RESEARCH ARTICLE

16S rRNA deep sequencing for the characterization of healthy human pharyngeal microbiome

Gioula G, Melidou A, Siasios P, Minti F, Malisiovas N

Microbiology Department, School of Medicine, Faculty of Health Science, Aristotle University of Thessaloniki, Thessaloniki, Greece

Abstract

Background: The recent advent of high-throughput sequencing methods enabled the study of the composition of the upper respiratory tract (URT) microbial ecosystem and its relationship with health and disease in immense detail. The aim of the present study was the characterization of the human pharyngeal microbiome of healthy individuals in Greece.

Materials and Methods: We obtained ten pharyngeal specimens from healthy volunteers, Greek resident, with Greek nationality, who were eligible to the selection criteria. The construction of DNA libraries was performed by using two primer sets that amplify selectively the corresponding hypervariable regions of the 16s region in bacteria (V2-V9). The Ion Torrent PGM platform was used for the performance of next-generation sequencing.

Results: In the study samples, twelve phyla were identified. The most abundant ones were *Firmicutes*, *Proteobacteria, Bacteroidetes,* followed by *Actinobacteria* and *Fusobacteria.* Seventy-nine families, 79 genera and 137 species were identified and characterized. Moreover, 17 unique differentially abundant families, 30 unique differentially abundant genera and 24 unique differentially abundant species were identified among healthy subgroups with adjusted p-values <0.05. At the genus level, *Moraxella (Proteobacteria)* and *Gemella (Firmicutes)* were detected with a statistical significance in non-smokers, while *Bifidobacterium (Actinobacteria), Alloscardovia (Actinobacteria), Dialister (Firmicutes)* and *Filifactor (Firmicutes*) were present mostly in smokers.

Conclusions: The URT is colonized by a variety of protective and potentially pathogenic bacteria. This microbiome system is highly diverse and varies significantly between individuals. Geographic location and ethnicity are considered to be a strong determinants and factors affecting the diversity and abundance of the URT microbiome. Although some of the most abundant families are common irrespective of these factors, the dominance patterns are usually different between the study subjects and between the studies from other geographic locations. Unique differentially abundant families, genera and species were identified, and further studies are needed to elucidate their role. Further studies should focus on the investigation of the URT microbiome dynamics and the interaction with the host in health and disease.
HIPPOKRATIA 2018, 22(1): 29-36.

Keywords: Pharyngeal microbiome, 16srRNA, healthy, sequencing, Greece

Corresponding author: Georgia Gioula, Associate Professor of Medical Microbiology, Laboratory Department, School of Medicine, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece, tel: +302310999121, e-mail: ggioula@med.auth.gr

Introduction

The upper respiratory tract (URT) is the natural niche for respiratory bacterial and viral microorganisms and the cause of respiratory tract infections (RTIs). The URT is an interesting area from an ecological perspective: the continuous exposure to the external environment and internal host mechanisms, like respiration and gastrointestinal processes, form different habitats for a wide variety of microorganisms. Potentially pathogenic bacteria are embedded in a community of commensals, forming together the nasopharyngeal microbiome. Many pathogenic species, including *Streptococcus pneumoniae*, *Haemophilus influenza, Moraxella catarrahlis, Staphylococcus aureus,* and *Neisseria meningitidis*, exist in the nasopharynx of apparently healthy individuals. Environmental factors, bacterial acquisition during the first years of life and host factors profoundly influence the microbial colonization succession. Unraveling the relationships between microbes of the URT during perturbations is anticipated to provide insights into the pathogenesis of respiratory infections[1]. Discovery of the microbiome of both the upper and the lower respiratory tracts created the opportunity for understanding the disease onset, exacerbation, and progression of chronic respiratory diseases which may be directly associated with dysbiosis in the microbiome[2].

The recent advent of high-throughput sequencing methods enabled the study of these communities and their relationship with health and disease in immense detail. Extensive research has focused on the gut microbiome and its role in metabolism, immune maturation, mucosal bar-

rier functions, and colonization resistance. Although much less extensively studied, the URT microbiome is equally important and represents a strong determinant of respiratory health[3]. Deep sequencing methods have enabled the evaluation of the composition of the whole URT microbial ecosystem, giving the opportunity to explore its diversity and relative abundance down to the species level and to determine its relationship to human health and disease[4].

Recent studies have focused on the differences in the microbiome composition between healthy individuals from different races, ethnicity and geographic location that are important determinants of the microbiome composition. Various factors are responsible for the geographical differences: host genetics, innate/adaptive immunity, environmental factors, social and behavioral features (hygiene, diet, parasitic load)[5]. Fewer studies have focused on such differences in the composition of the URT microbiome and generally, it is considered to be a less diverse system, compared to the gut, skin or oral microbiome. The URT microbiome has been characterized in geographically diverse populations from South Korea and USA, the Netherlands and Canada[6-8].

