The ISME Journal (2016) 10, 97–108
© 2016 International Society for Microbial Ecology All rights reserved 1751-7362/16
www.nature.com/ismej

ORIGINAL ARTICLE

Dysbiosis of upper respiratory tract microbiota in elderly pneumonia patients

Wouter AA de Steenhuijsen Piters[1], Elisabeth GW Huijskens[2,3], Anne L Wyllie[1], Giske Biesbroek[1], Menno R van den Bergh[1,4], Reinier H Veenhoven[4,*], Xinhui Wang[1], Krzysztof Trzciński[1], Marc J Bonten[5], John WA Rossen[2,6], Elisabeth AM Sanders[1] and Debby Bogaert[1]

[1]*Department of Paediatric Immunology and Infectious Diseases, The Wilhelmina Children's Hospital/ University Medical Center Utrecht, Utrecht, The Netherlands;* [2]*Laboratory of Medical Microbiology and Immunology, St Elisabeth Hospital, Tilburg, The Netherlands;* [3]*Department of Medical Microbiology, Albert Schweitzer Hospital, Dordrecht, The Netherlands;* [4]*Linnaeus Institute, Spaarne Hospital, Hoofddorp, The Netherlands;* [5]*Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands and* [6]*Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands*

Bacterial pneumonia is a major cause of morbidity and mortality in elderly. We hypothesize that dysbiosis between regular residents of the upper respiratory tract (URT) microbiome, that is balance between commensals and potential pathogens, is involved in pathogen overgrowth and consequently disease. We compared oropharyngeal microbiota of elderly pneumonia patients ($n=100$) with healthy elderly ($n=91$) by 16S-rRNA-based sequencing and verified our findings in young adult pneumonia patients ($n=27$) and young healthy adults ($n=187$). Microbiota profiles differed significantly between elderly pneumonia patients and healthy elderly (PERMANOVA, $P<0.0005$). Highly similar differences were observed between microbiota profiles of young adult pneumonia patients and their healthy controls. Clustering resulted in 11 (sub)clusters including 95% (386/405) of samples. We observed three microbiota profiles strongly associated with pneumonia ($P<0.05$) and either dominated by lactobacilli ($n=11$), *Rothia* ($n=51$) or *Streptococcus (pseudo)pneumoniae* ($n=42$). In contrast, three other microbiota clusters (in total $n=183$) were correlated with health ($P<0.05$) and were all characterized by more diverse profiles containing higher abundances of especially *Prevotella melaninogenica*, *Veillonella* and *Leptotrichia.* For the remaining clusters ($n=99$), the association with health or disease was less clear. A decision tree model based on the relative abundance of five bacterial community members in URT microbiota showed high specificity of 95% and sensitivity of 84% (89% and 73%, respectively, after cross-validation) for differentiating pneumonia patients from healthy individuals. These results suggest that pneumonia in elderly and young adults is associated with dysbiosis of the URT microbiome with bacterial overgrowth of single species and absence of distinct anaerobic bacteria. Whether the observed microbiome changes are a cause or a consequence of the development of pneumonia or merely coincide with disease status remains a question for future research.

The ISME Journal (2016) **10,** 97–108; doi:10.1038/ismej.2015.99; published online 7 July 2015

Introduction

Lower respiratory tract infections (LRTI), including pneumonia, are the most important cause of morbidity in adults, accounting for 6.2% of the total disability-adjusted life years (Mathers *et al.*, 2008). Moreover, LRTIs are responsible for 7.1% of deaths worldwide each year and mortality rates of up to 44% in elderly aged >85 years have been reported (Kaplan *et al.*, 2003; Tichopad *et al.*, 2013).

The incidence of LRTI peaks in young infants and in adults over 60 years of age (Jansen *et al.*, 2009; Millett *et al.*, 2013). In children, this high incidence can be explained by the naivety of their immune system; however, in elderly, the mechanisms of the heightened susceptibility to pneumococcal disease are still poorly understood. Immunosenescence, defined as age-related deterioration of both innate and adaptive immunity, seems to impair elderly individuals to elicit effective immune responses against invading pathogens, at least partly explaining their vulnerability to LRTI (Krone *et al.*, 2014).

