JOURNAL OF ORAL MICROBIOLOGY
2023, VOL. 15, 2266655
https://doi.org/10.1080/20002297.2023.2266655

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Oral microbiome and ischemic stroke risk among elderly Chinese women

Cong Wang[a*], Yaohua Yang[a,b*], Qiuyin Cai[a], Yutang Gao[c], Hui Cai[a], Jie Wu[a], Wei Zheng[a], Jirong Long[a] and Xiao-Ou Shu[a]

[a]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA; [b]Center for Public Health Genomics, Department of Public Health Sciences, UVA Comprehensive Cancer Center, School of Medicine, University of Virginia, Charlottesville, VA, USA; [c]Shanghai Cancer Institute, Shanghai Jiao Tong University Renji Hospital, Shanghai, China

ABSTRACT

Background: Stroke, a leading cause of disability worldwide, has been associated with periodontitis. However, whether stroke risk is related to oral microbiota remains unknown. This study aims to evaluate the associations between the oral microbiome and ischemic stroke risk.

Methods: In a case-control study of 134 case-control pairs nested within a prospective cohort study, we examined pre-diagnostic oral microbiome in association with stroke risk via shotgun metagenomic sequencing. The microbial sub-community and functional profiling were performed using Latent Dirichlet Allocation and HUMAnN2. Associations of microbial diversity, sub-community structure, and individual microbial features with ischemic stroke risk were evaluated via conditional logistic regression.

Results: Alpha and beta diversities differ significantly between cases and controls. One genus- and two species-level sub-communities were significantly associated with decreased ischemic stroke risk, with odds ratios (95% confidence intervals) of 0.52 (0.31–0.90), 0.51 (0.31–0.84), and 0.60 (0.36–0.99), respectively. These associations were potentially driven by the representative taxa in these sub-communities, *i.e.*, genus *Corynebacterium* and *Lautropia*, and species *Lautropia mirabilis* and *Neisseria elongate* ($p < 0.05$). Additionally, 55 taxa, 1,237 gene families, and 90 metabolic pathways were associated with ischemic stroke risk at $p < 0.05$.

Conclusion: Our study highlights the role of oral microbiota in the etiology of ischemic stroke and calls for further research.

ARTICLE HISTORY
Received 21 April 2023
Revised 14 September 2023
Accepted 29 September 2023

KEYWORDS
Ischemic stroke; oral microbiome; nested case-control study; microbial sub-community; Latent Dirichlet Allocation

Introduction

Stroke is a leading cause of disability and death worldwide, including in China [1]. In 2019, there were approximately 3.9 million incident stroke cases, with over 70% of cases being ischemic stroke, and 2.2 million stroke-related deaths in China [1]. Multiple risk factors have been identified for ischemic stroke, including hypertension, hyperglycemia, hyperlipidemia, obesity, type 2 diabetes (T2D), and unhealthy lifestyles (*e.g.* smoking, physical inactivity, and the Western diet) [2]. In addition, several inflammatory biomarkers, such as C-reactive protein (CRP) and interleukin 6, have also been associated with ischemic stroke risk [3]. However, 1 in 4 ischemic stroke patients have none of these known risk factors [4].

A recent meta-analysis revealed a possible link between periodontitis and stroke risk [5]. Periodontitis, the result of inflammatory responses to bacterial infections caused by oral microbiota dysbiosis [6], has been found to be associated with an elevated serum CRP level. Evidence has suggested associations between oral microbiota and some known risk factors for stroke, such as T2D and hypertension [7–9]. However, few studies have directly investigated the relationship between oral microbiota and stroke [10,11].

A study among stroke patients showed that *Streptococcal* species, mainly *Streptococcus salivarius*, were the most common bacterial phylotype in the oral cavity of patients shortly after an incidence of stroke [10]. A case-control study found that saliva samples from stroke patients had a higher frequency of *Streptococcus mutans* strains than samples from healthy subjects [11]. These studies suggest a possible role of oral microbiota in stroke etiology. However, they were limited by the cross-sectional study design and the small number of oral bacteria investigated. Herein, we report a prospective and

CONTACT Xiao-Ou Shu xiao-ou.shu@vanderbilt.edu Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, 2525 West End Ave, Suite 600, Nashville, TN 37203
*These authors contributed equally to this work.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/20002297.2023.2266655.

comprehensive evaluation of the association between oral microbiota and ischemic stroke risk using resources from a population-based cohort study – the Shanghai Women's Health Study (SWHS).

Methods

Study population

The SWHS is a prospective cohort study of 74,940 Chinese women recruited from seven urban communities in Shanghai from 1996 to 2000 [12]. Briefly, an in-person interview was conducted by trained study interviewers for each participant at study enrollment to collect demographics, diet, disease history, and other health information. At baseline, 8,934 participants donated a mouth rinse sample according to the following modified mouthwash protocol performed by a trained professional: 1) rinsing the mouth twice with water; 2) brushing both cheeks with a soft toothbrush more than 10 times and expectorating into a cup; 3) rinsing the mouth and expectorating into the same cup, then repeating once; and 4) collecting the mouth rinse from the cup to a 50 mL collection tube containing 10 mL 99% isopropyl alcohol. Samples were kept at 4°C in transportation and processed within six hours of collection. After centrifuge at 3,000×g for 10 minutes, the supernatant was discarded. 2 mL 0.9% saline solution was added to the tube and resuspended. The sample was then transferred into a storage tube and stored at −70°C until DNA isolation.

The case-control sampling process is shown in Figure 1. Incident ischemic stroke was defined as the first nonfatal stroke. Details of stroke ascertainment has been reported [13]. Briefly, during study follow-ups (in 2000–2002 and 2008–2010) 495 study participants self-reported having an incident stroke. For these cases, information on the date, hospital, and stroke subtypes (*i.e.*, ischemic stroke or intracerebral hemorrhage) was collected, and medical charts and brain imaging, including computed tomography or magnetic resonance imaging report, were requested from diagnostic hospitals. Stroke diagnosis was adjudicated by a cardiologist [13]. Only the adjudicated cases were included in the current study. Individually matched controls (1:1 ratio) were selected from participants who donated mourthwash samples and were free of stroke, acute myocardial infarction, and cancer at the time of diagnosis of the index case using the incidence-density sampling method. Cases and controls were matched according to age (± two years), cigarette smoking status (never/ever), alcohol consumption status (never/ever), fasting time (± two hours), and date (± 30 days) and time (morning/afternoon) of mouthwash sample collection. A total of 134 ischemic stroke cases and 134 individually matched controls were included in the study.

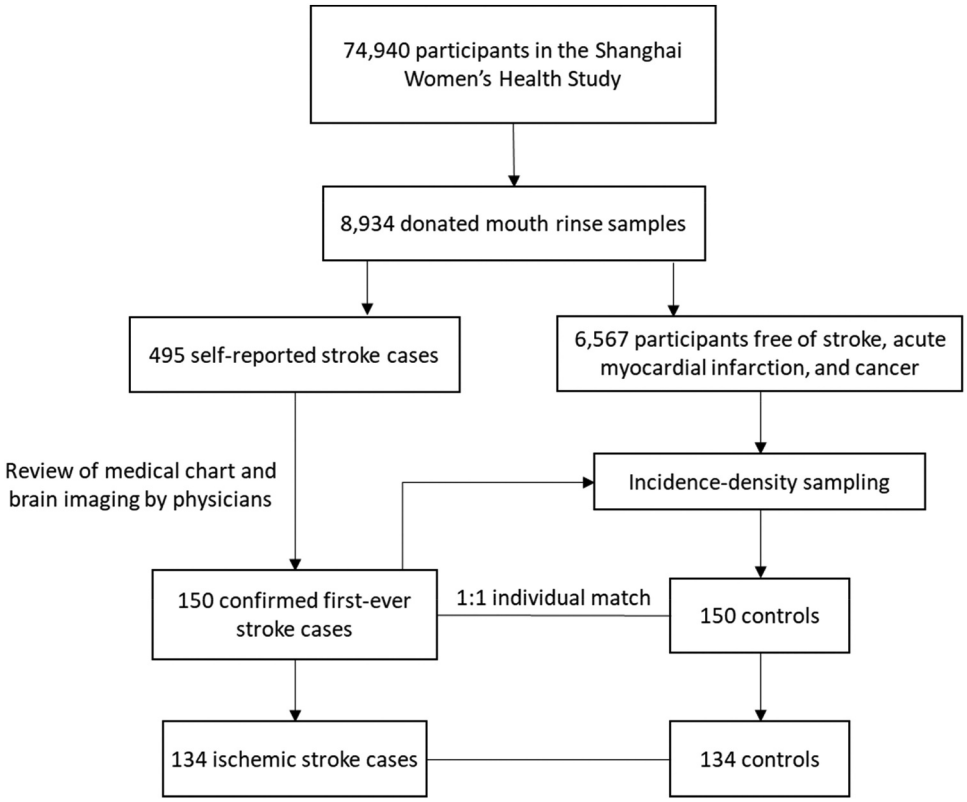

Figure 1. Flowchart of study sampling.

