Correlation Between Salivary Microbiology and H2S Concentration of Oral Cavity

Jiarui Jiang
China Tobacco Yunnan Industrial Corporation https://orcid.org/0000-0001-7985-1198
Yufeng Huang
China Tobacco Yunnan Industrial Corporation
Na Luo
Kunming Medical University
Qili Mi
China Tobacco Yunnan Industrial Corporation
Xuemei Li
China Tobacco Yunnan Industrial Corporation
Wei Zhang
China Tobacco Yunnan Industrial Corporation
Baokun Zhu
China Tobacco Yunnan Industrial Corporation
Qian Gao (gaoqian840905@hotmail.com)
China Tobacco Yunnan Industrial Corporation

Research Article

Keywords: salivary microbiology, H2S, oral cavity, halitosis, metagenome

Posted Date: September 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-919525/v1

Correlation between salivary microbiology and H_2S concentration of oral cavity

Jiarui Jiang[1#], Yufen Huang[1#], Na Luo[2], Qili Mi[1], Xuemei Li[1], Wei Zhang[1], Baokun Zhu[1*], Qian Gao[1*]

Abstract

Background: Halitosis is caused by metabolites produced by oral microorganisms. Hydrogen sulfide is the most important compound that leads to the oral malodor, and is thought to be closely correlated with the activity of oral microorganism. Therefore, it is important to clarify the correlation between oral microbes and metabolites.

Methods: Based on the 16S rRNA gene amplicon and shotgun metagenomic sequencing of oral microorganism, and oral malodor test, this study attempted to explain the contribution of oral microorganisms to the hydrogen sulfide of oral malodor.

Results: The data shows that microbial taxa consisted in the H_2S low and high groups are different, and most of the enriched taxa in the H_2S high group are genus that correlated with H_2S concentration. The two species *Fusobaeterium periodonticum* and *Prevotella nanceiensis* are significant different in both coverage breadth and depth and LPS biosynthesis contributions in two groups. According to KEGG metabolism pathways detected by HUMAnN2, subjects of the H_2S high group may have a high risk to bacterial infection, since the LPS biosynthesis is enrichment. The contribution of *F. periodonticum* to sulfur metabolism between two groups is significantly different, and the relative abundance of *F. periodonticum* is higher in the H_2S high group as well.

Conclusions: The H_2S content, is significantly associated with the composition and abundance of microorganisms in the oral cavity. The increase of microbial abundance and metabolism of some sulfide products are the main causes of halitosis. The most of the enriched microorganisms enriched in people with high H_2S are associated with oral diseases such as caries and periodontal diseases, indicating that the diseases associated with oral microbes are not independent of each other and have some associations between some oral diseases.

Keywords: salivary microbiology; H_2S; oral cavity; halitosis; metagenome

*Correspondence:
Baokun Zhu, zhubk@ynzy-tobacco.com
Qian Gao, gaoqian840905@hotmail.com

Author Contributions:
#Jiarui Jiang and Yufen Huang contributed equally to the article.

Background

Halitosis is mainly divided into transient halitosis, extra-oral halitosis (EOH), and intra-oral halitosis (IOH)[1]. IOH is mainly caused by the putrefactive actions of microorganisms such as bacteria, fungi, viruses, and protozoa[2]. The number of microbial communities in human oral cavity is second only to that of gastrointestinal tract[3].The main cause of IOH is volatile sulfur

compounds (VSCs) produced by oral bacteria. The most important VSCs are hydrogen sulfide(H_2S), dimethyl sulfide $(CH_3)_2S$, dimethyl disulfide ($C_2H_6S_2$), and methyl mercaptan (methanethiol, CH_3SH)[4]. Previous studies have revealed that bacteria to produce of VSCs, such as *Solobacterium moorei*, *Prophyromonas gingivalis*, *Treponema denticola*, *Prevotella intermedia*, *Streptococcus oralis*, and *Tannerella forsythia*, are associated with IOH[4-7].

In particular, hydrogen sulfide (H_2S), methyl mercaptan (methanethiol, CH_3SH), and dimethyl sulfide $(CH_3)_2S$ are considered important markers of IOH[8]. H_2S is the most common VSCs formed by bacterial degradation of the sulfur-containing amino acid cysteine in the oral cavity. It is a low-molecular weight and volatile gas compound detected in halitosis (bad breath) patients and in periodontal pockets in patients with periodontitis [9–11]. H_2S is regarded as one of the most toxic metabolites produced in the periodontal pocket. In vitro studies have shown that H_2S can damage epithelial cells [12], enhance permeability of the oral mucosa [13], and cause apoptosis of gingival fibroblasts [14]. Likewise, the pathogenesis of periodontal disease is poorly understood but it is usually accepted that bacterial metabolites in general, and H_2S in particular, are of importance in the development and activity of the disease[15].

Traditional oral microbial studies usually collect samples from a certain part of the oral cavity for culture, and identify microorganisms through morphological, physiological and biochemical characteristics. This method has great advantages for analyzing the classification and function of microorganisms, but its disadvantage is that the culture process is complicated and susceptible to environmental [16]. Emerging microbiome approaches, such as metagenomics, metaproteinomics, and metatranscripts, are targeted at entire microbial populations in environmental samples, with screening and sequencing of functional genes, and accessing to all microbial genomes, proteomics and transcription information in environmental samples through high-throughput sequencing. Moreover, they can effectively explain the microbial diversity, community structure, evolutionary relationships, functional activity, collaborative relationships and relationships with the environment [17-18]. Differences of composition and functions of oral microorganism group in patient and healthy people can be compared and analyzed by oral metagenomic research, which plays an important role in the study of oral microorganism group, and is beneficial to the prevention and treatment of disease as well. Living habits and diet rules also have a great influence on the oral microbiome.

Exogenous factors include tobacco, alcohol, onion, garlic and other foods and seasonings, directly from the digestive tract and mouth issued a temporary odor. Tobacco and alcohol may give rise to distinct halitosis that can last a few hours, and the odour can be almost continuous if the person has a persistent habit [19]. Halitosis is the consequence of microbial putrefaction of food debris, cells, saliva, and blood. However, no obvious association exists between halitosis and any specific bacterial infection, suggesting that halitosis reflects complex interactions between several oral bacterial species [19].We expect to probe into the correlation between oral microbes and oral H_2S,,and analyze the relationship between the ecological balance and halitosis of the oral microbiota and study microorganisms with positive and negative effects on halitosis.

Methods

Ethics Statement

Sample survey questionnaires

The questionnaire mainly involved three parts: living habits, smoking behavior and oral health. The lifestyle part is mainly exclude the influence of age, gender, BMI, drinking on oral microorganisms. In the investigation of smoking behavior, draw on the experience of the design of nicotine dependence test scale (FTND), mainly including cigarette age and daily smoking, which can be applicable to refined cohort analysis. In the oral health part, oral diseases related to oral microorganisms mainly include: periodontal disease, dental caries, halitosis and oral ulcers. All procedures were followed in accordance with relevant guidelines. Exclusion criteria are as follows:

a. Age is less than 18 years or older than 60 years;
b. Suffering from heart disease, high blood pressure, serious liver and kidney insufficiency, serious oral diseases and other systemic diseases;
c. A history of medication and basic periodontal treatment existed within the first 3 months of oral saliva samples;
d. Periodontal bleeding, pregnancy, or lactation;
e. The number of remaining teeth in the mouth is less than 28;
f. Influenza, fever-like symptoms;
g. Antibiotics were used within 14 days before sampling.