Correspondence: D Bogaert, Department of Paediatric Immunology and Infectious Diseases, The Wilhelmina Children's Hospital, University Medical Center Utrecht, PO Box 85090, 3508 AB Utrecht, The Netherlands.
E-mail d.bogaert@umcutrecht.nl
*Deceased.
Received 25 November 2014; revised 3 March 2015; accepted 4 May 2015; published online 7 July 2015

Although the etiology of LRTIs in many cases remains unknown, the most important causative bacterial pathogen for LRTI and particularly pneumonia is *Streptococcus pneumoniae* (File, 2003). Typically, this pathogen populates the upper respiratory tract (URT) asymptomatically in coherence with a vast and diverse community of other potential pathogenic and commensal bacteria, fungi and viruses. In general, during health, the so-called microbiome is presumed to be at an equilibrium, providing beneficial functions to the host, such as pathogen resistance (Cho and Blaser, 2012). However, acquisition of new bacterial or viral pathogens, environmental factors or immunological perturbations can potentially disrupt the equilibrium, leading to dysbiosis, pathogen overgrowth and dissemination, resulting in symptomatic infections like pneumonia (Bogaert *et al.*, 2011; Dethlefsen *et al.*, 2007).

Only few studies have addressed the possible etiologic influence of the URT microbiome on the development of LRTIs. The available studies typically include only small numbers of subjects without a healthy control population (Zhou *et al.*, 2010; Bousbia *et al.*, 2012). Also, different microbial niches are studied (Sakwinska *et al.*, 2014) and research is focused on specific risk groups like HIV (human immunodeficiency virus)-infected patients (Iwai *et al.*, 2012) or patients infected with the pandemic H1N1 influenza virus (Chaban *et al.*, 2013; Leung *et al.*, 2013). Nevertheless, microbiota studies in chronic respiratory diseases like chronic obstructive pulmonary disease (COPD) and asthma suggests an important role for bacterial communities, both commensals and potential pathogens, in the pathogenesis of respiratory diseases (Hansel *et al.*, 2013; Molyneaux *et al.*, 2013).

As the oropharynx is the niche that can be regarded as the reservoir from which potentially pathogenic bacteria may spread to the lower respiratory tract, we hypothesized that pneumonia in adults and elderly is associated with imbalanced oropharyngeal bacterial communities leading to lack of containment and overgrowth of specific bacterial species (Morris *et al.*, 2013). Therefore, we investigated the oropharyngeal microbiome in a large group of elderly with pneumonia ($n = 100$) using 16S-based sequencing and compared this with healthy community-dwelling controls ($n = 91$). To validate our findings and to explore the effect of age, we sought to replicate our results in a cohort of young adult pneumonia patients ($n = 27$) and healthy controls ($n = 187$).

Materials and methods

Study population

Patients aged $\geqslant 18$ years with a suspected community-acquired pneumonia (CAP) presenting in two general hospitals in Tilburg, The Netherlands, were invited to participate in the study (Central Committee on Research involving Human Subjects, NL18747.041.07). In- and exclusion criteria were previously described (Huijskens *et al.*, 2012), see also Supplementary Information S1. We studied the microbiota composition of the oropharynx in 100 patients aged $\geqslant 60$ years ('elderly pneumonia') and 27 patients aged <60 years ('adult pneumonia'). For all pneumonia patients, the pneumonia severity index (PSI) was assessed and used to stratify patients in mild (PSI 1–3), moderate (PSI 4) and severe pneumonia (PSI 5; Fine *et al.*, 1997). We additionally studied the microbiota composition of oropharyngeal samples obtained from 91 healthy community-dwelling elderly living in the western part of the Netherlands who participated in a pneumococcal carriage study ('healthy elderly'; https://clinicaltrials.gov, NCT00744263).

Finally, we studied the bacterial community profiles of 187 healthy young adult parents ('healthy adults') participating in a bacterial carriage study and living in the western part of the Netherlands (https://clinicaltrials.gov, NCT00189020; van Gils *et al.*, 2009). All samples were taken after the introduction of the seven-valent pneumococcal vaccine in the national immunization program in newborns (June 2006) and before the start of the CAPiTA-trial evaluating the efficacy of the 13-valent pneumococcal vaccine in elderly for prevention of CAP. Furthermore, all four groups were matched for season (that is, all inclusions took place between November and February).

