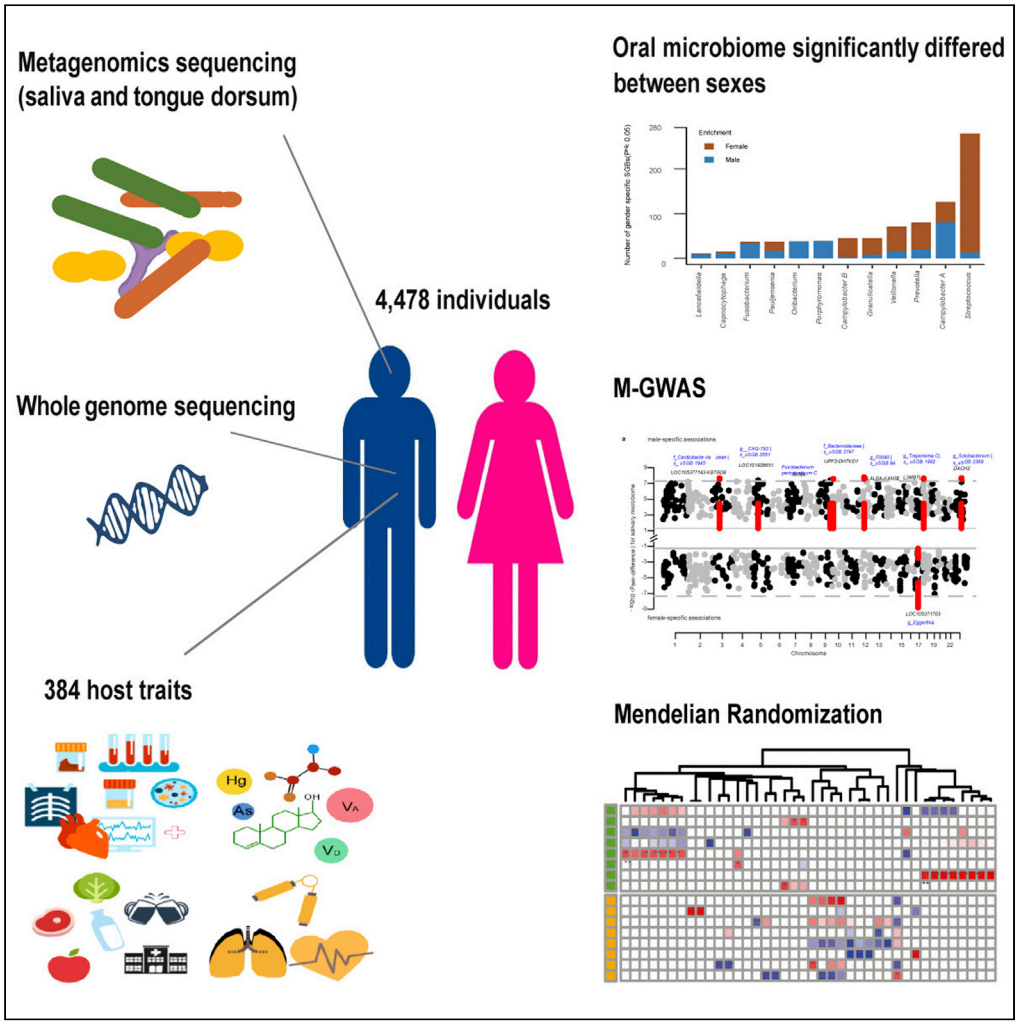
iScience

CellPress
OPEN ACCESS

Article

Sex differences in the oral microbiome, host traits, and their causal relationships



Xiaomin Liu, Xin Tong, Zhuye Jie, ..., Karsten Kristiansen, Huijue Jia, Tao Zhang

karsten@genomics.cn (K.K.)
jiahuijue@genomics.cn (H.J.)
tao.zhang@genomics.cn (T.Z.)

Highlights

The human oral microbiome significantly differed between sexes

Host factors contributing to the oral microbiome composition differed between sexes

M-GWAS identifies sex-specific genetic associations for the oral microbiome

MR reveals sex-specific causal relationships between the oral microbiome and host traits

Liu et al., iScience 26, 105839
January 20, 2023
https://doi.org/10.1016/j.isci.2022.105839

iScience

CellPress
OPEN ACCESS

Article

Sex differences in the oral microbiome, host traits, and their causal relationships

Xiaomin Liu,[1,2,8] Xin Tong,[2,8] Zhuye Jie,[2,3,8] Jie Zhu,[2] Liu Tian,[2] Qiang Sun,[2,4] Yanmei Ju,[2] Leying Zou,[2] Haorong Lu,[5] Xuemei Qiu,[2] Qiang Li,[2] Yunli Liao,[2] Heng Lian,[2] Yong Zuo,[2] Xiaomin Chen,[2] Weiqiao Rao,[2] Yan Ren,[2] Yuan Wang,[2] Jin Zi,[2] Rong Wang,[2] Xun Xu,[2] Huanming Yang,[2,6,7] Jian Wang,[2,6,7] Yang Zong,[2] Weibin Liu,[2] Yong Hou,[2] Xin Jin,[2] Liang Xiao,[2] Karsten Kristiansen,[2,3,*] Huijue Jia,[2,5,*] and Tao Zhang[2,3,9,*]

SUMMARY

The oral microbiome has been implicated in a growing number of diseases; however, determinants of the oral microbiome and their roles remain elusive. Here, we investigated the oral (saliva and tongue dorsum) metagenome, the whole genome, and other omics data in a total of 4,478 individuals and demonstrated that the oral microbiome composition and its major contributing host factors significantly differed between sexes. We thus conducted a sex-stratified metagenome-genome-wide-association study (M-GWAS) and identified 11 differential genetic associations with the oral microbiome ($p_{sex\text{-}difference} < 5 \times 10^{-8}$). Furthermore, we performed sex-stratified Mendelian randomization (MR) analyses and identified abundant causalities between the oral microbiome and serum metabolites. Notably, sex-specific microbes-hormonal interactions explained the mostly observed sex hormones differences such as the significant causalities enrichments for aldosterone in females and androstenedione in males. These findings illustrate the necessity of sex stratification and deepen our understanding of the interplay between the oral microbiome and serum metabolites.

INTRODUCTION

The gut microbiome has been referred to as our "other genome", due to its integral role in human health. The oral microbiome can colonize the intestines[1] and has also been implicated in a growing number of diseases[2–4] beyond dental caries[5] and periodontitis.[6] Many studies have indicated that host factors, such as sex, body mass index (BMI), lifestyle, and sociodemographic differences, may influence the gut microbiome composition.[7–10] However, there are limited studies on the various contributing factors to the oral microbiome variability,[11–13] and most of them were conducted on small collection samples and utilized 16S rRNA gene amplicon sequencing for the oral microbiota, whereas previous studies suggested that 16S rRNA cannot achieve the taxonomic resolution afforded by sequencing the entire gene.[14] Therefore, unraveling the factors that shape and define the oral microbiome in a well-designed cohort is crucial for the understanding of both oral and broader systemic health.

The 4D-SZ cohort is a well-designed multi-omics cohort,[15–18] with shotgun data for the metagenome across multiple body sites and the host genome, as well as metabolic traits, detailed questionnaires, and clinical information. Based on the cohort, we found host genetics significantly contributed to the gut microbiome[15] and identified 58 causal relationships between blood metabolites and the gut microbiome[16] in a total of 3,432 individuals. In addition to the gut microbiome-related studies, we have constructed a high-quality oral genome catalog[17] and further confirmed the impact of the host genome on the oral microbiome[18] by aligning metagenome data to the oral genome catalog. Notably, multiple studies including our recent paper showed evidence for the sex-specific effect on the gut microbiome.[15] Sex differences in the gut microbiome were proposed to mainly be driven by sex hormones,[19,20] which in turn contribute to sex differences in immunity and susceptibility to diseases.[7,21–23] Therefore, based on our previous good research results and the evidence of sex explaining the variance of oral microbiome composition,[13,17,18] we aimed to comprehensively dissect the sex differences in the oral microbiome, host traits (mainly blood metabolites, such as sex hormones), and their causal relationships in the 4D-SZ cohort.

[1]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[2]BGI-Shenzhen, Shenzhen 518083, China

[3]Department of Biology, University of Copenhagen, Universitetsparken 13, 2100 Copenhagen, Denmark

[4]Department of Statistical Sciences, University of Toronto, Toronto, ON M5G 1Z5, Canada

[5]China National Genebank, BGI-Shenzhen, Shenzhen 518120, China

[6]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China

[7]James D. Watson Institute of Genome Sciences, Hangzhou 310058, China

[8]These authors contributed equally

[9]Lead contact

*Correspondence: karsten@genomics.cn (K.K.), jiahuijue@genomics.cn (H.J.), tao.zhang@genomics.cn (T.Z.)

https://doi.org/10.1016/j.isci.2022.105839

This paper analyzed a whole-genomic and metagenomic shotgun sequencing dataset in a total of 4,478 individuals (3,504 saliva and 3,694 tongue dorsum) from the 4D-SZ cohort, along with host factors (anthropometric measurements, diet, periodontal conditions, routine blood tests, serum metabolome, and other metadata), to investigate the effect of sex on the oral microbiome composition (Figure S1; a study overview panel). We first identified sex-specific oral microbiome and host major factors that were significantly associated with the oral microbiome. We then performed a sex-stratified metagenome-genome-wide-association study (M-GWAS) analysis for identifying sex-specific host genetic loci. Using significant genetic variants identified by sex-stratified M-GWAS as instrumental variables, we further took Mendelian randomization (MR) analysis to excavate sex-specific causal relationships between the host factors, especially serum metabolites, and oral microbes. To the best of our knowledge, this paper is the first to use M-GWAS and MR to study the sex effect on the oral microbiome. This study not only provides a reference for the subsequent mechanism research on sex-specific differences in the oral microbiota but also helps to develop microbiota-targeted therapeutic interventions according to sex.

RESULTS

The oral microbiome differed between sexes

The oral microbial species profile for the 4D-SZ (a multi-omics cohort[15–18] sequenced in Shenzhen with 3,504 saliva and 3,694 tongue dorsum samples, among which 3,165 individuals had both sample types; Table S1), according to alignment to a previously well-constructed high-quality oral genome catalog, was highly representative of the oral microbiome including uncultured strains as well as potentially parasitic bacteria of the candidate phyla radiation (CPR).[17] Among the 384 host traits (anthropometric and serum metabolites, etc.), sex was a significant factor associated with the oral microbiome composition (PERMANOVA[24] test, $p = 2 \times 10^{-4}$ for both saliva and tongue dorsum, Table S2). 1,393 (38.834%) of the total 3,589 species-level genome bins (SGBs) in saliva and 1,416 (39.476%) SGBs in tongue dorsum were significantly different between males and females (multivariate linear model, false discovery rate [FDR] < 0.05) after adjusting for age, BMI, medication, supplements, and mouthwash frequency (see STAR Methods). 903 of the sex-differential SGBs were shared by saliva and tongue dorsum samples (Table S3). In saliva, women exhibited more *Streptococcus* (20.028%, SGBs number of female/male enrichment = 266:13), *Prevotella* (5.743%, female:male = 60:20), and *Granulicatella* (3.230%, female:male = 37:8). In contrast, men's saliva showed more *Campylobacter A* (9.045%, female:male = 46:80), *Veillonella* (5.096%, female:male = 56:15), *Porphyromonas*(2.799%, female:male = 1:38), and *Oribacterium* (2.727%, female:male = 1:37) (Figures 1A and 1B). In the tongue dorsum, some genera extended the sex differences. Notably, women showed more *Streptococcus* (12.711%, female:male = 146:34), *Lancefieldella* (7.768%, female:male = 107:3), *Prevotella* (6.850%, female:male = 81:16), *Granulicatella* (3.531%, female:male = 38:12), and *Pauljensenia* (4.802%, female:male = 59:9), while men's tongue dorsum showed more *Campylobacter A* (10.734%, female:male = 60:92), *Porphyromonas* (3.178%, female:male = 0:45), *Capnocytophaga* (2.966%, female:male = 0:42), *Fusobacterium* (5.014%, female:male = 1:70), and *Oribacterium* (2.471%, female:male = 6:29) (Figures 1C and 1D). These sex-differential SGBs showed good discriminated efficiency for sex groups, with the average area under the receiver operating characteristic (ROC) curve (AUC) = 0.878 (95% confidence interval [CI]: 0.866–0.890; Figure S2) and AUC = 0.818 (95% CI: 0.804–0.833) in saliva and tongue dorsum samples, respectively. When splitting into the discovery and replication cohorts (Table S1), the differentially abundant species between males and females identified in the discovery cohort can also discriminate the sex groups in the independent replication cohort with the same, even slightly higher, accuracy (AUC; Figure S3 and Table S4), again suggesting the robustness and the critical roles of the oral microbiome in the discriminating sex.