Sample collection

Sample collection mainly consisted into three parts: volunteer invitation, questionnaire screening and sample collection. The first stage focuses on obtaining population samples for 16S rRNA sequencing, and the second stage focuses on selecting representative populations for metagenome sequencing based on the first stage study results. Samples not meeting above project requirements were excluded in the invitation and questionnaire section. Overall, 300 current smokers (CS) and 150 nonsmokers (NS) were included in the study. Before collecting samples, subjects were required not to drink alcohol within 48 hours, not to brush their teeth and not to take food in 1 hour after getting up in the morning.

Saliva collection: A GeneFix^{TM} saliva microbiome DNA collector (MFX, Isohelix company) was used for saliva collection (2 mL, no pigment, lipstick, blood, food residue, conserved in liquid nitrogen).

Oral malodor test: A OralChroma^{TM} portable gas chromatograph (CHM-2, FISG company, Japen) was used to make a relatively accurate qualitative and quantitative analysis of H_2S, CH_3SH and $(CH_3)_2S$, the three main gases in oral malodor. Subjects closed their mouth for 3 minutes, insert the collector into the mouth about 3 cm, push and pull the collector 3 times, remove the collector and push the gas up to the 1 mL scale, insert the instrument inlet into the gas for detection, repeat detection 3 times and take the average.

Extraction of oral microbial DNA

The total DNA of saliva microbial was extracted using the NEB microbiome DNA enrichment kit

(New England Biolabs, Ipswich, MA, US). The testing of DNA samples concentration use Qubit fluorescence quantiimeter. The 1% agarose gel electrophoresis detects the integrity of the DNA samples. Qualified samples can only be used for library preparation.

16S rRNA gene amplification, sequencing and bioinformatics

The V3-V4 regions of bacterial 16S rRNA gene was amplified using primers 341F (5'-ACTCCTACGGGAGGCAGCAG-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3'). PCR amplification was performed in a 50 μL reaction containing 30 ng template, fusion PCR primer and PCR master mix. PCR cycling conditions were as follows: 94℃ for 3 minutes, 30 cycles of 94℃ for 30 seconds, 56℃ for 45 seconds, 72℃ for 45 seconds and final extension for 10 minutes at 72℃ for 10 minutes. The PCR products were purified with AMPure XP beads (Backman Coulter, USA) and eluted in Elution buffer. Libraries were qualified by the Agilent 2100 bioanalyzer (Agilent, USA). The validated libraries were sequenced on Illumina Hiseq 2500 platform following the standard pipelines of Illumina (Illumina, USA), and generated 2 x 300 bp paired-end reads.

After removing the reads containing low quality and fuzzy bases from the raw sequencing data, a total of 12,920,173 paired-end sequencing reads were generated from 203 samples, with an average of 63,646 ± 356 per sample. The paired-end sequencing reads were merged by FLASH[20], and clustered into Operational Taxonomy Units (OTUs) at 97% identity by UPARSE[21]. Finally, a total of 769 OTUs were obtained, with an average of 321 ± 51 per sample. The obtained OTU representative sequences were taxonomically classified with a 0.8 threshold by RDP Classifier (v2.2)[22] based on Greengene database (v201305)[23]. Alpha and beta diversity were calculated using MOTHUR[18] and QIIME[24].

Whole-genome shutgun sequencing and bioinformatics

1μg genomic DNA was randomly fragmented by Covaris and an average size of 200-400bp fragmented was selected by Magnetic beads. The selected fragments were through end-repair, 3' adenylated, adapters-ligation, PCR amplifying and the products were purified by the Magnetic beads. The double stranded PCR products were heat denatured and circularized by the splint oligo sequence. The single strand circle DNA (ssCir DNA) were formatted as the final library and qualified by QC. The qualified libraries were sequenced on DNBSEQ platform (BGI-Shenzhen, China) and 150 bp paired-end sequencing data was generated.

Reads with low quality, ambiguous bases (N bases), adaptor contamination were filtered from the sequencing raw data using SOAPnuke[25]. A total of 434.90 GB high quality data were obtained from 31 samples, with an average of 14.03 GB per sample. In order to remove Human DNA contamination, sequencing reads were aligned to hg19 using Bowtie2[26] (default parameters), and average rate of contamination for each sample was 81.46%. KEGG metabolism pathways profiling was preformed using HUMAnN2[27]. Reference genome sequences of *F. periodonticum* (strain 2_1_31, GCA_003019755.1) and *P. nanceiensis* (strain DSM 19126, GCA_000379965.1) were downloaded from National Center for Biotechnology Information (NCBI), and strain diversity analysis of these two species in samples was performed using inStrain[28].

Statistical analysis

All statistical analyses in this study were performed using R software (v3.5.1). Principal coordinate analysis (PCoA) was performed based on unweighted Unifrac, weighted Unifrac and Bray Curtis distance matrices and the significant was determined by permutational multivariate analysis of variance (PERMANOVA) ('vegan' package of R) with 9,999 permutations. Differential taxa between H_2S high and low groups were identified by LEfSe[29] and taxa with |LDA score| (log10) > 2 and p value < 0.05 were considered significant different. Differential metabolic pathways between the two groups were determined by Wilcoxon rank-sum test and p values corrected by Benjamini-Hochberg method < 0.05 were considered as significant. Correlation between genera and H_2S concentration was determined by Spearman rank correlation coefficients.

Results

Baseline characteristics

Statistical analysis of the basic information showed that there were no significant differences in other phenotypes except H_2S content between the CS and NS (Table S1). It was worth noting that H_2S content of the NS was significantly higher than the CS, while there were no significant differences in CH_3SH and $(CH_3)_2S$ between smokers and non-smokers. In order to further investigate the reason, the 450 samples were screened at H_2S and CH_3SH concentrations. The baseline characteristics, dietary status and oral status are all considered. Samples with the H_2S concentration more than 1,000 ppb and the ratio of CH_3SH/ H_2S less than 0.5 were selected as the H_2S high group, while with H_2S and CH_3SH concentration less than 100 ppb as the H_2S low group (Figure 1). There are 101 and 102 samples for the H_2S high and low group, respectively (Table 1). 31 samples for metagenome sequencing from the total 203 samples, including 15 from the high H_2S group and 16 from the low H_2S group. The CH_3SH concentration in the two groups is significant different (p < 0.01). Despite the differences of smokers between the two groups, no statistically significant differences were observed. Other characteristics in the table 1 have no statistically significant differences.

