Sputum Microbiome Composition in Patients With Squamous Cell Lung Carcinoma

E. D. Baranova
Kemerovo State University
V. G. Druzhinin
Kemerovo State University
L. V. Matskova (liudmila.matskova@ki.se)
Kant Baltic Federal University
P. S. Demenkov
Institute of Cytology and Genetics SB RAS
V. P . Volobaev
Center for Genetics and Life Science, Sirius University of Science and Technology
V. I. Minina
Federal Research Center of Coal and Coal Chemistry of Siberian Branch of the Russian Academy of Sciences
A. V. Larionov
Kemerovo State University
V. A. Titov
Kemerovo Regional Oncology Center

Research Article

Keywords: lung cancer, squamous cell lung carcinoma, sputum microbiome, taxonomic composition

Posted Date: October 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-902707/v1

Abstract

Recent findings indicate that the microbiome can have a significant impact on the development of lung cancer by inducing inflammatory responses, causing dysbiosis and generating genome damage. The aim of this study was to search for bacterial markers of squamous cell carcinoma (LUSC). In the study, the taxonomic composition of the sputum microbiome of 40 men with untreated LUSC was compared with 40 healthy controls. Next Generation sequencing of bacterial 16S rRNA genes was used to determine the taxonomic composition of the respiratory microbiome. There was no differences in alpha diversity between the LUSC and control groups. Meanwhile, differences in the structure of bacterial communities (β diversity) among patients and controls differed significantly in sputum samples (pseudo-F = 1.65; p = 0.026). Only *Streptococcus*, *Bacillus*, *Gemella* and *Haemophilus* were found to be significantly increased in patients with LUSC compared to the control subjects, while 19 bacterial genera were significantly reduced, indicating a decrease in beta diversity in the microbiome of patients with LUSC. From our study, *Streptococcus* (*Streptococcus agalactiae*) emerges as the most likely LUSC biomarker, but more research is needed to confirm this assumption.

Introduction

Lung cancer or bronchogenic carcinoma refers to tumors that arise in the lung parenchyma or bronchi. Lung cancer (LC) is diagnosed in approximately 1.2 million people worldwide each year, and more than 1 million people die from it during this period [1]. Although all forms of lung cancer originate from epithelial cells of the airway mucosa, the current pathomorphological classification of LC includes several different variants of this disease [2]. LC is usually divided into small cell lung cancer and non-small cell lung cancer (NSLC), which accounts for 85% of all bronchogenic tumors [3]. NSLC is further subdivided into large cell lung cancer, adenocarcinoma of the lung (AD), and lung squamous cell carcinoma (LUSC). LUSC accounts for about 30% of all NSLC cases. It is associated with a poor clinical prognosis, and no targeted therapy is as yet available [4].

The mortality rate from LUSC remains high, partly due to the lack of early detection of diagnostic biomarkers, including metagenomic ones. However, the search for bacteria associated with the risk of developing LC has intensified markedly in recent years, especially using NGS sequencing technology [5, 6]. Several previous studies have shown that changes in the number of specific microbiota taxa in bronchoalveolar lavage fluid (BALF), lung tissue, and saliva samples may be associated with LC, but these studies are largely inconsistent as far as specific taxa are concerned [7 - 15]. Another source of information on the composition of the respiratory tract microbiota is sputum, which has been poorly studied in patients with LC in general and particularly in those with LUSC diagnosis [16 - 19]. However, the use of sputum, while not reflecting the microbiome of any particular part of the respiratory tract, can be very useful as a metagenomic biomarker, since it does not require complex invasive procedures.

It is known that different pathological types of lung cancer are characterized by different biological patterns, molecular markers, and treatment strategies [20]. However, only very few studies have so far

examined the relationship between the respiratory tract microbiome and individual histological forms of lung cancer. Recent findings in this area suggest differences in the taxonomic composition of the respiratory microbiota between patients with AD or LUSC [21].

In this report we present, for the first time, the results of a comparison of the taxonomic composition of the sputum microbiome in two cohorts of residents of the Kuzbass region of Western Siberia: patients with LUSC and control donors.

Methods

Cohort information. The composition of the sputum bacterial microbiome was studied in 40 patients with newly-diagnosed LUSC (male only, average age 59.9±6.9 years) who were admitted to the Kemerovo Regional Oncology Center (Kemerovo, Russian Federation) and 40 healthy male donors, residents of Kemerovo (average age 54.0±5.3 years). There were differences in mean age between patients and control ($p < 0.05$). Among LUSC patients, 75% were active smokers, among the control subjects 55%. In addition, for each patient, the stage of the disease was determined in accordance with the TNM classification [22]. In accordance with this, 18 patients (45%) had stage I-II, and 22 patients (55%) stage III-IV of the disease. A summary of the information regarding LUSC and control subjects is shown in Table 1. A questionnaire was filled out for each survey participant, containing information on place and date of birth, occupation, exposure to occupational hazards, health status, dietary habits, and intake of medications (use of antibiotics three months prior to the study), X-ray procedures, smoking and drinking status.

Table 1. Characteristics of the study cohorts

Variables	Squamous cell lung carcinoma, n=40	Healthy Kemerovo residents (control), n=40
Age (years) (mean ± SD)	59.9 ± 6.9	54.0 ± 5.3
Smoking status (%):		
Smokers	75.0	55.0
Non-smokers	25.0	45.0
TNM* (%):		-
I, II	45	
III, IV	55	
Distant metastasis (%):		-
Yes	5	
No	95	
Central lung cancer	30	-
Peripheral lung cancer	42.5	
Mixed lung cancer	7.5	
Bronchial cancer	20	

For patients with LUSC, the results of clinical and histological analyses were additionally taken into account.

Ethics statement. All procedures undertaken were in accordance with the ethical standards of the Helsinki Declaration (1964 and amended 2008) of the World Medical Association. All participants (patients and control subjects) were informed about the aim, methodology and possible risks of the study; informed consent was signed by each donor. The design of this study was approved by the Ethics Committee of the Kemerovo State University.

Sample collection, process and storage. To analyze the composition of the microbiome of the respiratory tract, sputum samples obtained from LC patients and control subjects were used. The sputum and whole blood samples from patients were obtained prior to all diagnostic or therapeutic procedures. Sputum was collected on the first day of hospitalization. Before sputum collection, patients were asked to rinse their mouth. Sputum samples were collected non-invasively through participant-induced coughing (i.e., without induction) and represent the oropharyngeal secretion. The obtained samples were immediately placed in sterile plastic vials and frozen (−20°C). Frozen samples were transported to the laboratory and stored at −80°C.