The present study aimed to characterize the relative abundance of microbial communities of the pharynx at the phylum, family, genus and species levels, by deep sequencing the 16S rRNA gene of bacteria in pharyngeal swab samples of healthy individuals from Greece. This study will form the basis for further studies in our geographical area that will focus on the relationship of the URT microbiome with disease.

Materials and Methods

Ten volunteers were eligible according to the selection criteria and enrolled in this prospective study. Ethical approval was obtained from the Bioethics Committee of the Medical Faculty of Aristotle University and all study subjects consented for their participation in the study. Their mean age was 32.3 years (range 5-61 years), while 80 % of these volunteers were female. All individuals had Greek nationality and were residents of areas of northern Greece at the time of sample collection (January to March 2017). None of them were vaccinated against influenza virus or suffered any known respiratory infection at least during the preceding year prior to sample collection or required antibiotic treatment for the preceding six months before sample collection. Their medical history is shown in Table 1. For the calculation of the sample size, G * Power 3.1.9.2v was used, and *a priori* power analysis was calculated to achieve a statistical power of 80 % with a level of significance α =5 %. Specifically, a repeated measures ANOVA was used and taxon level as the within-subject factor with three measurements was considered. The same subjects were used for family, genus, and species level.

Pharyngeal swab samples were obtained from volunteers using Sterile Catch-AII™ sample collection swabs. We utilized a consistent methodology for storage and processing of all samples. A single-sample was collected at any time of the day. Samples were immediately placed at -80ºC until DNA extraction and were only subjected to a single free-thaw cycle.

DNA extraction

For each subject, pharyngeal aspirates were collected, and total genomic DNA was extracted from 200 µl of sample according to the manufacturer's recommended procedure using the Qiagen DNA mini kit (Qiagen, Hilden, Germany). The DNA was eluted in 100 µl of AE elution buffer (Qiagen, Hilden, Germany) and stored at -20 °C until further processing. The total DNA concentration of each DNA extract was measured using Qubit fluorometer (Life Technologies Corporation, Carlsbad, CA, USA) with the Qubit dsDNA HS assay (Life Technologies Corporation, Carlsbad, CA, USA).

16S rRNA amplicon preparation, sequencing, and analysis
Library preparation

DNA libraries for next-generation sequencing were constructed using two primer sets that amplify selectively the corresponding hypervariable regions of the 16s region in bacteria (V2, 3, 4, 6, 7, 8, 9). We used for amplification primers that contain adaptors for Ion Torrent sequencing and Ion Xpress barcodes so that the polymerase chain reaction (PCR) products could be pooled and sequenced directly. For the library preparation, the end repair was followed by purification of pooled amplicons with AMPure beads, ligation of Ion Xpress barcode adapters and nick-repair, using Ion Plus Fragment Library kit (Life Technologies Corporation, Carlsbad, CA, USA). Purification steps were performed using AMPure beads (Agencourt), eluted in low Tris- ethylenediaminetetraacetic acid (TE) buffer, and quantified by using a Qubit dsDNA HS (high sensitivity) kit (Life Technologies Corporation, Carlsbad, CA, USA). The library concentration was determined using the Ion Universal Library Quantification kit (Life Technologies Corporation, Carlsbad, CA, USA) in an ABI7500 real-time PCR system. The two resulting PCR libraries were equimolarly pooled after DNA purification.

Template preparation and sequencing

Before emulsion PCR, the library concentration was adjusted to 26 pM. Template preparation was performed in the Ion 400 Template One-Touch 2 following enrichment of the amplified Ion Sphere Particles using Dynalbeads MyOne Streptavidin C1 beads (Life Technologies Corporation, Carlsbad, CA, USA) using the Ion One-Touch ES system according to the manufacturer's protocol.

The Ion Xpress barcoded library was sequenced through PGM, Ion Torrent (Life Technologies Corporation, Carlsbad, CA, USA). A total of 10 barcoded samples were pooled and loaded into each 316 chip (Life Technologies, MA, USA) using the Ion PGM 400 Sequencing kit (Life Technologies Corporation, Carlsbad, CA, USA), according to the recommended protocol.

Table 1: Demographic characteristics (age, sex, nationality, origin) and clinical characteristics (smoking, influenza vaccination, respiratory system infection, and medical history) of the study population.

SampleID	Age	Sex	Nationality	Origin	Smoking	Influenza Vaccination	Respiratory System Infection	Medical_History
S01	35	Female	Greek	Kozani	No	No	No	AT, thyroidectomy, SDII
S02	44	Female	Greek	Thessaloniki	Yes	No	No	No
S03	9	Female	Greek	Thessaloniki	No	No	No	No
S04	35	Female	Greek	Kavala	No	No	No	No
S05	31	Female	Greek	Kozani	No	No	No	No
S06	43	Female	Greek	Kozani	No	No	No	irritable bowel syndrome
S07	11	Female	Greek	Thessaloniki	No	No	No	No
S08	5	Female	Greek	Thessaloniki	No	No	No	No
S09	61	Male	Greek	Thessaloniki	Yes	No	No	No
S10	49	Male	Greek	Thessaloniki	Yes	No	No	No

AT: Ataxia-Telangiectasia, SDII: Sudden Death in Infancy.