Microbiome profiling

DNA extraction, shotgun metagenomic sequencing, and data processing

DNA from buccal samples was extracted using the DNeasy PowerSoil Kit (Qiagen) following the manufacturer's instructions. Libraries for shotgun metagenomic sequencing were built using TruePrep DNA Library Prep Kit V2 or Nextera XT DNA Library Preparation Kit (Illumina). Sequencing was conducted at paired-end 150bp using the Illumina HiSeq System at BGI Americas. Raw sequencing reads were processed by Trimmomatic [14] to trim low-quality bases and discard reads that were shorter than 105 nucleotides (70% of original read length). Potential human contaminants were identified and discarded through mapping trimmed reads onto the human genome (GRCh38) using Bowtie2 [15] (v2.3.0). We then utilized a combination of Kraken [16] (v2.1.1) and Bracken [17] (v2.6) to perform taxonomic classification from clean reads against a customized reference database including high-quality genomes of oral bacteria from the expanded Human Oral Microbiome Database (eHOMD, v3.1) [18]. Within each sample, only taxa with a relative abundance of > 0.001% [19] were considered detected. Finally, functional profiling of the oral microbiome was performed using HUMAnN [20] v3.0.0 with the UniRef90 comprehensive protein database [21] as reference.

Statistical analysis

To estimate the richness/evenness and overall composition of the oral microbiota, the species level absolute abundance data were rarefied to 169,000 reads per sample. Alpha and beta diversity were measured by Faith's Phylogenetic Diversity index and unweighted UniFrac distance matrix, respectively, as calculated by the R *vegan* package (v2.5–7) [22]. The associations of alpha and beta diversity with ischemic stroke risk were examined via conditional logistic regression and the Microbiome Regression-based Kernel Association Test (MiRKAT, v1.1.2) [23], respectively.

To capture the complex inter-species interactions on human health and diseases, we utilized Latent Dirichlet Allocation (LDA), a three-level Bayesian probabilistic generative model that uncovers latent structures in unlabeled data [24], to infer sub-communities in oral microbiota. This unsupervised machine learning method determines the underlying subgroups of microbiota by taxa distribution in the observations and has been utilized in several microbiome studies [25,26]. After removing taxa that were detected in < 20% of all samples [26], unrarefied absolute abundance data from 80 genera and 296 species were included in the LDA analyses using the R *MetaTopics* package (v1.0) [27]. In our study, the oral microbiota was assigned to 10 genus-level and 30 species-level subgroups. The number of subgroups was determined following a recently described strategy [28]. Briefly, we fitted a series of LDA models with a number of subgroups ranging from 5 to 50, plotted model performance against the number of subgroups, and determine the optimal subgroup number according to the first jump in model performance [28]. Given a specific number of subgroups, the LDA algorithm estimates the probability that a taxon belongs to each subgroup, and the probability that each subgroup belongs to a participant [28]. The probability values of all subgroups for a participant and those of all taxa for a subgroup both add up to 1. We then evaluated the associations between probabilities of subgroups and the risk of ischemic stroke risk using conditional logistic regression.

Finally, we investigated the abundance of individual microbial taxa, gene families, and metabolic pathways in association with ischemic stroke risk. Taxa observed in < 10% of all participants, gene families, and pathways with a median relative abundance < 0.0001% among all participants were excluded from analyses. Unrarefied absolute abundance values of taxa were normalized using centered log-ratio (clr) transformation at each level, with zeros replaced by the minimum read count value of the whole dataset. Relative abundance values of gene families and metabolic pathways were normalized using arcsine square-root (asr) transformation. Associations of transformed abundances of taxa, gene families, and metabolic pathways with ischemic stroke risk were tested by conditional logistic regression.

All these analyses were adjusted for age, highest level of education attainment, baseline body mass index (BMI, kg/m^2), and baseline physical activity in the metabolic equivalent of task (model 1), and they were additionally adjusted for dietary intake, including fiber, fruit, vegetable, fat, cholesterol, and carbohydrate (model 2), and hypertension and T2D (model 3). All statistical tests were two-sided with $p < 0.05$ considered statistically significant.

Results

The mean time interval between the sample collection and stroke incidence was 4.4 years among cases, with a range of 0.5 to 8.7 years. Demographic characteristics of the 134 ischemic stroke case-control pairs are shown in Table 1. Compared to controls, ischemic stroke cases had a slightly younger age at enrollment (62.5 vs 63.3 years old, $p < 0.001$), a higher proportion of hypertension (54% vs 36%, $p = 0.005$), and a higher proportion of T2D (12% vs 3%, $p = 0.01$) at

Table 1. Characteristics of study participants from the Shanghai women's health study (SWHS).

Characteristics	Case ($N = 134$)	Control ($N = 134$)	p[a]
Age at enrollment	62.5 ± 6.50	63.3 ± 6.51	<0.001
Body mass index (BMI)	24.9 ± 3.72	24.5 ± 3.71	0.51
Highest level of education			0.72
<High school	93 (69%)	92 (69%)	
High school-college	35 (26%)	33 (25%)	
≥College	6 (4%)	9 (7%)	
Annual per capita income (¥)			0.92
Low (<4,000)	33 (25%)	31 (23%)	
Middle (4,000–8,000)	93 (69%)	96 (72%)	
High (≥8,000)	8 (6%)	7 (5%)	
Physical activity			0.18
Yes	67 (50%)	55 (41%)	
No	67 (50%)	79 (59%)	
Metabolic equivalent of task (MET)[b]	2.3 (2.0)	2.0 (1.5)	0.17
Menopausal status			1.00
Postmenopause	124 (93%)	125 (93%)	
Premenopause	10 (7%)	9 (7%)	
Hypertension at enrollment			0.005
Yes	72 (54%)	48 (36%)	
No	62 (46%)	86 (64%)	
Type 2 Diabetes at enrollment			0.01
Yes	16 (12%)	4 (3%)	
No	118 (88%)	130 (97%)	
Dietery intake[c]			
Total energy (kcal/day)	1599.8 ± 365.3	1569.9 ± 391.4	0.53
Fiber (g/day)	10.3 ± 3.5	10.1 ± 3.7	0.70
Fruits (g/day)	205.2 ± 139.7	229.5 ± 168.7	0.20
Vegetables (g/day)	259.7 ± 138.6	269.2 ± 150.1	0.60
Fat (g/day)	27.2 ± 12.8	25.4 ± 11.3	0.25
Cholesterol (g/day)	261.8 ± 168.1	254.7 ± 152.3	0.70
Carbohydrate (g/day)	276.3 ± 62.3	273.9 ± 67.4	0.76

Mean ± standard deviation (SD) is presented for continuous variables; N (%) is presented for categorical variables.
[a]Continuous and categorical variables were compared by paired t-test and χ2 test, respectively.
[b]MET among subjects with physical activities.
[c]Dietary intakes were adjusted for total energy intake using the residual method.

enrollment. No significant differences between cases and controls were found for BMI, physical activity, or dietary intake.

Alpha diversity was inversely associated with ischemic stroke risk, with one standard deviation (SD) increase in Faith's Phylogenetic Diversity index associated with a nearly 30% decreased risk of stroke [odds ratio (OR) = 0.71, 95% confidence interval (CI) = 0.51–0.99, $p = 0.04$]. There was a significant difference in beta diversity between ischemic stroke cases and controls ($p = 0.02$), with the case/control status explaining ~ 2% of variations in the unweighted UniFrac distance.