DNA extraction, 16S rRNA gene amplification and sequencing. DNA was extracted from each sample using FastDNA Spin Kit For Soil (MP Biomedicals) based on the manufacturer's recommendation. Eighty 16S rRNA gene amplicon libraries were prepared by PCR amplification of an approximate 467 bp region within the hypervariable (V3-V4) region of the 16S rRNA gene in bacteria, from 50 ng of each of the extracted and purified DNA from sputum samples, respectively, according to the Illumina 16S metagenomic sequencing library protocol. PCR was initially performed with broad-spectrum 16S rRNA primers (forward primer: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3', and reverse primer: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3'), using BioMaster Hi-Fi LR 2X ReadyMix DNA polymerase (BiolabMix company, Novosibirsk, Russia). Cycle conditions were 94°C (3 min 30 s), then 25 cycles of 94°C (30 s), 55°C (30 s), 68°C (40 s), then a final extension of 68°C (5 min). Libraries were purified using Agencourt AMPure XP beads (Beckman Coulter, Bray, USA) according to the Illumina 16 S metagenomic sequencing library protocol. Dual indices and Illumina sequencing adapters from the Illumina Nextera XT index kits v2 B and C (Illumina, San Diego, USA) were added to the target amplicons in a second PCR step using BioMaster Hi-Fi LR 2Ч 2Ч ReadyMix DNA polymerase (BiolabMix company, Novosibirsk, Russia). Cycle conditions were 94°C (3 min 30 s), then 8 cycles of 94°C (30 s), 55°C (30 s), 68°C (40 s), then a final extension of 68°C (5 min). Libraries were again purified using Agencourt AMPure XP beads (Beckman Coulter, Bray, USA) according to the Illumina 16 S metagenomic sequencing library protocol. Sample PCR products were then pooled in equimolar amounts, purified using AMPure XP Beads (Beckman Coulter), and then quantified using a fluorometer (Quantus Fluorometer dsDNA (Promega, Madison, WI, USA). Molarity was then brought to 4 nM, the libraries were denatured, and then diluted to a final concentration of 8 pM with a 10% PhiX spike buffer for sequencing on the Illumina MiSeq platform [23].

Taxonomy quantification using 16S rRNA gene sequences and statistical methods. The processing of the resulting sequence data was conducted using the program QIIME2 [24]. A quality check was carried out and a sequence library was generated. The sequences were combined into operational taxonomic units (OTUs) based on a 99% nucleotide similarity threshold using the Greengenes reference sequences library (versions 13–8) and SILVA (version 132), followed by the removal of singletons (OTUs containing only one sequence). The total diversity of prokaryotic communities (alpha diversity) of sputum was estimated by the number of allocated OTU (analogue of species richness) and Shannon indices ($H = \Sigma pi \ln pi$, pi – part of i-sh species in community). When calculating sample diversity indices, 328 sequences were normalized (the minimum number of received sequences per sample). The variation in the structure of the bacterial community of different samples (beta diversity) was analyzed using Bray-Curtis dissimilarity metrics [25] – a method common in microbial ecology that estimates the difference between communities based on the abundance relationships of the taxa present in the samples.

In addition, to assess the significance of differences in the relative percentage of individual bacterial taxa in sputum, the Mann-Whitney U test was used. Spearman's correlation coefficient was used to calculate correlations. The False Discovery Rate (FDR) correction was used to assess the significance of differences in the relative percentages of individual bacterial taxa taking into account multiple comparisons. Calculations were performed using the software package STATISTICA.10, Statsoft, USA.

Results

In the current study, the composition of the sputum bacterial microbiome was studied in 40 patients with LUSC and 40 healthy male donors, residents of Kemerovo. We have used a large scale approach to sequence the 16S rRNA V3–V6 region of the bacterial genomes purified from the sputum samples from the compared groups in the study.

For the LUSC group, the average number of analyzed sequences was 76776 (9694, 181146). For the healthy control group, the average number of analyzed sequences was 72613 (12537, 160232). We identified a total of 11 bacterial phyla with relative frequencies above 0.1%. The prevailing phyla in our dataset were *Firmicutes*, *Bacteroidetes*, *Actinobacteria* and *Proteobacteria* (Fig. 1), as could be expected from previous studies [16,17, 26, 27].

Regarding alpha diversity, neither the number of allocated OTUs nor the Shannon indices, showed significant differences between LUSC and control groups of people. Overall, the bacterial communities were fairly diverse in the two groups as indicated by Shannon index at genus level (5.267 in LUSC vs 5.439 in control groups). This suggests that any changes in the sputum microbiome in the LUSC malignancy are not large-scale shifts in the bacterial community.

Differences in the structure of bacterial communities in sputum samples of lung cancer patients and healthy subjects are shown in Fig. 2. The PERMANOVA (Adonis) test using the difference matrix, constructed by the Bray-Curtis method, showed a significant difference in the prokaryotic communities in sputum from healthy subjects and patients with LUSC (pseudo-F = 1.65; p = 0.026).

A comparison of the frequencies of the major bacteria phyla in sputum revealed a significant increase in the representatives of *Firmicutes* in patients with LUSC compared to control subjects (56.77±15.29 vs 47.34±10.65 %, respectively; p = 0.004). In contrast, the other four bacterial types (*Bacteroidetes*, *Fusobacteria*, *TM7* and *Spirochaetes*), were more represented in the sputum of the control group than in the the patient group (Fig. 3).

In total, the average percentage of representatives of 67 genera and 32 species of bacteria in the sputum of patient and control groups was compared. The tables include only those genera and species for which there were reliable (taking into account the FDR procedure) differences between LUSC and control. Sequencing statistics are summarised in Table 2 (for 23 genera) and in Table 3 (for 17 species), together with the corresponding U-rank Mann-Whitney p values. As can be seen from these tables, there is considerable variation in the relative percentages of all genera and species represented.