16s rRNA sequence data preprocessing

Base calling and run demultiplexing were performed by using Torrent Server software. The sequences were analyzed using the metagenomics workflow in the Ion Reporter™ Software that enables the identification of microbes present in complex multi-bacterial samples, at the genus or species level, and uses both the curated Greengenes database and the premium curated MicroSEQ™ ID 16S rRNA reference database.

Sequencing of the ten human nasopharyngeal samples on the Ion Torrent PGM instrument generated 1.791.402 total raw reads. We utilized the Ion Reporter software to process the raw reads, which performs demultiplexing and denoising, quality filtering, alignment against a reference database of 16SrRNA gene sequences and clustering into operational taxonomic units (OTUs) at 97 % similarity. OTUs were assigned taxonomy using the Greengenes database and analysis was performed using Quantitative Insights into Microbial Ecology (QIIME).

16S rRNA sequencing data analysis

Statistical analysis was conducted with open source programming R Language Version 3.3.1. Additionally, the following R packages were imported to further explore the microbial data: vegan Version 2.4.2, phyloseq Version 1.19.1 and DESeq2 Version 1.14.11[9-11]. Bar plots were designed to explore the relative abundance of the taxa at phylum, family, genus, and species level. To compare the microbial diversity within each sample, richness was evaluated, and the a-diversity Shannon index which examines both richness and evenness was calculated. The healthy subjects were categorized into subgroups in accordance with gender, origin, age group, smoking, and medical history. To detect differences in a-diversity among subgroups of the healthy subjects, the non-parametric Kruskal-Wallis test was implemented. Differential Expressions analysis for Sequence data (DESeq) was performed for the computation of the q-values, with the Benjamini & Hochberg method to detect differentially abundant taxa between the healthy subgroups, at a 5 % False Discovery Rate (FDR) threshold[12-14].

Furthermore, to consider b-diversity between samples, a distance matrix was calculated with the Bray-Curtis distance metric. Principal Coordinate Analysis (PCoA) was applied to the calculated dissimilarity matrix for finding out taxa that were count more in the total variance and for pattern visualization. To confirm significant differences between healthy subgroups, Permutational Multivariate Analysis of Variance (PERMANOVA) was implemented together with PERMutational analysis of multivariate DISPersion (PERMDISP) to strengthen the results of the PERMANOVA test further.

Results

In total 1173 OTUs were generated and were available for further analysis. Twelve phyla were identified in the ten study samples. The most abundant phyla detected, were *Firmicutes* (53.08 %), *Proteobacteria* (22.3 %), *Bacteroidetes* (14.77 %), followed by *Actinobacteria* (5.08 %), and *Fusobacteria* (4.41 %). The remaining seven phyla contributed less than 0.1 % to the total abundance. Rare phyla with low abundance levels were *Cyanobacteria, Tenericutes, Chloroflexi, Spirochaetes, Synergistetes, Deinococcus-Thermus*, and *Nitrospirae.*

At the family level, 79 families were identified, grouped into 12 phyla. Out of the 79 families, sixty (76 %) belong to the five most abundant phyla. The five most abundant families were *Streptococcaceae* (33.47 %) (*Firmicutes* phylum*)*, *Prevotellaceae* (13.5 %) (*Bacteroidete phylum), Pasteurellaceae* (12.87 %) (*Proteobacteria* phylum), *Bacillales incertae sedis* (9.1 %), and *Veillonel-*

laceae (7.21 %), (*Firmicutes* phylum*).* The family *Pasteurellaceae* was prevalent in nine subjects, whereas the other four were prevalent in all samples. Moreover, the families *Leptotrichiaceae* (3.05 %) and *Fusobacteriacea* (1.36 %) (*Fusobacteria* phylum), *Carnobacteriaceae* (1.32 %) *and Lactobacillaceae* (0.33 %) (Firmicutes phylum)*, Enterobacteriaceae* (1 %) (*Proteobacteria* phylum), and *Porphyromonadaceae* (0.60 %) *in Bacteroidetes* were present in all study samples, but were less abundant.

In total, 79 genera were detected. The most abundant ones were *Streptococcus* (39.96%), *Gemella* (11.33 %) and *Veillonella* (7.41 %) that belong to phylum *Firmicutes* and genus *Prevotella* (16.34 %) that belong to phylum *Bacteroidetes*. Moreover, *Haemophilus* (3.67 %) and *Neisseria* (3.08 %) derived from phylum *Proteobacteria* and *Leptotrichia* (3.33 %) and *Fusobacterium* (1.68 %) from *Fusobacteria*. Out of the most abundant taxa at the genus level, *Streptococcus, Prevotella, Gemella, Veillonella, Fusobacterium* were present in all individuals. Additionally, genus *Delftia* from *Proteobacteria* was presented in high abundance (1.63 %) in only one sample (S06). Percentages are expressed in percent of total reads. In total, 137 species were identified in the subjects. The ten most abundant species detected, are presented in Table 2.