A total of 489 taxa were identified, including 9 phyla, 18 classes, 30 orders, 40 families, 82 genera, and 310 species. At the genus level, the oral microbiome community was classified into 10 subgroups (Figure 2), with each subgroup having a genus with the probability being > 20%. Compared to controls, cases had a lower probability of subgroup 9 after multivariable adjustment in model 3 (OR = 0.52, 95% CI = 0.31–0.90), where the five genera with the highest probabilities were *Actinomyces* (26.3%), *Streptococcus* (20.6%), *Corynebacterium* (7.2%), *Lautropia* (6.8%), and *Selenomonas* (5.9%) (Table 2). Among them, a significant association with stroke risk was observed for *Corynebacterium* (OR = 0.71, 95% CI = 0.50–0.99, model 1), *Lautropia* (OR = 0.54, 95% CI = 0.31–0.93, model 1), and *Selenomonas* (OR = 0.69, 95% CI = 0.47–1.00, model 2) in the analyses of individual taxa (Table 2).

At the species level, there were 30 heterogeneous subgroups. Some subgroups had a dominant species with an occurrence probability of > 80% (*e.g.*, 81% for *Rothia mucilaginosa* in subgroup 8 and 93% for *Aggregatibacter actinomycetemcomitans* in subgroup 22), while some had a more evenly distributed composition where each species made up < 10% of the subgroup (*e.g.*, subgroups 2 and 27) (Figure 3). Subgroups 3 and 26 were significantly associated with decreased ischemic stroke risk. Compared to controls, cases had a 49% lower probability of subgroup 3 [representative taxa: *Lautropia mirabilis* (16.9%), *Rothia aeria* (13.7%), and *Neisseria elongata* (6.3%)] and a 40% lower probability of subgroup 26 [representative taxa: *Actinobaculum sp. oral taxon 183* (4.9%), *Selenomonas sputigena* (4.5%), and *Actinomyces dentalis* (4.1%)], with an OR (95% CI) of 0.51 (0.31–0.84) and 0.60 (0.36–0.99), respectively (Table 2). When species were analyzed individually, two representative taxa in subgroup 3, *L. mirabilis* (OR = 0.54, 95% CI = 0.31–0.93, model 1) and *N. elongata* (OR = 0.69, 95% CI = 0.49–0.97, model 1; OR = 0.67, 95% CI = 0.46–0.97, model 2) were

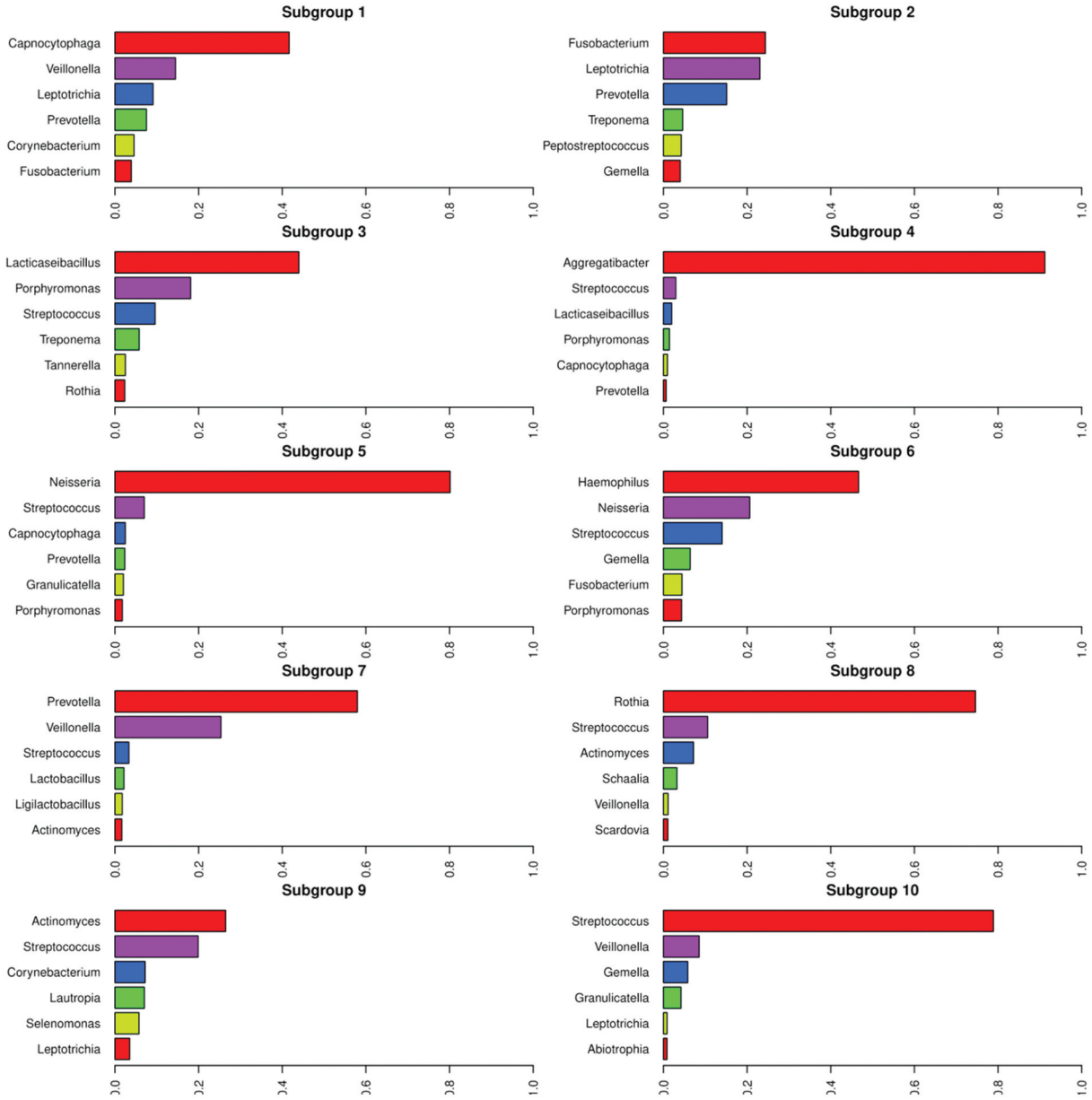

Figure 2. Genus-level sub-communities of oral microbiota.

[a]Taxa with highest probability of occurrence in each subgroup, using Latent Dirichlet allocation.

significantly associated with ischemic stroke risk. No significant association of individual taxa was observed in the top five taxa in subgroup 26. The associations between all genus- and species-level subgroups and ischemic stroke are shown in **Supplemental Table S1**.

Among all 489 taxa, we found 55 significant associations at $p < 0.05$ in at least one of the three models (**Supplemental Table S2**). Abundances of all 55 significant taxa were associated with a decreased stroke risk. Taxa with the smallest p-values from model 3 were *Neisseria bacilliformis* (OR = 0.49, 95% CI = 0.30–0.81), *Prevotella dentalis* (OR = 0.32, 95% CI = 0.14–0.72), *Actinomyces sp. oral taxon 169* (OR = 0.55, 95% CI = 0.36–0.85), *Kocuria* (OR = 0.32, 95% CI = 0.13–0.77), and *Aggregatibacter sp. oral taxon 458* (OR = 0.57, 95% CI = 0.36–0.89).