Table 2
Average percentage abundance of genera present in «core» microbiome

Genus	Squamous cell lung cancer, n=40	Controls, n=40	P*
Streptococcus	36.26±20.02	18.93±10.43	0.00001
Prevotella (f.Prevotellfceae)	10.86±7.03	17.89±7.37	0.00003
Veillonella	6.6±7.45	11.17±6.22	0.00009
Anaerosinus	6.57±7.69	10.91±6.33	0.0003
Gemella	3.6±2.89	2.01±2.01	0.004
Bacillus	3.55±2.96	1.84±1.93	0.003
Haemophilus	2.2±10.32	0.13±0.4	0.003
Selenomonas	1.68±2.82	4.36±3.49	0.00003
Megasphaera	1.38±2.91	2.44±2.27	0.005
Streptobacillus	1.14±1.46	2.85±2.69	0.002
Atopobium	1.22±1.79	1.66±1.47	0.02
Leptotrichia	1.13±1.35	2.63±2.49	0.008
Treponema	0.43±0.76	0.73±1.01	0.002
Lachnoanaerobaculum	0.39±0.53	0.65±0.65	0.03
Porphiromonas	0.35±0.75	0.93±1.3	0.002
Parvimonas	0.42±0.83	0.83±1.06	0.002
Stomatobaculum	0.39±0.42	0.9±0.81	0.003
Vestibaculum	0.35±1.1	0.86±1.45	0.005
Catonella	0.09±0.32	0.14±0.25	0.03
Filifactor	0.06±0.16	0.21±0.33	0.003
Mycoplasma	0.04±0.09	0.28±0.72	0.02
Moriella	0.05±0.18	0.36±0.83	0.02
Cardiobacterium	0.02±0.1	0.04±0.1	0.03
Note: * - P Value lesser than FDR corrected P			

Table 3
Average percentage abundance of species present in «core» microbiome

Species	Squamous cell lung cancer, n=40	Controls, n=40	P*
Streptococcus agalactiae	35.47±20.19	19.11±10.06	0.00004
Anaerosinus glycerini	4.8±6.31	10.19±6.81	0.0003
Selenomonas bovis	1.44±2.62	4.36±4.46	0.00001
Prevotella histicola F0411	1.42±2.27	2.84±3.2	0.02
Atopobium rimae	1.29±1.77	1.68±1.46	0.03
Megasphaera Micronuciformis	1.23±2.88	2.39±2.34	0.001
Lachnoanaerobaculum orale	0.36±0.5	0.69±0.64	0.01
Vestibaculum illigatum	0.29±0.98	0.91±1.46	0.003
Rothia dentocariosa ATCC 17931	0.23±0.6	0.53±0.8	0.02
Prevotella sp.oral clone DO014	0.16±0.44	0.52±0.6	0.0007
Porphyromonas endodontalis	0.14±0.34	0.93±1.3	0.003
Prevotella intermedia	0.14±0.34	0.47±0.96	0.04
Moryella indoligenes	0.06±0.19	0.4±0.895	0.03
Prevotella nigrescens	0.09±0.22	0.43±1.11	0.02
Oribacterium Sinus	0.07±0.22	0.26±0.51	0.03
Leptotrichia sp. oral clone EI013	0.06±0.18	0.24±0.54	0.0008
Filifactor alocis ATCC 35896	0.06±0.16	0.21±0.33	0.007
Note: * - P Value lesser than FDR corrected P			

In the sputum of patients with LUSC, as compared to control subjects, there was a significant increase in the abundance of the following genera (by percentage): *Streptococcus* (36.26±20.02 vs 18.93±10.43; p = 0.00001); *Bacillus* (3.55±2.9 vs 1.84±1.93; p = 0.003); *Gemella* (3.6±2.89 vs 2.01±2.01; p = 0.004) and *Haemophilus* (1.27±8.07 vs 0.11±0.36; p = 0.003). At the same time, members of the 19 genera in Table 2 were significantly more represented in the microbiome of the control group in comparison to the patients with LUSC.

At the species level, only one bacterial species, *Streptococcus agalactiae*, was significantly higher in the sputum of patients compared to control subjects (35.47±20.19 vs 19.11±10.06; p = 0.00004). Representatives of the other 17 species were significantly more common in the microbiome of healthy controls in comparison to the patients with LUSC, as shown in Table. 3.

We found no specific association of any bacterial taxon in the sputum with the age of patients or control donors participating in the study.

The influence of smoking status on the microbiota composition in patients with LUSC and control subjects was studied separately. For patients, no significant difference was found in the bacterial genera or species in sputum between smokers and nonsmokers. Controls differing in smoking status revealed a significant difference in the occurrence of several genera and species in the sputum. Control group smokers (Fig. 4) had less *Neisseria* than non-smokers (0.56±1.16% vs 3.94±5.63%; p = 0.00006); *Fusobacterium* (1.4±1.55% vs 3.39±3.01%; p = 0.02); *Prevotella nigrescens* (0.35±1.38% vs 0.52±0.68%; p = 0.01) and *Peptostreptococcus Anaerobius* (0.04±0.1% vs 0.39±0.71%; p = 0.02). At the same time, control group smokers had more *Streptobacillus* in their sputum compared to nonsmokers (3.62±2.8% vs 1.92±2.28%; p = 0.03).

Comparison of the total composition of the microbiome in patients with different stages of LUSC (I-II and III-IV), as well as between subgroups with different localization of the primary tumor, revealed no differences between them.

Discussion

The respiratory tract microbiome is closely linked to the occurrence of lung diseases, including lung cancer. It has been previously confirmed that there are changes in the microecology of the lungs in patients with lung cancer compared to healthy subjects. In addition, the abundance of some bacterial species correlates with pathology, suggesting their potential use as microbial markers for the detection of lung cancer. However, until now, the composition of the lung microbiome in patients with different histological types of lung cancer has not been determined.

In this study, we examined the difference between spontaneous sputum samples from patients with LUSC and healthy controls. In general, the sputum microbiota in men with malignant lung tumors had a significant decrease in beta diversity, which is consistent with the results of previous studies [8, 16, 28, 29]. At the level of bacterial types, the most notable finding in our patients with LUSC was an abundance of *Firmicutes* to the detriment of *Proteobacteria*. The dominance of *Proteobacteria* in healthy lung microbiota was also detected earlier [13, 30].