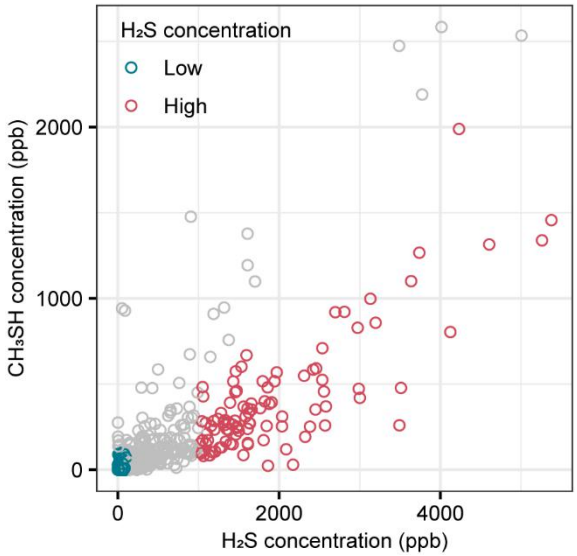

Figure 1 Hydrogen sulfide (H_2S) and methyl mercaptan (CH_3SH) concentration in the oral malodor of 450 samples. Samples with H_2S concentration more than 1,000 ppb and the ratio of

CH_3SH/ H_2S less than 0.5 were selected as the H_2S high group (red, n = 101), while these with H_2S and CH_3SH concentration less than 100 ppb as the H_2S low group (green, n = 102).

Table 1 Baseline characteristics, dietary status and oral status of 203 individuals.

Characteristics	Low H_2S (n = 102)	High H_2S (n = 101)	*p* value*
Age, mean (s.d.)	38.3 (9.5)	35.6 (8.9)	1
BMI, mean (s.d.)	21.9 (2.9)	21.8 (2.7)	1
Male, n (%)	71 (69.61)	68 (67.33)	0.727
Dietary status			
Diet, n (%)			
Mixed meat and vegetables	74 (72.55)	75 (75.00)	
Preference for meat	14 (13.73)	10 (10.00)	0.709
Preference for vegetables	14 (13.73)	15 (15.00)	
Drinking, n (%)	58 (56.86)	56 (55.45)	0.839
Smoking, n (%)	76 (74.51)	64 (63.37)	0.086
Oral status			
CH_3SH, mean (s.d.)	23.8 (26.2)	400.2 (334.0)	< 0.001
$(CH_3)_2S$, mean (s.d.)	157.8 (565.4)	66.1 (224.8)	0.805
Periodontitis, n (%)	10 (9.80)	9 (8.91)	0.827
Caries, n (%)	51 (50.00)	44 (43.56)	0.358
Bleeding gums, n (%)	31 (30.39)	37 (36.63)	0.346
Halitosis, n (%)	36 (35.29)	35 (34.65)	0.924

*p values was calculated by Wilcoxon rank-sum test (non-normal distribution) or one-way ANOVA test (normal distribution) for continuous variables, and Fisher's exact test (expected frequency < 5) or Chi-squared test (expected frequency > 5) for categorical variables.

Alteration of microbial community in the high H_2S group

The distributions of richness and evenness at the OTU level are not noticeable(Figure S1). All methods used in the current study, the unweighted UniFrac (PCoA=17.63%), the weighted UniFrac (PCoA=47.08%), and the Bray Curtis distance matrix (PCoA=20.41%), showed that oral microbial communities clustered well into high and low H_2S groups, with the supporting of PERMANOVA analyses (Figure 2), indicating that the oral microbial community of high and low H_2S conditions were significant different.

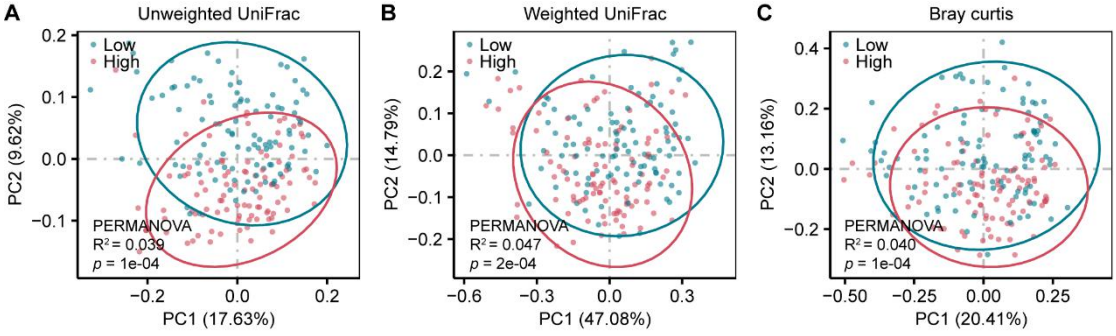

Figure 2 Different oral microbial compositions between the high and low H_2S groups. A) PCoA based on the unweighted UniFrac distance matrix. B) PCoA based on the weighted UniFrac

distance matrix. C) PCoA based on the Bray Curtis distance matrix. 3 index distance matrices were calculated based on OTU level and the PERMANOVA test was used to determine significance.

Genera enriched in high H_2S group were associated with H_2S concentration

A total of 77 differential taxa between the high and low H_2S groups were detected using LEfSe, including 4 phyla, 8 classes, 15 orders, 23 familes and 27 genera (Table S2). In phylum level, the abundance of *Bacteroidetes* was increased significantly in the high H_2S group, while *Actinobacteria* was decreased. In genus level, 19 genera were enriched in the low H_2S group, while 8 genus were enriched in the high H_2S group (Figure 3A). The LDA scores of *Rothia*, *Streptococcus*, *Leptotrichia*, *Oribacterium*, *Campylobacter*, *Azospirillum*, *Selenomonas*, *Sphingomonas*, *Actinomyces* and *Lautropia* were exceed 3 in the low group, while that of *Prevotella*, *Fusobacterium*, *Aggregatibacter*, *Porphyromonas*, *Peptostreptococcus, Parvimonas* and *Moraxella* were exceed 3 in the high group. According to the association analysis between differential genera profile and H_2S concentration, the relative abundance of 7 genera (*Bulleidia*, *Parvimonas*, *Peptostreptocoeeus*, *Moraxella*, *Aggregatibacter*, *Fusobaeterium*, and *Prevotella*) that enriched in H_2S high group were significantly positively correlated with H_2S concentration, while 13 genera that enriched in H_2S low group was negatively correlated with H_2S concentration (Figure 3B). In the H_2S high group, the relative abundance of the genera *Peptostreptococcus* had the strongest correlation coefficient with H_2S concentration (r = 0.529, p = 1.99e-14, Spearman's rank correlation).