A total of 7,656 microbial gene families and 444 microbial metabolic pathways were investigated in our study. Among them, 1,237 and 90 showed a significant association with the risk of ischemic stroke at $p < 0.05$. The top five gene families with the smallest p-values from model 3 were reported for the OR < 1 and OR > 1 groups (Table 3). Four of the top five gene families associated with reduced ischemic stroke risk were related to enzyme activity, namely gene ontology (GO):0106008 (2-oxoglutaramate amidase activity), GO:0016040 [glutamate synthase (NADH) activity], GO:0033539 (fatty acid beta-oxidation using acyl-CoA dehydrogenase), and GO:0009042 (valine-pyruvate transaminase activity), with ORs (95% CIs) of 0.42 (0.25–0.71), 0.38 (0.21–0.69), 0.40 (0.22–0.71), and 0.34 (0.17–0.67), respectively. The top five GO terms associated with

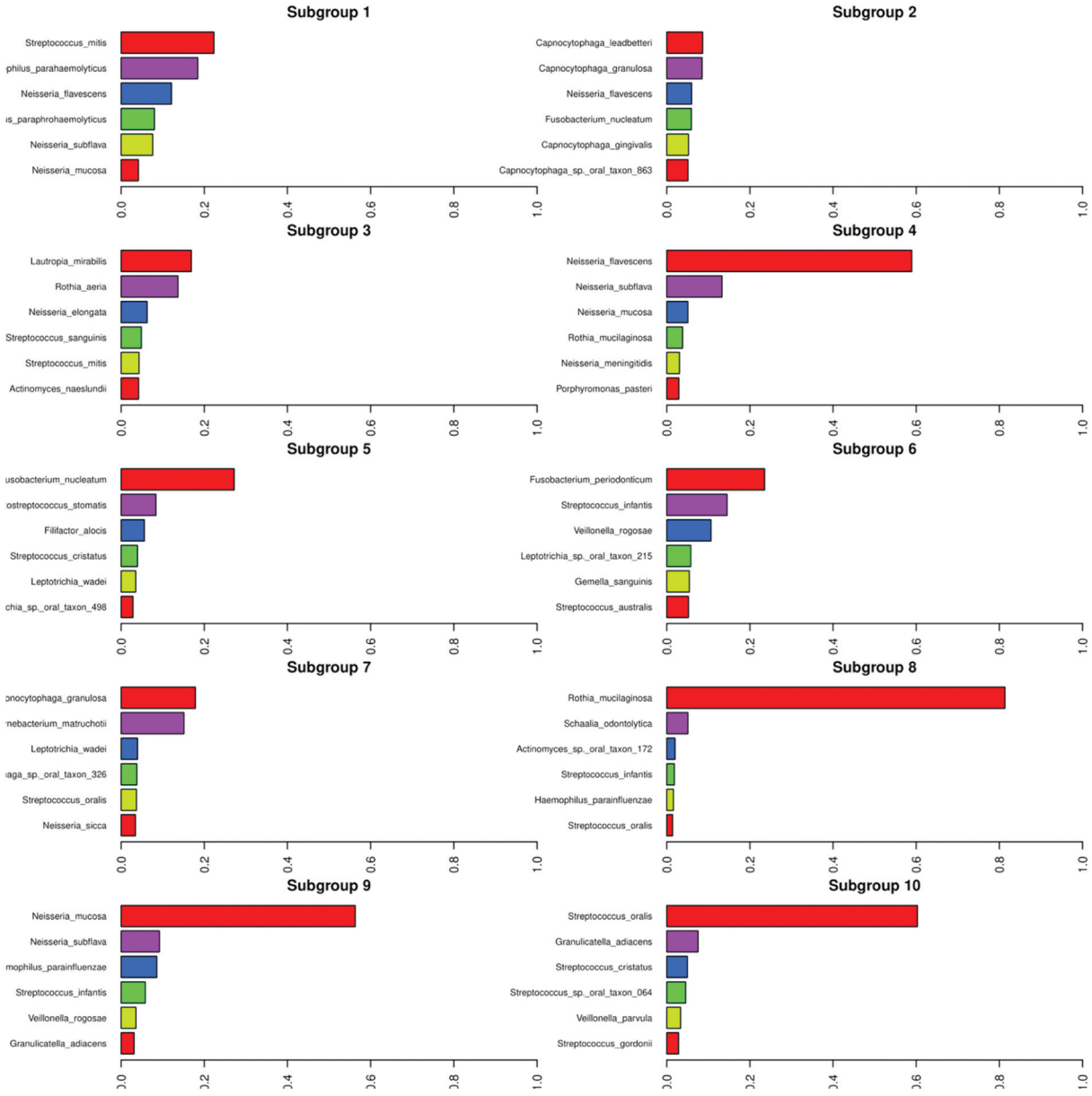

Figure 3. Species-level sub-communities of oral microbiota.

[a]Taxa with highest probability of occurrence in each subgroup, using Latent Dirichlet allocation.

increased stroke risk were GO:0001664 (G protein-coupled receptor binding), GO:0055131 (C3HC4-type RING finger domain binding), GO:0005770 (late endosome), GO:0005776 (autophagosome), and GO:0061635 (regulation of protein complex stability), with ORs ranging from 2.37 to 2.98. Table 4 presents the top 10 pathways with the smallest *p*-values, all of which were associated with a lower ischemic stroke risk. Consistent with gene family associations, the pathways were involved in glutamate metabolism [PWY-4321 (L-glutamate degradation IV), OR = 0.36, 95% CI = 0.18–0.70], gamma-aminobutyric acid (GABA) metabolism [GLUDEG-I-PWY (GABA shunt), OR = 0.47, 95% CI = 0.27–0.82], and tricarboxylic acid (TCA) cycle and variation/upstream pathways, including ARGDEG-PWY, (Superpathway of L-arginine, putrescine, and 4-aminobutanoate degradation; OR = 0.45, 95% CI = 0.26–0.79); ORNARGDEG-PWY (Superpathway of L-arginine and L-ornithine degradation; OR = 0.45, 95% CI = 0.26–0.79); GLYOXYLATE-BYPASS (Glyoxylate cycle; OR = 0.47, 95% CI = 0.26–0.84); and TCA-GLYOX-BYPASS (Superpathway of glyoxylate bypass and TCA; OR = 0.48, 95% CI = 0.27–0.86). Species-specific pathway annotation showed that several of the top taxa in the three subgroups associated with stroke risk are contributors to the pathways. For example, *Corynebacterium variabile* (*Corynebacterium* in genus subgroup 9), *L. mirabilis*, and *N. elongate* (both in species subgroup 3) are main contributors to the GLYOXYLATE-BYPASS (Glyoxylate cycle) pathway.