Major factors associated with the oral microbiome differed between sexes

To identify the contributing factors to the microbiome composition, we first examined the variance of the microbiome composition explained by each factor. We identified 158 and 129 host factors, including host physiological, lifestyle, and blood metabolites, which were significantly associated with saliva microbiome and tongue dorsum, respectively (FDR < 0.05) (Table S2 and Figure S4). Dental calculus (explained 1.3% of the variance for saliva microbial composition, 1.24% for tongue dorsum), the bleeding frequency of gums (0.87% for saliva, 0.3% for tongue dorsum), and high-fat and high-sugar diet frequency (0.88% for saliva, 1.3% for tongue dorsum) were among the most important factors associated with the oral microbiome compositions (Figure 2 and Table S2). We further investigated the correlation between the relative abundances of each oral microbial taxa and each host metabolic trait using multivariate linear regression. After adjustment for sex, age, and other confounders (see STAR Methods), we observed 3,843 and 3,963

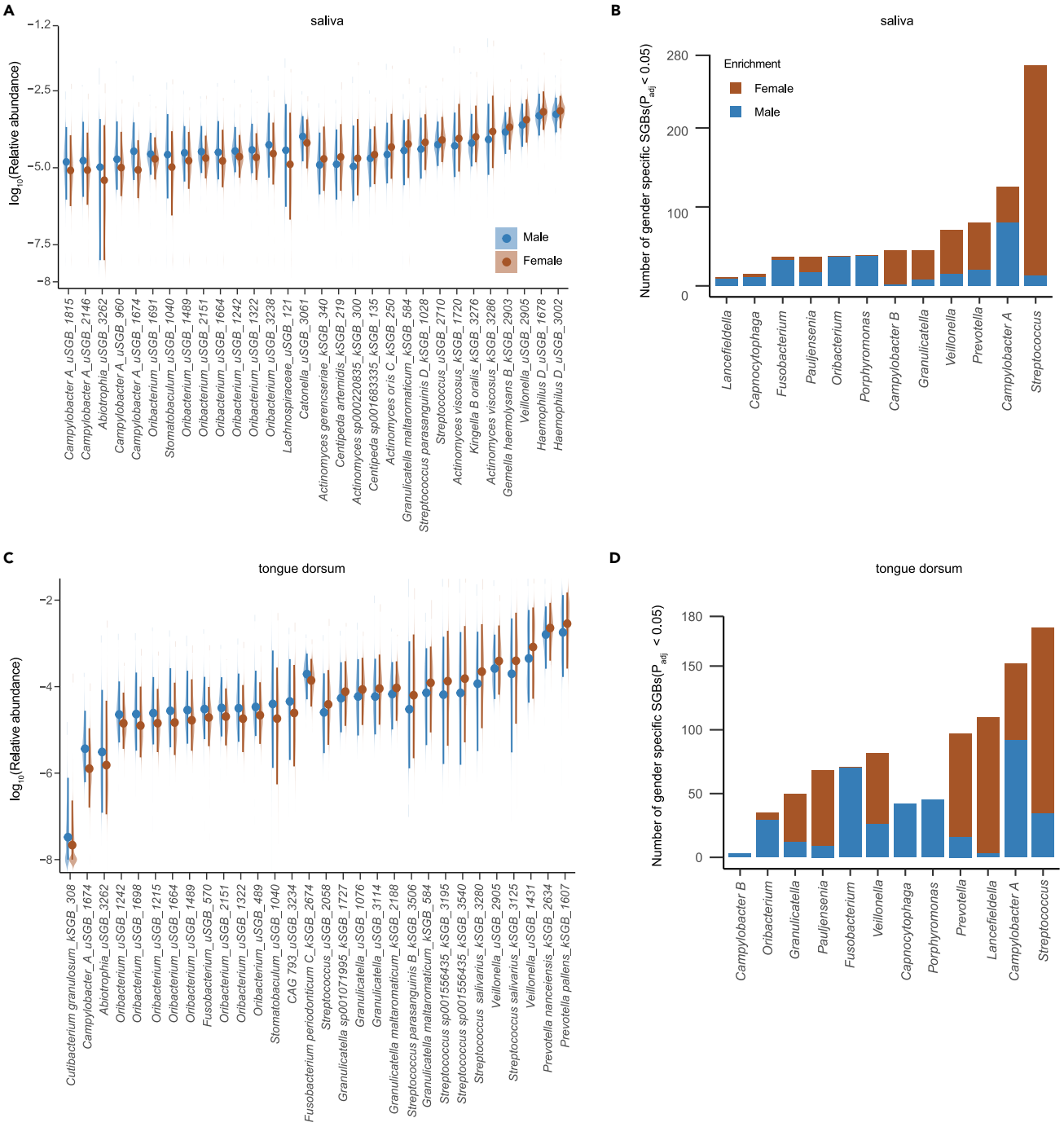


Figure 1. The oral microbiome differed between sexes

(A) Boxplot of represented sex-specific SGBs for the salivary microbiome. The top 15 important (ranked by five-fold random forest classifier for sex) females- or males-enriched SGBs with $p < 1 \times 10^{-8}$ (GLM test with adjusting confounders) were shown.

(B) SGBs number of female/male enrichment for the salivary microbiome.

(C) Boxplot of represented sex-specific SGBs for tongue dorsum microbiome. The top 15 important (ranked by five-fold random forest classifier for sex) females- or males-enriched SGBs with $p < 1 \times 10^{-8}$ (GLM test with adjusting confounders) were shown.

(D) SGBs number of female/male enrichment for tongue dorsum microbiome.

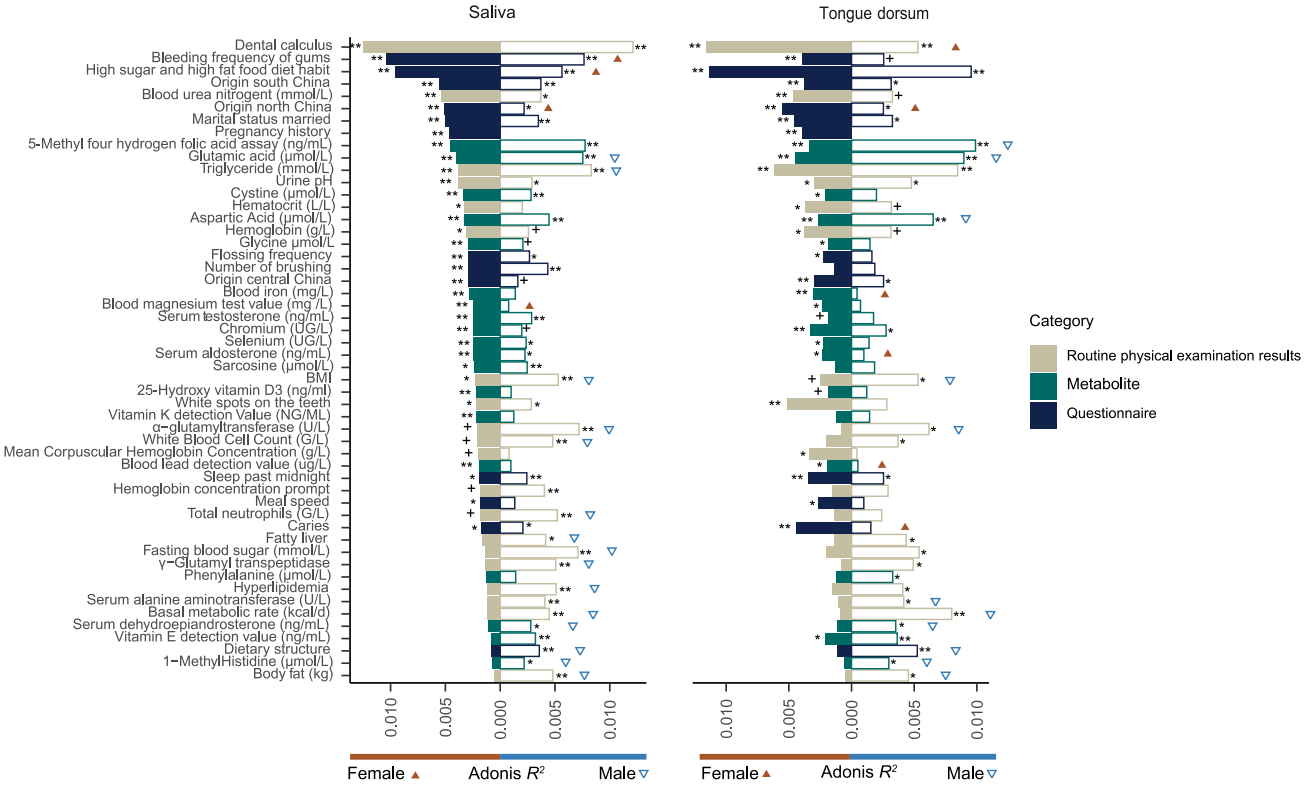


Figure 2. Host major factors associated with the oral microbiome differed between sexes

The top 52 host major factors correlated with microbiome composition for each oral niche between sexes were shown using the PERMANOVA test (based on bray distance, 4,999 permutations). Major factors were ranked based on variance explanation (R^2) from the females' saliva profiles. Stars on the bar represented the major factors well explained the microbiome composition in the PERMANOVA test. ** for Benjamini-Hochberg (BH) significance Q < 0.01, * for Q < 0.05, + for Q < 0.1. Triangle represented the significant differences between beta-diversity-based F statistics of females and those of males at Q < 0.001, by applying a bootstrapping method to estimate the dispersion of differences in the microbiota community (Methods). 27 of the top 52 host major factors explained the variance of microbiome compositions largely differed between sexes, with the red upper triangle representing stronger correlations in females, while the blue lower triangle representing stronger correlations in males.

significant correlations between host metabolic traits and salivary and tongue dorsum microbiota, respectively (FDR < 0.05, Table S5), which as a prerequisite suggested a possible strong causality in the subsequent MR analyses.

When stratified by sex, dental calculus explained a larger variance of the tongue dorsum microbial composition in females than males; however, the bleeding frequency of gums and high-fat and high-sugar diet frequency explained larger variances of the salivary microbial composition in females (n = 2,509) than males (n = 1,955) (Figure 2 and Table S6). Blood urea nitrogen (BUN) was among the top 10 factors (0.6% for saliva, 0.8% for tongue) and explained a larger variance of microbial composition in females for both saliva and tongue dorsum samples. Urea entering the mouth is hydrolyzed to carbon dioxide and ammonia by bacterial ureases.[25] The positive association between BUN and *Haemophili* strains and the negative association between BUN and *Streptococcus* strains supported the ureolytic activity of *Streptococcus*[25] (Table S5). Factors such as folic acid, triglyceride, BMI, hyperlipidemia, fatty liver, fasting blood sugar, and basal metabolic rate had stronger associations in males than in females (Figure 2). Besides, female-specific factors (age of menarche, the menstrual cycle, pregnancy, lactation, menopause; 0.1%–0.7% for saliva, 0.1%–1.2% for tongue), dietary structure (0.1% for saliva, 0.1% for tongue), and oral lifestyle (number of brushing, flossing frequency; 0.3%–0.4% for saliva, 0.2%–0.3% for tongue) were all found to be important factors for the oral microbiome (Figure S4 and Table S6). Serum steroid metabolites (aldosterone, progesterone, corticosterone, testosterone; 0.1%–0.4% for saliva, 0.1%–0.4% for tongue) and biomarkers of female-specific factors were also found to be important and will be described in detail in the MR analyses.

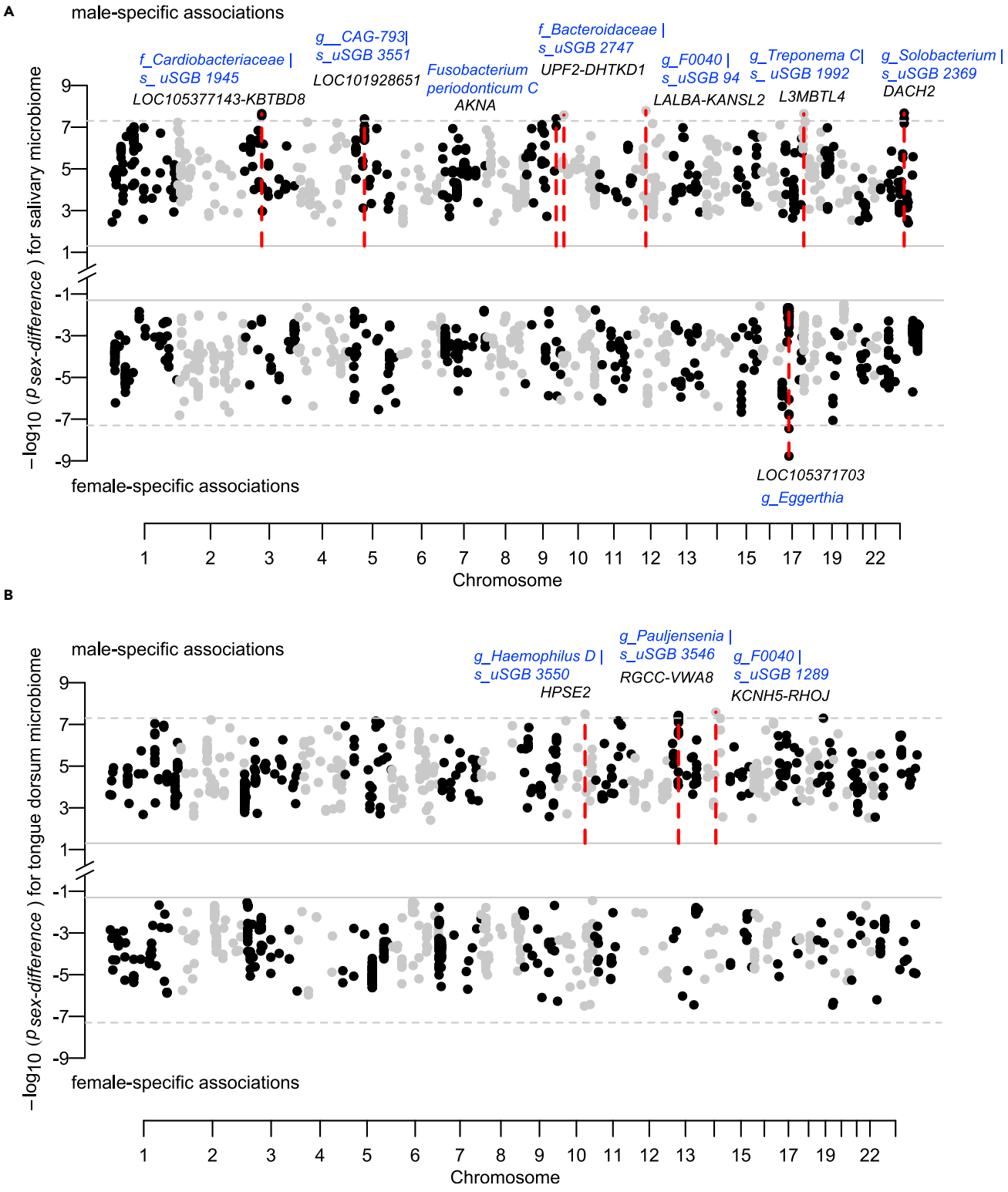


Figure 3. Sex-stratified M-GWAS identified sex-specific signals for the oral microbiome

(A and B) Sex-stratified M-GWAS identified sex-specific signals for the salivary (A) and tongue dorsum (B) microbiome. Male-specific associations represented the M-GWAS results with $p < 5 \times 10^{-8}$ in males but $p > 0.05$ in females and $p_{sex\text{-}difference} < 0.05$. Female-specific associations represented the

Figure 3. *Continued*

M-GWAS results with $p < 5 \times 10^{-8}$ in females, $p > 0.05$ in males, and $p_{sex\text{-}difference} < 0.05$. The most significant sex-differential M-GWAS associations ($p < 5 \times 10^{-8}$ in one sex but $p > 0.05$ in the other sex, and $p_{sex\text{-}difference} < 5 \times 10^{-8}$) were listed using genes and associated microbial taxa and marked in red lines. The gray solid line represented the sex-differential p cut-off of 0.05 and the gray dotted line represented the sex differential *p* cut-off of 5×10^{-8}.