Exploration of a-diversity

The number of families observed at each individual and the Shannon diversity index was calculated in order to explore the microbial composition. Sample S09 had the most diverse microbial community, while samples S03, S05, and S07 had communities with smaller diversity. Nevertheless, these a-diversity metrics did not denote sizeable differences between them. The Kruskal Wallis non-parametric test, that was used to compare a-diversity in subgroups separated by gender, origin, smoking, medical history, and age, did not denote statistically significant differences (p-values >0.05).

DESeq was implemented, to detect statistically significant differentially abundant taxa between healthy individuals, at a 5 % False FDR threshold, by using the Benjamini & Hochberg method. In total, we detected seventeen unique differentially abundant families, thirty unique differentially abundant genera and twenty-four unique differentially abundant species, among healthy subgroups with adjusted p-values <0.05 (Figure 1, Figure 2).

More specifically, from the 17 unique differentially abundant families detected, *Bacillales Incertae Sedis* (*Firmicutes*) was present with a statistically significant difference in non-smokers, while *Acetobacteraceae* (*Proteobacteria*), *Aerococcaceae* (*Firmicutes*) and *Phormidiaceae* (*Cyanobacteria*) were present only in smokers. According to gender, the family *Acetobacteraceae* (*Proteobacteria*) *was* detected only in males.

At the genus level, from the thirty unique differentially abundant genera, *Gemella* (*Firmicutes*) and *Moraxella* (*Proteobacteria*) were present with a statistical significance in non-smokers (padj =6.64E-03, and 5.06E-10, respectivelly), while *Alloscardovia* (*Actinobacteria*), *Bifidobacterium* (*Actinobacteria*), *Filifactor* (*Firmicutes*), and *Dialister* (*Firmicutes*) were detected with a statistically significant difference mostly in smokers (padj =5.76E-04, 4.29E-04, 5.06E-10, and 1.52E-02, respectively).

Exploration of b-diversity

A dissimilarity matrix was calculated with the Bray-Curtis distance metric. PCoA that was used for visualization did not denote clearly clusters either for the subgroups that were separated by gender, origin, smoking, medical history, and age (plots not shown) or for the species (Figure 3). PERMANOVA test with 1,000 permutations was performed, and statistically significant differences were detected between smokers and non-smokers (p =0.04), while no statistically significant differences were detected between the centroids of the remaining healthy subgroups (p >0.05). PERMDISP2 procedure that was implemented to compare subgroup dispersions led to insignificant results for the origin (p =0.11) and medical history (p =0.5). However, it showed statistically significant results for subgroups based on gender (p =0.0009), age (p =0.0009), and smoking (p =0.0009),

Table 2: The ten most abundant species and their taxonomy (phylum, class, order, family, genus, and species) detected in the study samples.

Total Reads	*Percentages*	*Phylum*	*Class*	*Order*	*Family*	*Genus*	*Species*
35537	*16.64%*	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillalesincertaesedis*	*Gemella*	*haemolysans*
26760	*12.53%*	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Prevotellaceae*	*Prevotella*	*histicola*
19564	*9.16%*	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Prevotellaceae*	*Prevotella*	*melaninogenica*
13074	*6.12%*	*Fusobacteria*	*Fusobacteriia*	*Fusobacteriales*	*Leptotrichiaceae*	*Leptotrichia*	*genomosp.*
12609	*5.90%*	*Firmicutes*	*Negativicutes*	*Selenomonadales*	*Veillonellaceae*	*Veillonella*	*atypica*
11942	*5.59%*	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Prevotellaceae*	*Prevotella*	*salivae*
6597	*3.09%*	*Proteobacteria*	*Gammaproteobacteria*	*Pasteurellales*	*Pasteurellaceae*	*Haemophilus*	*parainfluenzae*
5100	*2.39%*	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	*Actinomycetaceae*	*Actinomyces*	*odontolyticus*
4910	*2.30%*	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Streptococcaceae*	*Streptococcus*	*salivarius*