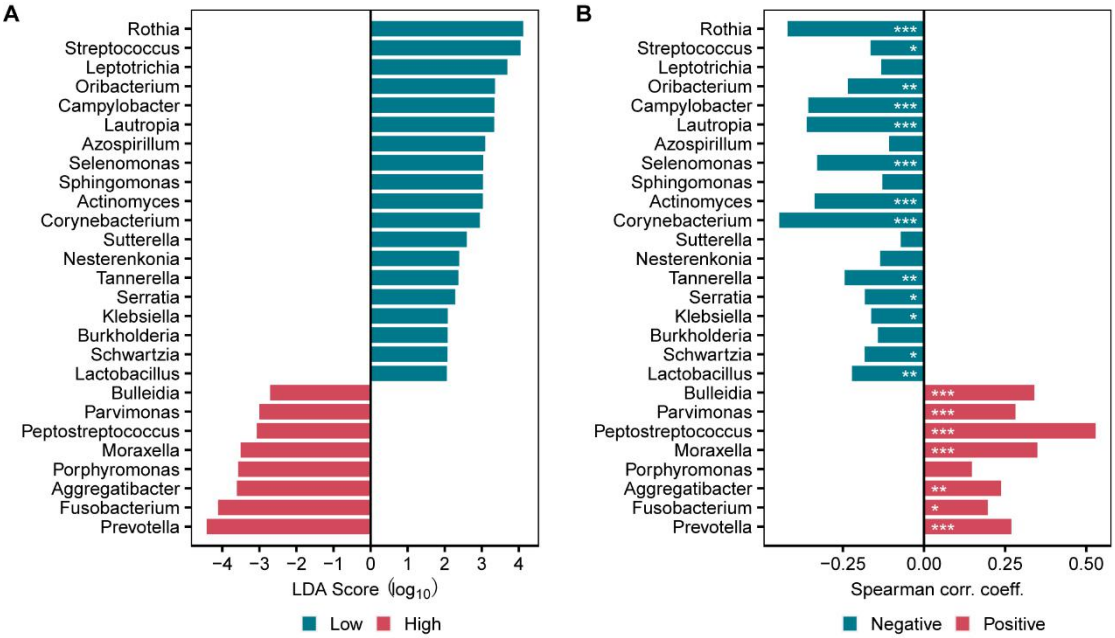

Figure 3 Differential genera between the high and low H_2S groups. A) Differential genera detected by LEfSe between the high and low H_2S groups. The genera with LDA scores (log 10) > 2 and $p < 0.05$ are shown. Negative LDA scores indicate enrichment in the high H_2S group (red), while positive LDA scores indicate enrichment in the low H_2S group (green). The length of the bar indicates the effect size of each genus. B) Association analysis between differential genus profile and H_2S concentration. The correlation was calculated using Spearman's rank correlation coefficient. The negative values (blue) indicate that the relative abundance of the genera is

negatively correlated with H_2S concentration, while positive values (red) indicate positive correlations. The length of the bar indicates the correlation coefficient. Asterisks indicate significant correlation. *, adjusted $p < 0.05$; **, adjusted $p < 0.01$; ***, adjusted $p < 0.001$.

Alterations of microbial functions in high H_2S group

In order to investigate the alterations of the microbial functions between high and low H_2S groups, KEGG metabolism pathways profiling was obtained using HUMAnN2 and Wilcoxon rank-sum test used to detected the significantly difference pathways. 18 differential KEGG metabolism pathways were detected in these two groups, of which nine pathways were enriched in the high H_2S group and the remaining nine were enriched in the low group (Figure 4A). It should be noted that the lipopolysaccharide (LPS) biosynthesis was enriched in the high H_2S group. LPS is an endotoxin derived from the outer membrane of Gram-negative bacteria, especially, the relative abundance of two main contributors of LPS biosynthesis, *F. periodonticum* and *P. nanceiensis,* were significantly different higher in the high H_2S group than the low groups (Figure 4C). Moreover, the relative abundance of *F. periodonticum,* a main contributor of sulfur metabolism was also significantly higher in the high H_2S group than the low groups (Figure 4B).

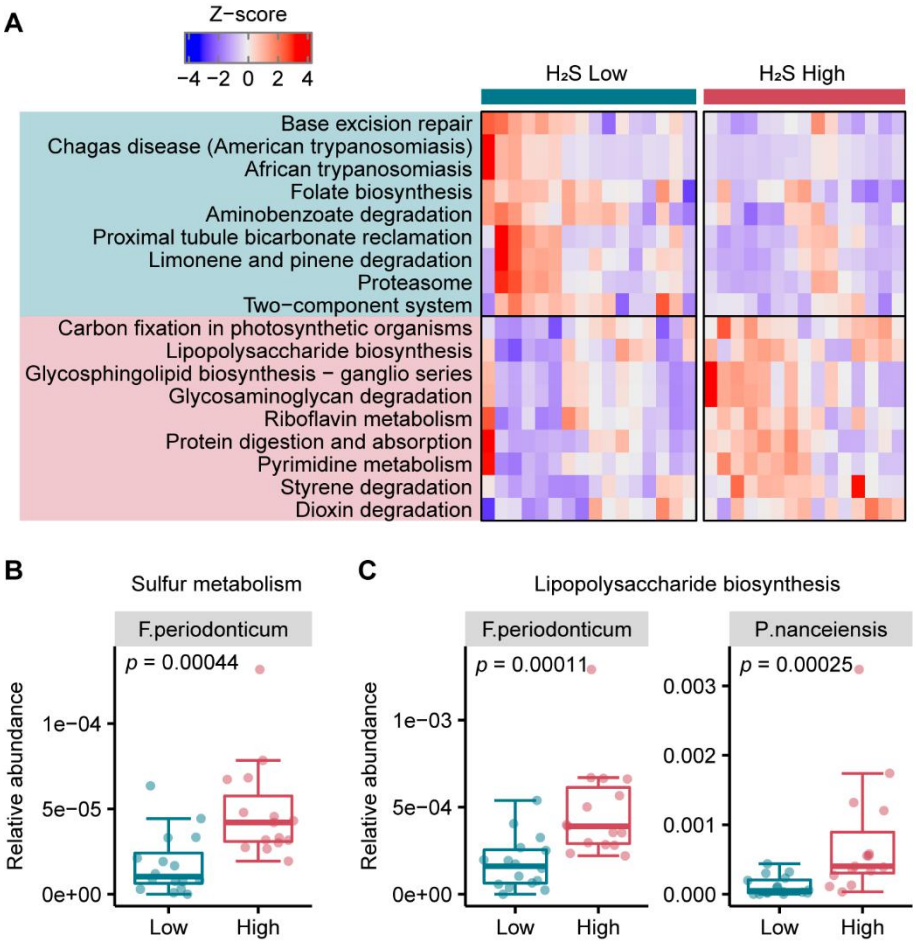

Figure 4 Differential microbial functions between high and low H_2S groups. A) Heatmap of 18 differential KEGG metabolism pathways detected by HUMAnN2. The relative abundances of pathways were normalized by row z-score. Pathways with green background were enriched in the low H_2S group, while red background were enriched in the high H_2S group. B) Contribution of *F.*

periodonticum to sulfur metabolism between two groups. C) Contributions of *F. periodonticum* and *P. nanceiensis* to lipopolysaccharide biosynthesis between two groups. Wilcoxon rank-sum test was used to determine significance.