Many of the differences in beta-diversity metrics (40%) between sexes were supported by both saliva and tongue dorsum samples (Figures 2 and S5A; Table S6). These data were further analyzed by using the generalized linear model (Table S5). This model confirmed the sexual dimorphic associations of metabolites with microbiota. The metabolite showing the greatest difference in the association with the oral microbiome between males and females was glutamic acid (Figure S5B). The average concentration of glutamic acid in males is 1.39 times higher than that in females when calculating sex difference of host traits (Table S7). These results indicated sexual dimorphism of oral microbial biodiversity as measured by the variance of beta diversity explained by host traits.

Human genetic associations with the oral microbiome include associations with the sex chromosomes

Sex-combined M-GWAS analyses of the 4D-SZ cohort identified 340 and 374 independent loci significantly associated with tongue dorsum and salivary microbiome ($p < 5 \times 10^{-8}$), respectively.[18] 17 of the loci associated with the tongue dorsum microbiome and 14 of the loci associated with the salivary microbiome were located in the X chromosome (Table S8). For example, in M-GWAS for tongue dorsum samples, the abundance of genus *Leptotrichia A* was associated with the *WDR44* gene, which encodes a protein that interacts with the small GTPase rab11. *Haemophilus influenzae*, the most common pathogen causing asthma,[26] was associated with a deletion near the *FAM9C* gene, which has been reported to be associated with asthma.[27] In M-GWAS for salivary samples, *Veillonella parvula*, related to the autoimmune disease rheumatoid arthritis (RA)[2] which is more prevalent in females, was associated with SNPs near the *SMARCA1* gene.

Although the sample size became smaller, stratifying the individuals by sex could reduce heterogeneity[28,29] and help obtain sex-differential signals from M-GWAS. Sex-stratified M-GWAS for salivary samples identified 327 male-specific independent associations (Table S9 and Figure 3A; $p < 5 \times 10^{-8}$ in males; $p > 0.05$ in females and $p_{sex\text{-}difference} < 0.05$; independent loci defined as distance <1MB and $r^2 < 0.1$) and 277 female-specific associations ($p < 5 \times 10^{-8}$ in females; $p > 0.05$ in males and $p_{sex\text{-}difference} < 0.05$) in the discovery dataset. Apart from the associations involving the variants not found in the low-depth dataset (Table S1), 13 of the 242 male-specific and 14 of the 216 female-specific associations could be replicated in the same direction of the minor allele in the same sex ($p < 0.05$). The strongest female-specific association was observed for rs1410336242 near the *LOC105371703* gene ($\beta_{female} = 0.22$, $p_{female} = 1.30 \times 10^{-12}$ compared with $\beta_{male} = -0.06$, $p_{male} = 0.08$; $p_{sex\text{-}difference} = 1.68 \times 10^{-9}$, Figure 3A) associated with genus *Eggerthia*. Seven male-specific associations were observed with marginal $p_{sex\text{-}difference} < 5 \times 10^{-8}$: rs142374260 at *LALBA-KANSL2* with *F0040 uSGB 94*; rs370266368 at *L3MBTL4* with *Treponema C uSGB 1992*; rs756149092 at *UPF2-DHTKD1* with *uSGB 2747* belonging to family Bacteroidaceae; rs147250707 at *AKNA* with *Fusobacterium periodonticum C*; rs11739353 at *LOC101928651* with *CAG-793 uSGB 3551*; rs1206782373 at *DACH2* with *Solobacterium spp.*; and rs74694053 at *LOC105377143-KBTBD8* with *uSGB 1945* belonging to family Cardiobacteriaceae. Both *Solobacterium* and Carnobacteriaceae in the oral cavity showed associations with a reduced risk of colorectal cancer (CRC).[17]

For the tongue dorsum, we identified 339 male-specific and 269 female-specific independent associations (Table S9 and Figure 3B; $p < 5 \times 10^{-8}$ in one sex but $p > 0.05$ in other sex; $p_{sex\text{-}difference} < 0.05$). Except for the not found variants in the low-depth replication dataset, 18 of the 247 (18/247) available male-specific associations and 15/197 female-specific associations could be replicated in the same direction of the minor allele in the same sex ($p < 0.05$), showing a stronger replication power than for saliva. Three male-specific associations were observed with marginal $p_{sex\text{-}difference} < 5 \times 10^{-8}$ but none for female-specific associations. They are rs192945350 at *KCNH5-RHOJ* associated with *F0040 uSGB 1289*; chr10:99050952 at *HPSE2* associated with *Haemophilus D spp.* that were depleted in individuals with RA[2]; and rs9532851 at *RGCC-VWA8* associated with *Pauljensenia* spp., a genus recently renamed from *Actinomyces*,[30] which was previously identified as highly predictive of RA and a decreased risk of CRC.[17]

In addition to the sex-specific associations including the 11 top associations with $p_{sex\text{-}difference} < 5 \times 10^{-8}$, we also investigated sex-specific pathways based on sex-specific loci (Table S10). For the salivary

microbiome, the male-specific associated SNPs were mainly enriched in 6 pathways in KEGG or REACTOME database ($p < 0.05$ in DAVID tools). These pathways were hsa00510: N-Glycan biosynthesis, hsa04514: cell adhesion molecules (CAMs), hsa01100: metabolic pathways (for example carbohydrate metabolic process from GO database), hsa04015: Rap1 signaling pathway, hsa01230: biosynthesis of amino acids, and R-HSA-205017: NFG and proNGF binds to p75NTR. These results supported previous findings that the males exhibited more abundant N-glycans[31] and tend to oxidize more total carbohydrate[32] than females. The female-specific associated SNPs enriched pathway included R-HSA-167060: NGF (nerve growth factor) processing, GO:0045596~negative regulation of cell differentiation, and GO:0007507~heart development that indicated the sex-specific modification on cardiac gene expression.[33] For the tongue dorsum microbiome, the male-specific associated SNPs were clustered into heparan sulfate/heparin-glycosaminoglycan (HS-GAG) metabolism-related pathway, such as R-HSA-2024096: HS-GAG degradation, R-HSA-2022928: HS-GAG biosynthesis, R-HSA-1971475: a tetrasaccharide linker sequence is required for GAG synthesis, and hsa04722: neurotrophin signaling pathway. Studies have indicated that some functions or action mechanisms of brain-derived neurotrophic factors vary in a sex-dependent manner and may be modulated by sex hormones.[34,35] The female-specific associated SNPs were enriched in 6 pathways including hsa04261: adrenergic signaling in cardiomyocytes, hsa04024: cAMP signaling pathway, hsa05032: morphine addiction, hsa05033: nicotine addiction, hsa04973: carbohydrate digestion and absorption, and hsa04725: cholinergic synapse. Sex-specific adrenergic signaling in cardiomyocytes again suggested sex dimorphism in cardiac functions.[33,36] The cholinergic synapse transmission also showed sex dimorphism in adult *C. elegans*.[37]

MR identified causal relationships between the oral microbiome and host traits

With the strong genetic associations with the oral microbiome, we next performed MR to identify potential causal relationships between host traits and oral microbes. Sex-combined one-sample MR analyses identified 502 and 290 significant causal relationships between host traits and the salivary or tongue dorsum microbiome, respectively (Table S11; *Bonferroni* adjusted $p < 0.05$ for the 3,843 and 3,963 observationally significant correlations between host traits and the salivary and tongue dorsum microbiota, respectively). 85.5% (429/502) of the host traits-salivary microbiome causalities and 57.6% (167/290) of the host traits-tongue dorsum microbiome causalities were replicated in the MR analysis on the replication cohort (in the same direction with $p < 0.05$; Table S11), reinforcing these causal effects. A majority of the causalities lay in the bleeding frequency of gums and dental calculus, in agreement with our above finding that they explained most for oral microbial compositions. Our MR identified the causal effects of periodontal pathogens in both niches such as *Porphyromonas endodontalis* (Figure S6; $\beta = 0.285$, $p = 2.69 \times 10^{-10}$ for saliva; $\beta = 0.266$, $p = 1.13 \times 10^{-11}$ for tongue), *Prevotella intermedia* ($\beta = 0.196$, $p = 1.04 \times 10^{-6}$ for saliva; $\beta = 0.201$, $p = 2.92 \times 10^{-7}$ for tongue), *Treponema B denticola* ($\beta = 0.226$, $p = 8.75 \times 10^{-7}$ for saliva; $\beta = 0.211$, $p = 1.14 \times 10^{-7}$), and *Fusobacterium nucleatum* ($\beta = 0.168$, $p = 1.21 \times 10^{-3}$ for saliva; $\beta = 0.139$, $p = 2.43 \times 10^{-3}$ for tongue) on the bleeding frequency of gums and dental calculus, and the reverse effects were also true. These causal relationships were robust and well replicated in the replication dataset (Figure S6). In addition, we found several cardiometabolic-related factors, such as chromium,[38] creatine,[39] cystine,[40] and glutamic acid[41] causally linked to the abundances of a few oral microbiota (Table S11). For example, chromium ($\beta = -0.158$, $p = 4.25 \times 10^{-6}$), cystine ($\beta = -0.247$, $p = 1.14 \times 10^{-12}$), and cystathionine ($\beta = -0.252$, $p = 3.50 \times 10^{-10}$) negatively correlated with the abundance of *Lactobacillus*, while creatine ($\beta = 0.327$, $p = 1.52 \times 10^{-13}$) and glutamic acid ($\beta = 0.345$, $p = 2.46 \times 10^{-17}$) positively correlated with the abundance of *Lactobacillus* in salivary samples, which suggested that the human circulating metabolites were also important for *Lactobacillus* residing in the oral cavity. These causalities inferences were in line with the experimental verification that blood chromium could be reduced by using gut *lactobacillus*,[42] which directly catalyzed the generation of L-cysteine persulfide from L-cystine and generation of L-cysteine from cystathionine.[43]

Sex-stratified one-sample MR identified 62 male-specific and 45 female-specific causal relationships between the salivary microbiome and host traits (Figures 4A and 4B; Table S11), as well as 27 male-specific and 36 female-specific causal relationships between the tongue dorsum microbiome and host traits (Figures 4C and 4D; Table S11). These sex-specific causalities were those significant in one sex (*Bonferroni* adjusted $p < 0.05$) but not significant in the other ($p > 0.05$) and exhibiting significant differences between sexes ($p_{sex\text{-}difference} < 0.05$). Oral hygiene practices and diets showed different effects between males and females. Mouthwash exhibited a negative effect on *Lachnoanaerobaculum uSGB 1301* and *Oribacterium spp.* in saliva only for males, and species belonging to these two genera were associated with low uric

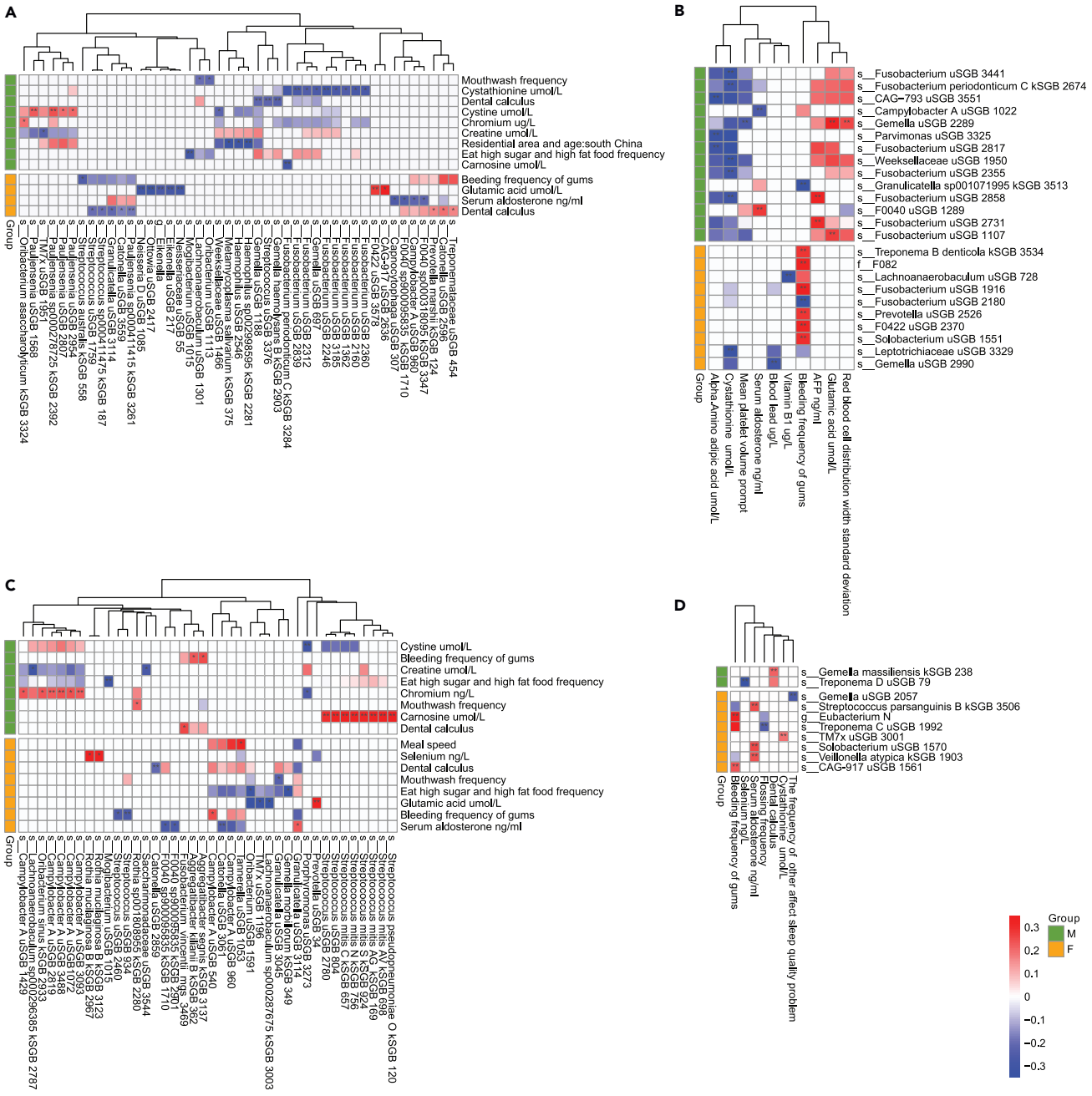


Figure 4. Sex-stratified MR results were identified for the oral microbiome

(A–D) Left panels showed the causal effects of host phenotypic traits (serum metabolites etc.) on the salivary (A) and tongue dorsum microbiota (C), respectively. Right panels showed the causal effects of the salivary (B) and tongue dorsum microbiota (D), respectively, on host phenotypic traits (serum metabolites, etc.). The cells marked with "**" represented the p value of sex difference less than 0.01, while "*" represented the p value of sex difference less than 0.05. Each part had two groups, representing significant MR results in males (blue) or females (orange), respectively.