In a pairwise comparison, representatives of four bacterial types (*Bacteroidetes*, *Fusobacteria*, *TM7* and *Spirochaetes*) and the 19 genera represented in Table 2 were significantly more represented in healthy control samples compared to samples from patients with LUSC. On the other hand, we found that *Streptococcus*, belonging to the *Firmicutes* type, was represented significantly more in patients and had the greatest effect on the difference between LUSC and control samples. Two more genera (*Bacillus* and *Gemella*) from the *Firmicutes* type, and one representative of *Proteobacteria* - *Haemophilus*, were more represented in the sputum of patients compared to controls. We believe that all four genera can be considered as potential bacterial biomarkers of LUSC.

An increased prevalence of *Streptococcus* in the sputum of patients with lung cancer has previously been reported in several publications [16, 17, 21]. However, no increases in the *Bacillus*, *Gemella* and *Haemophilus* bacterial types were reported previously. Indeed, a recent study using ddPCR found a significant increase in *Streptococcus* load ($p = 0.0042$ determined by the Mann - Whitney U test) in the sputum of seven patients with LUSC compared with ten control patients. [31]. Interestingly, at the same time, a significant increase in *Veillonella* was found in the sputum of the same patients in comparison to control participants. In our study, however, representatives of this bacterial genus, were more represented in controls than in the patients with LUSC (Table 2). Finally, the amount of *Haemophilus* in the sputum of patients and controls was alsmost equal, while in our cohort of patients this facultative anaerobe was significantly more common than in healthy donors. Another recent study of the respiratory microbiome (saliva and bronchial biopsy specimens) in 25 patients with central lung cancer from Spain [32] found a significant increase in *Streptococcus*, *Rothia*, *Gemella* and *Lactobacillus*, which partially agrees with our results (for *Streptococcus* and *Gemella*). Thus, it appears that *Streptococcus* is a major bacterial marker in the airway associated with lung cancer, although there may be mixed results for different histopathological types of this bronchogenic tumor and at different stages of the disease. For example, Z. Ran and colleagues reported that *Streptococcus* and *Neisseria* were predominant in the sputum of patients with adenocarcinoma, while *Streptococcus* and then *Veillonella* predominated in patients with LUSC, and *Neisseria* and then *Streptococcus* in the small cell lung cancer group [21].

Comparison of sputum microbiome composition in subgroups of patients with LUSC differing in TNM disease stages, central or peripheral tumor localization, and smoking status, revealed no significant differences in bacterial content. However, in the group of healthy donors, we found a clear decrease in *Neisseria* in the sputum of smokers compared to nonsmokers (Figure 4), which is consistent with previously published results for the upper respiratory tract [33]. It should be noted that the effect of smoking on the sputum microbiota remains unclear, according to the latest published data [34] and this issue requires further study.

Streptococcus agalactiae, as shown in Table 3, was the only bacterial species significantly increased in patient sputum, according to sequencing data and analysis of two databases (Greengenes and SILVA). In the future, we plan to confirm the presence of this bacterial species using ddPCR. Previous studies have reported the prevalence of *Streptococcus viridans* in the sputum of patients with lung cancer [17].

Streptococcus agalactiae is the most frequently represented species in the sputum of both LUSC patients and controls, and its significant increase in LC patients suggests its utility as a possible biomarker for this cancer, similar to *Streptococcus gallolyticus subsp*. in colorectal carcinoma [35]. *Streptococcus agalactiae* (also known as GBS) is an important opportunistic species that can cause pneumonia, sepsis and meningitis in newborns and in immunocompromised patients [36, 37]. Cases of invasive GBS infections are frequently reported in the elderly and immunocompromised adults, including patients with diabetes mellitus, alcoholism and cancer [38]. In the respiratory tract, GBS sometimes contributes to community-acquired pneumonia and empyema in adults [39]. When GBS causes a pulmonary infection, it is usually defined as part of polymicrobial pneumonia [40]. GBS bacteria effectively attach to pulmonary

epithelial cells and are capable of invasion. This is initiated by attachment to extracellular matrix molecules such as agglutinin, fibronectin, fibrinogen and laminin, which facilitates their attachment to host cell surface proteins, such as integrins. Thus, the invasive potential of GBS is influenced by changes in the surface proteome of host cells, which can be caused by various lung pathologies [41]. The molecular mechanisms of cytopathology caused by GBS bacteria in patients are currently being intensively studied. It was shown that GBS induces generation of reactive oxygen species (ROS) and loss of mitochondrial membrane potential [42]. In human endothelial cells, ROS species are generated via the NADPH oxidase pathway, which is accompanied by cytoskeletal reorganisation through the PI3K/Akt pathway, and is generally associated with pathogen penetration, providing evidence for the involvement of oxidative stress in the pathogenesis associated with *S. agalactiae* [43].

Conclusion

In this report, we used mass parallel sequencing of bacterial 16S ribosomal genes to compare the taxonomic composition of the sputum microbiome of patients with LUSC and healthy donors. It was found that the bacterial taxonomic groups detected in the microbiome of patients were significantly different compared to controls. The sputum of patients with LUSC contains significantly more members of the genera *Streptococcus*, *Bacillus*, *Gemella* and *Haemophilus*. *Streptococcus* (*Streptococcus agalactiae*) is the most likely LUSC biomarker from this list, but more research is still required to validate this assumption.

In order to consider these bacteria as biomarkers for the risk of LUSC development, it is necessary to have information about their population dynamics in the respiratory microbiome from health to lung malignancy. This can be solved, for example, by forming a database of the respiratory microbiome in healthy individuals over a long period of time. Another possible and more accessible approach is to study the composition of the microbiome in the sputum of patients with chronic inflammatory diseases of the lungs. A recent study showed increased numbers of *Streptococci* in airway microbiome samples from patients with idiopathic pulmonary fibrosis and COPD. Thus, future studies to establish the role of bacteria as biomarkers of lung cancer should examine the composition of the sputum microbiome in these and other non-malignant lung diseases.

Declarations

Funding

This study was funded by Russian Science Foundation Grant No. 18-14-00022p.

Acknowledgements

The authors wish to thank the physicians and staff of the Kemerovo Regional Oncology Center, all of the surveyed individuals who voluntarily participated in this study and the employees of the Kemerovo State

University and the Institute of Human Ecology who participated in the organization and conducting of this research.

References

1. Cheng, T.Y. et al. The International Epidemiology of Lung Cancer: Latest Trends, Disparities, and Tumor Characteristics. *J. Thorac. Oncol.* **11**, 1653-1671. https://doi.org/10.1016/j.jtho.2016.05.021 (2016).