Diversity analysis of *F. periodonticum* and *P. nanceiensis*

There were significant differences in function between the two bacteria, and in order to analyze in depth whether the abundance and diversity varied significantly at the strain level, the coverage depth and width of the samples were analyzed by inStrain software, and dendrogram tree was constructed through the ANI of the strains between the samples. The coverage breadth and depth of samples to reference *F. periodonticum* and *P. nanceiensis* were significantly different in two groups (Figure 5A-B; Figure S2A-B). The values were lower in both coverage breadth and depth in the H_2S low group, indicating that the abundance of *F. periodonticum* is difference in two groups. With analyzing the genomic data of *F. periodonticum*, the dendrogram tree of all samples based on popANI with a higher ANI value indicates that samples in the tow group have higher similarity (Figure 5C). Samples in H_2S high groups have a higher ANI value, and most samples cluster in the dendrogram tree, indicating that higher H_2S samples have higher similarity in the genome sequence and are distinguishable from samples in the low H_2S group. However, such obvious result is not revealed in the genomic data of *P. nanceiensis* (Figure S2C).

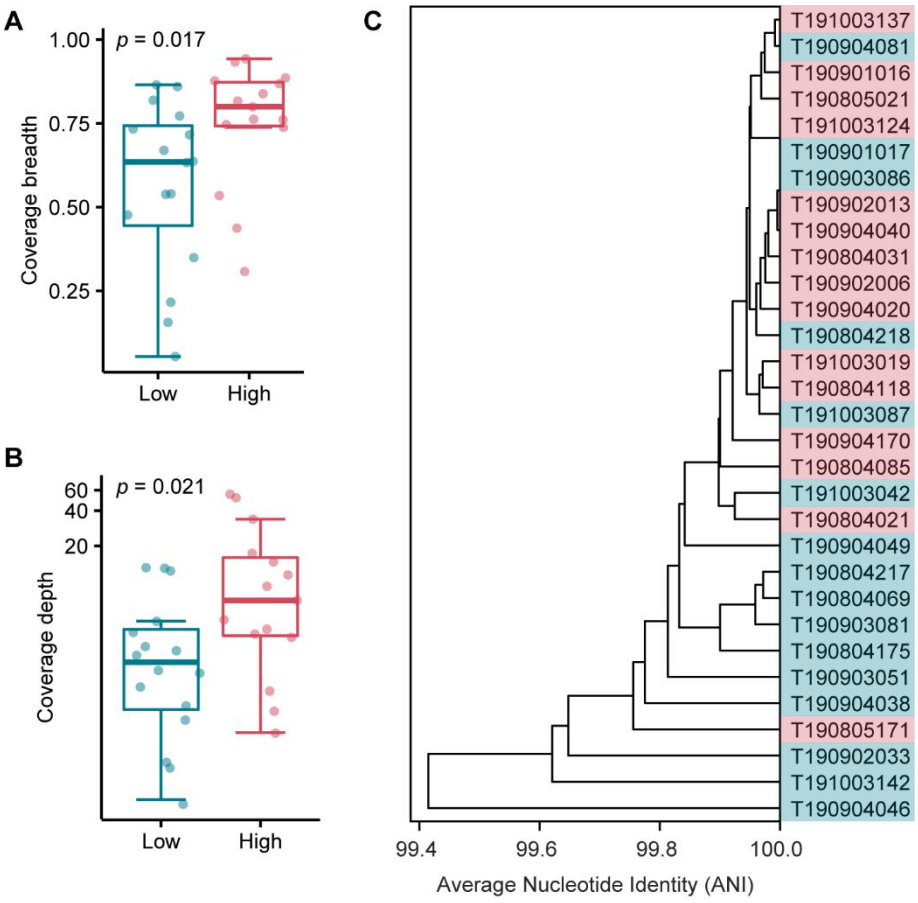

Figure 5 *F. periodonticum* diversity in the H_2S oral metagenomics cohort. A) Coverage breadth of samples to reference *F. periodonticum.* B) Coverage depth of samples to reference *F. periodonticum.* C) Dendrogram tree of all samples based on popANI.

Discussion

The high H_2S samples had a diverse microbial colonization than the low H_2S samples and may caused by altered bacterial acquisition. Many oral bacteria can produce H_2S and their proportions were generally highly correlated with H_2S in mouth air. Studies have shown that the microbiota of the subjects in the H_2S group were characterized by lower proportions of *Streptococcus* and *Granulicatella* and higher proportions of *Leptotrichia*, *Peptostreptococcus*, *Eubacterium*, and *Fusobacterium* compared with the no-odor group, while *Neisseria* and *Porphyromonas* were detected in higher proportions in the H_2S group too [30]. We also find that 8 genera (*Porphyromonas*, *Prevotella*, *Peptostreptococcus*, *Parvimonas*, *Bulleidia*, *Fusobacterium*, *Aggregatibacter*, *Moraxella*) were enriched in the high H_2S group, while *Streptococcus* and *Leptotrichia* were enriched in the low H_2S group.

The results of the current study revealed that the H_2S concentration of oral malodor was strongly associated with the composition of oral microorganism. It is known that 2 species of *Peptostreptocoeeus* (*P. magnus* and *P. micros*) and 5 species of *Fusobaeterium (F. nueleatum subsp. fusiforme*, *F. nueleatum subsp. nucleatum*, *F. nueleatum subsp. polymorphum*, *F. nueleatum subsp. vincentii* and *F. periodonticum*) could produce high amounts of hydrogen sulfide from glutathione [31]. Moreover, study has shown that *Prevotella tannerae* and *Porphyromonas* spp. could produce hydrogen sulfide [32]. In the current study, All of the above genera are enriched in the high H_2S group. Meanwhile, the correlation between the relative abundance of genera *Peptostreptocoeeus* and H_2S concentration is positive and the strongest (Figure 3B). The increase of microbial abundance and metabolism of some sulfide production are the main causes of bad breath, and halitosis-related microorganisms have certain associations with caries and periodontal disease, which improve the oral microenvironment from the perspective of inhibiting bad breath-related microorganisms.

The most of the enriched microorganisms enriched in people with high H_2S are associated with oral diseases such as caries and periodontal diseases, indicating that the diseases associated with oral microbes are not independent of each other and have some associations between some oral diseases. In the H_2S high group, *Prevotella* is reported as a pathogenic bacteria that associated with adult periodontitis, acute necrotizing ulcerative gingivitis, pregnancy gingivitis [33]. Heparin enzymes produced by *Prevotella* can degrade heparin effectively, which may improve the epithelial penetration by degrading heparin sulfate located in the cell gap, connecting the epithelial basement membrane, and allowing further microbial toxic factors to enter [34]. *Bulleidia* is associated with periodontits, and Down et al. isolated *Bulleidia Kloesel* from the mouth of patients with periodontitis and alveolar abscess [35]. *Parvimonas* is prevalent in subgingival biofilms in patients with chronic and aggressive periodontitis, and the presence of *Parvimonas* in root canal infection has been confirmed [36]. *Porphyromonas* is the major pathogen leading to the development of chronic periodontitis. It produces a variety of toxicity factors such as LPS (an important component of bacterial outer membrane), surface protein HA associated with bacterial adhesion to host cells and red blood cell coagulations, extracellular vesicles rich in microbial components, toxins, and toxicity factors, guivivin, Toll receptors, etc, involved in pro-inflammatory response and regulating the immune response of host cells.