acid determinator *SLC2A9* in our recent study.[18] Mouthwash was also positively correlated with *Rothia sp001808955* in males and negatively associated with *Granulicatella uSGB 3045* in females for the tongue dorsum microbiome. Bleeding frequency of gums had a negative causal effect on female-enriched *Streptococcus spp.* in females for both niches and a positive causal effect on male-enriched *Aggregatibacter spp.* (*A. kilianii B* and *A. segnis*) in males for the tongue dorsum microbiome. When looking in the opposite direction, increased periodontal pathogens *Treponema B denticola*, *Fusobacterium uSGB 1916*, family

F082, *Prevotella uSGB 2526*, *F0422 uSGB 2370*, and *Solobacterium uSGB 1551* all caused increased bleeding frequencies of gums, but increased *Fusobacterium uSGB 2180* caused decreased bleeding frequencies of gums. High-fat and high-sugar diet causally linked to lower *Mogibacterium uSGB 1015* abundance for males in both oral niches and lower abundances of *Gemella morbillorum* and *Oribacterium uSGB 1591* in females for tongue dorsum microbiome. Increased *Gemella massiliensis kSGB 238* was causally related to an increased risk of dental calculus.

Serum amino acids as basic metabolic regulators also showed abundant sex-specific causal relationships with the oral microbiome (Figure 4). Increased carnosine, which marked sexual dimorphism in skeletal muscles and was remarkably higher in males than in females,[44] was causally related to decreased *Fusobacterium periodonticum C* level in saliva and increased abundances of *Streptococcus spp.* (mainly *S. mitis and S. pseudopneumoniae O*) in tongue dorsum, only for males (Figure S7). Cystathionine had a causal effect on lower *Fusobacterium spp.* abundance in males' tongue dorsum. In turn, *Fusobacterium spp.* such as *Fusobacterium periodonticum C* significantly correlated with a lower cystathionine level in males' saliva, suggesting the negative correlation between them is reciprocal when combined with the finding in the previous literature[45,46] that *Fusobacterium spp.* can efficiently utilize cystathionine to produce hydrogen sulfide (H2S), a toxic foul-smelling gas. However, *TM7x uSGB 3001* had a positive causal effect on cystathionine in females' tongue dorsum. When investigating metagenomics functional modules[47] related to cystathionine, we found that K01739: cystathionine gamma-synthase (metB: 2.5.1.48) and K01758: cystathionine gamma-lyase (CTH: 4.4.1.1) were enriched in cystathionine-associated genus like *Fusobacterium* (Figure S8). Increased glutamic acid was causally related to decreased abundances of *Neisseriaceae spp.*, *Ottowia uSGB 2417*, and *Eikenella* but increased abundance of *F0422 uSGB 1710* (family Veillonellaceae) in females' saliva. Increased glutamic acid also had a causal effect on lower abundances of *TM7x uSGB 1196* and *Lachnoanaerobaculum sp000287675* but had a positive correlation with *Prevotella uSGB 34*. We found glutamate-related modules (eg. ornithine biosynthesis, proline biosynthesis, heme biosynthesis, and glutathione biosynthesis) were enriched in *Neisseriaceae*, *Ottowia*, *Eikenella*, *Lachnoanaerobaculum sp000287675*, and *Prevotella uSGB 34* genomes. Cystine had a positive causal effect on genus *Pauljensenia spp.* such as *Pauljensenia sp000278725* in saliva but a negative causal effect on *Porphyromonas uSGB 3273* in tongue dorsum only for males. Cystine-related modules (eg. cysteine biosynthesis, methionine salvage, methionine biosynthesis, and methionine degradation) were also enriched in *Pauljensenia* and *Porphyromonas uSGB 3273* (Figure S8). Increased *Fusobacterium uSGB 2817*, *Parvimonas uSGB 3325*, and *CAG−793 uSGB 3551* were causally linked to a lower level of alpha-aminoadipic acid, which was found to be a marker of diabetes risk,[48] autoimmunity, and age-associated changes in human collagen.[49,50]

Sex hormones also showed sex-specific causalities enrichments. The hormone aldosterone, whose level is higher in females than in males in this cohort (Figure 5A and Table S7), exhibited significant causalities enrichments in females for both niches, with 139 causalities significant in females ($p < 0.01$) compared with 52 causalities significant in males for saliva (Figure 5B; chi-squared test, $p = 7.39 \times 10^{-14}$) and 139 causalities significant in females compared with 52 causalities significant in males for tongue dorsum ($p < 2.2 \times 10^{-16}$). For example, aldosterone showed causal effects on females' oral microbiome such as decreased *F0040 spp.* abundance in both niches. However, species *Streptococcus parasanguinis B*, *Solobacterium uSGB 1570*, and *Veillonella atypica* all had positive causal effects on serum aldosterone in females' tongue dorsum samples (Figure 5C). When further investigating the other sex hormones, we found cortisone and dehydroepiandrosterone, whose levels were higher in females than in males (Table S7), both showed significant causalities enrichments in females for tongue dorsum ($p = 0.03$ and $p = 0.002$, respectively). Consistently, one of several "male" sex hormones, androstenedione as an androgen showed causalities enrichments in males for tongue dorsum ($p = 1.17 \times 10^{-3}$). Androstenedione was genetically positively correlated with the abundance of the genus *Pauljensenia* (renamed from *Actinomyces*) and its species such as *P. cellulosilytica* and *P. odontolyticus*. Moreover, its other two species, *Pauljensenia uSGB 1578* and *uSGB 682* showed reciprocal positive correlations with androstenedione. All these causal relationships between androstenedione and *Pauljensenia* only existed in males but not in females for tongue dorsum ($p_{\text{sex-difference}}<0.05$; Figure 5D). The above findings suggest that sex-specific microbes-hormonal interplays explain the mostly observed sex-specific differences in the oral microbial composition and serum hormones.

In addition, the microelement chromium was causally and positively correlated with male-enriched *Campylobacter A* in males' tongue dorsum, and selenium had a positive causal effect on *Rothia mucilaginosa* in females' tongue dorsum. *Lachnoanaerobaculum uSGB 728* associated with low uric acid

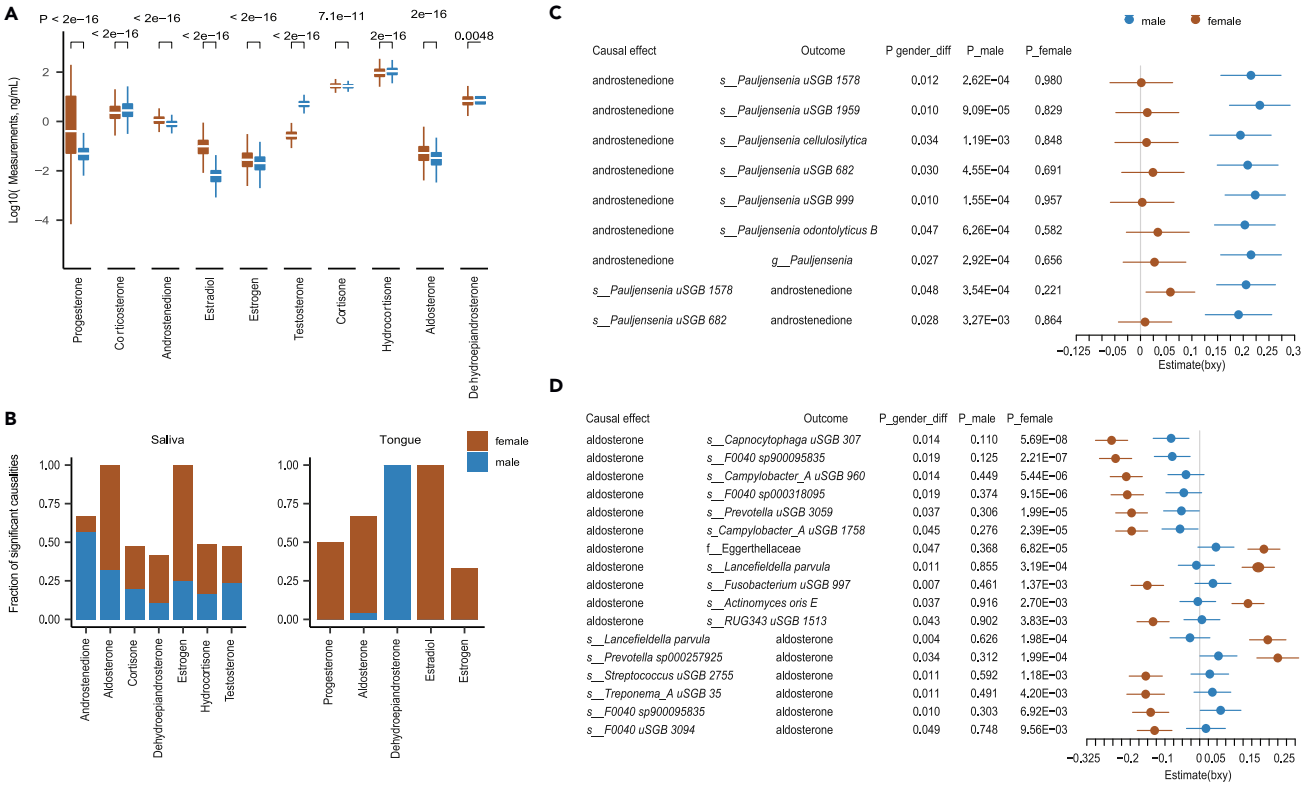


Figure 5. Sex-specific microbes-hormonal interplays explained the observed sex-specific differences in serum hormones

(A) Barplot showed the comparisons of sex hormone levels in males and females, and significant sex differences were observed (see data details in Table S7). A chi-square test was used, and a p value was marked.

(B) Several sex hormones showed consistent MR enrichments with oral microbiota in specific sex. The length of the bar represented the proportion of significant causalities ($p < 0.01$) in all causalities associated with the specific sex hormone that were calculated in males and females, respectively.

(C) The Forest plot showed androstenedione (higher levels in males than in females) was genetically positively correlated with the abundance of genus *Pauljensenia* and its species in males but not in females for salivary samples ($p < 0.01$ in males; $p > 0.05$ in females; and $p_{sex\text{-}difference} < 0.05$). The MR estimates and 95% CI values were shown in the plot. Corresponding p values of MR analysis in males, in females, and their compared T statistics were listed.

(D) The Forest plot showed aldosterone (higher levels in females than in males) had causal relationships with abundances of several oral bacteria in females but not in males for salivary samples ($p < 0.01$ in females; $p > 0.05$ in males; and $p_{sex\text{-}difference} < 0.05$). The MR estimates and 95% CI values were shown in the plot. Corresponding p values of MR analysis in males, in females, and their compared T statistics were listed.

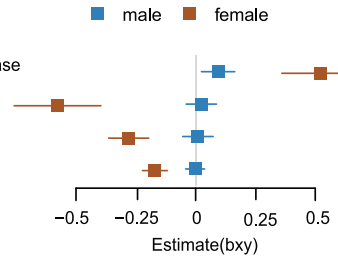
determinator SLC2A9 in our M-GWAS study exhibited a causal effect on a lower vitamin B1 level. *Treponema D uSGB 79* was causally related to a lower selenium level. A previous study reported *Treponema spp.* showed a growth dependence on selenium.[51] *Gemella uSGB 2990* was causally linked to a blood lead level.

