and infectious risk: Bacterial and viral pathogens in saliva. Periodontol *55*, 48–69. https://doi.org/10.1111/j.1600-0757.2010.00361.x.
38. Arbøll, T.P., and Rasmussen, S.L. (2023). The ancient history of kissing. Science *380*, 688–690. https://doi.org/10.1126/science.adf0512.
39. J. Vonk and T. Shackelford, eds. (2022). Encyclopedia of Animal Cognition and Behavior, 1st ed. (Springer International Publishing). https://doi.org/10.1007/978-3-319-55065-7.
40. Liao, C., Rolling, T., Djukovic, A., Fei, T., Mishra, V., Liu, H., Lindberg, C., Dai, L., Zhai, B., Peled, J.U., et al. (2024). Oral bacteria relative abundance in faeces increases due to gut microbiota depletion and is linked with patient outcomes. Nat. Microbiol. *9*, 1555–1565. https://doi.org/10.1038/s41564-024-01680-3.
41. Manghi, P., Antonello, G., Schiffer, L., Golzato, D., Wokaty, A., Beghini, F., Mirzayi, C., Long, K., Gravel-Pucillo, K., Piccinno, G., et al. (2025). Meta-analysis of 22,710 human microbiome metagenomes defines an oral-to-gut microbial enrichment score and associations with host health and disease. Nat. Commun. *17*, 196. https://doi.org/10.1038/s41467-025-66888-1.
42. Shao, Y., Wang, S., Gichuki, B.M., Stares, M.D., Rozday, T.J., Kumar, N., Browne, H.P., Dawson, N.J.R., Njunge, J.M., Tigoi, C., et al. (2026). Genomic atlas of Bifidobacterium infantis and B. longum informs infant probiotic design. Cell *189*, 1854–1873.e17. https://doi.org/10.1016/j.cell.2026.01.007.
43. Manghi, P., Filosi, M., Zolfo, M., Casten, L.G., Garcia-Valiente, A., Mattevi, S., Heidrich, V., Golzato, D., Perini, S., Thomas, A.M., et al. (2024). Large-scale metagenomic analysis of oral microbiomes reveals markers for autism spectrum disorders. Nat. Commun. *15*, 9743. https://doi.org/10.1038/s41467-024-53934-7.
44. Feehily, C., O'Neill, I.J., Walsh, C.J., Moore, R.L., Killeen, S.L., Geraghty, A.A., Lawton, E.M., Byrne, D., Sanchez-Gallardo, R., Nori, S.R.C., et al. (2023). Detailed mapping of Bifidobacterium strain transmission from mother to infant via a dual culture-based and metagenomic approach. Nat. Commun. *14*, 3015. https://doi.org/10.1038/s41467-023-38694-0.
45. Zepeda-Rivera, M., Minot, S.S., Bouzek, H., Wu, H., Blanco-Míguez, A., Manghi, P., Jones, D.S., LaCourse, K.D., Wu, Y., McMahon, E.F., et al. (2024). A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. Nature *628*, 424–432. https://doi.org/10.1038/s41586-024-07182-w.
46. Andreu-Sánchez, S., Blanco-Míguez, A., Wang, D., Golzato, D., Manghi, P., Heidrich, V., Fackelmann, G., Zhernakova, D.V., Kurilshikov, A., Valles-Colomer, M., et al. (2025). Global genetic diversity of human gut microbiome species is related to geographic location and host health. Cell *188*, 3942–3959.e9. https://doi.org/10.1016/j.cell.2025.04.014.
47. Chen-Liaw, A., Aggarwala, V., Mogno, I., Haifer, C., Li, Z., Eggers, J., Helmus, D., Hart, A., Wehkamp, J., Lamousé-Smith, E.S.N., et al. (2025). Gut microbiota strain richness is species specific and affects engraftment. Nature *637*, 422–429. https://doi.org/10.1038/s41586-024-08242-x.
48. Ricci, L., Heidrich, V., Punčochář, M., Armanini, F., Ciciani, M., Nabinejad, A., Fazaeli, F., Piperni, E., Servais, C., Pinto, F., et al. (2026). Baby-to-baby strain transmission shapes the developing gut microbiome. Nature *651*, 191–200. https://doi.org/10.1038/s41586-025-09983-z.
49. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

50. Chklovski, A., Parks, D.H., Woodcroft, B.J., and Tyson, G.W. (2023). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat. Methods *20*, 1203–1212. https://doi.org/10.1038/s41592-023-01940-w.

51. Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ *3*, e1029. https://doi.org/10.7717/peerj.1029.

52. Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat. Commun. *11*, 2500. https://doi.org/10.1038/s41467-020-16366-7.

53. Aton, M., McDonald, D., Cañardo Alastuey, J., Azom, R., Batra, P., Bezshapkin, V., Bolyen, E., Cagle, A., Caporaso, J.G., Debelius, J.W., et al. (2026). Scikit-bio: a fundamental Python library for biological omic data analysis. Nat. Methods *23*, 274–276. https://doi.org/10.1038/s41592-025-02981-z.

54. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

55. Shaw, J., and Yu, Y.W. (2023). Fast and robust metagenomic sequence comparison through sparse chaining with skani. Nat. Methods *20*, 1661–1665. https://doi.org/10.1038/s41592-023-02018-3.

56. Weimann, A., Mooren, K., Frank, J., Pope, P.B., Bremges, A., and McHardy, A.C. (2016). From genomes to phenotypes: Traitar, the microbial trait analyzer. mSystems *1*, e00101-16. https://doi.org/10.1128/mSystems.00101-16.

57. Bengtsson-Palme, J., Angelin, M., Huss, M., Kjellqvist, S., Kristiansson, E., Palmgren, H., Larsson, D.G.J., and Johansson, A. (2015). The human gut microbiome as a transporter of antibiotic resistance genes between continents. Antimicrob. Agents Chemother. *59*, 6551–6560. https://doi.org/10.1128/AAC.00933-15.

58. Costea, P.I., Coelho, L.P., Sunagawa, S., Munch, R., Huerta-Cepas, J., Forslund, K., Hildebrand, F., Kushugulova, A., Zeller, G., and Bork, P. (2017). Subspecies in the global human gut microbiome. Mol. Syst. Biol. *13*, 960. https://doi.org/10.15252/msb.20177589.

59. Ghensi, P., Manghi, P., Zolfo, M., Armanini, F., Pasolli, E., Bolzan, M., Bertelle, A., Dell'Acqua, F., Dellasega, E., Waldner, R., et al. (2020). Strong oral plaque microbiome signatures for dental implant diseases identified by strain-resolution metagenomics. NPJ Biofilms Microbiomes *6*, 47. https://doi.org/10.1038/s41522-020-00155-7.