Sampling and storage methods

Oropharyngeal samples of patients were taken before hospitalization at the emergency ward and samples of controls were taken during routine visits. Samples were obtained and transported according to protocol (Huijskens *et al.*, 2012). Samples were subsequently frozen at $-70\,^{\circ}C$ until further analyses.

Microbiota composition analysis

Bacterial DNA isolation and quantification. DNA was isolated from 200 µl sample as previously described (Wyllie *et al.*, 2014), eluded in two aliquots of 25 µl elution buffer and stored at $-20\,^{\circ}C$ until further analyses. Bacterial DNA was quantified for each sample by quantitative PCR (qPCR) as previously described (Bogaert *et al.*, 2011; Biesbroek *et al.*, 2012).

Amplicon library preparation. Generation of a PCR amplicon library was performed by amplification of the 16S ribosomal RNA gene V5–V7 hypervariable region as previously described (Bogaert *et al.*, 2011), except templates containing $\leqslant 10\,pg\,\mu l^{-1}$ DNA were cycled 35 times instead of 30 times for adequate amplicon recovery. The library was sequenced in three runs with the 454 GS-FLX-Titanium Sequencer

(Life Sciences, Roche, Branford, USA). These sequence data have been submitted to the NCBI GenBank database under accession number SRP055536.

Data processing. The raw sequences obtained were processed and classified using QIIME version 1.7 (Caporaso *et al*., 2010). Barcodes and primers were trimmed off, chimeric sequences were identified and removed using chimeraslayer, and the sequences were checked for their quality. Sequences with $\geqslant 1$ error in the primer or >1 error in the barcode, a sequence length of <200 and >1000 base pairs (bp), average quality window score <25, >6 ambiguous bp or >6 homopolymers were removed from the database. Sequences were clustered into operational taxonomic units (OTUs) using UCLUST at 97% similarity. Aligned sequences were taxonomically classified using the Bayesian RDP classifier. For each sample, α-diversity was determined by calculating diversity indices (number of observed species and Shannon) at a rarefaction depth of 1100 sequences (Magurran, 2004).

Nucleic acids extraction and qPCR for viruses and bacteria

qPCR was used to test for the presence of respiratory viruses and bacteria, including human bocavirus, polyomaviruses WU and KI, respiratory syncytial virus A and B, human influenza virus A and B, parainfluenza virus 1–4, human rhinoviruses, adenoviruses, human coronavirus OC43, NL63, HKU and 229E and human metapneumovirus. Nucleic acids were extracted from the oropharyngeal samples as previously described (van de Pol *et al*., 2007; 2009). Details on the probes, primers and PCR assay are described elsewhere (Huijskens *et al*., 2012). For the identification of *Streptococcus pneumoniae*, a qPCR was used targeting a conserved region of the autolysin (*lytA*) gene (Carvalho *et al*., 2007).

Statistical analysis

Differences in microbial composition between (sub) groups were assessed using significance analysis of microarrays in TIGR MeV (Tusher *et al*., 2001). Subsequent analyses were performed in SPSS or R using either Student's *t*-tests or Mann–Whitney *U*-tests for continuous data and χ^2-tests for categorical data. A multivariate linear regression model was used to explore the independent association between microbiome composition and pneumonia, correcting for potential drivers of microbiota changes including comorbidity (COPD, heart failure and immunocompromised status (ICS)) and smoking (Wang *et al*., 2012). The correlation between relative abundance of OTUs and age was assessed using corrected Spearman rank coefficients.

Nonmetric multidimensional scaling plots based on various distance measures were used to visualize differences between groups and statistical significance of these differences in the overall microbial community composition was calculated using PERMANOVA. Hierarchical clustering, average linkage dendrograms were built and visualized using iTol (Letunic and Bork, 2011) and the optimal number of clusters was estimated using the Silhouette index (Maechler *et al*., 2014).

Random forest analysis was performed to select microbial community members most important in discriminating between health and disease. Selected OTUs were included in a decision tree model; specificity and sensitivity were assessed over all study groups, for the elderly and adult cohorts separately and a *k*-fold cross-validation was performed.

In order to study significant, nonrandom copresence or mutual exclusion of various OTUs, a network was inferred by CoNet (Cytoscape plugin; Faust *et al*., 2012), the network was subsequently clustered using the Markov clustering algorithm in ClusterMaker (Cytoscape plugin) and analyzed using NetworkAnalyzer (Cytoscape plugin). For details on the statistical analysis, see Supplementary Information S1.