MR inferred causalities between the oral microbiome and diseases

We inferred the causal relationships between the oral microbiome and diseases by performing two-sample MR analyses using the oral microbiome GWAS summary data in this cohort, together with diseases GWAS summary statistics from Japan Biobank.[52] MR analysis identified abundant causal effects between the oral microbiome and diseases (Table S12). Both *Pauljensenia odontolyticus* (also named *Actinomyces odontolyticus*) group C and B ($p = 6.49 \times 10^{-7}$ and $p = 3.90 \times 10^{-4}$, respectively) from saliva were causally related to increased risk of type 2 diabetes (T2D). *Neisseria meningitidis* is a major cause of meningococcal disease,[53] and its relative abundance in the tongue dorsum was causally linked to an increased risk of arrhythmia ($p = 4.44 \times 10^{-6}$). In addition, we also found 97 causalities shared by both salivary and tongue dorsum microbiomes with diseases (Table S12). Notably, the causality between *Veillonella rogosae* and asthma was reciprocal, with *Veillonella rogosae* reducing the risk of asthma, and asthma correlated with a lower abundance of *Veillonella rogosae*. Increased abundances of *Fusobacterium periodonticum C*, *Streptococcus cristatus B*, *Pauljensenia pyogenes*, and *Pauljensenia uSGB 2911* were causally linked to

A

Causal effect	Outcome	P_obs_diff	P_MR_diff	P_male	P_female
Fusobacterium periodonticum C kSGB 3157	Peripheral artery disease	6.39E−05	1.47E−02	0.185	1.18E−03
Veillonella rogosae kSGB 3494	COPD	1.24E−11	1.77E−03	0.727	1.40E−03
Oribacterium sinus kSGB 2108	Colorectal Cancer	1.74E−07	5.92E−03	0.903	7.56E−04
Leptotrichia A sp001274535 kSGB 2052	Arrhythmia	4.71E−09	1.05E−02	0.924	9.79E−04



B

Causal effect	Outcome	P_obs_diff	P_MR_diff	P_male	P_female
Cerebral aneurysm	*Streptococcus uSGB 564*	4.40E−16	3.86E−04	9.64E−05	0.206
Cerebral aneurysm	*Actinomyces uSGB 1511*	1.20E−04	3.70E−02	1.20E−03	0.976
Cerebral aneurysm	*Streptococcus cristatus kSGB 1269*	9.80E−05	5.10E−03	1.19E−03	0.363
Cerebral aneurysm	*Streptococcus oralis kSGB 1087*	2.06E−15	2.78E−03	9.18E−04	0.268
Cerebral aneurysm	*Streptococcus pseudopneumoniae M kSGB 576*	3.75E−06	2.85E−03	7.47E−04	0.312
Cerebral aneurysm	*Catonella uSGB 2784*	2.16E−16	2.66E−02	8.00E−04	0.947
Cerebral aneurysm	*f__Anaerovoracaceae|g__UBA6648 uSGB 45*	6.55E−04	4.12E−03	1.25E−03	0.315
Cerebral aneurysm	*Fusobacterium uSGB 2161*	9.11E−07	3.33E−02	7.16E−04	0.846

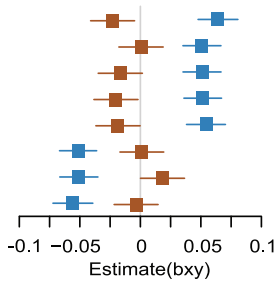


Figure 6. Sex-specific causal relationships between the oral microbiome and diseases

(A) Forest plots represented the significant causal effects of four salivary bacteria, which were more prevalent in females than in males, on diseases in females but not in males. The MR estimates and 95% CI values were shown.

(B) Forest plots represented the MR estimates and 95% CI values of the causal effect of the cerebral aneurysm on salivary microbiota in males and females, respectively. P_obs_diff represented the differential significance of relative abundances between sexes in observational analyses (see details in Table S3). P_MR_diff represented the differential significance of MR estimates between sexes (see details in Table S13). P_male and P_female represented the significance of causalities as inferred by MR in males and females, respectively.

increased risk of T2D. Decreased abundance of *Treponema C uSGB 1992* was causally related to an increased risk of T2D. These results were in line with previous findings[54] that the oral microbiome *Streptococcus* and *Fusobacterium* had higher abundances but *Treponema* had a lower abundance in T2D patients compared to control individuals. Corroborating results from the 4D-SZ cohort, two species, from male-enriched *Campylobacter A* (*uSGB 1156 and uSGB 3418*), increased the risk of arrhythmia. Consistently, *Campylobacter A uSGB 2510* and *Campylobacter A uSGB 2316* increased risks of congestive heart failure and ischemic stroke, respectively. These confirmed that *Campylobacter A species* contributed to cardiovascular disease as previously reported.[55]

MR also shows sex-specific causal relationships between the oral microbiome and diseases (Table S13). *Fusobacterium periodonticum C*, causally negatively influenced by carnosine and cystathionine in males by our above MR analysis, was causally related to increased risk of peripheral artery disease in females ($p = 1.19 \times 10^{-3}$; Figure 6A). Three taxa, *Veillonella rogosae*, *Oribacterium sinus*, and *Leptotrichia A sp001274535*, which all exhibited higher abundance in females than males in observational studies, were causally related to decreased risks of chronic obstructive pulmonary disease, colorectal cancer, and arrhythmia, respectively, only in females (Figure 6A). Epilepsy had causal effects on the increased abundance of *Lancefieldella* spp. such as *Lancefieldella sp000564995* in both tongue and saliva, which exists only in females but not in males. In addition, epilepsy was causally linked to increased abundances of uSGBs from *Solobacterium*, *Catonella*, *Veillonella*, and *TM7x* but decreased abundances of uSGBs from Weeksellaceae, *Gemella*, and *Haemophilus D*. Cerebral aneurysm, which is more prevalent in women, had significant causal effects on the decreased abundance of *Streptococcus* species such as *S. oralis*, *S. australis*, and *S. mitis* in females' tongue dorsum and was causally related to increased abundance of *Streptococcus* and its species such as *S. oralis*, *S. cristatus*, and *S. sp001553685* in males' saliva (Figure 6B). These results help to understand whether specific oral microbiota plays a causal role or a consequence in the sexual dimorphism of cardiometabolic diseases and certain cancers.

DISCUSSION

In this study, we generated a complete resource database including abundant sex-combined and sex-specific observational correlations and further causal relationships between the oral microbiome and host factors (mainly blood metabolites). We underline the necessity of sex stratification in investigating the contributions of host genetic variants and other major factors to the oral microbiome composition, in agreement with the increasing sex-stratified genetic association studies.[29,56–59] We summarize several key findings here. We first observed that the oral microbiome was sex dimorphic, and the major factors determining the oral microbiome composition differed between sexes. We identified sex-specific genetic loci associated with oral microbiota and further found abundant sex-differential causal relationships between the oral microbiome and serum metabolites, by applying one-sample bidirectional MR in 2,984 individuals from the discovery cohort and replicated in an independent cohort of 1,494 individuals. Sex-dimorphic MR analyses found both increased carnosine and cystathionine in males were causally linked to decreased abundance of *Fusobacterium periodonticum C* which was then found to have causal effects on T2D in disease MR analyses. Androstenedione as an androgen showed significant MR enrichments in males for the tongue dorsum. Likewise, aldosterone whose level was higher in females exhibited significant MR enrichments in females for both niches. These results were consistent with the fact that the differential sex hormone levels between males and females began *in utero* and continued throughout the lifetime of the organism.[60] Overall, this valuable resource offers potential microbes-metabolites casual relationships that underlie sex differences for mechanistic and clinical studies. Future studies are expected to replicate the causalities in more samples and populations, as well as to investigate the underlying mechanism using germ-free mice and reference bacteria strains.

By applying this MR analysis to explore causalities, our results provided further support for several previously reported microbiome-metabolites relationships. For example, four taxa, *Fusobacterium nucleatum*, *Prevotella intermedia*, *Treponema denticola*, and *Porphyromonas endodontalis* have been widely reported to be associated with periodontitis diseases.[61–63] Our data-driven MR confirmed these four taxa in both niches causally contributed to periodontitis problems such as bleeding frequencies of gums and dental calculus (Figure S6). Moreover, the four taxa also showed causal effects on multiple diseases in our disease MR analyses ($p < 0.05$): *Prevotella intermedia in saliva* genetically increased risks of a dozen of diseases, including CRC, T2D, ischemic stroke, chronic obstructive pulmonary disease, and hepatocellular carcinoma; *Porphyromonas endodontalis* in saliva genetically increased risks of nephrotic syndrome and arrhythmia; *Treponema B denticola* in saliva increased risks of lung cancer and endometriosis; and *Fusobacterium nucleatum* in tongue genetically increased risks of lung cancer and esophageal cancer. These pieces of evidence indicated that the periodontal pathogens could cause chronic inflammation of the gum and even contribute to diseases by inducing local inflammation and aggravating insulin resistance,[64,65] in line with previous findings that chronic periodontitis is linked with systemic disorders like diabetes,[66] cardiovascular disease,[67] and certain cancers.[68] Moreover, targeted therapies against specific periodontal pathogens (such as *P. intermedia* and *P. endodontalis*) may help ameliorate specific periodontitis-related comorbidities, such as T2D and arrhythmia. Flossing was also a simple way to reduce abundances of periodontal pathogens. We observed that flossing frequency was indeed negatively connected with the periodontal pathogen *Prevotella intermedia* ($\beta = -0.199$, $p = 6.07 \times 10^{-6}$ for saliva) in the MR analysis and *Porphyromonas endodontalis* and *Treponema B denticola* in the observational analysis.

Androstenedione was genetically positively correlated with the abundance of genus *Pauljensenia* species such as *P. odontolyticus* in our metabolites' MR analysis. *Pauljensenia odontolyticus* was causally linked to increased risk of T2D in diseases' MR analysis. These indicated sex hormone alters microbial community and may regulate autoimmune disease fate through interaction with microbiota that are not only limited to the intestinal tract but also existed in the sex dimorphism.[7,20] The impacts of oral and intestinal microbial dysbiosis on autoimmune disease are increasingly recognized,[69] including the causal effects illustrated by previous MR studies.[70,71] Both *Streptococcus parasanguinis B* and *Veillonella (V. dispar, V. parvula A)* were positively correlated with triglyceride and hyperlipidemia in observation analysis ($p < 1 \times 10^{-8}$), which supported the connection between the oral microbiome and cardiovascular disease, and these associations were also observed between the gut microbiome and cardiovascular disease.[72] *Streptococcus parasanguinis B* was also positively correlated with the frequency of mouthwash ($p < 2.48 \times 10^{-5}$), high-fat and high-sugar diet ($p < 6.52 \times 10^{-13}$), and sleep past midnight ($p < 3.30 \times 10^{-7}$) (Table S5), which suggested that a health manager of lifestyle may lower abundance of this taxon to reduce cardiovascular disease risk instead of mouthwash.[72] MR further revealed that *Streptococcus parasanguinis B* had a positive causal effect on the

serum aldosterone in females' tongue dorsum samples (Figure 5C, $p < 6.12 \times 10^{-6}$), and high aldosterone was associated with the development and severity of hypertension, congestive heart failure, coronary artery disease, chronic kidney disease, and metabolic syndrome.[73] These causal understandings of host microbiome-metabolites interactions and their roles in disease may help to translate toward the future development of microbiome-targeted therapeutic interventions.

In summary, our work provides pieces of evidence on the potential interplay between the oral microbiome and blood metabolites and has potential implications for disease. Our results also called on the necessity to consider sex-aware analyses for future studies to shed light on possible sex-specific microbes-metabolites interplay mechanisms.

Limitations of the study

There are two limitations in our current study. First, we defined the 3,589 SGBs as "species-level" according to the two most frequently used criteria[74–76]: i) at least 95% average nucleotide identity (ANI) and ii) at least 30% overlap of the aligned metagenome-assembled genome (MAG). This may be a little inaccurate since two or more SGBs could be annotated to the same most representative species. Despite this, we do not rule out that they come from different species but not subspecies (commonly defined $\geq$ 95% ANI[77,78]), especially for those who show low abundance correlations and negative beta effects of associations in the population. There are 16.7% (598 out of the total 3,589 SGBs) SGBs that could be annotated to the same representative species, which will not significantly decrease the number of taxa for analysis. 485 (35%) out of the total 1,395 pairs of SGBs that were annotated to the same representative species showed a less moderate correlation (spearman $r < 0.6$), especially for several *Veillonella rogosae* SGBs (id: 2362; 3333; 1190; 3494). They all showed a negative correlation with each other (spearman $r < 0$); such conditions were also observed for the *Aggregatibacter sp000466335* SGBs and *Pauljensenia sp000278725* SGBs, suggesting the possible inaccurate taxonomic annotations to the same species. In combination with these listed facts, we thought the SGBs level exhibited better resolution and refrained from pooling the different SGBs into the same species id by directly adding up their abundances, but the accurate taxonomy of "species-level" merits further studies. We hope to achieve higher-quality assembled, binning, and assigned/taxonomic results with the increasingly developed metagenomic technology, tools, and pipelines in the future. Secondly, to reduce the complexity of correlations, associations, and causalities, we used a greedy algorithm to obtain a minimal set of core taxa (Table S14). We filtered highly correlated bacteria and only kept the more representative one among all highly correlated bacteria to reduce the number of GWAS tests, as done by many relevant studies.[70,79–81] However, the filtered-out bacteria may be biologically more meaningful. The identification of driver species controlling the ecological network in the microbial community may help to explain easier. We also hope to take the microbial functional network as the study objective in the future but not only limited to the microbial taxa.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study subjects
- METHOD DETAILS
  - Whole genome sequence and quality control
  - Oral metagenomic sequencing and profiling
  - Sex difference/comparison in the oral microbiome
  - Major factors associated with the oral microbiome
  - Sex-stratified GWAS analysis for oral microbial features
  - GWAS analysis for host traits
  - Functional analysis of sex-specific significant loci
  - Independent instrumental variables (IVs) selection
  - One-sample MR analysis

- Two-sample MR analysis for diseases in Japan biobank
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105839.

ACKNOWLEDGMENTS

We sincerely thank the support provided by the China National Gene Bank. We thank all the volunteers for their time and self-collecting the oral samples using our kit. We thank for the support from the National Natural Science Foundation of China (No. 32200548).

AUTHOR CONTRIBUTIONS

H.J. and T.Z. conceived and organized this study. J.W. initiated the overall health project. X.X., H.Y., X.J., Y.H., Y. Zong, and W.L. contributed to the organization of the cohort, the sample collection, and the questionnaire collection. H. Lu led the DNA extraction and sequencing. X.L. and X.T. processed the whole-genome data. Z.J., J.Z., L.T., Q.S., Y.J., and L.Z. processed the metagenome data. Z.J. performed the sex-differential analysis for the oral microbiome. X.L. and X.T. performed the sex-stratified M-GWAS and MR analyses. X.L. and H.J. wrote the manuscript. All authors contributed to the data and texts in this manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 5, 2022
Revised: August 9, 2022
Accepted: December 16, 2022
Published: January 20, 2023

REFERENCES

1. 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.

2. Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., et al. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat. Med. *21*, 895–905. https://doi.org/10.1038/nm.3914.

3. Ruff, W.E., Greiling, T.M., and Kriegel, M.A. (2020). Host-microbiota interactions in immune-mediated diseases. Nat. Rev. Microbiol. *18*, 521–538. https://doi.org/10.1038/s41579-020-0367-2.

4. Belstrøm, D. (2020). The salivary microbiota in health and disease. J. Oral Microbiol. *12*, 1723975. https://doi.org/10.1080/20002297.2020.1723975.

5. Wang, Y., Wang, S., Wu, C., Chen, X., Duan, Z., Xu, Q., Jiang, W., Xu, L., Wang, T., Su, L., et al. (2019). Oral microbiome alterations associated with early childhood caries highlight the importance of carbohydrate metabolic activities. mSystems *4*. 004500-19. https://doi.org/10.1128/mSystems.00450-19.

6. Hajishengallis, G. (2015). Periodontitis: from microbial immune subversion to systemic inflammation. Nat. Rev. Immunol. *15*, 30–44. https://doi.org/10.1038/nri3785.

7. Markle, J.G.M., Frank, D.N., Mortin-Toth, S., Robertson, C.E., Feazel, L.M., Rolle-Kampczyk, U., von Bergen, M., McCoy, K.D., Macpherson, A.J., and Danska, J.S. (2013). Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. Science *339*, 1084–1088. https://doi.org/10.1126/science.1233521.

8. Valeri, F., and Endres, K. (2021). How biological sex of the host shapes its gut microbiota. Front. Neuroendocrinol. *61*, 100912. https://doi.org/10.1016/j.yfrne.2021.100912.

9. Lynch, S.V., and Pedersen, O. (2016). The human intestinal microbiome in health and disease. N. Engl. J. Med. *375*, 2369–2379. https://doi.org/10.1056/NEJMra1600266.

10. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227. https://doi.org/10.1038/nature11053.

11. Willis, J.R., González-Torres, P., Pittis, A.A., Bejarano, L.A., Cozzuto, L., Andreu-Somavilla, N., Alloza-Trabado, M., Valentín, A., Ksiezopolska, E., Company, C., et al. (2018). Citizen science charts two major "stomatotypes" in the oral microbiome of adolescents and reveals links with habits and drinking water composition. Microbiome *6*, 218. https://doi.org/10.1186/s40168-018-0592-3.

12. David, L.A., Materna, A.C., Friedman, J., Campos-Baptista, M.I., Blackburn, M.C., Perrotta, A., Erdman, S.E., and Alm, E.J. (2014). Host lifestyle affects human microbiota on daily timescales. Genome Biol. *15*, R89. https://doi.org/10.1186/gb-2014-15-7-r89.

13. Ma, Z.S., and Li, W. (2019). How and why men and women differ in their microbiomes: medical ecology and network analyses of the microgenderome. Adv. Sci. *6*, 1902054. https://doi.org/10.1002/advs.201902054.

14. Johnson, J.S., Spakowicz, D.J., Hong, B.Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat. Commun. *10*, 5029. https://doi.org/10.1038/s41467-019-13036-1.

15. Liu, X., Tang, S., Zhong, H., Tong, X., Jie, Z., Ding, Q., Wang, D., Guo, R., Xiao, L., Xu, X., et al. (2021). A genome-wide association study for gut metagenome in Chinese adults illuminates complex diseases. Cell Discov. *7*,

9. https://doi.org/10.1038/s41421-020-00239-w.

16. Liu, X., Tong, X., Zou, Y., Lin, X., Zhao, H., Tian, L., Jie, Z., Wang, Q., Zhang, Z., Lu, H., et al. (2022). Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. Nat. Genet. *54*, 52–61. https://doi.org/10.1038/s41588-021-00968-y.

17. Zhu, J., Tian, L., Chen, P., Han, M., Song, L., Tong, X., Sun, X., Yang, F., Lin, Z., Liu, X., et al. (2022). Over 50, 000 metagenomically assembled draft genomes for the human oral microbiome reveal new taxa. Dev. Reprod. Biol. *20*, 246–259. https://doi.org/10.1016/j.gpb.2021.05.001.

18. Liu, X., Tong, X., Zhu, J., Tian, L., Jie, Z., Zou, Y., Lin, X., Liang, H., Li, W., Ju, Y., et al. (2021). Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome. Cell Discov. *7*, 117. https://doi.org/10.1038/s41421-021-00356-0.

19. Neuman, H., Debelius, J.W., Knight, R., and Koren, O. (2015). Microbial endocrinology: the interplay between the microbiota and the endocrine system. FEMS Microbiol. Rev. *39*, 509–521. https://doi.org/10.1093/femsre/fuu010.

20. Yoon, K., and Kim, N. (2021). Roles of sex hormones and gender in the gut microbiota. J. Neurogastroenterol. Motil. *27*, 314–325. https://doi.org/10.5056/jnm20208.

21. Rizzetto, L., Fava, F., Tuohy, K.M., and Selmi, C. (2018). Connecting the immune system, systemic chronic inflammation and the gut microbiome: the role of sex. J. Autoimmun. *92*, 12–34. https://doi.org/10.1016/j.jaut.2018.05.008.

22. Maffei, S., Forini, F., Canale, P., Nicolini, G., and Guiducci, L. (2022). Gut microbiota and sex hormones: crosstalking players in cardiometabolic and cardiovascular disease. Int. J. Mol. Sci. *23*, 7154. https://doi.org/10.3390/ijms23137154.

23. Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. Nat. Rev. Immunol. *16*, 626–638. https://doi.org/10.1038/nri.2016.90.

24. Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. Austral Ecol. *26*, 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.

25. Mora, D., and Arioli, S. (2014). Microbial urease in health and disease. PLoS Pathog. *10*, e1004472. https://doi.org/10.1371/journal.ppat.1004472.

26. Essilfie, A.T., Simpson, J.L., Horvat, J.C., Preston, J.A., Dunkley, M.L., Foster, P.S., Gibson, P.G., and Hansbro, P.M. (2011). Haemophilus influenzae infection drives IL-17-mediated neutrophilic allergic airways disease. PLoS Pathog. *7*, e1002244. https://doi.org/10.1371/journal.ppat.1002244.

27. Ferreira, M.A.R., Mathur, R., Vonk, J.M., Szwajda, A., Brumpton, B., Granell, R., Brew, B.K., Ullemar, V., Lu, Y., Jiang, Y., et al. (2019). Genetic architectures of childhood- and adult-onset asthma are partly distinct. Am. J. Hum. Genet. *104*, 665–684. https://doi.org/10.1016/j.ajhg.2019.02.022.

28. Khramtsova, E.A., Davis, L.K., and Stranger, B.E. (2019). The role of sex in the genomics of human complex traits. Nat. Rev. Genet. *20*, 173–190. https://doi.org/10.1038/s41576-018-0083-1.

29. Bernabeu, E., Canela-Xandri, O., Rawlik, K., Talenti, A., Prendergast, J., and Tenesa, A. (2021). Sex differences in genetic architecture in the UK Biobank. Nat. Genet. *53*, 1283–1289. https://doi.org/10.1038/s41588-021-00912-0.

30. Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J.P., Woyke, T., Kyrpides, N.C., Pukall, R., Klenk, H.P., Goodfellow, M., and Göker, M. (2018). Genome-based taxonomic classification of the phylum actinobacteria. Front. Microbiol. *9*, 2007. https://doi.org/10.3389/fmicb.2018.02007.

31. Li, H., Patel, V., DiMartino, S.E., Froehlich, J.W., and Lee, R.S. (2020). An in-depth comparison of the pediatric and adult urinary N-glycomes. Mol. Cell. Proteomics *19*, 1767–1776. https://doi.org/10.1074/mcp.RA120.002225.

32. Wismann, J., and Willoughby, D. (2006). Gender differences in carbohydrate metabolism and carbohydrate loading. J. Int. Soc. Sports Nutr. *3*, 28–34. https://doi.org/10.1186/1550-2783-3-1-28.

33. InanlooRahatloo, K., Liang, G., Vo, D., Ebert, A., Nguyen, I., and Nguyen, P.K. (2017). Sex-based differences in myocardial gene expression in recently deceased organ donors with no prior cardiovascular disease. PLoS One *12*, e0183874. https://doi.org/10.1371/journal.pone.0183874.

34. Sardar, R., Zandieh, Z., Namjoo, Z., Soleimani, M., Shirazi, R., and Hami, J. (2021). Laterality and sex differences in the expression of brain-derived neurotrophic factor in developing rat hippocampus. Metab. Brain Dis. *36*, 133–144. https://doi.org/10.1007/s11011-020-00620-4.

35. Chan, C.B., and Ye, K. (2017). Sex differences in brain-derived neurotrophic factor signaling and functions. J. Neurosci. Res. *95*, 328–335. https://doi.org/10.1002/jnr.23863.

36. Murphy, E., Amanakis, G., Fillmore, N., Parks, R.J., and Sun, J. (2017). Sex differences in metabolic cardiomyopathy. Cardiovasc. Res. *113*, 370–377. https://doi.org/10.1093/cvr/cvx008.

37. Yan, Z., Cheng, X., Li, Y., Su, Z., Zhou, Y., and Liu, J. (2021). Sexually dimorphic neurotransmitter release at the neuromuscular junction in adult Caenorhabditis elegans. Front. Mol. Neurosci. *14*, 780396. https://doi.org/10.3389/fnmol.2021.780396.

38. Farrokhian, A., Mahmoodian, M., Bahmani, F., Amirani, E., Shafabakhsh, R., and Asemi, Z. (2020). The influences of chromium supplementation on metabolic status in patients with type 2 diabetes mellitus and coronary heart disease. Biol. Trace Elem. Res. *194*, 313–320. https://doi.org/10.1007/s12011-019-01783-7.

39. Wyss, M., and Kaddurah-Daouk, R. (2000). Creatine and creatinine metabolism. Physiol. Rev. *80*, 1107–1213. https://doi.org/10.1152/physrev.2000.80.3.1107.

40. Elshorbagy, A.K., Church, C., Valdivia-Garcia, M., Smith, A.D., Refsum, H., and Cox, R. (2012). Dietary cystine level affects metabolic rate and glycaemic control in adult mice. J. Nutr. Biochem. *23*, 332–340. https://doi.org/10.1016/j.jnutbio.2010.12.009.

41. Park, S.Y., Kim, J., Son, J.I., Rhee, S.Y., Kim, D.-Y., Chon, S., Lim, H., and Woo, J.-T. (2021). Dietary glutamic acid and aspartic acid as biomarkers for predicting diabetic retinopathy. Sci. Rep. *11*, 7244. https://doi.org/10.1038/s41598-021-83165-5.

42. Wu, G., Xiao, X., Feng, P., Xie, F., Yu, Z., Yuan, W., Liu, P., and Li, X. (2017). Gut remediation: a potential approach to reducing chromium accumulation using Lactobacillus plantarum TW1-1. Sci. Rep. *7*, 15000. https://doi.org/10.1038/s41598-017-15216-9.

43. Matoba, Y., Noda, M., Yoshida, T., Oda, K., Ezumi, Y., Yasutake, C., Izuhara-Kihara, H., Danshiitsoodol, N., Kumagai, T., and Sugiyama, M. (2020). Catalytic specificity of the Lactobacillus plantarum cystathionine γ-lyase presumed by the crystallographic analysis. Sci. Rep. *10*, 14886. https://doi.org/10.1038/s41598-020-71756-7.

44. Peñafiel, R., Ruzafa, C., Monserrat, F., and Cremades, A. (2004). Gender-related differences in carnosine, anserine and lysine content of murine skeletal muscle. Amino Acids *26*, 53–58. https://doi.org/10.1007/s00726-003-0034-8.

45. Basic, A., Blomqvist, M., Dahlén, G., and Svensäter, G. (2017). The proteins of Fusobacterium spp. involved in hydrogen sulfide production from L-cysteine. BMC Microbiol. *17*, 61. https://doi.org/10.1186/s12866-017-0967-9.

46. Mothersole, R.G., and Wolthers, K.R. (2019). Structural and kinetic insight into the biosynthesis of H2S and l-lanthionine from l-cysteine by a pyridoxal l-phosphate-dependent enzyme from Fusobacterium nucleatum. Biochemistry *58*, 3592–3603. https://doi.org/10.1021/acs.biochem.9b00487.

47. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314. https://doi.org/10.1093/nar/gky1085.

48. Wang, T.J., Ngo, D., Psychogios, N., Dejam, A., Larson, M.G., Vasan, R.S., Ghorbani, A., O'Sullivan, J., Cheng, S., Rhee, E.P., et al. (2013). 2-Aminoadipic acid is a biomarker for diabetes risk. J. Clin. Invest. *123*, 4309–4317. https://doi.org/10.1172/jci64801.

49. Sell, D.R., Strauch, C.M., Shen, W., and Monnier, V.M. (2008). Aging, diabetes, and renal failure catalyze the oxidation of lysyl residues to 2-aminoadipic acid in human skin collagen: evidence for metal-catalyzed oxidation mediated by alpha-dicarbonyls. Ann. N. Y. Acad. Sci. *1126*, 205–209. https://doi.org/10.1196/annals.1433.065.

50. Chen, J., Wright, K., Davis, J.M., Jeraldo, P., Marietta, E.V., Murray, J., Nelson, H., Matteson, E.L., and Taneja, V. (2016). An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. Genome Med. *8*, 43. https://doi.org/10.1186/s13073-016-0299-7.

51. Rother, M., Böck, A., and Wyss, C. (2001). Selenium-dependent growth of Treponema denticola: evidence for a clostridial-type glycine reductase. Arch. Microbiol. *177*, 113–116. https://doi.org/10.1007/s002030100351.

52. Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S., Matoba, N., Low, S.K., Okada, Y., et al. (2020). Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat. Genet. *52*, 669–679. https://doi.org/10.1038/s41588-020-0640-3.

53. Caugant, D.A., and Brynildsrud, O.B. (2020). Neisseria meningitidis: using genomics to understand diversity, evolution and pathogenesis. Nat. Rev. Microbiol. *18*, 84–96. https://doi.org/10.1038/s41579-019-0282-6.

54. Willis, J.R., and Gabaldón, T. (2020). The human oral microbiome in health and disease: from sequences to ecosystems. Microorganisms *8*, 308. https://doi.org/10.3390/microorganisms8020308.

55. Kaakoush, N.O., Castaño-Rodríguez, N., Mitchell, H.M., and Man, S.M. (2015). Global epidemiology of Campylobacter infection. Clin. Microbiol. Rev. *28*, 687–720. https://doi.org/10.1128/CMR.00006-15.

56. Zeng, Y., Nie, C., Min, J., Chen, H., Liu, X., Ye, R., Chen, Z., Bai, C., Xie, E., Yin, Z., et al. (2018). Sex differences in genetic associations with longevity. JAMA Netw. Open *1*, e181670. https://doi.org/10.1001/jamanetworkopen.2018.1670.

57. Liu, X., Song, Z., Li, Y., Yao, Y., Fang, M., Bai, C., An, P., Chen, H., Chen, Z., Tang, B., et al. (2021). Integrated genetic analyses revealed novel human longevity loci and reduced risks of multiple diseases in a cohort study of 15, 651 Chinese individuals. Aging Cell *20*, e13323. https://doi.org/10.1111/acel.13323.

58. Rask-Andersen, M., Karlsson, T., Ek, W.E., and Johansson, Å. (2019). Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. Nat. Commun. *10*, 339. https://doi.org/10.1038/s41467-018-08000-4.

59. Graham, S.E., Nielsen, J.B., Zawistowski, M., Zhou, W., Fritsche, L.G., Gabrielsen, M.E., Skogholt, A.H., Surakka, I., Hornsby, W.E., Fermin, D., et al. (2019). Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis. Nat. Commun. *10*, 1847. https://doi.org/10.1038/s41467-019-09861-z.

60. Alonso, L.C., and Rosenfield, R.L. (2002). Oestrogens and puberty. Best Pract. Res. Clin. Endocrinol. Metabol. *16*, 13–30. https://doi.org/10.1053/beem.2002.0177.

61. Galimanas, V., Hall, M.W., Singh, N., Lynch, M.D.J., Goldberg, M., Tenenbaum, H., Cvitkovitch, D.G., Neufeld, J.D., and Senadheera, D.B. (2014). Bacterial community composition of chronic periodontitis and novel oral sampling sites for detecting disease indicators. Microbiome *2*, 32. https://doi.org/10.1186/2049-2618-2-32.

62. Patini, R., Staderini, E., Lajolo, C., Lopetuso, L., Mohammed, H., Rimondini, L., Rocchetti, V., Franceschi, F., Cordaro, M., and Gallenzi, P. (2018). Relationship between oral microbiota and periodontal disease: a systematic review. Eur. Rev. Med. Pharmacol. Sci. *22*, 5775–5788. https://doi.org/10.26355/eurrev_201809_15903.

63. Chen, C., Hemme, C., Beleno, J., Shi, Z.J., Ning, D., Qin, Y., Tu, Q., Jorgensen, M., He, Z., Wu, L., and Zhou, J. (2018). Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. ISME J. *12*, 1210–1224. https://doi.org/10.1038/s41396-017-0037-1.

64. Blasco-Baque, V., Garidou, L., Pomié, C., Escoula, Q., Loubieres, P., Le Gall-David, S., Lemaitre, M., Nicolas, S., Klopp, P., Waget, A., et al. (2017). Periodontitis induced by Porphyromonas gingivalis drives periodontal microbiota dysbiosis and insulin resistance via an impaired adaptive immune response. Gut *66*, 872–885. https://doi.org/10.1136/gutjnl-2015-309897.

65. Tian, J., Liu, C., Zheng, X., Jia, X., Peng, X., Yang, R., Zhou, X., and Xu, X. (2020). Porphyromonas gingivalis induces insulin resistance by increasing BCAA levels in mice. J. Dent. Res. *99*, 839–846. https://doi.org/10.1177/0022034520911037.

66. Lalla, E., and Papapanou, P.N. (2011). Diabetes mellitus and periodontitis: a tale of two common interrelated diseases. Nat. Rev. Endocrinol. *7*, 738–748. https://doi.org/10.1038/nrendo.2011.106.

67. Sanz, M., Del Castillo, A.M., Jepsen, S., Gonzalez-Juanatey, J.R., D'Aiuto, F., Bouchard, P., Chapple, I., Dietrich, T., Gotsman, I., Graziani, F., et al. (2020). Periodontitis and cardiovascular diseases. Consensus report. Glob. Heart *15*, 1. https://doi.org/10.5334/gh.400.

68. Hajishengallis, G., and Chavakis, T. (2021). Local and systemic mechanisms linking periodontal disease and inflammatory comorbidities. Nat. Rev. Immunol. *21*, 426–440. https://doi.org/10.1038/s41577-020-00488-6.

69. Zheng, D., Liwinski, T., and Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. Cell Res. *30*, 492–506. https://doi.org/10.1038/s41422-020-0332-7.

70. Sanna, S., van Zuydam, N.R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Võsa, U., Mujagic, Z., Masclee, A.A.M., Jonkers, D.M.A.E., et al. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat. Genet. *51*, 600–605. https://doi.org/10.1038/s41588-019-0350-x.

71. Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C.I., Raygoza Garay, J.A., Finnicum, C.T., Liu, X., et al. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat. Genet. *53*, 156–165. https://doi.org/10.1038/s41588-020-00763-1.

72. Jie, Z., Xia, H., Zhong, S.-L., Feng, Q., Li, S., Liang, S., Zhong, H., Liu, Z., Gao, Y., Zhao, H., et al. (2017). The gut microbiome in atherosclerotic cardiovascular disease. Nat. Commun. *8*, 845. https://doi.org/10.1038/s41467-017-00900-1.

73. Gaddam, K.K., Pimenta, E., Husain, S., and Calhoun, D.A. (2009). Aldosterone and cardiovascular disease. Curr. Probl. Cardiol. *34*, 51–84. https://doi.org/10.1016/j.cpcardiol.2008.10.002.

74. Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., and Finn, R.D. (2019). A new genomic blueprint of the human gut microbiota. Nature *568*, 499–504. https://doi.org/10.1038/s41586-019-0965-1.

75. Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. *9*, 5114. https://doi.org/10.1038/s41467-018-07641-9.

76. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., et al. (2019). Extensive unexplored human microbiome diversity revealed by over 150, 000 genomes from metagenomes spanning age, geography, and lifestyle. Cell *176*, 649–662.e20. https://doi.org/10.1016/j.cell.2019.01.001.

77. Brooks, B., Olm, M.R., Firek, B.A., Baker, R., Thomas, B.C., Morowitz, M.J., and Banfield, J.F. (2017). Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nat. Commun. *8*, 1814. https://doi.org/10.1038/s41467-017-02018-w.

78. He, C., Keren, R., Whittaker, M.L., Farag, I.F., Doudna, J.A., Cate, J.H.D., and Banfield, J.F. (2021). Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. Nat. Microbiol. *6*, 354–365. https://doi.org/10.1038/s41564-020-00840-5.

79. Lopera-Maya, E.A., Kurilshikov, A., van der Graaf, A., Hu, S., Andreu-Sánchez, S., Chen, L., Vila, A.V., Gacesa, R., Sinha, T., Collij, V., et al. (2022). Effect of host genetics on the gut microbiome in 7, 738 participants of the Dutch Microbiome Project. Nat. Genet. *54*, 143–151. https://doi.org/10.1038/s41588-021-00992-y.

80. Qin, Y., Havulinna, A.S., Liu, Y., Jousilahti, P., Ritchie, S.C., Tokolyi, A., Sanders, J.G., Valsta, L., Brożyńska, M., Zhu, Q., et al. (2022). Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. Nat. Genet. *54*, 134–142. https://doi.org/10.1038/s41588-021-00991-z.

81. Hughes, D.A., Bacigalupe, R., Wang, J., Rühlemann, M.C., Tito, R.Y., Falony, G., Joossens, M., Vieira-Silva, S., Henckaerts, L., Rymenans, L., et al. (2020). Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. Nat. Microbiol. *5*, 1079–1087. https://doi.org/10.1038/s41564-020-0743-8.

82. Fang, C., Zhong, H., Lin, Y., Chen, B., Han, M., Ren, H., Lu, H., Luber, J.M., Xia, M., Li, W., et al. (2018). Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. Gigascience *7*, 1–8. https://doi.org/10.1093/gigascience/gix133.

83. Vujkovic-Cvijin, I., Sklar, J., Jiang, L., Natarajan, L., Knight, R., and Belkaid, Y. (2020). Host variables confound gut microbiota studies of human disease. Nature *587*, 448–454. https://doi.org/10.1038/s41586-020-2881-9.

84. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.

85. Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat. Protoc. *10*, 1556–1566. https://doi.org/10.1038/nprot.2015.105.

86. Jiao, X., Sherman, B.T., Huang, D.W., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2012). DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics *28*, 1805–1806. https://doi.org/10.1093/bioinformatics/bts251.

87. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826. https://doi.org/10.1038/s41467-017-01261-5.

88. Permutt, T., and Hebel, J.R. (1989). Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. Biometrics *45*, 619–622.

89. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., Wray, N.R., and Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat. Commun. *9*, 224. https://doi.org/10.1038/s41467-017-02317-2.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
summary statistics	Liu X, Tong X, Zhu J, et al. Cell discovery, 2021;7:117. https://doi.org/10.1038/s41421-021-00356-0	https://db.cngb.org/search/project/CNP0001664
Clean reads of metagenomics	Zhu, J. et al. Genomics, Proteomics & Bioinformatics, https://doi.org/10.1016/j.gpb.2021.05.001 (2021)	https://db.cngb.org/search/project/CNP0000687/
Biobank Japan	Ishigaki, K. et al. Nat Genet 52, 669–679, https://doi.org/10.1038/s41588-020-0640–3 (2020)	http://jenger.riken.jp/en/result
Software and algorithms		
metapi	This paper	https://github.com/ohmeta/metapi
Bowtie2(v2.3.564)	Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012, 9:357–359.	http://bowtie-bio.sourceforge.net/bowtie2
seqtk(v1.3)	Heng Li(lh3)	https://github.com/lh3/seqtk
PERMANOVA	Anderson M J. Wiley statsref: statistics reference online, 2014: 1–15.	https://doi.org/10.1002/9781118445112.stat07841
Fastp(v0.19.463)	Shifu Chen, et al; Bioinformatics, Volume 34, Issue 17, 1 September 2018	https://github.com/OpenGene/fastp
PLINK(V1.9)	Christopher Chang	https://www.cog-genomics.org/plink/
ANNOVAR	Wang K, et al; Nucleic Acids Research, 38:e164, 2010	http://www.openbioinformatics.org/annovar/
DAVID	Jiao, X. et al. Bioinformatics 28, 1805–1806, https://doi.org/10.1093/bioinformatics/bts251 (2012)	https://david.ncifcrf.gov/
FUMA	K. Watanabe,et al. Nat. Commun. 8:1826. (2017)	http://fuma.ctglab.nl/
AER(R package)	Christian Kleiber et al	https://cran.r-project.org/web/packages/AER/index.html
GCTA-GSMR	Zhu, Z. et al. (2018). Nature Communications, 9: 224.	https://yanglab.westlake.edu.cn/software/gsmr/

RESOURCE AVAILABILITY

Lead contact

Further information and request for resources and reagents should be directed to and would be fulfilled by the lead contact, tao.zhang@genomics.cn.

Materials availability

This study did not generate new unique reagents.

Data and code availability

All statistical results were available in supplementary tables. The summary statistics including the associations between host genetics and tongue dorsum microbiome, host genetics and saliva microbiome are publicly available from CNGBdb: https://db.cngb.org/search/project/CNP0001664. The release of these summary statistics data was approved by the Ministry of Science and Technology of China (Project ID: 2021BAT1539). According to the Human Genetic Resources Administration of China regulation and the institutional review board of BGI-Shenzhen related to protecting individual privacy, sequencing data are controlled-access and are available via the application on request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study subjects

All the adult Chinese individuals in this 4D-SZ cohort were recruited for a multi-omics study as previously reported.[15–18] This study at present included a total of 4,478 individuals, of which 3,504 had saliva samples

and 3,694 had tongue dorsum samples (3,165 individuals having both sample types) for whole metagenomic sequencing (Table S1). The cohort could be divided into two independent parts, namely discovery and replication. In the discovery cohort, 2,984 individuals were enlisted with blood samples in the city of Shenzhen for high-depth whole genome sequencing (a mean depth of 33X), of which 2,017 had tongue dorsum and 1,915 had salivary samples for whole metagenomic sequencing. In the replication cohort, blood samples were collected from 1,494 individuals with low-depth whole genome sequencing (a mean depth of 9X), out of which 1,333 had tongue dorsum samples and 1,299 had salivary samples for metagenomic sequencing (Table S1). The replication cohort was designed in the same manner but organized at smaller scales in multiple cities (Wuhan, Qingdao, etc.) in China. The protocols for blood and oral collection, as well as the whole genome and metagenomic sequencing, were similar to our previous literature.[15–18] For the blood sample, DNA was extracted using MagPure Buffy Coat DNA Midi KF Kit (no. D3537-02) according to the manufacturer's protocol. Tongue dorsum and salivary samples were collected with the MGIEasy kit. For the salivary sample, a 2x concentration of stabilizing reagent kit was used and 2 mL saliva was collected. DNA of oral samples was extracted using MagPure Stool DNA KF Kit B (no. MD5115-02B). The DNA concentrations from blood and oral samples were estimated by Qubit (Invitrogen). 500 ng of input DNA from blood and oral samples were used for library preparation and then processed for paired-end 100 bp sequencing using the BGISEQ-500 platform.[82]

The study was approved by the Institutional Review Boards (IRB) at BGI-Shenzhen, and all participants provided written informed consent at enrollment.