60. Schirmer, M., Franzosa, E.A., Lloyd-Price, J., McIver, L.J., Schwager, R., Poon, T.W., Ananthakrishnan, A.N., Andrews, E., Barron, G., Lake, K., et al. (2018). Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat. Microbiol. *3*, 337–346. https://doi.org/10.1038/s41564-017-0089-z.

61. Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature *569*, 655–662. https://doi.org/10.1038/s41586-019-1237-9.

62. Zhou, W., Sailani, M.R., Contrepois, K., Zhou, Y., Ahadi, S., Leopold, S.R., Zhang, M.J., Rao, V., Avina, M., Mishra, T., et al. (2019). Longitudinal multi-omics of host-microbe dynamics in prediabetes. Nature *569*, 663–671. https://doi.org/10.1038/s41586-019-1236-x.

63. Hall, A.B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., Lagoudas, G.K., Vatanen, T., Fornelos, N., Wilson, R., et al. (2017). A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. Genome Med. *9*, 103. https://doi.org/10.1186/s13073-017-0490-5.

64. Hansen, L.B.S., Roager, H.M., Søndertoft, N.B., Gøbel, R.J., Kristensen, M., Vallès-Colomer, M., Vieira-Silva, S., Ibrügger, S., Lind, M.V., Mærkedahl, R.B., et al. (2018). A low-gluten diet induces changes in the intestinal microbiome of healthy Danish adults. Nat. Commun. *9*, 4630. https://doi.org/10.1038/s41467-018-07019-x.

65. Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A.-M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., et al. (2016). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe *20*, 121. https://doi.org/10.1016/j.chom.2016.06.009.

66. Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541–546. https://doi.org/10.1038/nature12506.

67. Louis, S., Tappu, R.-M., Damms-Machado, A., Huson, D.H., and Bischoff, S.C. (2016). Characterization of the Gut Microbial Community of Obese Patients Following a Weight-Loss Intervention Using Whole Metagenome Shotgun Sequencing. PLoS One *11*, e0149564. https://doi.org/10.1371/journal.pone.0149564.

68. Mehta, R.S., Abu-Ali, G.S., Drew, D.A., Lloyd-Price, J., Subramanian, A., Lochhead, P., Joshi, A.D., Ivey, K.L., Khalili, H., Brown, G.T., et al. (2018). Stability of the human faecal microbiome in a cohort of adult men. Nat. Microbiol. *3*, 347–355. https://doi.org/10.1038/s41564-017-0096-0.

69. Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D.R., Gautier, L., Pedersen, A.G., Le Chatelier, E., et al. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat. Biotechnol. *32*, 822–828. https://doi.org/10.1038/nbt.2939.

70. Pehrsson, E.C., Tsukayama, P., Patel, S., Mejía-Bautista, M., Sosa-Soto, G., Navarrete, K.M., Calderon, M., Cabrera, L., Hoyos-Arango, W., Bertoli, M.T., et al. (2016). Interconnected microbiomes and resistomes in low-income human habitats. Nature *533*, 212–216. https://doi.org/10.1038/nature17672.

71. Raymond, F., Ouameur, A.A., Déraspe, M., Iqbal, N., Gingras, H., Dridi, B., Leprohon, P., Plante, P.-L., Giroux, R., Bérubé, È., et al. (2016). The initial state of the human gut microbiome determines its reshaping by antibiotics. ISME J. *10*, 707–720. https://doi.org/10.1038/ismej.2015.148.

72. Vatanen, T., Kostic, A.D., d'Hennezel, E., Siljander, H., Franzosa, E.A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T.D., Hämäläinen, A.-M., et al. (2016). Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell *165*, 842–853. https://doi.org/10.1016/j.cell.2016.04.007.

73. Vincent, C., Miller, M.A., Edens, T.J., Mehrotra, S., Dewar, K., and Manges, A.R. (2016). Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and Clostridium difficile colonization or infection. Microbiome *4*, 12. https://doi.org/10.1186/s40168-016-0156-3.

74. Yassour, M., Lim, M.Y., Yun, H.S., Tickle, T.L., Sung, J., Song, Y.-M., Lee, K., Franzosa, E.A., Morgan, X.C., Gevers, D., et al. (2016). Sub-clinical detection of gut microbial biomarkers of obesity and type 2 diabetes. Genome Med. *8*, 17. https://doi.org/10.1186/s13073-016-0271-6.

75. Yassour, M., Jason, E., Hogstrom, L.J., Arthur, T.D., Tripathi, S., Siljander, H., Selvenius, J., Oikarinen, S., Hyöty, H., Virtanen, S.M., et al. (2018). Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. Cell Host Microbe *24*, 146–154.e4. https://doi.org/10.1016/j.chom.2018.06.007.

76. Brito, I.L., Yilmaz, S., Huang, K., Xu, L., Jupiter, S.D., Jenkins, A.P., Naisilisili, W., Tamminen, M., Smillie, C.S., Wortman, J.R., et al. (2016). Mobile genes in the human microbiome are structured from global to individual scales. Nature *535*, 435–439. https://doi.org/10.1038/nature18927.

77. Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C.J., Fagerberg, B., Nielsen, J., and Bäckhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature *498*, 99–103. https://doi.org/10.1038/nature12198.

78. Forslund, S.K., Chakaroun, R., Zimmermann-Kogadeeva, M., Markó, L., Aron-Wisnewsky, J., Nielsen, T., Moitinho-Silva, L., Schmidt, T.S.B., Falony, G., Vieira-Silva, S., et al. (2021). Combinatorial, additive and

dose-dependent drug-microbiome associations. Nature *600*, 500–505. https://doi.org/10.1038/s41586-021-04177-9.

79. Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60. https://doi.org/10.1038/nature11450.

80. Xu, Q., Wu, C., Zhu, Q., Gao, R., Lu, J., Valles-Colomer, M., Zhu, J., Yin, F., Huang, L., Ding, L., et al. (2022). Metagenomic and metabolomic remodeling in nonagenarians and centenarians and its association with genetic and socioeconomic factors. Nat. Aging *2*, 438–452. https://doi.org/10.1038/s43587-022-00193-0.

81. Sankaranarayanan, K., Ozga, A.T., Warinner, C., Tito, R.Y., Obregon-Tito, A.J., Xu, J., Gaffney, P.M., Jervis, L.L., Cox, D., Stephens, L., et al. (2015). Gut microbiome diversity among Cheyenne and Arapaho individuals from western Oklahoma. Curr. Biol. *25*, 3161–3169. https://doi.org/10.1016/j.cub.2015.10.060.

82. Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. J. Educ. Stat. *6*, 107–128. https://doi.org/10.3102/10769986006002107.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
DNA Prep (M) Tagmentation Kit	Illumina	Catalog No. 20060060
DNA/RNA Shield buffer	Zymo	Catalog No. R1100
DNeasy PowerSoil Pro kit	QIAGEN	Catalog No. 47017
Nextera DNA Library Preparation Kit	Illumina	Catalog No. 20018705
Deposited data		
CM_MTB metagenomic data	This paper and Ricci et al.[48]	NCBI-SRA: PRJNA1140720
CM_MTC metagenomic data	This paper	NCBI-SRA: PRJNA1308818
B. longum genomes	This paper	https://doi.org/10.5281/zenodo.18155149
Software and algorithms		
Bowtie2 (v2.3.4.3)	Langmead and Salzberg[49]	https://github.com/BenLangmead/bowtie2
CheckM2 (v1.1.0)	Chklovski et al.[50]	https://github.com/chklovski/CheckM2
GraPhlAn (v1.1.3)	Asnicar et al.[51]	https://github.com/biobakery/graphlan
MetaPhlAn 4 (v4.1)	Blanco-Míguez et al.[21]	https://github.com/biobakery/MetaPhlAn
PhyloPhlAn (v3.1)	Asnicar et al.[52]	https://github.com/biobakery/phylophlan
PythonMeta (v1.26)	N/A	https://pypi.org/project/PythonMeta/
Python (v3.10.12)	Python Software Foundation	https://www.python.org/
SameStr (v1.2025.03)	Podlesny et al.[22]	https://github.com/danielpodlesny/samestr
scikit-bio (v0.5.9)	Aton et al.[53]	https://github.com/scikit-bio/scikit-bio
scipy (v1.10.1)	Virtanen et al.[54]	https://github.com/scipy/scipy
skani (v0.2.1)	Shaw and Yu[55]	https://github.com/bluenote-1577/skani
StrainPhlAn 4 (v4.1)	Blanco-Míguez et al.[21]	https://github.com/biobakery/MetaPhlAn
Traitar (v3.0.1)	Weimann et al.[56]	https://github.com/hzi-bifo/traitar
TrimGalore (v0.6.6)	N/A	https://github.com/felixkrueger/trimgalore
Other		
curatedFoodMetagenomicData (v1.1.0)	Carlino et al.[35]	https://github.com/SegataLab/cFMD
curatedMetagenomicData (v3.14 and v3.18)	Manghi et al.[41]	https://waldronlab.github.io/curatedMetagenomicData/
Human Reference Oral Microbiome database	Cha et al.[36]	https://www.decodebiome.org/HROM/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Description of newly sequenced cohorts

Among the seven datasets analyzed in this study, two of them represent newly sequenced cohorts: CM_MTB and CM_MTC. CM_MTB is a large longitudinal study aiming at investigating oral and gut strain transmission in a nursery setting among babies, their immediate family (parents, siblings, and pets), and educators[48]. Here, we repurpose CM_MTB parents and siblings' samples to study household inter-individual transmission (n = 38 households) and oral-to-gut intra-individual transmission (n = 70 individuals with matched oral-fecal samples), as well as leverage its longitudinal design to investigate strain replacement among parents. CM_MTB participants provided written informed consent before enrollment, with protocol approved by the Ethics Committee of the University of Trento (protocol nr. 2022-040) and by the Ethics Panel of the European Research Council Executive Agency upon evaluation of the project (microTOUCH Grant agreement ID 101045015). CM_MTC is a cross-sectional cohort of potentially dysbiotic cohabiting individuals (n = 48 households) for which oral and gut microbiome samples were collected. Households were composed of 1–3 celiac individuals living with 0–2 non-celiac (but genetically predisposed) individuals (household composition detailed in Figure S5A). After confirming the non-dysbiotic nature of oral and gut celiac microbiomes compared to non-celiac

subjects in the same cohort (Figure S5B, Table S14), we consider all participants in CM_MTC as healthy from the microbiome perspective and leverage their samples to study household transmission and oral-to-gut transmission (n = 73 individuals with matched oral-fecal samples). CM_MTC participants provided written informed consent before enrollment, with protocol approved by the Ethics Committee of the Fondazione Policlinico Gemelli IRCCS (protocol ID #6778).

Survey of public metagenomic datasets

Besides analyzing newly sequenced cohorts, we also incorporate five additional publicly available metagenomic datasets for oral-to-gut transmission evaluation (HMP_2012[1], FerrettiP_2018[18], BritoIL_2019[10], NagataN_2022[19], and KartalE_2022[20]). Those were identified by searching the literature for human metagenomic cohorts of studies that profiled simultaneously oral and gut microbiomes of healthy subjects. BritoIL_2019, due to consisting of a cohort of Fijian cohabiting individuals, was also leveraged for the household transmission analysis. After downloading of the metagenomes and homogenization of the metadata, we proceeded with the analysis only with metagenomes (regardless if public or in-house sequenced) having ≥0.15 Gbp (∼1M reads).

METHOD DETAILS

Sample collection, DNA extraction, and metagenomic sequencing

Newly sequenced samples were self-collected at home following detailed instructions. CM_MTB oral sample collection was based on HMP sampling guidelines (Human Microbiome Project Consortium, 2012b) utilizing a 2mm break-point sterile swab, while CM_MTC oral samples consisted of unstimulated saliva. All samples were kept in collection tubes containing DNA/RNA Shield buffer (Zymo), with CM_MTB samples being stored at room temperature and processed for DNA extraction within two weeks of sample collection, while CM_MTC samples were stored at -80°C until DNA extraction.

Following samples homogenization and resuspension (in the case of swabs) by vortexing, DNA was extracted using the DNeasy PowerSoil Pro kit (QIAGEN), following the manufacturer's instructions with slight modifications to allow extraction from swab samples[1], and quantified using the Qubit 2.0 fluorometer (Thermo Fisher Scientific) in the case of CM_MTB samples or the Varioskan LUX Microplate Reader (Thermo Fisher Scientific) in the case of CM_MTC samples. CM_MTB libraries were prepared using the Nextera DNA Library Preparation Kit (Illumina) and sequenced on the Illumina NovaSeq 6000 platform. CM_MTC libraries were prepared with the DNA Prep (M) Tagmentation Kit (Illumina) and sequenced on the Illumina NovaSeq X Plus platform. For both CM_MTB and CM_MTC target sequencing depth was set at 15 Gbp per sample.

Metagenomic sequences quality control

In-house sequenced metagenomes were preprocessed using our previously validated pipeline (available at https://github.com/SegataLab/preprocessing). In brief, reads considered of low quality (<Q20 or >2 ambiguous bases) or too short (<75 bp) were filtered out with TrimGalore (v0.6.6). Reads mapping against the human genome (hg19) or the phiX 174 Illumina spike-in as identified with Bowtie2 (v2.3.4.3) were also removed. Remaining high-quality reads were sorted and split to create standard forward, reverse, and unpaired preprocessed *fastq* files for each metagenome.