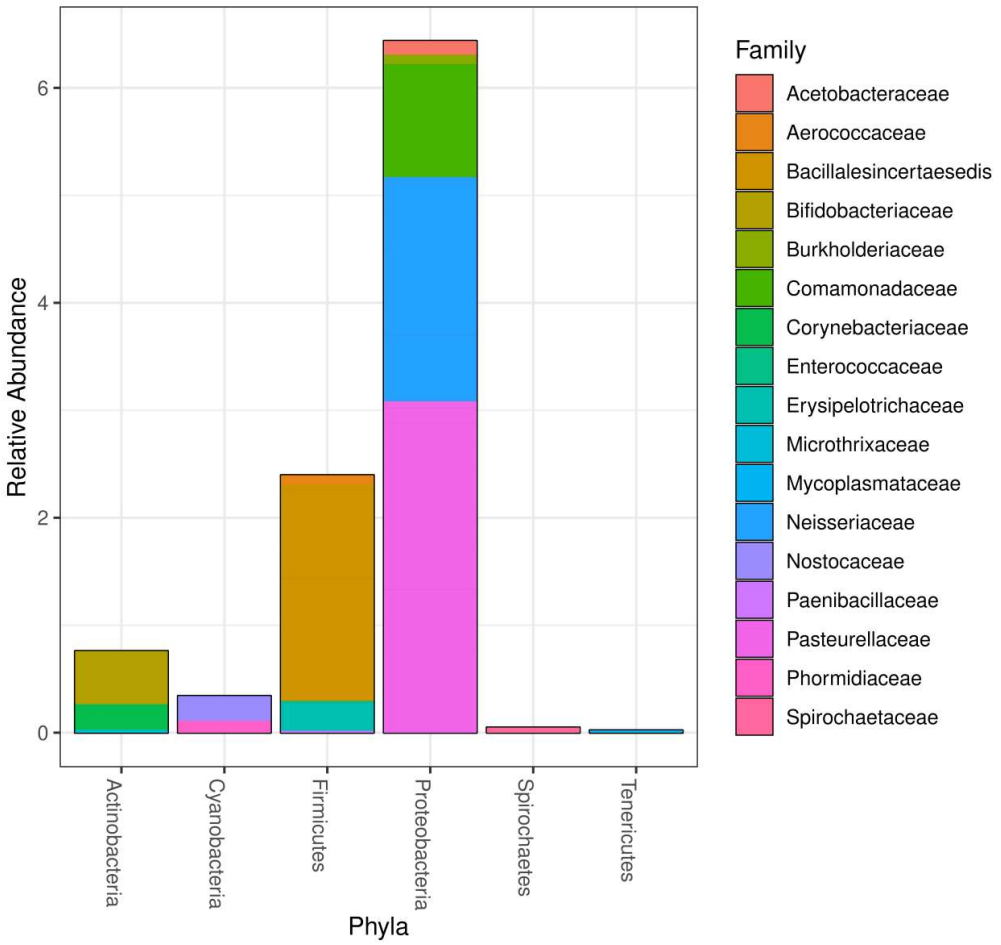

Figure 1: The seventeen unique differentially abundant families detected by using the Benjamini & Hochberg method.

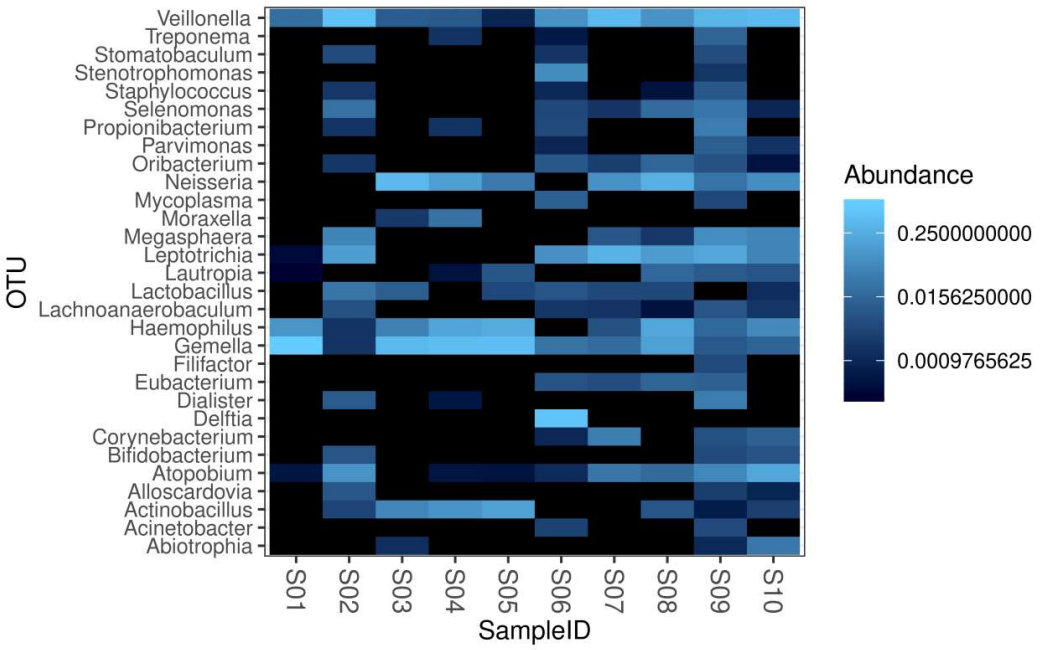

Figure 2: The unique differentially abundant genera detected by using the Benjamini & Hochberg method.
OUT: operational taxonomic units.