Conclusions

The results of 16S show that the amount of sulfide in the oral malodor, especially the H_2S content, is significantly associated with the composition and abundance of microorganisms in the oral cavity. The increase of microbial abundance and metabolism of some sulfide products are the main causes of halitosis. The H_2S content, is significantly associated with the composition and abundance of microorganisms in the oral cavity. The increase of microbial abundance and metabolism of some sulfide products are the main causes of halitosis. The most of the enriched microorganisms enriched in people with high H_2S are associated with oral diseases such as caries and periodontal diseases, indicating that the diseases associated with oral microbes are not independent of each other and have some associations between some oral diseases.

Abbreviations

LPS: Lipopolysaccharide; KEGG: ;EOH: Extra-oral halitosis; IOH: Intra-oral halitosis; FTND: Fagerström Test for Nicotine Dependence; VSCs: Volatile sulfur compounds; CS: Current smokers; NS: Non-smokers; OTUs: Operational Taxonomy Units; ssCir DNA: Single strand circle DNA; NCBI: National Center for Biotechnology Information; PCoA: Principal coordinate analysis.

Authors' contributions

Conception and design: QG, BKZ and XML. Acquisition of data: NL, QLM and WZ. Processing of specimens and generation of data: YFH, Analysis and interpretation of data: JRJ and YFH; Drafting or revising of manuscript: JRJ and YFH; Final approval of manuscript: JRJ, YFH, QG and XML. JRJ, YFH and QG have access to all study data and takes responsibility for the data integrity and accuracy. All authors read and approved the fnal manuscript.

Funding

This work was supported by the Research Foundation of China Tobacco Company under Grant (110201901021 (JY-08)).

Availability of data and materials

The data are all published. The data that support the fndings of this study are available on request from the corresponding author.

Ethics approval and consent to participate

This study was approved by the Biomedical Ethics Committee of the Joint Institute of Tobacco and Health, Yunnan Academy of Tobacco Science, No. 41 Keyi Road, Kunming, China (approval number: 2020-001), and all participants provided written informed consent.

Consent for publication

There are no individual person identifers in this manuscript. Consent for publication was not sought.

Competing interests

All authors declare they have no competing interests.

Author details

[1]Joint Institute of Tobacco and Health, Yunnan Academy of Tobacco Science, No. 41 Keyi Road, Kunming 650106, China. [2]School of Pharmaceutical Science &Yunnan Key Laboratory of Pharmacology for Natural Products, Kunming Medical University, No. 1168 West Chunrong Road, Kunming 650504, China

References

1. Tangerman A. Halitosis in medicine: a review. International dental journal. 2002; 52(S5P1): 201-206.

2. Kleinberg I, Codipilly M. Modeling of the oral malodor system and methods of analysis. Quintessence international. 1999; 30.

3. Mok SF, Chinna K, Cheah YK, et al. The oral microbiome community variations associated with normal, potentially malignant disorders and malignant lesions of the oral cavity. The Malaysian journal of pathology. 2017; 39: 1-15.

4. Hampelska K, Jaworska MM, Babalska ZŁ, et al. The role of oral microbiota in intra-oral halitosis. Journal of clinical medicine. 2020; 9: 2484.

5. Haraszthy VI, Zambon JJ, Sreenivasan PK, et al. Identification of oral bacterial species associated with halitosis. The Journal of the American Dental Association. 2007; 138: 1113-1120.

6. Awano S, Gohara K, Kurihara E, et al. The relationship between the presence of periodontopathogenic bacteria in saliva and halitosis. International dental journal. 2002; 52: 212-216.

7. Tetz G, Vikina D, Brown S, et al. Draft Genome Sequence of Streptococcus halitosis sp. nov., Isolated from the Dorsal Surface of the Tongue of a Patient with Halitosis. Microbiology resource announcements. 2019; 8: e01704-18.

8. Tonzetich J. Production and origin of oral malodor: a review of mechanisms and methods of analysis. Journal of periodontology. 1977; 48: 13-20.

9. Rizzo AA. The possible role of hydrogen sulfide in human periodontal disease. I. Hydrogen sulfide production in periodontal pockets. Periodontics. 1967; 5: 233-236.

10. Persson S. Hydrogen sulfide and methyl mercaptan in periodontal pockets. Oral microbiology and immunology. 1992; 7: 378-379.

11. Persson S. Hydrogen sulfide and methyl mercaptan in periodontal pockets. Oral microbiology and immunology. 1992; 7: 378-379.

12. Morhart RE, Mata LJ, Sinskey AJ, et al. A microbiological and biochemical study of gingival crevice debris obtained from Guatemalan Mayan Indians. Journal of periodontology. 1970; 41: 644-649.

13. Ng W, Tonzetich J. Effect of hydrogen sulfide and methyl mercaptan on the permeability of oral mucosa. Journal of dental research. 1984; 63: 994-997.

14. Yaegaki K, Qian W, Murata T, et al. Oral malodorous compound causes apoptosis and genomic DNA damage in human gingival fibroblasts. Journal of periodontal research. 2008; 43: 391-399.

15. Basic A, Blomqvist M, Dahlén G, et al. The proteins of *Fusobacterium* spp. involved in hydrogen sulfide production from L-cysteine. BMC microbiology. 2017; 17: 1-10.