Species- and strain-level microbiome profiling

Metagenomes were profiled at the resolution of species-level genome bins (SGBs) with MetaPhlAn 4 (v4.1)[21] and strain-level phylogenies were built for each SGB detected in our study with StrainPhlAn 4 (v4.1)[21], both using the vJun23_202307 marker genes database.

Strain-sharing assessment

To detect strain-sharing events, it is necessary to build strain-level phylogenies and define strain-level boundaries for each SGB detected in our study. This starts by augmenting the aforementioned strain-level SGB phylogenies with strains from additional public longitudinal metagenomic data. To do so, we queried the curatedMetagenomicData v3.18 for oral and gut metagenomes (≥1 Gbp) from healthy human individuals (≥2 years of age). After disregarding subjects sampled only cross-sectionally and datasets having less than 3 subjects, we managed to incorporate 20 longitudinal metagenomic datasets including 3,030 samples (≥1 Gbp) from 1,117 individuals: Bengtsson-PalmeJ_2015[57], CosteaPI_2017[58], GhensiP_2019[59], HMP_2012[1], HMP_2019_ibdmdb[60,61], HMP_2019_t2d[62], HallAB_2017[63], HansenLBS_2018[64], KosticAD_2015[65], LeChatelierE_2013[66], LouisS_2016[67], MehtaRS_2018[68], NielsenHB_2014[69], PehrssonE_2016[70], RaymondF_2016[71], SchmidtTSB_2019[13], VatanenT_2016[72], VincentC_2016[73], YassourM_2016[74], YassourM_2018[75]. These were then used together with the aforementioned 8 datasets analyzed as part of this study to launch the construction of StrainPhlAn 4 phylogenetic trees for 1,768 SGBs present in the latter. In the case of SGBs whose genome was previously reconstructed from food microbiomes[35], we also included food-derived microbial genomes into the SGBs phylogenies to be able to disentangle acquisition from food from human-to-human transmission. Among them, 365 phylogenies were not built due to not enough samples having the SGB and/or not enough marker genes being available for the SGB according to default StrainPhlAn v4.1 parameters. A further 121 SGBs were discarded due to their marker genes concatenated alignment being shorter than 1000 nt. From the remaining 1282 SGBs, 24 were previously identified as food-associated[35], and we removed strains too close to food-derived strains (avg. nucleotide identity on marker genes > 99.85%) from 12 of them. In 8 such

cases, >20% of strains were discarded, causing the removal of the SGB altogether. Two additional SGBs were discarded due to their phylogeny not having at least two strains belonging to one of the 8 datasets of interest.

We then calculated the pairwise phylogenetic distances between strains for the remaining 1,272 SGBs. This is defined as the length of the shortest path between strains along the tree branches, which is then divided by the median pairwise distance across all pairs in a tree to provide normalized distance distributions for each tree. To define strain-sharing events we focus on distances between strains profiled in longitudinal samples from the same subject (sampled $\leq$6 months apart, using only one sample pair per subject) and those between strains profiled in unrelated subjects (from different datasets, using only one sample per subject). In 286 cases the SGBs were discarded due to not having $\geq$50 pairs in the across-subject pairwise distances distribution. For the remaining 986 SGBs, SGB-specific strain-sharing boundaries were identified by finding the threshold that best separates within-subject and across-subject pairwise distances distributions. In case there were $\geq$20 pairs in the within-subject pairwise distance distribution, the strain-sharing threshold was calculated as the distance maximizing the Youden's index (n = 362 SGBs), unless the expected false positive rate (FPR, see below) exceeded 5%, in which case the threshold is set at the 5th percentile of the across-subject pairwise distance distribution (n = 40 SGBs). This expected FPR is defined as the percentage of across dataset pairs for which a same strain event would be called with the threshold that was chosen for the SGB, while the strain-sharing expected false negative rate (FNR) is defined as the percentage of same-subject pairs for which a different strain event would be called with the threshold that was chosen for the SGB For the cases where there were <20 pairs in within-subject pairwise distance distribution, the threshold is set at the 3rd percentile of the across-subject pairwise distance distribution (n = 584 SGBs).

Strain-sharing events were then identified by evaluating the phylogenetic distance between conspecific strains (i.e. belonging to the same SGB) profiled in different samples and comparing this to the strain-sharing threshold defined for the SGB (available in Table S15).

QUANTIFICATION AND STATISTICAL ANALYSIS

Strain-sharing rate, strain replacement rate, and SGB transmissibility

The extent of strain-sharing between a pair of microbiome samples is summarized in SSRs, defined as the number of strains shared between them over the number of SGBs typed at strain-level present in both of them. The strain replacement rate is defined as one minus the SSR between longitudinal samples of the same individual. Household SGB transmissibility refers to the ratio between the number of strain-sharing events for a given SGB across all pairs of cohabiting individuals having the SGB typed at strain-level (minimum of 10 pairs), which can also be computed among specific participant category pairs (e.g., household SGB transmissibility among partners; minimum of 3 pairs) or at the population-level (minimum of 10 pairs). In the case of mother-offspring pairs, especially in the case of young offspring, household SGB transmissibility should be considered a combination of horizontal and vertical transmission. High household transmissibility SGBs are defined as the ones having a TS > $^1/_3$ while being significantly more transmissible in the household versus in the general population according to a Fisher's exact test. Similarly, the oral-to-gut transmissibility score (Figure 4A) refers to the ratio between the number of strain-sharing events for a given SGB across all intra-individual oral-gut sample pairs having the SGB typed at strain-level.

Evaluation of strain-sharing with SameStr

To validate the strain-sharing results with an alternative approach, we applied SameStr (v1.2025.03)[22]. SameStr, similarly to StrainPhlAn, relies on clade-specific MetaPhlAn marker genes, but it infers shared strains by comparing directly SNV profiles rather than through phylogenetic inference. SameStr was run using the vJun23_202307 marker genes database developed originally for MetaPhlAn. MetaPhlAn marker gene alignments were converted to nucleotide variant profiles, merged, filtered, and compared pairwise between samples using the default SameStr parameters.

Definition of oral-typical SGBs

To define oral-typical SGBs we leverage five publicly available studies[1,18–20,76] and consider only healthy individuals with matched oral (either saliva or tongue dorsum) and stool microbiome samples (n = 495 individuals; Table S16). Similarly to Piccinno et al.[30], oral-typical SGBs were defined according to the following criteria: (i) exclusively present in the oral cavity of at least 20% of the individuals; (ii) being present in both the oral and fecal samples of fewer participants than those in which is present exclusively in the oral cavity; and (iii) being exclusively found in fecal samples in less than 5% of the individuals, resulting in a list of 226 oral-typical SGBs. As the datasets used to define oral-typical SGBs partially overlap with the datasets used in our oral-to-gut transmission evaluation, we only use datasets not used in the definition of oral-typical SGBs for the oral-typical SGBs analysis reported in Figure 3E. A sensitivity analysis with alternative oral-typical SGBs lists generated by altering the prevalence cutoff of the first rule from 20% either to 10 (n = 350 SGBs) or 30% (n = 172 SGBs) confirmed the pattern observed in Figure 3E (Figure S6).