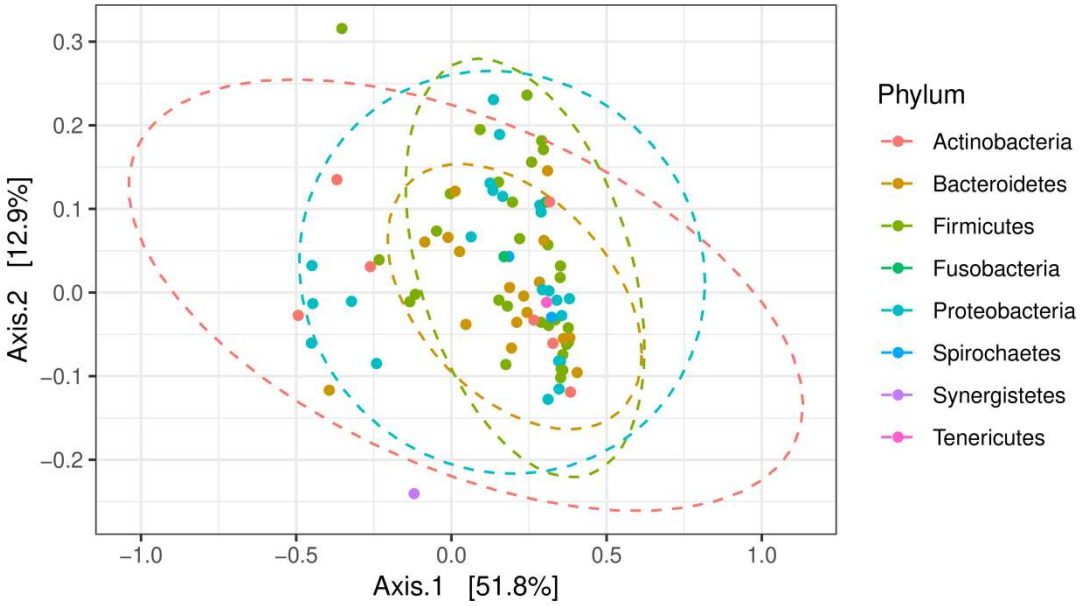

Figure 3: Principal Coordinate Analysis (PCoA) that was used for visualization did not denote clearly clusters either for the subgroups that were separated by gender, origin, smoking, medical history, and age or for the species.

which means that these subgroups had differences due to dispersions of the data at the species level. However, due to the aforementioned PERMANOVA results, we cannot infer real significant differences in the categories of healthy subgroups.

Discussion

The URT is colonized by a large variety of bacteria that constitute the URT microbiome. Although its largest proportion is composed of microorganisms without any pathogenic role, there is a number of bacteria that have the potential to cause disease[15]. Understanding the complex microbiome relationships in the URT is anticipated to provide insights into the pathogenesis of infections in this system. The potential role of commensal flora in the regulation of respiratory health was not evaluated until recently when high-throughput sequencing methods have made it possible to evaluate the whole composition of the URT microbial ecosystem[16].

In the present study, 12 phyla in total were identified, with the most abundant ones detected, being *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, followed by *Actinobacteria* and *Fusobacteria*. Similar to our findings, other studies have also shown that the microbial ecosystem of the URT is typically enriched by members of these phyla, although not always in the same order[5,16]. The relative abundance of each phylum was highly variable across the ten study subjects; similarly, in another study conducted by the Centre of Public Health Genomics in the USA, across all ten subjects studied, the dominant phyla were first *Firmicutes*, followed by *Actinobacteria* and *Proteobacteria*, while almost 20 % remained unclassified. In accordance with our results, the relative abundance of each phylum was found to be highly variable[15]. The same dominant phyla were also identified in a study of the nasopharyngeal microbiota of healthy Chinese adults[17]. Geography and ethnicity appear to be determining factors, although the URT microbiome seems to be the less divergent across the different geographic locations compared to the gut, skin, or oral microbiome.

At the family level, 79 families were identified in the present study, grouped into 12 phyla. The five most abundant families were *Streptococcaceae*, *Bacillales incertae sedis* and *Veillonellaceae*, *Prevotellaceae* and *Pasteurellaceae.* At the genus level, similar to other studies, *Moraxella*, *Corynebacterium*, *Propionibacterium*, Staphylococcus *spp,* and *Streptococcus* were the most frequently carried genera, irrespective of the ethnical background and geographical location[5]. On the other hand, the dominance patterns seem to be variable. The microbiome of the adult nasopharynx can generally be sub-classified in one of four distinct microbial profiles or 'types', which are characterized by the predominance of either *Corynebacterium*, *Propionibacterium*, *Moraxella* or *Staphylococcus* spp., suggesting that complex synergistic and antagonistic relationships between these bacteria are driving within-niche variation. Interestingly, *Corynebacterium*- and *Propionibacterium*-enriched profiles seem to be more 'tolerant' to members of other genera, like *staphylococci*, in contrast to *Moraxella*-dominated bacterial communities[18].