2. Tsao, M.S., Yoon, J.Y. The eighth TNM classification for lung cancer-What is next? *Lung Cancer.* **121**, 97-98. https://doi.org/10.1016/j.lungcan.2018.04.018 (2018).

3. Molina, J.R. et al. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc*. **83**, 584-594. https://doi.org/10.4065/83.5.584 (2008).

4. Man, J. et al. Screening and identification of key biomarkers in lung squamous cell carcinoma by bioinformatics analysis. *Oncol. Lett*. **18**, 5185-5196. https://doi.org/10.3892/ol.2019.10873 (2019).

5. Mao, Q. et al. Interplay between the lung microbiome and lung cancer. *Cancer Lett*. **415**, 40-48. https://doi.org/10.1016/j.canlet.2017.11.036 (2018).

6. Maddi, A. et al. The microbiome and lung cancer. *J. Thorac. Dis.* **11**, 280–291. https://doi.org/10.21037/jtd.2018.12.88 (2019).

7. Hasegawa, A. et al. Detection and identification of oral anaerobes in intraoperative bronchial fluids of patients with pulmonary carcinoma. *Microbiol. Immunol*. **58**, 375-381. https://doi.org/10.1111/1348-0421.12157 (2014).

8. Lee, S.H. et al. Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. *Lung Cancer*. **102**, 89-95. https://doi.org/10.1016/j.lungcan.2016.10.016 (2016).

9. Liu, H.X. et al. Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. *Int. J. Cancer*. **142**, 769-778. https://doi.org/10.1002/ijc.31098 (2018).

10. Tsay, J.J. et al. Airway Microbiota Is Associated with Upregulation of the PI3K Pathway in Lung Cancer. *Am. J. Respir. Crit. Care. Med.* **198**, 1188-1198. https://doi.org/10.1164/rccm.201710-2118OC (2018).

11. Peters, B.A. et al. The microbiome in lung cancer tissue and recurrence-free survival. Cancer Epidemio.l Biomarkers Prev. **28**,731-740. https://doi.org/10.1158/1055-9965.EPI-18-0966 (2019).

12. Zhang, W. et al. Salivary Microbial Dysbiosis is Associated with Systemic Inflammatory Markers and Predicted Oral Metabolites in Non-Small Cell Lung Cancer Patients. *J. Cancer*. **10**, 1651-62.

https://doi.org/10.7150/jca.28077 (2019).

13. Wang, K. et al. A preliminary study of microbiota diversity in saliva and bronchoalveolar lavage fluid from patients with primary bronchogenic carcinoma. Med. Sci. Monit. **25**, 2819-2834. https://doi.org/10.12659/MSM.915332 (2019).

14. Cheng, C. et al. Characterization of the lung microbiome and exploration of potential bacterial biomarkers for lung cancer. *Transl. Lung Cancer Res*. **9**, 693-704. https://doi.org/10.21037/tlcr-19-590 (2020).

15. Zhuo, M. et al. Characterization of Microbiota in Cancerous Lung and the Contralateral Non-Cancerous Lung Within Lung Cancer Patients. *Front. Oncol*. **10**,1584. https://doi.org/10.3389/fonc.2020.01584 (2020).

16. Hosgood, H.D. 3rd et al. The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. *Env.Mol. Mutagen.* **55**, 643–651. https://doi.org/10.1002/em.21878 (2014).

17. Cameron, S.J.S. et al. A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. *PLoS One*. **12**,e0177062. https://doi.org/10.1371/journal.pone.0177062 (2017).

18. Hosgood, H.D. 3rd et al. The respiratory tract microbiome and its relationship to lung cancer and environmental exposures found in rural china. *Environ. Mol. Mutagen*. **60**, 617-623. https://doi.org/10.1002/em.22291 (2019).

19. Druzhinin, V.G. et al. Taxonomic diversity of sputum microbiome in lung cancer patients and its relationship with chromosomal aberrations in blood lymphocytes. *Sci. Rep*. **10**(1):9681. https://doi.org/10.1038/s41598-020-66654-x (2020).

20. Herbst, R.S. et al. Lung cancer. *N. Engl. J. Med*. **359**, 1367-1380. https:// doi. org/0.1056 /NEJMra0802714 (2008).

21. Ran, Z. et al. [Analysis of Pulmonary Microbial Diversity in Patients with Advanced Lung Cancer Based on High-throughput Sequencing Technology]. *Zhongguo Fei Ai Za Zhi*. 23,1031-1038. Chinese. https://doi.org/10.3779/j.issn.1009-3419.2020.103.16 (2020).

22. Goldstraw, P. New staging system: how does it affect our practice? *J. Clin. Oncol.* **31**, 984–991. https://doi.org/10.1200/JCO.2012.42.7922 (2013).

23. Caporaso, J.G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, **7**, 335–336. https://doi.org/10.1038/nmeth.f.303 (2010).

24. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.,* **37**, 852–857. https://doi.org/10.1038/s41587-019-0209-9 (2019).

25. Sorenson, T. A method of establishing groups of equal amplitudes in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish Commons. *Biologiske Skrifter*, **5**, 1–34. (1948).

26. Huang D, et al. The characterization of lung microbiome in lung cancer patients with different clinicopathology. *Am. J. Cancer Res*. **9**, 2047-2063. (2019).

27. Sommariva, M. et al. The lung microbiota: role in maintaining pulmonary immune homeostasis and its implications in cancer development and therapy. *Cell Mol. Life Sci*. **77**, 2739-2749. https://doi.org/10.1007/s00018-020-03452-8 (2020).

28. Greathouse, K.L, et al. Interaction between the microbiome and TP53 in human lung cancer. *Genome Biol*. **19**, 123. https://doi.org/0.1186/s13059-018-1501-6 (2018).

29. Xu, N. et al. (2020) Microbiota dysbiosis in lung cancer: evidence of association and potential mechanisms. *Transl. Lung Cancer Res*., **9**, 1554-1568. https://doi.org/10.21037/tlcr-20-156 (2020).

30. Segal, L.N. et al. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat .Microbiol.* **1**, 16031. https://doi.org/10. 1038/ nmicrobiol .2016.31 (2016).

31. Leng, Q. et al. Microbiota Biomarkers for Lung Cancer. *Diagnostics (Basel)*. **11**, 407. https://doi.org/10.3390/diagnostics11030407 (2021).