Bifidobacterium longum (SGB17248) analysis

Deeper characterization of *B. longum* (SGB17248) was performed by strain-level phylogeny reconstruction with StrainPhlAn 4 (v4.1)[21]. This included all strains from metagenomes profiled in this study and additional strains from all food metagenomes (n = 50, 30 of which with strains reconstructed) containing *B. longum* (SGB17248) from curatedFoodMetagenomicData (v1.1.0)[35]

and from a random subset of oral metagenomes (n = 112, 97 of which with strains reconstructed) from Manghi et al.[43]. We also downloaded from NCBI RefSeq and added to the phylogeny up to five reference genomes for each of the six previously described/proposed *B. longum* subspecies (subsp. *longum* (n = 5), subsp. *infantis* (n = 5), subsp. *suis* (n = 5), subsp. *suillum* (n = 4), subsp. *iuvenis* (n = 1), and subsp. *nexus* (n = 1); Table S17). The resulting phylogenetic tree was rerooted at midpoint and annotated and visualized with GraPhlAn (v1.1.3)[51]. To further validate the existence of an oral-specific clade, we queried the microbial genomic database MetaRefSGB (vJan25) for strains assigned to *B. longum* (SGB17248) coming from the same metagenomes included in the StrainPhlAn 4 phylogeny of Figure 4B, retrieving 120 metagenome-assembled genomes (MAGs) of at least medium-quality according to CheckM2 (v1.1.0)[50] (Table S13). We also considered additional *B. longum* (SGB17248) MAGs (n = 12 from 7 metagenomic saliva datasets; Table S13) from the human reference oral microbiome (HROM) database[36], as assigned with PhyloPhlAn (v3.1)[52]. The average nucleotide identity between the 132 MAGs and the 21 reference genomes also added to the tree were assessed with skani (v0.2.1)[55] and visualized with a heatmap.

Type 2 diabetes meta-analysis

In order to find gut microbiome SGBs associated with T2D we leveraged five public T2D metagenomic datasets[77–81] available in curatedMetagenomicData v3.14[41] and ran meta-analysis (n = 775 T2D and 968 controls). For each cohort, we filtered their MetaPhlAn 4 (v4.1, vJun23_202307 marker genes database) taxonomic profiles to contain only SGBs for which we had strain-sharing data and that had a prevalence > 1%. We then added a pseudocount equal to the minimum non-zero relative abundance and log2-transformed the data. Next, linear regressions models were fitted within each cohort, with SGBs as dependent variables and T2D (yes/no), age, sex and BMI as independent variables. Finally, the regression estimates for each SGB across the different cohorts were meta-analyzed using fixed-effects models. The results of the meta-analysis are available in Table S18.

Traitar SGB profiling and association with household transmissibility

In order to infer phenotypic traits of the 986 SGBs included in our strain-sharing analysis we pulled the core genome (genes present in at least 50% of the genomes of the SGB in MetaRefSGB vJun23) of each SGB and profiled it with Traitar (v3.0.1)[56]. We consider an SGB to have a phenotype when it is reported as present by both phypat and phypat + Phenotype Gains and Losses (PGL) classifiers (Table S19). We then associate SGB phenotypes with SGB transmissibility by applying a Mann-Whitney U test on household TS (between SGBs having versus not having the phenotype) followed by Benjamini-Hochberg multiple testing correction.

Statistical analysis

Statistical analyses were performed in Python (v3.10.12) using libraries scikit-bio (v0.5.9)[53] and scipy (v1.10.1)[54]. Cross-sectional comparisons were performed using the Mann-Whitney U test (two groups) or the Kruskal-Wallis test with post-hoc Dunn tests (multiple groups), while comparisons of longitudinal or dependent observations were performed using the Wilcoxon signed-rank test. Contingency tables were analyzed using the Pearson's chi-squared test or the Fisher's exact test, as appropriate according to the sample size. The meta-analysis was performed in Python using the library PythonMeta (v1.26) using arcsine square root transformed relative abundances as input and computing standardized mean differences as effect sizes using Hedges' g correction[82]. Microbiomes of celiac versus non-celiac individuals of CM_MTC were compared by running PERMANOVA models on multiple beta-diversity metrics while taking into account sex, age, and diet (gluten versus gluten-free).

CPBLUE, Volume *1*

Supplemental information

Strain transmission links human microbiomes along the oral-gut axis and across cohabiting individuals

Vitor Heidrich, Gloria Fackelmann, Liviana Ricci, Roan Spadazzi, Gabriel Baldanzi, Michal Punčochář, Giulia Catassi, Paolo Marchi, Monica Modesto, Gianmarco Piccinno, Serena Porcari, Debora Rondinella, Francesco Asnicar, Mireia Valles-Colomer, Paola Mattarelli, Gianluca Ianiro, and Nicola Segata