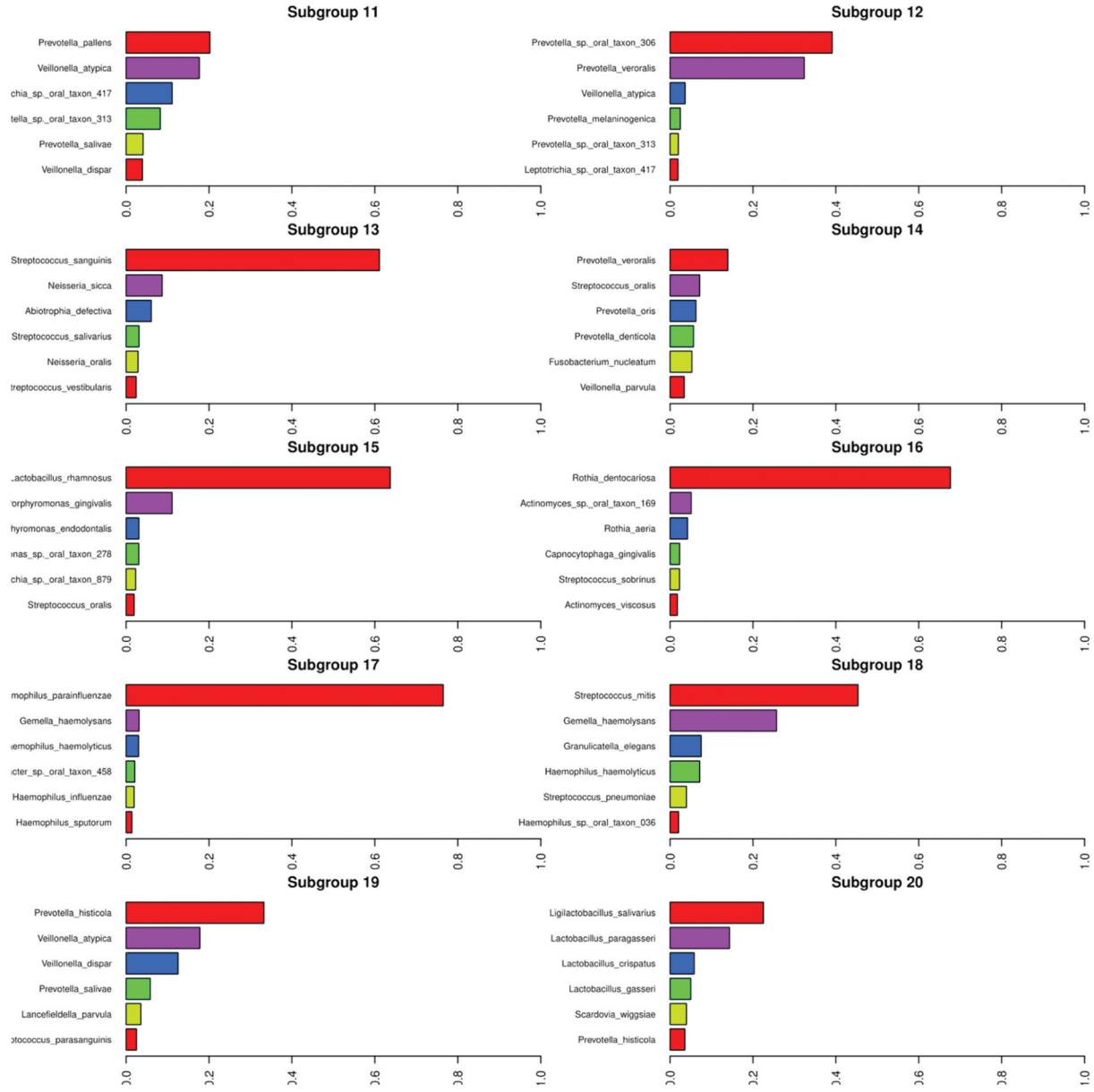

Figure 3. Continued.

We further evaluated correlations of the relative abundances of the top 10 pathways and found that the Pearson correlation coefficient was > 0.7 among the four pathways of PWY-4321, ARGDEG-PWY, ORNARGDEG-PWY, and GLUDEG-I-PWY (0.74–1.00), between GLYOXYLATE-BYPASS and TCA-GLYOX-BYPASS (0.99), and between PWY3O–355 and PWY0–881 (0.92).

Discussion

In our study of 134 ischemic stroke case-control pairs of elderly Chinese women, oral microbial richness was associated with decreased stroke risk, with 2% of the variations in overall oral microbial composition possibly explained by the case/control status. When classifying oral microbiota into subgroups, a genus-level sub-community and two species-level sub-communities were inversely associated with the risk of ischemic stroke. Four top representative taxa in the two aforementioned sub-communities were significantly associated with stroke risk when analyzed individually. A total of 55 taxa, 1,237 gene families, and 90 pathways also showed nominal significant associations with ischemic stroke risk.

The associations identified in our study between oral microbial sub-communities and stroke support previous findings implicating oral microbiota in oral and systemic health. The genus subgroup 9 that was inversely associated with ischemic stroke risk consists of genera related to oral and systemic diseases. Four out of five representative genera in this subgroup – *Actinomyces*, *Streptococcus*, *Lautropia*, and *Corynebacterium* – have been found enriched in

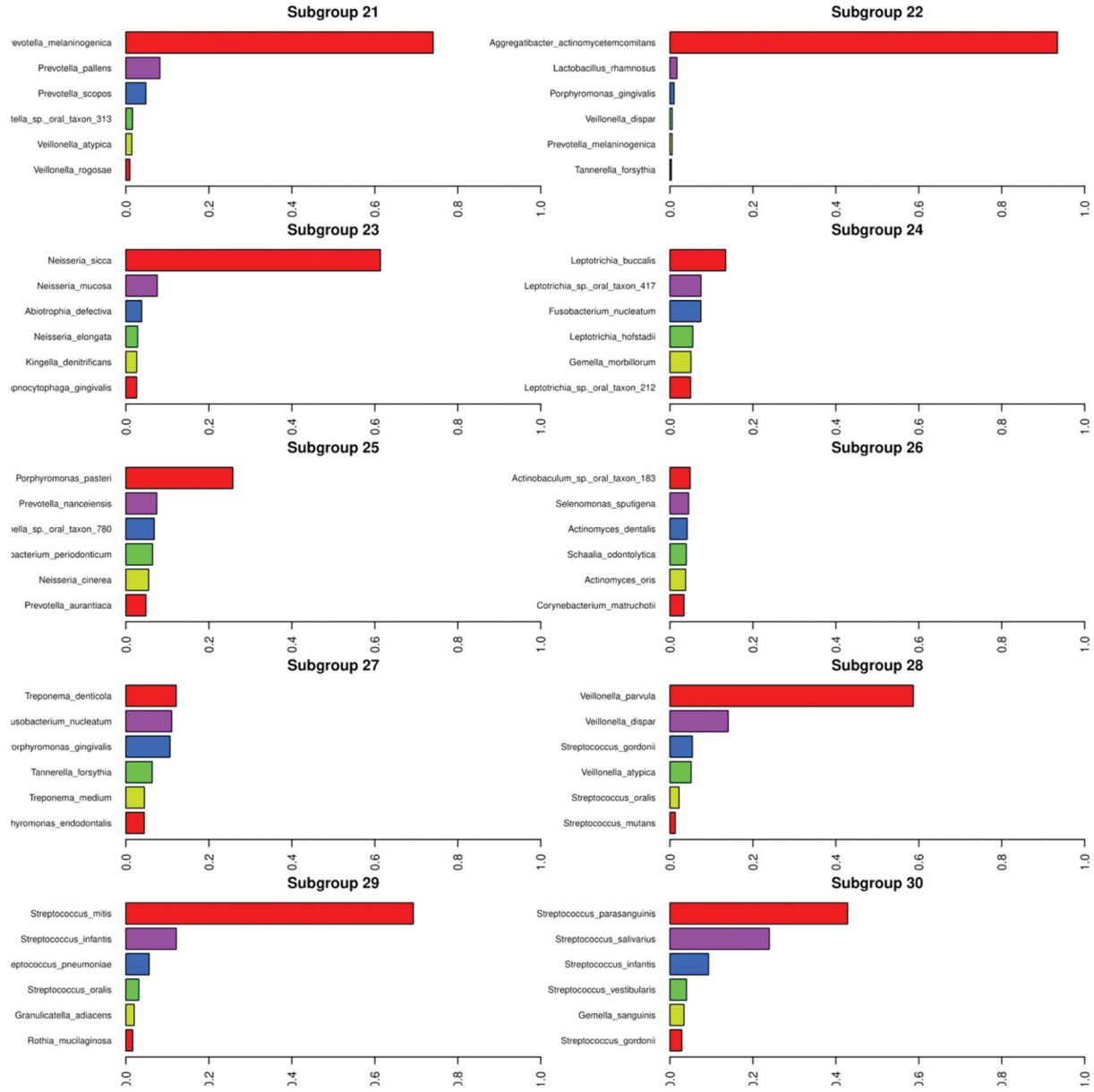

Figure 3. Continued.

subgingival samples of periodontally-healthy individuals, compared to those of periodontitis patients [29–31], while *Selenomonas* spp. are associated with aggressive periodontitis [32,33]. In a study of stroke patients, a 5.1-fold higher load of *Streptococcus* species, mainly *S. mitis*, was detected in aspirated thrombi samples than in arterial blood samples, indicating its potential role in the progression of cerebrovascular diseases [34]. However, our study found no association with stroke risk for *Streptococcus* or *S. mitis* in oral samples. In species subgroup 3, *L. mirabilis* has been reported in higher abundance in healthy individuals and patients with better response to periodontal treatment [33], compared with chronic periodontitis patients [35]. On the other hand, *R. aeria* and *N. elongata* have been indicated in case reports as rare causes of infective endocarditis [36–38]. However, our study found an inverse but non-significant association between stroke risk and these two species in oral microbiota after full adjustment. In species subgroup 26, several leading taxa, including *S. sputigena*, *A. dentalis*, and *A. oris*, were commonly detected in the subgingival plaque of subjects with periodontitis [39,40].