METHOD DETAILS

Whole genome sequence and quality control

In the discovery cohort, 2,984 individuals with blood samples were sequenced to a mean of 33x for the whole genome. For replication, 1,494 individuals were sequenced to a mean of 9x for the whole genome. Full study designs including inclusion and exclusion criteria are described elsewhere.[18] In brief, after variant and sample quality control, 2,984 individuals (out of which 2,017 had matched tongue dorsum and 1,915 had matched salivary samples) with 10 million common and low-frequency (MAF $\geq$0.5%) variants in the discovery cohort were left for M-GWAS and MR analyses. 1,494 individuals (out of which 1,333 had matched tongue dorsum and 1,299 had matched salivary samples) with 8.6 million common and low-frequency variants (MAF $\geq$0.5%) in the replication cohort were left for validation analysis of M-GWAS and MR discovery results.

Oral metagenomic sequencing and profiling

Metagenomic sequencing was done on the BGISEQ-500 platform, with 100 bp of paired-end reads for all samples and four libraries were constructed for each lane. We generated 19.18 $\pm$ 7.90 Gb (average $\pm$standard deviation) and 19.90 $\pm$ 7.73 Gb raw bases per sample for tongue dorsum samples in discovery and replication cohorts, respectively (Table S1). We also generated 13.64 $\pm$ 2.91 Gb and 13.66 $\pm$ 2.80 Gb raw bases per sample for salivary samples in discovery and replication cohorts, respectively. After using the quality control module of the metapi pipeline followed by reads filtering and trimming with strict filtration standards (not less than a mean quality Phred score of 20 and not shorter than 51 bp read length) using the fastp v0.19.463, host sequences contamination removing using the Bowtie2 v2.3.564 (hg38 index) and seqtk65 v1.3, we finally got an average of 3.1Gb (host rate:77%) and 9.9Gb (host rate:31%) raw bases per sample for salivary and tongue dorsum samples, respectively.

The high-quality oral genome catalogue was constructed in a previous study.[17] The oral metagenomic sequencing reads were mapped to the oral genome catalogue (http://ftp.cngb.org/pub/SciRAID/Microbiome/human_oral_genomes/bowtie2_index) using Bowtie2 with parameters: "–end-to-end –very-sensitive –seed 0 –time -k 2 –no-unal –no-discordant -X 1200", and the normalized contigs depths were obtained by using jgi_summarize_bam_contig_depths, then based on the correspondence of contigs and genome, the normalized contig depth was converted to the relative abundance of each species for each sample. Finally, we merged all representative species' relative abundance to generate a taxonomic profile for oral samples. The profiling workflow was implemented in the metapi jgi_profiling module (https://github.com/ohmeta/metapi/blob/v1.1.0/metapi/rules/profiling.smk#L585). The complete metapi pipeline could be found on this website: https://github.com/ohmeta/metapi/tree/v1.1.0/metapi/rules.

Sex difference/comparison in the oral microbiome

A generalized linear model (the 'glm' function in R) was used to compare the sex difference of each SGB or each host factor (mainly the metabolic parameter) after adjusting for the potential confounders (age, sex, BMI, drugs, mouthwash frequency, and whether suffering from respiratory infection disease or not). The results with Benjamini–Hochberg adjusted $p < 0.05$ was reported.

The modified RFCV classifier function from the randomForest package in R was used to evaluate the sex classification performance based on the oral microbiome. That was classified with varying numbers of SGBs (1, 3, 7, 14, 28, 56, 112, 223, 446, 892, 1784, 3569). The best ROC was achieved with 223 SGBs for saliva and 112 SGBs for tongue dorsum. ROC was plotted with the pROC package in R.

Major factors associated with the oral microbiome

The connections between microbiota composition and host metabolic traits were identified utilizing the PERMANOVA[24] analysis (adonis test, based on SGB-level Bray–Curtis dissimilarity, 4999 permutations) in the vegan R package. To remove potential confounding effects, age, sex, BMI, drugs, mouthwash frequency, and respiratory infection disease were included in the PERMANOVA regression model. To test the sex bias metric measured by beta-diversity-based F statistics of the metabolic parameter, a bootstrapping method was used to assess the dispersion of differences in the microbiota community as described previously.[83] More specifically, we bootstrapped F statistic values by resampling the sample 25 times for each sex. For all analyses, a Mann–Whitney U-test was performed to compare the bootstrapping F statistic from male cohorts to that from female cohorts, after which Benjamini–Hochberg false discovery rate corrections were applied.

Sex-stratified GWAS analysis for oral microbial features

As previously described,[18] we filtered the bacteria to keep those with an occurrence rate over 90% and average relative abundance over 1×10^{-5} (The represented genus of these microbial taxa covered between 99.63% (tongue dorsum) and 99.76% (saliva) of the whole community) in the cohort. As many oral microbial taxa are highly correlated, we performed many Spearman correlation tests and kept only one member of pairs of bacteria showing a >0.8 correlation coefficient. Spearman's correlations were calculated pairwise between all taxa, and the correlations were used to generate an adjacency matrix where correlations of >0.8 represented an edge between taxa. A graphical representation of this matrix was then used for the greedy selection of representative taxa. Nodes (microbiota taxa) were sorted by degree and the one with the highest degree was then chosen as a final taxon (selecting at random in the case of a tie). The taxon and its connected nodes were then removed from the network and the process was repeated until a final set of taxa sets was found such that each of the discarded taxa was correlated with at least one taxon. This filtering resulted in a final set of 1,583 and 1,685 independent microbial taxa for tongue dorsum and saliva, respectively, that were used for association analyses. These remaining independent taxa and their highly correlated taxa were listed in Table S14.

We performed sex-specific GWAS analysis in males and females separately. To test the associations between host genetics and oral bacteria, we used a linear model via PLINK v1.9[84] based on the relative abundance of oral bacteria. Specifically, the relative abundance was transformed by the natural logarithm that performed better than CLR transformation (Figure S9). the outlier individual who was located away from the mean by more than four standard deviations was removed so that the abundance of bacteria could be treated as a quantitative trait. Given the effects of environmental factors such as diet and lifestyles on microbial features, we included all potential cofounders that were significantly associated with the β-diversity (Benjamini–Hochberg FDR < 0.05) estimates in the explained variance analysis, as well as the top four principal components (PCs) as covariates for the M-GWAS analysis.

Male-specific variants were identified as (i) significantly associated with taxa in males ($p_{male} < 5 \times 10^{-8}$) and not significant in females ($p_{female} > 0.05$), and (ii) had nominal significant sex difference (testing p-value for difference in sex-specific effect size estimated by beta value, $p_{difference} < 0.05$). Female-specific variants were identified as (i) significantly associated with taxa in females ($p_{female} < 5 \times 10^{-8}$) and not significant in males ($p_{male} > 0.05$), and (ii) had nominal significant sex difference ($p_{difference} < 0.05$, as explained below). For each variant and the phenotype (relative abundance of taxa), we computed p values ($p_{difference}$) testing for the difference between the male-specific and female-specific beta-estimates β_m and β_f using the t-statistic

$$t = \frac{\beta_m - \beta_f}{\sqrt{SE_m^2 + SE_f^2 - 2 * corr(\beta_m, \beta_f) * SE_m * SE_f}}$$

with SE_m and SE_f being the standard errors of β_m or β_f. The correlation between the sex-specific beta-estimates was computed as the Spearman rank correlation coefficient across all variants for each phenotype.

GWAS analysis for host traits

For each of the host phenotypic traits (anthropometric, diets, metabolic traits, etc.), if it's continuous data, the log10-transformed of the median-normalized values was used as a quantitative trait in a linear model implemented in PLINK. If the host trait (bleeding frequency of gums, whether having dental calculus or not, etc.) is discrete data, it's treated as a binary trait in a logistic model for association analysis. For a quantitative trait, samples with missing values and values beyond 4 s.d. from the mean were excluded from the association analysis. Each of the 10 million common and low-frequency variants identified in the discovery dataset and the 8.6 million common and low-frequency variants identified in the replication dataset was tested for association analysis independently. Age, sex, and the top four PCs were included as covariates. Likewise, the sex-stratified GWAS analyses were performed on males and females separately.

Functional analysis of sex-specific significant loci

The significant genetic variants identified in the association analysis were mapped to genes using ANNOVAR.[85] Given that some significant genetic variants were low-frequency in the M-GWAS results, it's most suitable to input gene lists for enrichment analysis. We mapped variants to genes based on physical distance within a 20 kb window and got the gene lists for the enrichment analysis. DAVID (https://david.ncifcrf.gov/) was utilized to perform functional and pathway enrichment analysis. DAVID is a systematic and integrative functional annotation tool for the analysis of the relevant biological annotation of gene lists and provides the functional interpretation of the GO enrichment and KEGG pathway analysis.[86] The p-value < 0.05 was considered statistically significant. In addition, the mapped genes were further investigated using the GENE2FUNC procedure in FUMA (http://fuma.ctglab.nl/), which provides hypergeometric tests for the list of enriched mapping genes in 53 GTEx tissue-specific gene expression sets, 7,246 MSigDB gene sets, and 2,195 GWAS catalog gene sets.[87] Using the GENE2FUNC procedure, we examined whether the mapped genes were enriched in specific diseases or traits in the GWAS catalog as well as whether showed tissue-specific expression. Significant results were selected if a false discovery rate (FDR)-corrected $p < 0.05$ was observed.

Independent instrumental variables (IVs) selection

For each whole-genome-wide association result of microbial features and host phenotypic traits (mainly metabolic traits), we selected genetic variants that showed association at $p < 1 \times 10^{-5}$ and then performed the linkage disequilibrium (LD) estimation with a threshold of LD $r^2 < 0.1$ for clumping analysis to get independent genetic variants as IVs. The p-value threshold of 1×10^{-5} was used for the selection of genetic IVs associated with microbial features by maximizing the strength of genetic instruments and the amount of the average genetic variance explained by the genetic predictors in an independent sample. For each microbial feature, we got genetic instruments in the discovery dataset using different *p* thresholds, including 5×10^{-8}, 1×10^{-7}, $1 \times 10^{-6,}$ and 1×10^{-5}. We tested the strength of these instruments under different *P* thresholds by checking whether they predicted corresponding microbial features in an independent sample (Figure S10), we observed that the mean value of instrumental F statistics is 3.57 and on average only 0.28% phenotype variance could be explained by instruments on microbial features when using 5×10^{-8} as an instrumental cut-off. Therefore, we used a more liberal threshold of $p < 1 \times 10^{-5}$ to select the instruments for microbial features, and the instrumental mean F statistics reached 22 for tongue dorsum microbiota and 18 for salivary microbiota (greater than 10) which indicated a strong instrument. For consistency, we used the same threshold and procedure for selecting genetic IVs of metabolic traits in both the discovery and the replication cohort. The LD estimation between variants was calculated in corresponding samples for the discovery cohort and the replication cohort, respectively.

One-sample MR analysis

To investigate the causal effects between microbial features and host phenotypic traits available from this multi-omics cohort, we first performed a one-sample bidirectional MR analysis in the discovery dataset. As above described, we specified a threshold of $p < 1 \times 10^{-5}$ to select SNP instruments and a threshold of LD

r^2 < 0.1 for clumping analysis to get independent genetic variants as IVs for MR analysis. Then, an unweighted polygenic risk score (PRS) was calculated for each individual using independent genetic variants from GWAS data. Each SNP was recoded as 0, 1, and 2, depending on the number of trait-specific risk-increasing alleles carried by an individual. We performed Instrumental variable (IV) analyses employing the two-stage least square regression (TSLS) method.[88] In the first stage, for each exposure trait, the association between the GRS and observational phenotype value was assessed using the linear regression model and predicted fitted values based on the instrument were obtained. In the second stage, linear regression was performed with outcome traits and genetically predicted exposure levels from the first stage. In both stages, analyses were adjusted for age, sex, and the top four principal components of population structure. For each trait, TSLS was performed using 'ivreg' command from the AER package in R. We next attempted to replicate the causal effects between traits in the replication dataset. When stratified by sex, we performed the same MR analyses in males and females separately as done above, except with no adjusting for sex.

Two-sample MR analysis for diseases in Japan biobank

To investigate the causal effect between microbial features and diseases, we performed a two-sample bidirectional MR analysis using summary statistics of oral microbiota from our cohort and disease information from Biobank Japan. We downloaded both sex-combined and sex-stratified summary statistics data for 42 diseases from Biobank Japan[52] (http://jenger.riken.jp/en/result). Both sex-combined and sex-stratified MR were performed for disease phenotypes. The two-sample bidirectional MR analysis was performed by applying the GCTA-GSMR (Generalised Summary-data-based Mendelian Randomization) method[89] using the HEIDI-outlier analysis to remove horizontal pleiotropic SNPs. For consistency, genetic variants with $p < 1 \times 10^{-5}$ and LD $r^2 < 0.1$ were selected as instrumental variables for oral microbiota in our cohort. For disease exposures, SNP instruments were selected at a genome-wide significant threshold ($p < 5 \times 10^{-8}$) in the Biobank Japan study.

QUANTIFICATION AND STATISTICAL ANALYSIS

A generalized linear model (the 'glm' function in R) was used to compare the sex difference of each SGB or each host factor. The modified RFCV classifier function from the randomForest package in R was used to evaluate the sex classification performance based on the oral microbiome. The connections between microbiota composition and host metabolic traits were identified utilizing the PERMANOVA[24] analysis (adonis test, based on SGB-level Bray–Curtis dissimilarity, 4999 permutations) in the vegan R package. In the statistics analysis, the results with Benjamini–Hochberg adjusted $p < 0.05$ was reported. Spearman correlation test was used to filter highly correlated taxa. A linear model was used to perform M-GWAS analysis based on the relative abundance of oral bacteria. TSLS model ('ivreg' from the AER package in R) was used to perform one-sample MR analysis. The two-sample bidirectional MR analysis was performed by applying the GCTA-GSMR method.