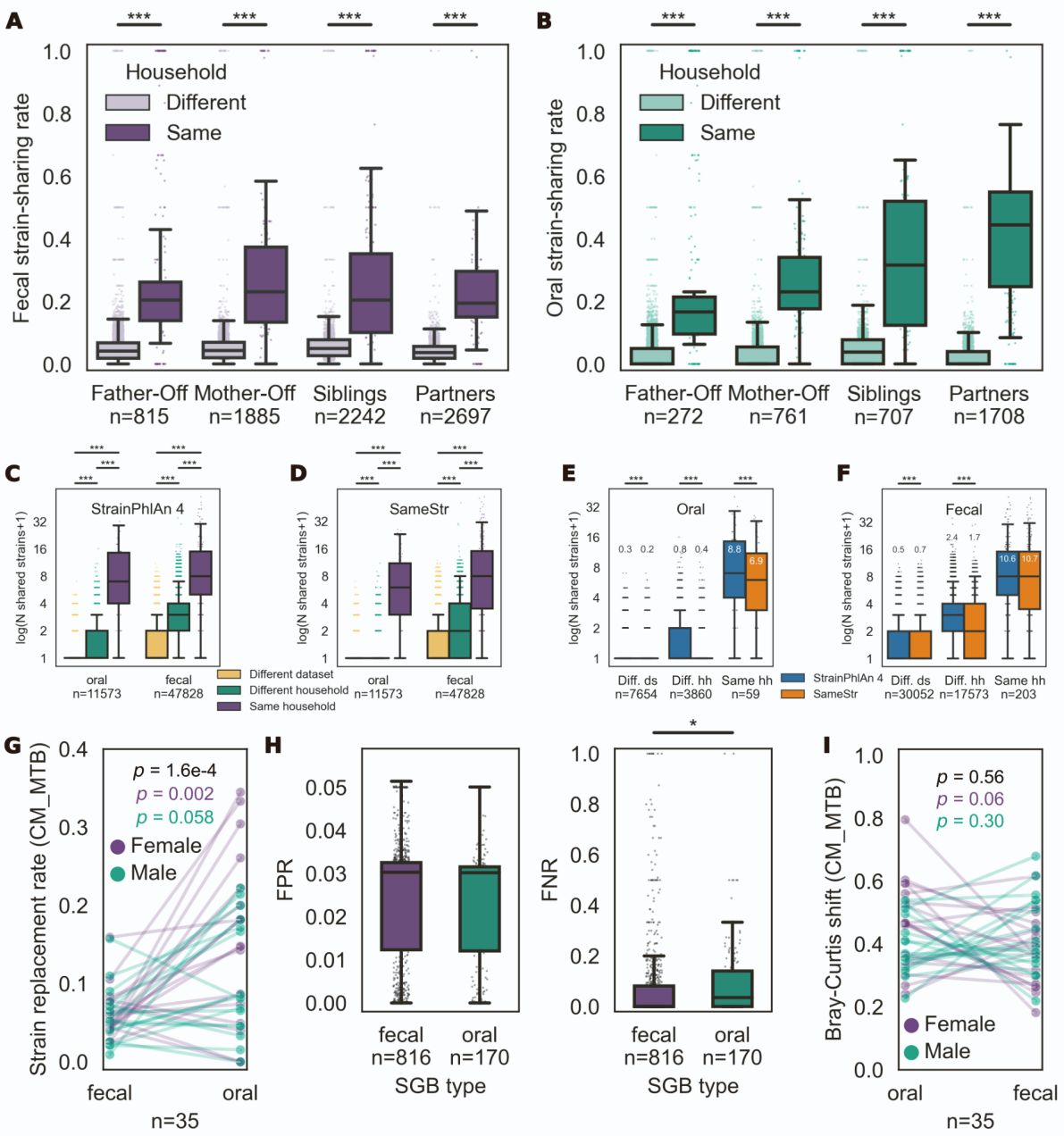


Figure S1. Additional validation of increased oral and gut microbiome strain-sharing rates among cohabitants.

(A) Gut microbiome strain-sharing rates for same vs. different households for different relationship types within households.

(B) Oral microbiome strain-sharing rates for same vs. different households for different relationship types within households.

(C) Number of strains shared between individuals from same versus different households versus different datasets for each body site according to our StrainPhlAn 4-based methodology.

(D) Number of strains shared between individuals from same versus different households versus different datasets for each body site according to our SameStr (**Methods**). Mann-Whitney U test was used.

(E) Comparison between the number of oral strains shared as reported by our StrainPhlAn 4-based methodology versus SameStr for each comparison category.

(F) Comparison between the number of fecal strains shared as reported by our StrainPhlAn 4-based methodology versus SameStr for each comparison category.

(G) Gut vs. oral microbiome strain replacement rate (**Methods**) among females (purple), males (green), or overall (black) in the CM_MTB dataset. Individuals undergoing antibiotic treatment

between sampling timepoints were excluded.
(H) False positive rate (FPR, left) and false negative rate for fecal vs. oral SGBs, defined based on the body site in which they have higher prevalence.
In A–F and H, the Mann-Whitney U test was used with statistical significance represented as: *, $p < 0.05$; ***, $p < 0.001$. The *n* below x-tick labels indicates the number of individual pairs (A–F) or SGBs (H).
(I) Gut vs. oral microbiome Bray-Curtis dissimilarity between consecutive timepoints among females (purple), males (green), or overall (black) in the CM_MTB dataset. Individuals undergoing antibiotic treatment between sampling timepoints were excluded.
In G and I, the Wilcoxon signed-rank test was used. The *n* below x-tick labels indicates the number of individuals.

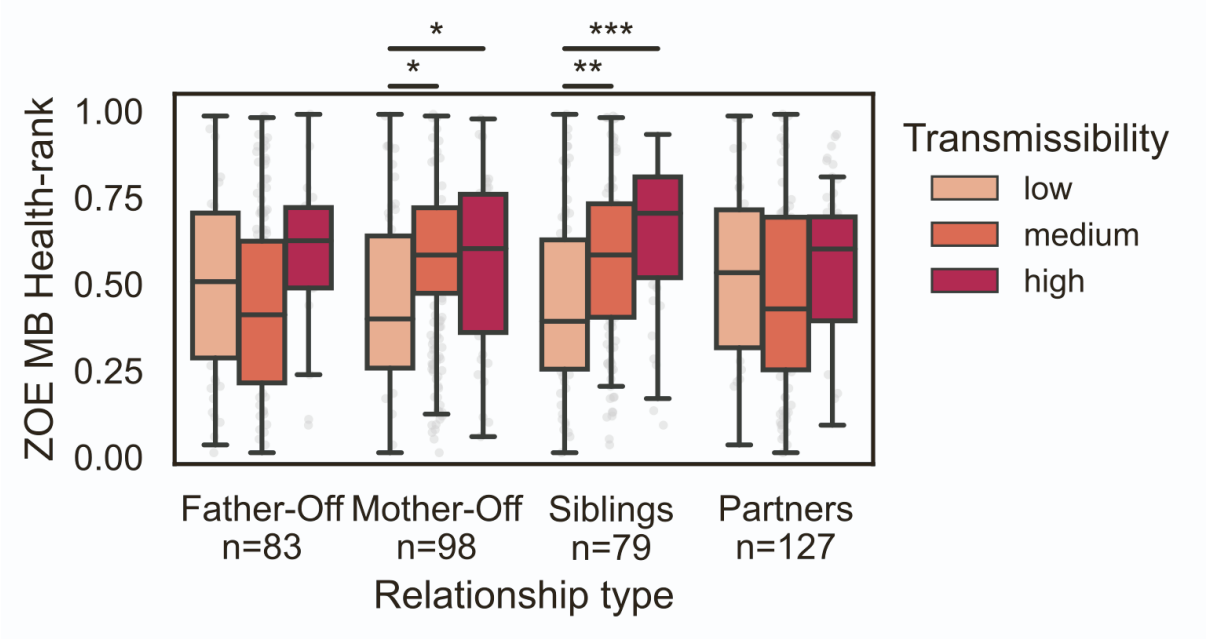


Figure S2. Association between ZOE MB Health-ranks and household transmissibility. ZOE MB health-ranks of SGBs with low vs. medium vs. high transmissibility (Methods) for different relationship types within households. The Mann-Whitney U test was used with statistical significance represented as: *, $p < 0.05$; **, $p < 0.05$, ***, $p < 0.001$. The *n* below x-tick labels indicates the number of SGBs.