Gram-negative anaerobes, such as *Prevotella* and *Veillonella* spp., were observed in younger ages; those bacteria that are encountered primarily in the oropharynx and oral cavity, are often observed in low abundance in the nasopharynx of young children, which might be explained by frequent nasal regurgitation of oral content in this age group[19].

Of the 79 detected genera, the most abundant were *Streptococcus*, *Gemella*, and *Veillonella* (Firmicutes phylum) and *Prevotella* (Bacteroidetes phylum). Moreover, *Haemophilus* and *Neisseria* (Proteobacteria phylum), as well as *Leptotrichia* and *Fusobacterium* (Fusobacteria phylum) were commonly present. Out of the most abundant taxa at the genus level, *Streptococcus, Prevotella, Gemella, Veillonella*, and *Fusobacterium* were present in all individuals. This is in accordance with the literature although, in a study conducted in the USA, the most abundant genera in the nasopharynx microbiota were *Alloiococcus*, followed by *Corynebacterium, Staphylococcus, Haemophilus, Propionibacterium*, and *Streptococcus*[16]. Genus *Delftia* (*Proteobacteria*) was presented in high abundance in only one sample that derived from a 61-year-old Greek female with no smoking history, who had an irritable bowel syndrome. There are no studies in the literature connecting the presence of the above genus with the specific syndrome, indicating, as it was already analyzed in many studies, that taxonomic composition varies widely among samples.

According to the literature, more stable patterns were marked by high prevalence of *Moraxella* and *Corynebacterium/Dolosigranulum* spp., beginning early in life, whereas the less stable profiles, characterized by high abundance of *Streptococcus* spp. and acquisition of *Haemophilus influenza* over time were associated with higher rates of respiratory infections and wheezing in the first years of life[19].

At the species level, 137 species in total were identified in the ten subjects of the present study, most of which were common but some were unique to the study subjects, indicating the heterogeneity of the microbiome composition among individuals.

Concerning the diversity of bacteria, the statistical analysis of the results of the study showed that the diversity of bacteria at the family level varied in the different samples; one sample that originated from the oldest study subject had the most diverse community, while three samples had communities with smaller diversity. Most of the samples with small diversity were derived from the younger study subjects, indicating that age may proportionally affect the microbiome diversity of the URT. There are only a few studies describing how the URT microbiota alters in older adults, but there is also increasing evidence that age-related changes may contribute to the increased susceptibility to respiratory infections. The nasopharyngeal microbiota of older adults appears to undergo profound changes and becomes dominated by *Streptococcus*, *Prevotella*, and *Veillonella*, while the abundance of different bacteria, increases in the elderly, especially in pneumonia patients[7,20].

In the present study, 17 unique differentially abundant families, 30 unique differentially abundant genera and 24 unique differentially abundant species were detected among healthy subgroups. More specifically, from the 17 unique differentially abundant families detected *Bacillales Incertae Sedis* (*Firmicutes*) was present with a statistically significant difference in non-smokers, while *Acetobacteraceae* (*Proteobacteria*), *Aerococcaceae* (*Firmicutes*), and *Phormidiaceae* (*Cyanobacteria*) were present only in smokers. According to the gender, *Acetobacteraceae* (*Proteobacteria*) and *Aerococcaceae* (*Firmicutes*) were detected only in males, although the male sample size was small and this result needs to be verified in future more extensive studies. At the genus level, from the 30 unique differentially abundant genera, *Moraxella* (*Proteobacteria*) and *Gemella* (*Firmicutes*) were present with a statistical significance in non-smokers, while *Bifidobacterium* (*Actinobacteria*), *Alloscardovia* (*Actinobacteria*), *Dialister* (*Firmicutes*) and *Filifactor* (*Firmicutes*) were detected with a statistically significant difference mostly in smokers. According to the literature, smokers have an increased microbial diversity compared to nonsmokers, having an increased likelihood of carrying pathogens, including *S. pneumoniae, Streptococcus pyogenes, H. influenza*, and *M. catarrhalis*. Consistent with this, smokers had less carriage of nonpathogenic *Streptococcus, Prevotella*, *and Peptostreptococcus* species, which have been shown to be inversely correlated with the presence of pathogens[21,22]. Since the size of the smokers' sample in the present study is quite small, no conclusions can be extracted and more data are needed.

A limitation of the present study is the small sample size. No conclusions can be drawn regarding differences between subgroups (smoking status, age, gender) and results need to be interpreted with caution.