16. Holben WE, Jansson JK, Chelm BK, et al. DNA probe method for the detection of specific

433 microorganisms in the soil bacterial community. Applied and environmental microbiology. 1988;
434 54: 703.
435 17. Peng Changwen, Yan Mei. Overview of Research Methods and Applications of Metagenomics.
436 Biology Teaching. 2009; 34: 7-8.
437 18. Ma Shuqi, Zhou Qin, Advances in oral microbiology and oral individualized medical
438 research.Journal of Clinical Stomatology. 2016; 32:633-635.
439 19. Porter SR, Scully C. Oral malodour (halitosis). Bmj. 2006; 333: 632-635.
440 20. Magoč T, Salzberg L. FLASH: fast length adjustment of short reads to improve genome
441 assemblies. Bioinformatics. 2011; 27: 2957-2963.
442 21. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature
443 methods. 2013; 10: 996-998.
444 22. Wang Q, Garrity GM, Tiedje JM, et al. Naive Bayesian classifier for rapid assignment of
445 rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology. 2007;
446 73: 5261.
447 23. DeSantis T Z, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene
448 database and workbench compatible with ARB. Applied and environmental microbiology. 2006;
449 72: 5069.
450 24. Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, et al. QIIME allows
451 analysis of high-throughput community sequencing data. MNature Methods. Nature Publishing
452 Group. 2010; 7 :335-6. http://dx.doi.org/10.1038/nmeth0510-335
453 25. Chen Y, Chen Y, Shi C, et al. SOAPnuke: a MapReduce acceleration-supported software for
454 integrated quality control and preprocessing of high-throughput sequencing data. Gigascience.
455 2018; 7: gix120.
456 26. Langmead B, Salzberg S L. Fast gapped-read alignment with Bowtie 2. Nature methods.
457 2012; 9: 357.
458 27. Franzosa E A, McIver L J, Rahnavard G, et al. Species-level functional profiling of
459 metagenomes and metatranscriptomes. Nature methods. 2018; 15: 962-968.
460 28. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF.
461 inStrain profiles population microdiversity from metagenomic data and sensitively detects shared
462 microbial strains. Nat Biotechnol. Springer US. 2021.
463 http://dx.doi.org/10.1038/s41587-020-00797-0
464 29. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation.
465 Genome biology. 2011; 12: 1-18.
466 30. Takeshita T, Suzuki N, Nakano Y, et al. Discrimination of the oral microbiota associated with
467 high hydrogen sulfide and methyl mercaptan production. Scientific reports. 2012; 2: 1-8.
468 31. Carlsson J, Larsen J T, Edlund M B. Peptostreptococcus micros has a uniquely high capacity
469 to form hydrogen sulfide from glutathione. Oral microbiology and immunology. 1993; 8: 42-45.
470 32. Basic A, Blomqvist S, Carlén A, et al. Estimation of bacterial hydrogen sulfide production in
471 vitro. Journal of oral microbiology. 2015; 7: 28166.
472 33. Tanaka S, Yoshida M, Murakami Y, et al. The relationship of *Prevotella intermedia*,
473 *Prevotella nigrescens* and *Prevotella melaninogenica* in the supragingival plaque of children,
474 caries and oral malodor. Journal of Clinical Pediatric Dentistry. 2008; 32: 195-200.
475 34. Watanabe M, Tsuda H, Yamada S, et al. Characterization of heparinase from an oral bacterium
476 *Prevotella heparinolytica*. The Journal of Biochemistry. 1998; 123: 283-288.

35. Downes J, Olsvik B, Hiom S J, et al. *Bulleidia extructa* gen. nov., sp. nov., isolated from the oral cavity. International journal of systematic and evolutionary microbiology. 2000; 50: 979-983.
36. Van Winkelhoff A J, Van Steenbergen T J M, De Graaff J. The role of black-pigmented Bacteroides in human oral infections. Journal of clinical periodontology. 1988; 15: 145-155.

Supplementary figure

Table S1 Baseline characteristics, dietary status and oral status of current smokers (CS) and nonsmokers (NS).

Characteristic	NS	CS	P value*
Age, mean (s.d.)	36.4 (10.0)	37.4 (8.7)	0.232
BMI, mean (s.d.)	21.9 (2.6)	22.2 (3.0)	0.500
Male, %	68.0	71.0	0.513
Antibiotic, %	11.3	12.7	0.684
Dietary status			
Diet, %			
Mixed meat and vegetables	76.7	73.6	
Preference for meat	10.0	15.1	0.308
Preference for vegetables	13.3	11.4	
Drinking, %	56.0	64.3	0.087
Oral status			
H_2S, mean (s.d.)	842.8 (984.2)	630.3 (854.4)	**0.003**
CH_3SH, mean (s.d.)	225.1 (359.5)	176.4 (324.7)	0.268
$(CH_3)_2S$, mean (s.d.)	55.5 (186.3)	97.5 (369.3)	0.601
Periodontitis, %	7.3	8.7	0.627
Caries, %	40.7	40.7	1.000
Bleeding gums	41.3	35.3	0.215
Halitosis	28.7	34.7	0.200

*p values was calculated by Wilcoxon rank-sum test (non-normal distribution) or one-way ANOVA test (normal distribution) for continuous variables, and Fisher's exact test (expected frequency < 5) or Chi-squared test (expected frequency > 5) for categorical variables.

Table S2 77 differential taxa between the high and low H_2S groups.

Level	Taxa	Mean abundance in high H_2S group (%)	Mean abundance in low H_2S group (%)	LDA score (log10)	*p* value	Enrichment
Plylum	*Actinobacteria*	3.28	6.22	4.17	1.95E-11	Low
	Bacteroidetes	30.11	24.01	4.46	2.89E-06	High
	GN02	0.18	0.10	2.79	4.97E-02	High
	SR1	0.98	0.25	3.54	5.24E-13	High
Class	*Actinobacteria*	2.50	5.52	4.17	1.95E-11	Low
	Bacteroidia	29.10	22.78	4.47	2.89E-06	High

	Bacilli	9.44	11.68	4.11	1.25E-02	Low
	Erysipelotrichi	0.29	0.15	2.84	5.88E-06	High
	BD1_5	0.18	0.10	2.79	4.97E-02	High
	Betaproteobacteria	14.30	18.55	4.26	2.11E-02	Low
	Epsilonproteobacteria	0.78	1.19	3.34	4.67E-05	Low
	Gammaproteobacteria	12.34	11.25	3.83	3.59E-02	High
Order	*Actinomycetales*	2.50	5.51	4.17	6.36E-13	Low
	Bacteroidales	29.10	22.78	4.47	2.89E-06	High
	Lactobacillales	8.54	10.56	4.07	9.52E-03	Low
	Erysipelotrichales	0.29	0.15	2.84	5.88E-06	High
	Rhodospirillales	0	3.88E-04	3.18	2.46E-02	Low
	Sphingomonadales	0	1.09E-04	2.96	4.50E-02	Low
	Burkholderiales	0.18	0.58	3.32	1.07E-05	Low
	Neisseriales	14.12	17.97	4.21	3.36E-02	Low
	Campylobacterales	0.78	1.19	3.34	4.67E-05	Low
	Aeromonadales	0	1.62E-04	2.66	1.36E-02	Low
	Enterobacteriales	0	0.84	3.62	1.55E-03	Low
	Pseudomonadales	0.81	0.20	3.49	1.14E-03	High
	CW040	0.51	0.42	2.68	1.96E-02	High
	I025	0.04	0.09	2.40	1.09E-02	Low
	RF39	0.07	0.03	2.35	1.50E-07	High
Family	*Actinomycetaceae*	0.26	0.47	3.05	3.90E-06	Low
	Corynebacteriaceae	0.07	0.23	2.95	1.12E-09	Low
	Micrococcaceae	2.17	4.80	4.12	2.74E-11	Low
	Prevotellaceae	19.42	15.72	4.42	1.58E-05	High
	Lactobacillaceae	0.00	0.03	2.05	6.52E-05	Low
	Streptococcaceae	7.73	9.66	4.05	1.03E-02	Low
	Lachnospiraceae	0.79	1.37	3.40	1.37E-02	Low
	Mogibacteriaceae	0.68	0.38	3.15	6.37E-06	High
	Peptostreptococcaceae	0.48	0.27	3.02	4.71E-05	High
	Tissierellaceae	0.29	0.11	3.00	1.32E-02	High
	Erysipelotrichaceae	0.29	0.15	2.84	5.88E-06	High
	Fusobacteriaceae	11.78	9.29	4.11	2.98E-03	High
	Rhodospirillaceae	0	3.88E-04	2.92	2.46E-02	Low
	Sphingomonadaceae	0	1.09E-04	3.00	4.50E-02	Low
	Alcaligenaceae	0	3.04E-04	2.60	4.50E-02	Low
	Burkholderiaceae	0.17	0.56	3.34	1.48E-05	Low
	Neisseriaceae	14.12	17.97	4.21	3.36E-02	Low
	Campylobacteraceae	0.78	1.19	3.34	4.34E-05	Low
	Aeromonadaceae	0	1.62E-04	2.68	1.36E-02	Low
	Enterobacteriaceae	3.59E-03	0.84	3.64	1.55E-03	Low
	Moraxellaceae	0.81	0.19	3.49	1.04E-04	High
	F16	0.26	0.20	2.46	1.27E-02	High