32. Bello, S. et al. Core Microbiota in Central Lung Cancer With Streptococcal Enrichment as a Possible Diagnostic Marker. *Arch. Bronconeumol*. (Engl Ed). **20**, 30192-30197. English, Spanish. https://doi.org/10.1016/j.arbres.2020.05.034 (2020).

33. Morris, A. et al. Microbiome Project. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am. J. Respir. Crit. Care Med*. **187**, 1067-75. https://doi.org/10.1164/rccm.201210-1913OC (2013).

34. Haldar, K. et al. The sputum microbiome is distinct between COPD and health, independent of smoking history. *Respir. Res*. **21**,183. https://doi.org/10.1186/s12931-020-01448-3 (2020).

35. Kumar, R. et al. Streptococcus gallolyticus subsp. gallolyticus promotes colorectal tumor development. *PLoS Pathog.***13**, e1006440. https://doi.org/10.1371/journal.ppat.1006440 (2017).

36. Manning, S.D. Molecular epidemiology of Streptococcus agalactiae (group B Streptococcus). *Front. Biosci*. **8**, s1-18. https://doi.org/10.2741/985 (2003).

37. Furfaro, L.L. et al. Perinatal Streptococcus agalactiae Epidemiology and Surveillance Targets. *Clin. Microbial. Rev*. **31**, e00049-18. https://doi.org/10.1128/CMR.00049-18 (2018).

38. Pimentel, B.A. et al. Streptococcus agalactiae infection in cancer patients: a five-year study. European journal of clinical microbiology & infectious diseases. **35**, 927-933. https://doi.org/10.1007/s10096-016-2617-9 (2016).

39. Phares, C.R. et al. Active Bacterial Core surveillance/Emerging Infections Program Network. Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005. *JAMA*. **299**, 2056-2065. https://doi.org/10.1001/jama.299.17.2056 (2008).

40. Farley, M.M. et al. A population-based assessment of invasive disease due to group B Streptococcus in nonpregnant adults. *N. Engl. J. Med*. **328**, 1807-1811. https://doi.org/10.1056/NEJM199306243282503 (1993).

41. Sharma, P. et al. Role of pilus proteins in adherence and invasion of Streptococcus agalactiae to the lung and cervical epithelial cells. *J. Biol. Chem*. **288**, 4023-4034. https://doi.org/10.1074/jbc.M112.425728 (2013).

42. Ferreira Eduardo da Costa, A. et al. Reactive oxygen species involved in apoptosis induction of human respiratory epithelial (A549) cells by Streptococcus agalactiae. *Microbiology (Reading)*. **162**, 94-99. https://doi.org/10.1099/mic.0.000202 (2013).

43. Oliveira, J.S.S. et al. Reactive oxygen species generation mediated by NADPH oxidase and PI3K/Akt pathways contribute to invasion of Streptococcus agalactiae in human endothelial cells. *Mem. Inst. Oswaldo Cruz*. **113**, e140421. https://doi.org/10.1590/0074-02760170421 (2018).

Figures

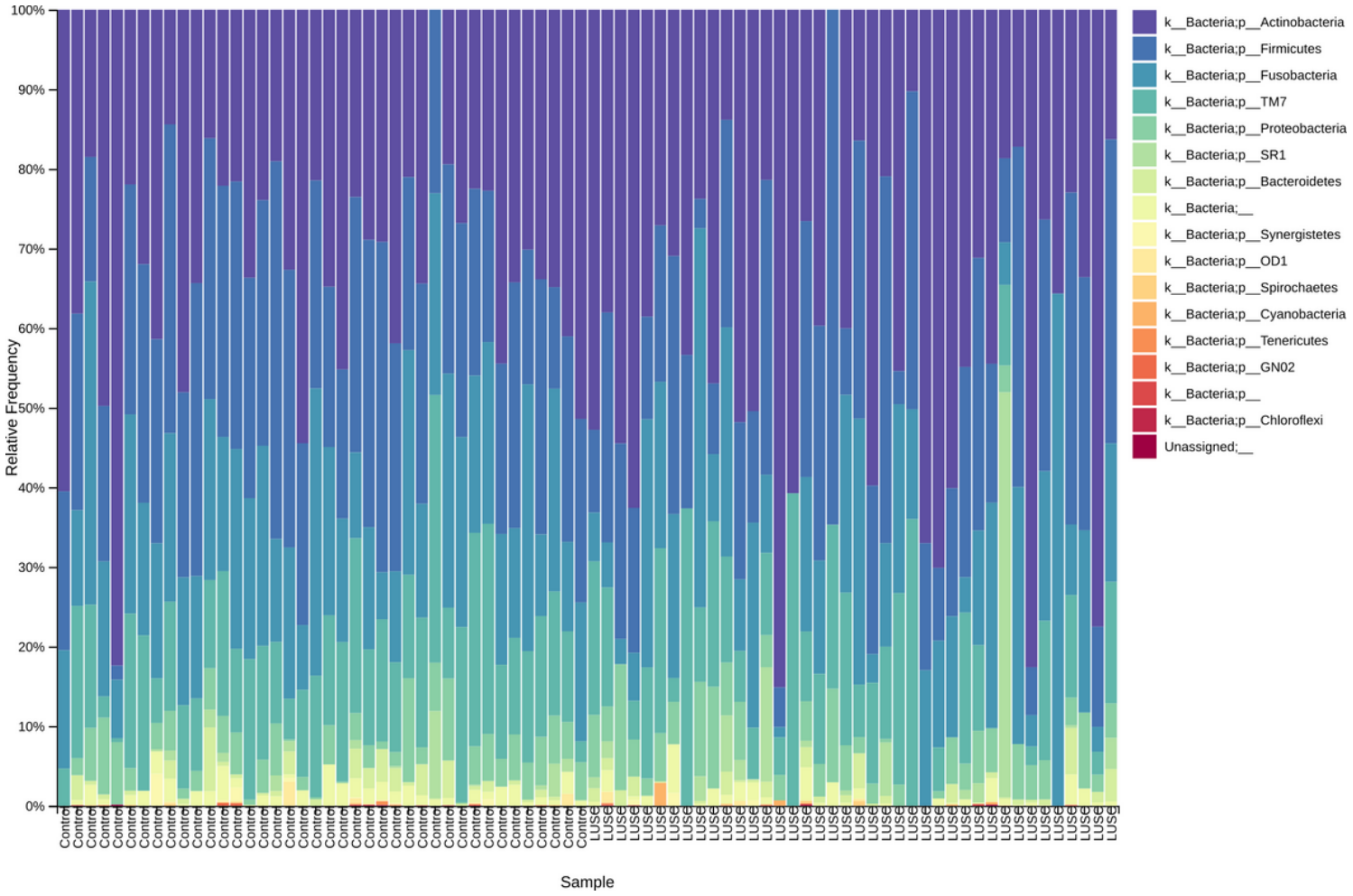

Figure 1

Taxonomic structure of the sputum microbiomes from LUSC patients and control subjects at the phyla level