In conclusion, the URT is colonized by a variety of protective and potentially pathogenic bacteria. This microbiome system is highly diverse and varies significantly between individuals. Geographical location and ethnicity are considered to be strong determinants and factors affecting the diversity and abundance of the URT microbiome. Although some of the most abundant families seem to be common irrespective of these factors, the dominance patterns are usually different between the study subjects and between the studies from other geographic locations. Unique differentially abundant families, genera and species were identified, and further studies are needed to elucidate their role. The URT microbiome has been significantly less studied compared to other systems, but recently collected data indicate that respiratory microbiota may contribute to health regulation and disease development in the respiratory tract. Such studies can form the basis for new approaches to diseases that respond poorly to traditional interventions. Next generation techniques will facilitate future research with high throughput, deep sequencing results that enable the identification and quantification of the microbiome at the species level. Further studies will aid better understanding of the dynamics of URT microbiota, and its interactions with the host in health and disease.

Conflict of interest

Authors declare no conflict of interest.

References

1. Koopen IJN, Bosch AATM, Sanders EAM, van Houten MA, Bogaert D. The respiratory microbiota during health and disease: a paediatric perspective. Pneumonia. 2015; 6: 90-100.
2. Martin C, Burgel PR, Lepage P, Andréjak C, de Blic J, Bourdin A, et al. Host-microbe interactions in distal airways: relevance to chronic airway diseases. Eur Respir Rev. 2015; 24: 78-91.
3. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012; 13: 260-270.
4. Vayssier-Taussat M, Albina E, Citti C, Cosson JF, Jacques MA, Lebrun MH, et al. Shifting the paradigm from pathogens to pathobiome: new concepts in the light of meta-omics. Front Cell Infect Microbiol. 2014; 4: 29.
5. Gupta VK, Paul S, Dutta C. Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity. Front Microbiol. 2017; 8: 1162.
6. Yi H, Yong D, Lee K, Cho YJ, Chun J. Profiling bacteria community in upper respiratory tracts. BMC Infect Dis. 2014; 14: 583.
7. de Steenhuijsen Piters WA, Huijskens EG, Wyllie AL, Biesbroek G, van den Bergh MR, Veenhoven RH, et al. Dysbiosis of upper respiratory tract microbiota in elderly pneumonia patients. ISME J. 2016; 10: 97-108.
8. Stearns JC, Davidson CJ, McKeon S, Whelan FJ, Fontes ME, Schryvers AB, et al. Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age. ISME J. 2015; 9: 1246-1259.
9. Oksanen J, Blanchet FJ, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Community Ecology Package. R package version 2.4-2. 2017. Available at: https://CRAN.R-project.org/package=vegan, last accessed: 21/01/18.
10. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE. 2013; 8: e61217.
11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol. 2014; 15: 550.
12. Benjamini, Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Statist Soc B. 1995; 57: 289-300.
13. Schenck LP, Surette MG, Bowdish DM. Composition and immunological significance of the upper respiratory tract microbiota. FEBS Lett. 2016; 590: 3705-3720.
14. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11: R106.
15. Allen EK, Koeppel AF, Hendley JO, Turner SD, Winter B, Sale MM. Characterization of the nasopharyngeal microbiota in health and during rhinovirus challenge. Microbiome. 2014; 2: 22.
16. de Steenhuijsen Piters WA, Sanders EA, Bogaert D. The role of the local microbial ecosystem in respiratory health and disease. Philos Trans R Soc Lond B Biol Sci. 2015; 370: 20140294.
17. Ling Z, Liu X, Luo Y, Yuan L, Nelson KE, Wang Y, et al. Pyrosequencing analysis of the human microbiota of healthy Chinese undergraduates. BMC Genomics. 2013; 14: 390.
18. De Boeck I, Wittouck S, Wuyts S, Oerlemans EFM, van den Broek MFL, Vandenheuvel D, et al. Comparing the Healthy Nose and Nasopharynx Microbiota Reveals Continuity As Well As Niche-Specificity. Front Microbiol. 2017; 8: 2372.
19. Biesbroek G, Tsivtsivadze E, Sanders EA, Montijn R, Veenhoven RH, Keijser BJ, et al. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. Am J Respir Crit Care Med. 2014; 190: 1283-1292.
20. Whelan FJ, Verschoor CP, Stearns JC, Rossi L, Luinstra K, Loeb M, et al. The loss of topography in the microbial communities of the upper respiratory tract in the elderly. Ann Am Thorac Soc. 2014; 11: 513-521.
21. Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R, et al. Disordered microbial communities in the upper respiratory tract of cigarette smokers. PLoS One. 2010; 20; 5: e15216.
22. Brook I, Gober AE. Recovery of potential pathogens and interfering bacteria in the nasopharynx of otitis media-prone children and their smoking and nonsmoking parents. Arch Otolaryngol Head Neck Surg. 2005; 131: 509-512.