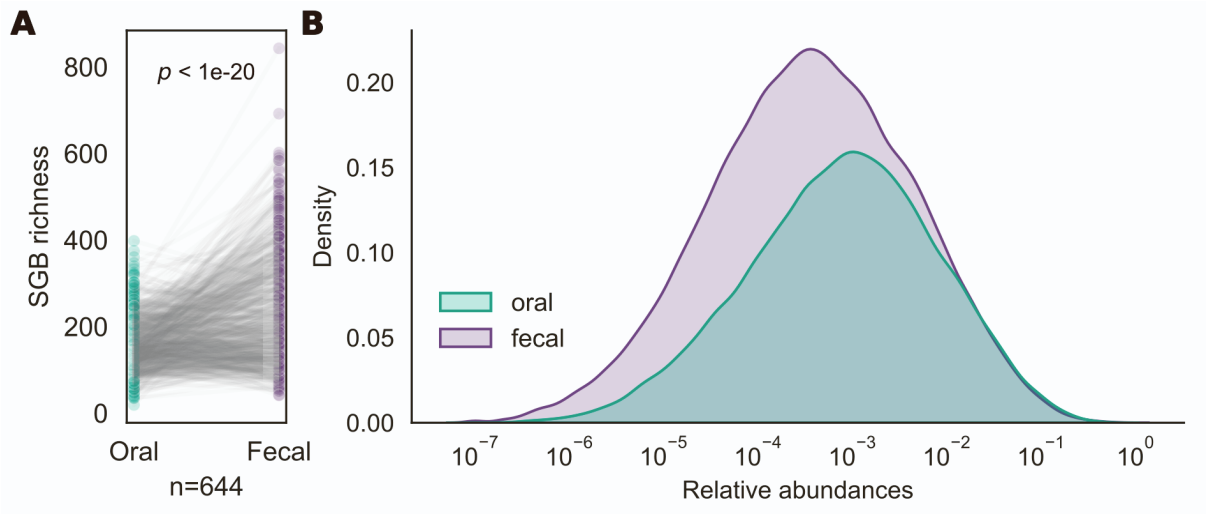


Figure S3. Comparison between oral and gut microbiome diversities.
(A) Oral vs. gut microbiome SGB richness for 644 individuals with paired samples available. The Wilcoxon signed-rank test was used.
(B) Distribution of SGB relative abundance values for oral vs. gut microbiomes.

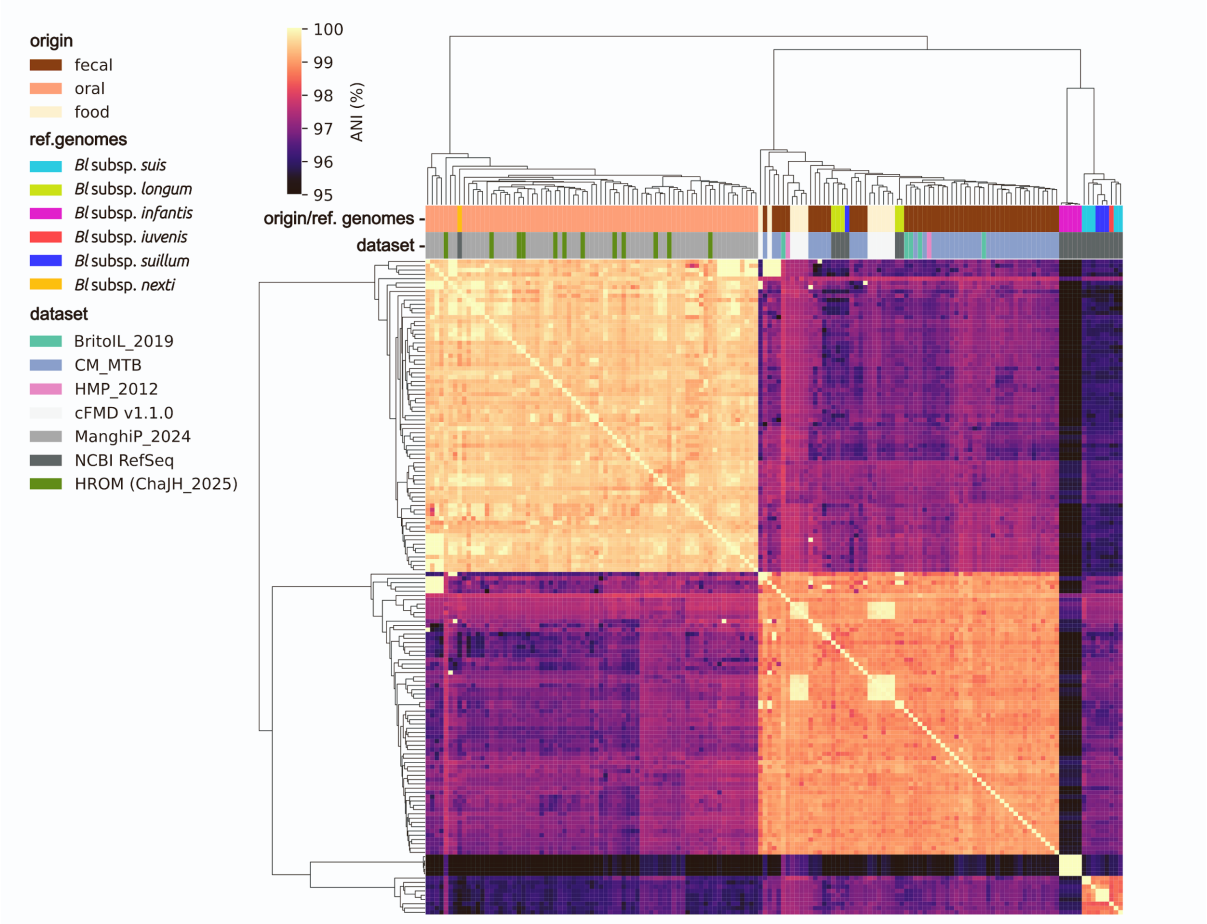


Figure S4. Genome analysis of *Bifidobacterium longum*.
All-vs-all clustermap representation of average nucleotide identities (ANIs) between *Bifidobacterium longum* (SGB17248) genomes reconstructed from metagenomes (i.e. metagenome-assembled genomes) and isolate genomes representative of each *B. longum* subspecies.