	Rs_045	0.04	0.09	2.40	1.09E-02	Low
Genus	*Actinomyces*	0.26	0.46	3.03	5.75E-06	Low
	Corynebacterium	0.07	0.23	2.95	1.12E-09	Low
	Nesterenkonia	1.89E-05	1.49E-04	2.39	3.01E-02	Low
	Rothia	2.17	4.80	4.12	2.74E-11	Low
	Porphyromonas	5.73	5.17	3.57	3.55E-02	High
	Tannerella	0.09	0.14	2.37	2.68E-03	Low
	Prevotella	22.96	17.20	4.42	1.58E-05	High
	Lactobacillus	3.89E-03	0.02	2.06	1.01E-04	Low
	Streptococcus	7.73	9.65	4.05	1.04E-02	Low
	Oribacterium	0.31	0.84	3.36	2.69E-03	Low
	Peptostreptococcus	0.30	0.07	3.07	2.34E-16	High
	Parvimonas	0.25	0.08	3.00	1.43E-04	High
	Schwartzia	0.02	0.05	2.07	1.09E-02	Low
	Selenomonas	0.21	0.43	3.04	2.51E-06	Low
	Bulleidia	0.21	0.11	2.71	6.92E-08	High
	Fusobacterium	11.78	9.29	4.11	2.98E-03	High
	Leptotrichia	2.31	3.31	3.69	4.98E-02	Low
	Azospirillum	0	3.88E-04	3.09	2.46E-02	Low
	Sphingomonas	0	1.09E-04	3.03	4.50E-02	Low
	Sutterella	0	3.04E-04	2.59	4.50E-02	Low
	Burkholderia	0	3.43E-04	2.08	1.17E-02	Low
	Lautropia	0.17	0.56	3.33	1.54E-05	Low
	Campylobacter	0.78	1.19	3.34	4.34E-05	Low
	Klebsiella	6.67E-04	0.03	2.08	1.06E-02	Low
	Serratia	2.60E-03	0.04	2.28	8.38E-03	Low
	Aggregatibacter	2.29	1.51	3.61	1.70E-04	High
	Moraxella	0.80	0.16	3.51	1.77E-07	High

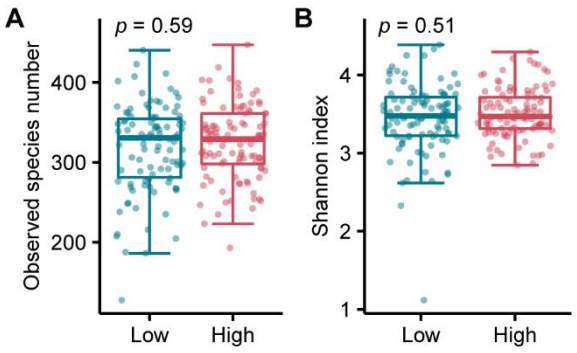

Figure S1 Alpha diversity of high and low H_2S groups. A) Box plot of distribution of richness at the OTU level. B) Box plot of distribution evenness at the OTU level.

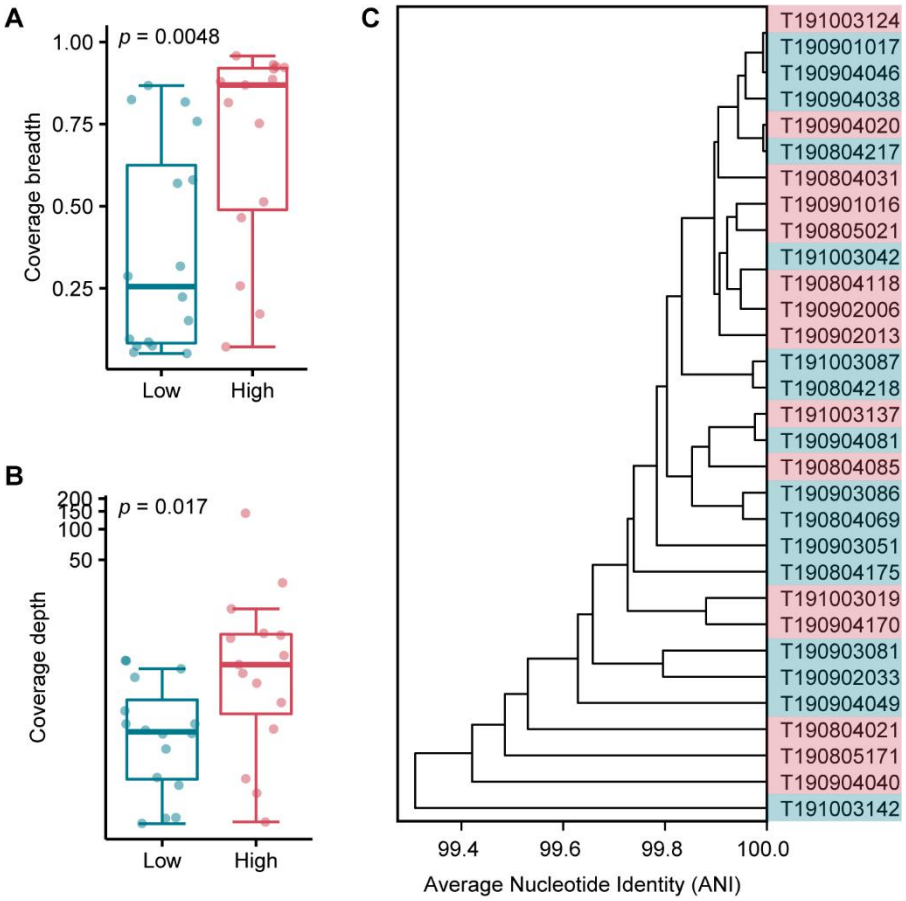

Figure S2 *P. nanceiensis* diversity in H_2S oral metagenomics cohort. A) Coverage breadth of samples to reference *P. nanceiensis.* B) Coverage depth of samples to reference *P. nanceiensis.* C) Dendrogram tree of all samples based on popANI.