Though direct evidence supporting the association between the oral microbial sub-community and ischemic stroke risk is limited, our study indicates a potential mechanism of the formation of sub-communities beneficial to periodontal health to reduce stroke risk. Active periodontal inflammation, caused by bacterial pathogenesis, may promote a prothrombotic state via recurrent

Table 2. Genus and species level subgroups 3 and 26 associated with ischemic stroke risk in SWHS.

	Probability/ Abundance (%)		Model 1[a]		Model 2[a]		Model 3[a]	
Subgroup and top 5 taxa	Cases	Controls	OR (95%CI)[b]	*p*	OR (95%CI)[b]	*p*	OR (95%CI)[b]	*p*
Genus Subgroup 9	**2.87**	**3.46**	**0.48 (0.29–0.79)**	**3.52E–03**	**0.52 (0.31–0.86)**	**0.01**	**0.52 (0.31–0.90)**	**0.02**
Actinomyces (26.3%)	1.60	1.41	0.71 (0.49–1.03)	0.07	0.69 (0.47–1.03)	0.07	0.70 (0.44–1.11)	0.13
Streptococcus (20.6%)	15.67	16.26	0.89 (0.60–1.32)	0.56	0.84 (0.55–1.28)	0.41	0.75 (0.46–1.21)	0.24
Corynebacterium (7.2%)	0.57	0.72	0.71 (0.50–0.99)	**0.04**	0.69 (0.48–1.01)	0.06	0.70 (0.45–1.08)	0.11
Lautropia (6.8%)	0.27	0.44	0.54 (0.31–0.93)	**0.03**	0.58 (0.32–1.03)	0.06	0.52 (0.26–1.02)	0.06
Selenomonas (5.9%)	0.44	0.59	0.80 (0.57–1.11)	0.19	0.69 (0.47–1.00)	**0.05**	0.71 (0.47–1.09)	0.12
Species Subgroup 3	**1.62**	**2.42**	**0.48 (0.31–0.73)**	**6.67E–04**	**0.50 (0.31–0.78)**	**2.42E–03**	**0.51 (0.31–0.84)**	**0.01**
Lautropia mirabilis (16.9%)	0.27	0.44	0.54 (0.31–0.93)	**0.03**	0.58 (0.32–1.03)	0.06	0.52 (0.26–1.02)	0.06
Rothia aeria (13.7%)	0.47	0.51	0.74 (0.52–1.06)	0.10	0.74 (0.50–1.09)	0.12	0.68 (0.45–1.04)	0.08
Neisseria elongata (6.3%)	0.45	0.62	0.69 (0.49–0.97)	**0.03**	0.67 (0.46–0.97)	**0.04**	0.71 (0.48–1.05)	0.09
Streptococcus sanguinis (4.9%)	0.80	1.01	0.99 (0.76–1.30)	0.96	1.02 (0.76–1.37)	0.89	1.04 (0.78–1.40)	0.79
Streptococcus mitis (4.3%)	4.08	4.73	1.01 (0.65–1.57)	0.96	0.97 (0.60–1.55)	0.89	0.87 (0.52–1.44)	0.59
Species Subgroup 26	**0.50**	**0.92**	**0.57 (0.36–0.90)**	**0.02**	**0.61 (0.38–0.97)**	**0.04**	**0.60 (0.36–0.99)**	**0.04**
Actinobaculum sp. oral taxon 183 (4.9%)	0.046	0.072	0.80 (0.50–1.28)	0.36	0.81 (0.48–1.37)	0.43	0.81 (0.47–1.40)	0.45
Selenomonas sputigena (4.5%)	0.10	0.16	0.86 (0.62–1.20)	0.37	0.74 (0.52–1.07)	0.11	0.74 (0.50–1.10)	0.13
Actinomyces dentalis (4.1%)	0.062	0.057	0.69 (0.45–1.06)	0.09	0.66 (0.41–1.05)	0.08	0.73 (0.42–1.28)	0.27
Schaalia odontolytica (3.9%)	0.22	0.26	1.05 (0.80–1.37)	0.74	1.03 (0.78–1.35)	0.85	1.01 (0.76–1.35)	0.92
Actinomyces oris (3.8%)	0.098	0.14	0.57 (0.31–1.04)	0.07	0.60 (0.31–1.14)	0.12	0.62 (0.31–1.25)	0.18

OR, odds ratio; CI, confidence interval.
[a]Model 1 is the conditional logistic regression model adjusted for age, education, body mass index, and physical activity; model 2 is additionally adjusted for intake of fiber, fruit, vegetable, fat, cholesterol, and carbohydrate; model 3 is additionally adjusted for hypertension and type 2 diabetes.
[b]OR and 95% CI per standard deviation increase in arcsine squared-root-transformed possibility of containing the sub-communities and in centered log-ratio transformed absolute abundance of taxa.

bacteremia, platelet activation, and elevated clotting factors [41], thus elevating the risk of ischemic stroke.

Results from the gene family and pathway analyses further revealed that the oral microbiome may affect stroke risk through involvement in metabolizing glutamate, GABA, and L-arginine, with a correlation between pathways. Those amino acids and their derivatives have been reported to be involved in responses during or after ischemic stroke [42–44]. These metabolic processes are repressed due to the inactivation of enzymes (*e.g.* glutamate synthase) by the oxygen- and nitrogen-centered radicals generated during ischemia, thus having an inverse association with stroke [44]. However, whether the alteration of metabolites in the development of stroke is similar to the processes after stroke attacks remains unknown. Direct investigations into levels of these metabolites, leveraging biospecimens collected prior to stroke onset, are needed to understand the causality of associations.

Table 3. Selected [a] gene families associated with ischemic stroke risk in SWHS (134 pairs).

		Model 1[b]		Model 2[b]		Model 3[b]	
Gene families	Annotation	OR (95%CI)[c]	*p*	OR (95%CI)[c]	*p*	OR (95%CI)[c]	*p*
OR <1							
GO:0106008	[MF] 2-Oxoglutaramate amidase activity	0.59 (0.41–0.85)	5.05E–03	0.54 (0.35–0.84)	6.45E–03	0.42 (0.25–0.71)	1.19E–03
GO:0016040	[MF] Glutamate synthase (NADH) activity	0.46 (0.30–0.70)	2.84E–04	0.40 (0.24–0.68)	5.71E–04	0.38 (0.21–0.69)	1.32E–03
GO:0009448	[BP] Gamma-aminobutyric acid metabolic process	0.47 (0.29–0.76)	2.05E–03	0.43 (0.25–0.72)	1.56E–03	0.36 (0.19–0.68)	1.72E–03
GO:0033539	[BP] Fatty acid beta-oxidation using acyl-CoA dehydrogenase	0.52 (0.35–0.79)	2.01E–03	0.46 (0.28–0.77)	3.06E–03	0.40 (0.22–0.71)	1.82E–03
GO:0009042	[MF] Valine-pyruvate transaminase activity	0.45 (0.27–0.75)	1.99E–03	0.43 (0.25–0.74)	2.68E–03	0.34 (0.17–0.67)	1.86E–03
OR >1							
GO:0001664	[MF] G protein-coupled receptor binding	1.57 (1.06–2.30)	0.02	1.79 (1.15–2.79)	9.93E–03	2.41 (1.40–4.14)	1.47E–03
GO:0055131	[MF] C3HC4-type RING finger domain binding	1.44 (0.97–2.14)	0.07	1.77 (1.10–2.85)	0.02	2.64 (1.44–4.84)	1.71E–03
GO:0005770	[CC] Late endosome	1.42 (0.95–2.12)	0.09	1.82 (1.11–2.99)	0.02	2.51 (1.40–4.47)	1.92E–03
GO:0005776	[CC] Autophagosome	1.38 (0.93–2.05)	0.11	1.74 (1.07–2.83)	0.03	2.98 (1.49–5.98)	2.10E–03
GO:0061635	[BP] Regulation of protein complex stability	1.40 (0.95–2.07)	0.09	1.70 (1.06–2.73)	0.03	2.37 (1.36–4.11)	2.18E–03

GO, gene ontology; OR, odds ratio; CI, confidence interval; MF, molecular function; BP, biological process; CC, cellular component.
[a]Selected from 1,237 gene families associated with cerebral infarction stroke risk at $p < 0.05$ in any of the three models, sorted by *p* value.
[b]Model 1: conditional logistic regression model adjusted for age, education, body mass index, and physical activity; model 2: additionally adjusted for intake of fiber, fruit, vegetable, fat, cholesterol, and carbohydrate; model 3: additionally adjusted for hypertension and type 2 diabetes.
[c]OR (95% CI) per standard deviation increase in arcsine squared-root transformed relative abundance of gene families.