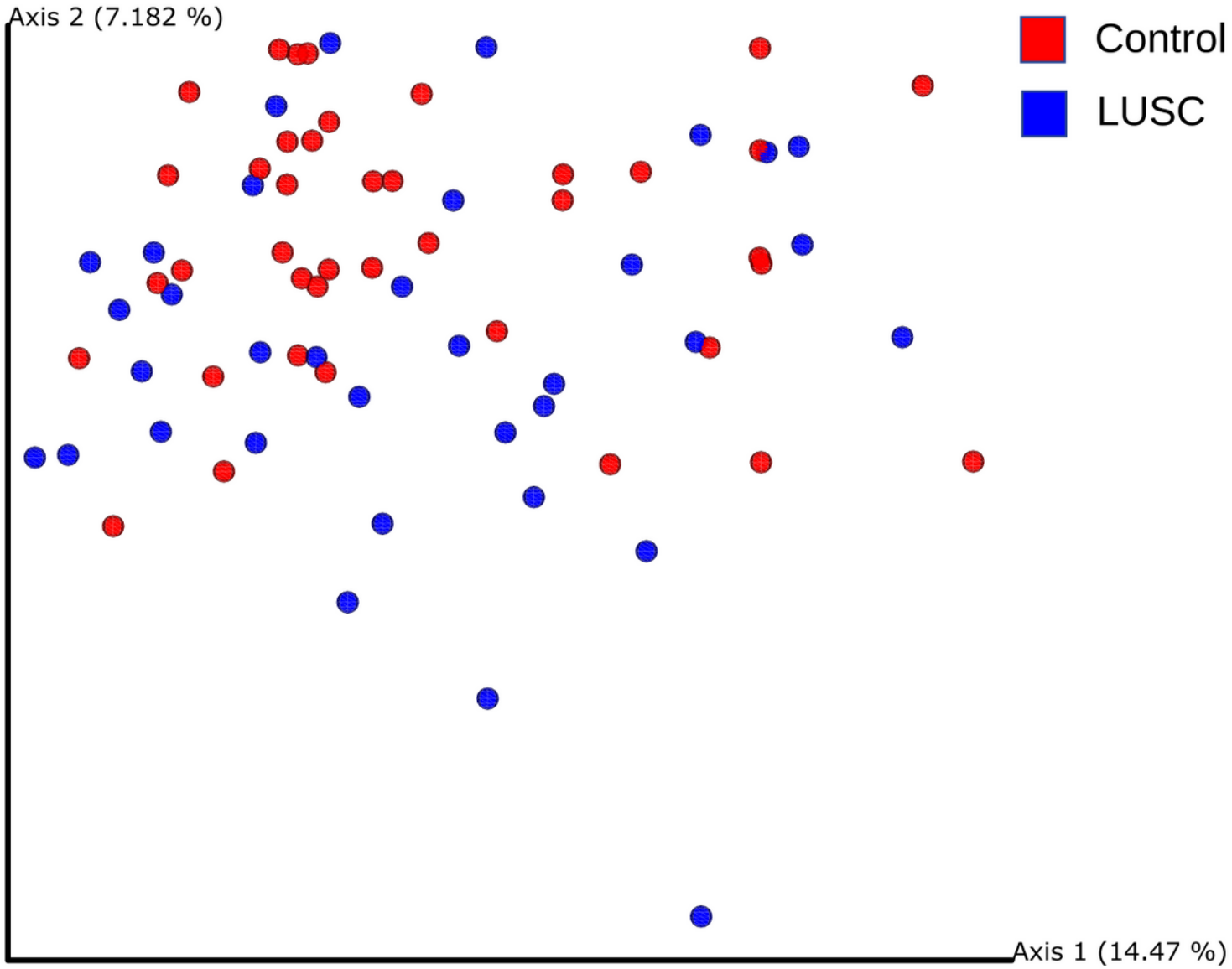

Figure 2

Two-dimensional diagram constructed by the method of principal components analysis demonstrating the phylogenetic similarity of prokaryotic sputum communities in LUSC patients and control subjects.