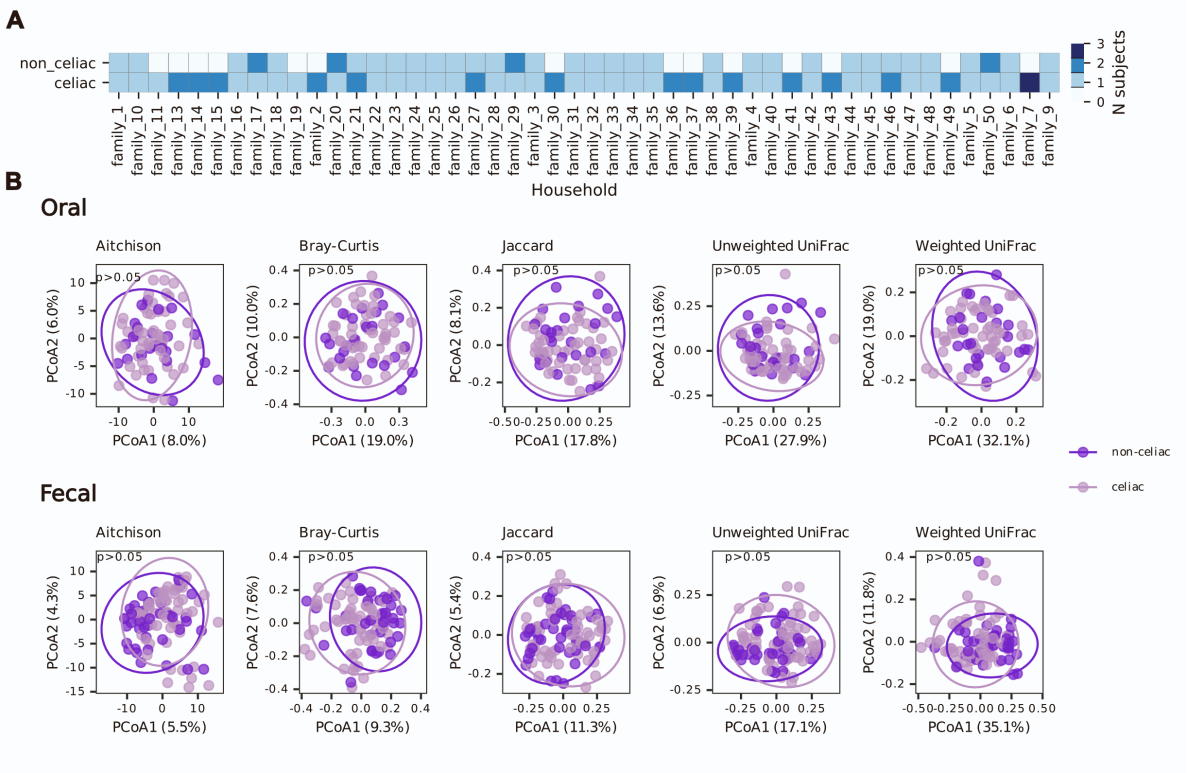


Figure S5. Evaluation of differences between celiac and non-celiac individuals from the MTC cohort.

(A) Tilemap representing the number of individuals with and without celiac disease for each household of the CM_MTC dataset included in this study.

(B) Principal Coordinate Analysis (PCoA) of oral (top row) and gut microbiomes (bottom row) in the CM_MTC dataset. Multiple beta-diversity metrics (indicated as subtitles) are shown. A multivariate PERMANOVA test was used (**Methods**) to assess differences between groups.

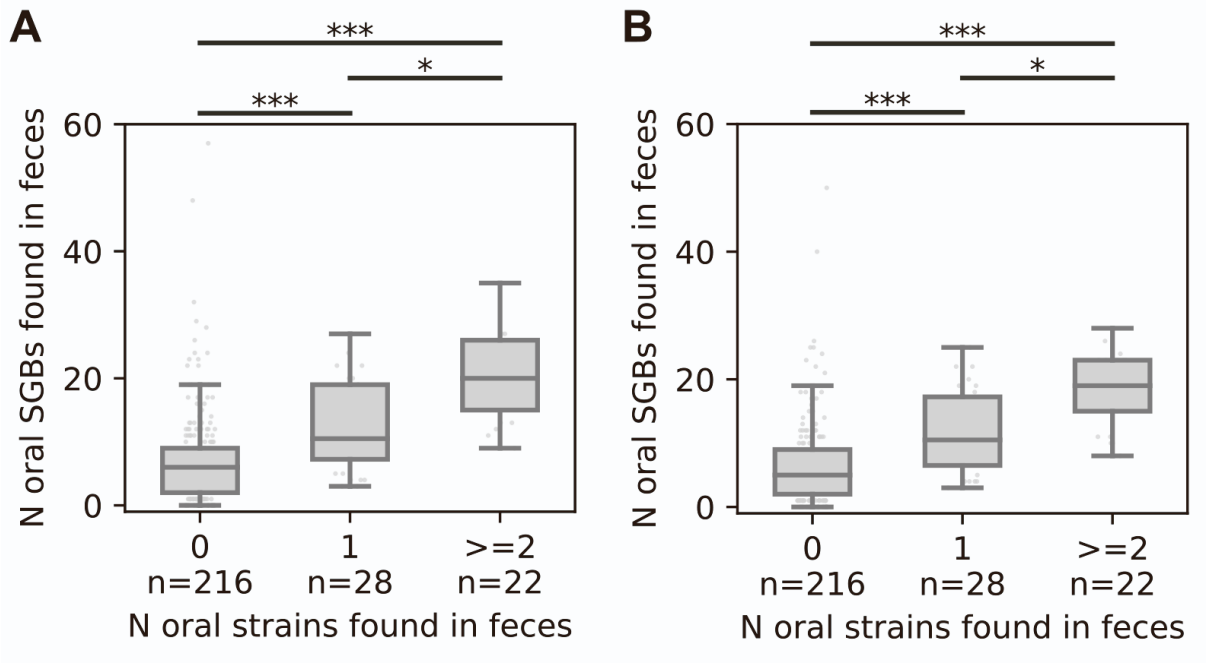


Figure S6. Sensitivity analysis of the oral-typical SGBs list definition.
(A) Number of oral-typical SGBs found in feces in cases in which 0 versus 1 versus ≥ 2 oral-derived strains are found in feces, using an alternatively generated oral-typical SGBs list considering a 10% prevalence cutoff for oral samples (**Methods**).
(B) Number of oral-typical SGBs found in feces in cases in which 0 versus 1 versus ≥ 2 oral-derived strains are found in feces, using an alternatively generated oral-typical SGBs list considering a 30% prevalence cutoff for oral samples (**Methods**).
In A and B, the post-hoc Dunn test for Kruskal-Wallis was used, with statistical significance represented as: **, $p < 0.01$; ***, $p < 0.001$. The *n* below x-tick labels indicates the number of individuals, restricting the analysis to those not used in the definition of oral-typical SGBs.