Table 4. Selected[a] MetaCyC pathways associated with ischemic stroke risk in SWHS (134 pairs).

		Model 1[b]		Model 2[b]		Model 3[b]	
Metabolic pathways	Annotation	OR (95%CI)[c]	*p*	OR(95%CI)[c]	*p*	OR(95%CI)[c]	*p*
PWY-4321	L-glutamate degradation IV	0.49 (0.32–0.75)	9.71E–04	0.42 (0.25–0.73)	2.14E–03	0.36 (0.18–0.70)	2.73E–03
ARGDEG-PWY	Superpathway of L-arginine, putrescine, and 4-aminobutanoate degradation	0.54 (0.35–0.82)	3.78E–03	0.51 (0.31–0.83)	7.11E–03	0.45 (0.26–0.79)	5.54E–03
ORNARGDEG-PWY	Superpathway of L-arginine and L-ornithine degradation	0.54 (0.35–0.82)	3.78E–03	0.51 (0.31–0.83)	7.11E–03	0.45 (0.26–0.79)	5.54E–03
GLUDEG-I-PWY	GABA shunt	0.60 (0.41–0.88)	8.30E–03	0.56 (0.36–0.87)	0.01	0.47 (0.27–0.82)	7.41E–03
PWY66–367	Ketogenesis	0.63 (0.42–0.93)	0.02	0.54 (0.34–0.86)	9.11E–03	0.50 (0.30–0.83)	7.83E–03
GLYOXYLATE-BYPASS	Glyoxylate cycle	0.43 (0.26–0.71)	8.67E–04	0.43 (0.25–0.75)	2.95E–03	0.47 (0.26–0.84)	0.01
PWY3O–355	Stearate biosynthesis III (fungi)	0.53 (0.34–0.83)	5.62E–03	0.54 (0.33–0.88)	0.01	0.50 (0.29–0.85)	0.01
PWY-6590	Superpathway of Clostridium acetobutylicum acidogenic fermentation | unclassified	0.63 (0.42–0.93)	0.02	0.51 (0.31–0.85)	9.24E–03	0.51 (0.31–0.86)	0.01
TCA-GLYOX-BYPASS	Superpathway of glyoxylate bypass and TCA	0.44 (0.27–0.72)	1.14E–03	0.44 (0.25–0.77)	3.70E–03	0.48 (0.27–0.86)	0.01
PWY0–881	Superpathway of fatty acid biosynthesis I	0.50 (0.31–0.81)	4.80E–03	0.51 (0.31–0.87)	0.01	0.51 (0.29–0.88)	0.02

OR, odds ratio; CI, confidence interval.
[a]Selected from 90 MetaCyC pathways associated with stroke risk at $p < 0.05$ in any of the three models, sorted by p-value.
[b]Model 1: conditional logistic regression model adjusted for age, education, body mass index, and physical activity; model 2: additionally adjusted for intake of fiber, fruit, vegetable, fat, cholesterol, and carbohydrate; model 3: additionally adjusted for hypertension and type 2 diabetes.
[c]OR (95% CI) per standard deviation increase in arcsine squared-root transformed relative abundance of MetaCyC pathways.

To our knowledge, this study is the first to prospectively investigate the association between oral microbiome and ischemic stroke risk. We have carefully adjusted for known/suspected risk factors for stroke including hypertension, T2D, obesity, and selected lifestyles in our analysis. The association between ischemic stroke and oral microbiome was assessed with and without adjustment for hypertension and T2D (model 3 and model 2), and the results were largely similar. This also holds true for adjustment of obesity and lifestyle as shown in the results of models 1 and 2. These results suggest that certain oral microbes might be risk factors of ischemic stroke that are independent of traditional stroke risk factors. Largely, previous investigations into the role of oral health in stroke focused on the improvement of oral health after stroke onset. In our study, we applied shotgun metagenomic sequencing data and LDA analyses, in addition to evaluation of individual taxa, to assess the role of the oral microbial community and oral microbes in stroke etiology. Microbial gene family and pathway analyses further deepened our understanding of potential mechanisms. Our findings suggest that oral microbial composites may function collectively as sub-communities, rather than individually, to modify the risk of ischemic stroke.

Limitations of the study include the one-time sample collection, preventing us from evaluating the influence of oral microbiota changes over time on stroke etiology. First, information on oral health was not available at the time of oral sample collection, preventing us from evaluating the possible mediation effect of the oral microbiome. Second, there might be unmeasured confounders in the association between oral microbiome and stroke risk, such as environmental exposures, oral health, and inflammation, preventing us from determining the causality of the association. In addition, despite our efforts to dissect the structures in oral microbiota with the help of LDA, the underlying structure of a community cannot be resolved [45]. To date, the sub-community structures of human oral microbiota have been understudied and further investigation with advanced tools to be developed in the future is needed. Lastly, our study has a small sample size, thus lacking the statistical power to control for multiple comparisons and more in-depth analyses.

In conclusion, in a prospective shotgun metagenomic study, alpha diversity, beta diversity, oral microbial sub-communities, and multiple individual microbial taxa, gene families, and pathways were identified as associated with the subsequent risk of ischemic stroke. Future studies with larger sample sizes, longitudinally collected oral samples, and information on oral diseases, particularly periodontitis, are needed to further evaluate the role of oral microbiota in the etiology of ischemic stroke and to assess the causality.

Acknowledgments

The authors would like to thank participants and research team members of SWHS for their contributions to the study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The SWHS was supported by the National Cancer Institute (NCI) grant UM1CA182910 (to W.Z). C.W. is supported by Ingram Cancer Research Professorship fund (to X-O.S.). Y.Y. is supported partially by the NCI grant K99/R00 CA248822. Sample preparation was conducted at the Survey and Biospecimen Shared Resources, which is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the study.

Author contributions

X-O.S. and J.L. conceived the study. C.W. and Y. Y. analyzed data and drafted the initial manuscript. W. Z. obtained the funding and contributed to study design. H.C. contributed to data management. Y.G. contributed to the field work. Q.C. and J.W. contributed to the lab work. All authors critically reviewed and proved the manuscript.

Data availability statement

Data are available from the corresponding author on request.

Statement of ethics

Written informed consent was provided by all participants, and the SWHS was approved by the institutional review boards of Vanderbilt University and the Shanghai Cancer Institute.

References

[1] Feigin VL, Stark BA, Johnson CO, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet Neurol. 2021;20 (10):795–820. doi: 10.1016/S1474-4422(21)00252-0

[2] Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics—2022 update: a report from the American Heart association. Circulation. 2022;145(8): e153–e639. doi: 10.1161/CIR.0000000000001052

[3] Esenwa CC, Elkind MS. Inflammatory risk factors, biomarkers and associated therapy in ischaemic stroke. Nat Rev Neurol. 2016;12(10):594–604. doi: 10.1038/nrneurol.2016.125

[4] Hart RG, Diener HC, Coutts SB, et al. Embolic strokes of undetermined source: the case for a new clinical construct. Lancet Neurol. 2014;13(4):429–438. doi: 10.1016/S1474-4422(13)70310-7

[5] Fagundes NCF, Almeida APCPSC, Vilhena KFB, et al. Periodontitis as a risk factor for stroke: a systematic review and meta-analysis. Vasc Health Risk Manag. 2019;15:519–532. doi: 10.2147/VHRM.S204097

[6] Chen C, Hemme C, Beleno J, et al. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. ISME J. 2018;12 (5):1210–1224. doi: 10.1038/s41396-017-0037-1