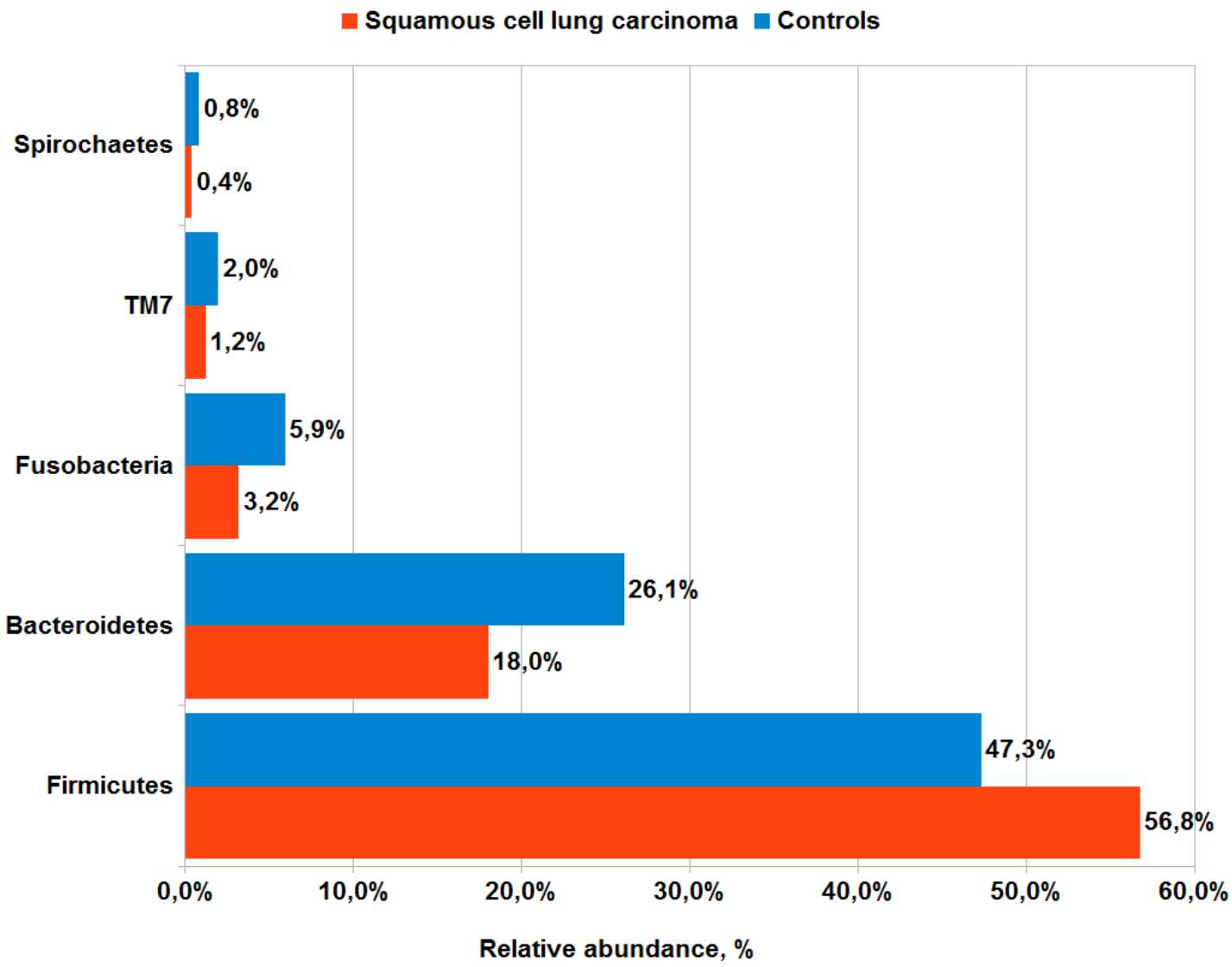

Figure 3

Occurrence frequencies of the main bacterial phyla in the sputum of LUSC patients compared to control subjects.

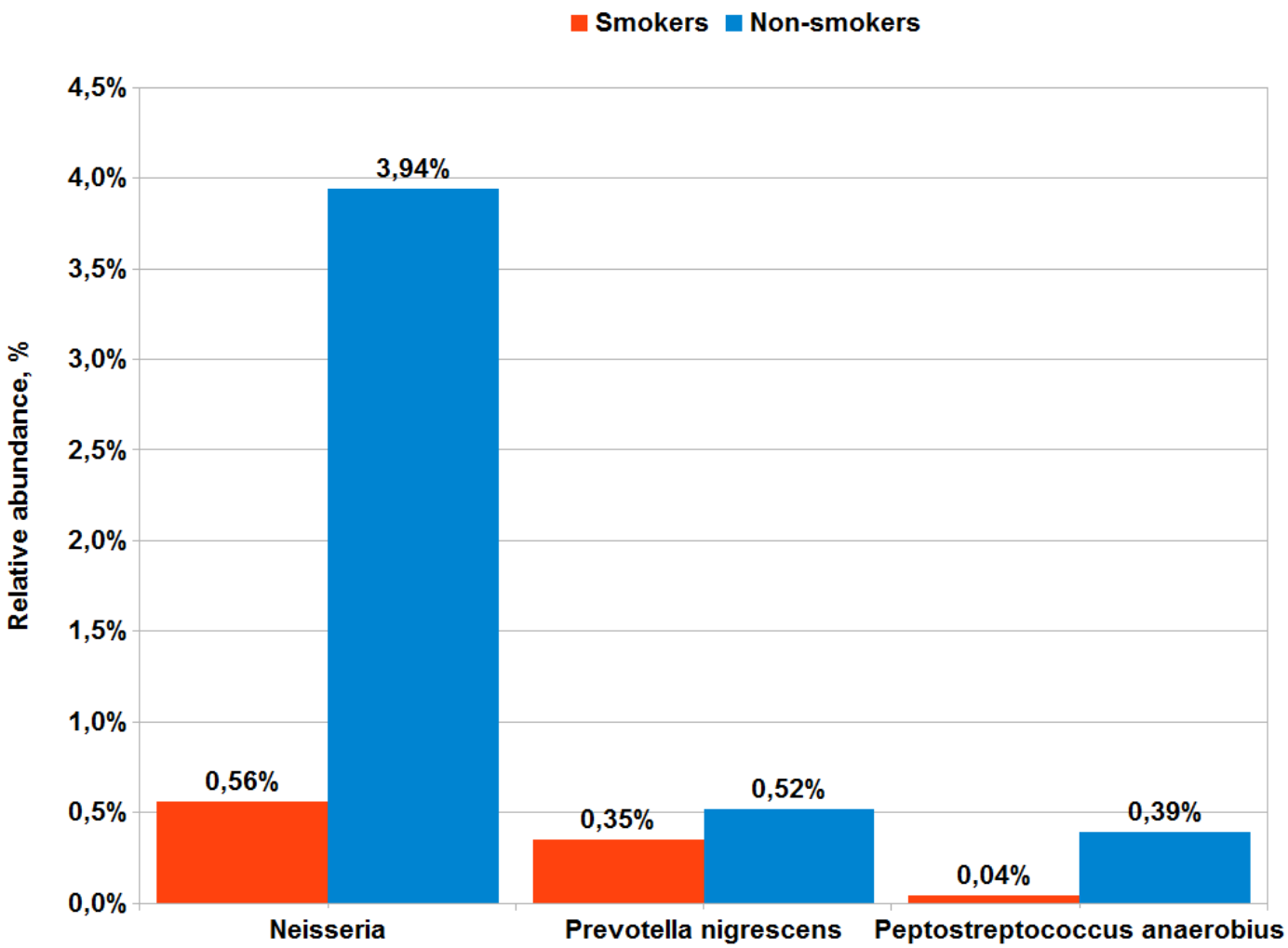

Figure 4

Differences in the representation of bacterial taxa in the sputum of smokers and non-smoker control subjects.