[7] Sampaio-Maia B, Caldas IM, Pereira ML, et al. The oral microbiome in health and its implication in oral and systemic diseases. Adv Appl Microbiol. 2016;97:171–210. doi: 10.1016/bs.aambs.2016.08.002

[8] Long J, Cai Q, Steinwandel M, et al. Association of oral microbiome with type 2 diabetes risk. J Periodontal Res. 2017;52(3):636–643. doi: 10.1111/jre.12432

[9] LaMonte MJ, Gordon JH, Diaz-Moreno P, et al. Oral microbiome is associated with incident hypertension among postmenopausal women. J Am Heart Assoc. 2022;11(6):e021930. doi: 10.1161/JAHA.121.021930

[10] Boaden E, Lyons M, Singhrao SK, et al. Oral flora in acute stroke patients: a prospective exploratory observational study. Gerodontology. 2017;34(3):343–356. doi: 10.1111/ger.12271

[11] Inenaga C, Hokamura K, Nakano K, et al. A potential New risk factor for stroke: streptococcus mutans with Collagen-Binding Protein. World Neurosurg. 2018;113:e77–e81. doi: 10.1016/j.wneu.2018.01.158

[12] Zheng W, Chow WH, Yang G, et al. The Shanghai women's health study: rationale, study design, and baseline characteristics. Am J Epidemiol. 2005;162 (11):1123–1131. doi: 10.1093/aje/kwi322

[13] Zhang X, Shu XO, Gao YT, et al. General and abdominal adiposity and risk of stroke in Chinese women. Stroke. 2009;40(4):1098–1104. doi: 10.1161/STROKEAHA.108.539692

[14] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120. doi: 10.1093/bioinformatics/btu170

[15] Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–359. doi: 10.1038/nmeth.1923

[16] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Bio. 2019;20(1):257. doi: 10.1186/s13059-019-1891-0

[17] Lu J, Breitwieser FP, Thielen P, et al. Bracken: estimating species abundance in metagenomics data. PeerJ Comput Sci. 2017;3:e104. doi: 10.7717/peerj-cs.104

[18] Escapa IF, Chen T, Huang Y, et al. New insights into human nostril microbiome from the expanded human oral microbiome database (eHOMD): a resource for the microbiome of the human aerodigestive tract. mSystems. 2018;3(6). doi: 10.1128/mSystems.00187-18

[19] Shao Y, Forster SC, Tsaliki E, et al. Stunted microbiota and opportunistic pathogen colonisation in caesarean section birth. Nature. 2019;574(7776):117–121. doi: 10.1038/s41586-019-1560-1

[20] Franzosa EA, McIver LJ, Rahnavard G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018;15 (11):962–968. doi: 10.1038/s41592-018-0176-y

[21] Suzek BE, Wang Y, Huang H, et al. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015;31 (6):926–932. doi: 10.1093/bioinformatics/btu739

[22] Oksanen J, Blanchet FG, Friendly M, et al. Vegan: community ecology package. Published online November 28, 2020. [cited 2021 Sep 6]; Available from: https://CRAN.R-project.org/package=vegan

[23] Zhao N, Chen J, Carroll IM, et al. Testing in microbiome-profiling studies with MiRKAT, the microbiome regression-based kernel association test. Am J Hum Genet. 2015;96(5):797–807. doi: 10.1016/j.ajhg.2015.04.003

[24] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res. 2003;3(Jan):993–1022.

[25] Hosoda S, Nishijima S, Fukunaga T, et al. Revealing the microbial assemblage structure in the human gut

microbiome using latent Dirichlet allocation. Microbiome. 2020;8(1):1–12. doi: 10.1186/s40168-020-00864-3
[26] Deek RA, Li H. A zero-inflated Latent Dirichlet allocation model for microbiome studies. Front Genet. 2021. doi: 10.3389/fgene.2020.602594
[27] Yan J, Chuai G, Qi T, et al. MetaTopics: an integration tool to analyze microbial community profile by topic model. BMC Genomics. 2017;18(Suppl 1):962. doi: 10.1186/s12864-016-3257-2
[28] Breuninger TA, Wawro N, Breuninger J, et al. Associations between habitual diet, metabolic disease, and the gut microbiota using latent dirichllocation. Microbiome. 2021;9(1):61. doi: 10.1186/s40168-020-00969-9
[29] Tsai CY, Tang CY, Tan TS, et al. Subgingival microbiota in individuals with severe chronic periodontitis. J Microbiol Immunol Infect. 2018;51(2):226–234. doi: 10.1016/j.jmii.2016.04.007
[30] Nibali L, Sousa V, Davrandi M, et al. Patterns of subgingival microbiota in different periodontal phenotypes. J Dent. 2022;117:103912. doi: 10.1016/j.jdent.2021.103912
[31] Meuric V, Le Gall-David S, Boyer E, et al. Signature of microbial dysbiosis in periodontitis. Appl environ microbiol. 2017;83(14):e00462–17. doi: 10.1128/AEM.00462-17
[32] Gonçalves LFH, Fermiano D, Feres M, et al. Levels of Selenomonas species in generalized aggressive periodontitis. J Periodontal Res. 2012;47(6):711–718. doi: 10.1111/j.1600-0765.2012.01485.x
[33] Nibali L, Sousa V, Davrandi M, et al. Differences in the periodontal microbiome of successfully treated and persistent aggressive periodontitis. J Clin Periodontol. 2020;47(8):980–990. doi: 10.1111/jcpe.13330
[34] Olli P, Juha-Pekka P, Sari T, et al. Oral bacterial signatures in cerebral thrombi of patients with acute ischemic stroke treated with thrombectomy. J Am Heart Assoc. 2019;8(11):e012330. doi: 10.1161/JAHA.119.012330
[35] Ikeda E, Shiba T, Ikeda Y, et al. Japanese subgingival microbiota in health vs disease and their roles in predicted functions associated with periodontitis. Odontology. 2020;108(2):280–291. doi: 10.1007/s10266-019-00452-4
[36] Tarumoto N, Sujino K, Yamaguchi T, et al. A first report of *Rothia aeria* endocarditis complicated by cerebral hemorrhage. Intern Med. 2012;51(23):3295–3299. doi: 10.2169/internalmedicine.51.7946
[37] Collarino R, Vergeylen U, Emeraud C, et al. Mitral endocarditis due to Rothia aeria with cerebral haemorrhage and femoral mycotic aneurysms, first French description. New Microbes New Infect. 2016;13:40–42. doi: 10.1016/j.nmni.2016.06.004
[38] Banjari M, Haddad E, Bonnet I, et al. Infective endocarditis due to Neisseria elongata: a case report and literature review. Infectious Diseases Now. 2021;51(7):622–626. doi: 10.1016/j.idnow.2021.01.013
[39] Faveri M, Mayer MPA, Feres M, et al. Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. Oral Microbiol Immunol. 2008;23(2):112–118. doi: 10.1111/j.1399-302X.2007.00397.x
[40] Vielkind P, Jentsch H, Eschrich K, et al. Prevalence of Actinomyces spp. In patients with chronic periodontitis. Int J Med Microbiol. 2015;305(7):682–688. doi: 10.1016/j.ijmm.2015.08.018
[41] Lourbakos A, Yuan Y, Jenkins AL, et al. Activation of protease-activated receptors by gingipains fromPorphyromonas gingivalis leads to platelet aggregation: a new trait in microbial pathogenicity. Blood. 2001;97(12):3790–3797. doi: 10.1182/blood.V97.12.3790
[42] Wu C, Sun D. GABA receptors in brain development, function, and injury. Metab Brain Dis. 2015;30(2):367–379. doi: 10.1007/s11011-014-9560-1
[43] Grosse GM, Schwedhelm E, Worthmann H, et al. Arginine derivatives in cerebrovascular diseases: mechanisms and clinical implications. Int J Mol Sci. 2020;21(5):1798. doi: 10.3390/ijms21051798
[44] Jeitner TM, Battaile K, Cooper AJL. Critical evaluation of the changes in glutamine synthetase activity in models of cerebral stroke. Neurochem Res. 2015;40(12):2544–2556. doi: 10.1007/s11064-015-1667-1
[45] Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLOS One. 2012;7(2):e30126. doi: 10.1371/journal.pone.0030126