Results

Characterization of study population

A total of 405 subjects were analyzed; characteristics of the study population are depicted in Table 1. In general, patient characteristics were similar between groups, except for several risk factors known to be associated with pneumonia, such as smoking (in elderly and young adults), COPD, heart failure (in elderly) and age (in adults; Jackson *et al*., 2004).

Characterization of microbiota analysis

A total of 1 528 288 sequences were obtained over all four cohorts (mean 3774 ± 1179 sequences per sample), classified into 243 OTUs (excluding singletons), representing 25 taxonomic phyla, with most prominent Firmicutes (58.8%), Actinobacteria (18.0%), Bacteroidetes (10.9%), Proteobacteria (8.4%) and Fusobacteria (3.1%), together accounting for 99.1% of all sequences. Sequence depth was sufficient given a Good's coverage of 0.938 (median 0.949, range 0.833–0.991). We observed a core microbiome (defined as OTUs present in $>80\%$ of individuals) of 12 out of 243 OTUs (Supplementary Table S1).

Microbiota composition in pneumonia patients and healthy controls

We observed significant differences in the overall microbiota composition between elderly patients with pneumonia and elderly controls (Bray–Curtis dissimilarity, unadjusted $R^2 = 0.096$, $P < 0.0005$), as

Table 1 Demographics and clinical parameters of elderly and adult controls and pneumonia cases

	Healthy elderly	*Elderly pneumonia*	*Healthy adults*	*Adult pneumonia*
No. of samples (% of total)	91 (22.5%)	100 (24.7%)	187 (46.2%)	27 (6.7%)
Inclusion period				
From	December 2007	November 2008	September 2007	November 2008
To	January 2008	February 2009	January 2008	February 2009
Age-mean (± s.d.) years	75.3 (6.3)[a]	75.7 (8.6)[a]	**34.4 (4.1)**[a]	**46.4 (10.0)**[a]
Male sex	49 (53.8%)[b]	61 (61.0%)[b]	**19 (10.2%)**[b]	**15 (55.6%)**[b]
Smoking				
No. smoking	**7 (7.7%)**[b]	**35 (35.0%)**[b]	**22 (11.8%)**[b]	**14 (51.9%)**[b]
No data	0 (0.0%)	5 (5.0%)	0 (0.0%)	0 (0.0%)
Comorbidity				
COPD	**16 (17.6%)**[b]	**42 (42.0%)**[b]	NA	9 (33.3%)
Heart failure	**6 (6.6%)**[b]	**39 (39.0%)**[b]	NA	2 (7.4%)
Diabetes mellitus	8 (8.8%)[b]	18 (18.0%)[b]	NA	1 (3.7%)
Malignancy	NA	18 (18.0%)	NA	1 (3.7%)
Disease for which radiation/chemotherapy <5 years	5 (5.5%)	NA	NA	NA
ICS[c]	7 (7.7%)[b]	8 (8.0%)[b]	NA	1 (3.7%)
Recent antibiotics usage[d]	0 (0.0%)[e]	0 (0.0%)[f]	11 (5.9%)[g]	0 (0.0%)[f]
Hospitalized after diagnosis	NA	96 (96.0%)	NA	26 (96.3%)

Abbreviations: COPD, chronic obstructive pulmonary disease; ICS, immunocompromised status; NA, not available. Statistically significant differences in variables between healthy elderly and elderly pneumonia patients as well as between healthy adults and adult pneumonia patients were calculated using [a]Student's *t*-test or [b]χ^2-test for continuous and categorical data, respectively. [c]ICS, defined as clinically suspected or proven immunodeficiency, the use of immunosuppressive therapy or immunomodulating medication in the past 3 months, or the use of more than 10 mg prednisone or equivalent each day for the past 3 months. [d]Recent antibiotics usage was slightly differently defined per cohort; [e]use of antibiotics at the moment of sample collection; [f]use of antibiotics <2 weeks before admission and [g]use of antibiotics <4 weeks before sampling.
Data reported are mean ± s.d. or number (percentage). Statistically significant differences defined as *P*-value <0.05 are depicted in bold text.

well as between adult pneumonia patients and their healthy controls (unadjusted $R^2 = 0.081$, $P < 0.0005$). These differences were also observed using Jensen–Shannon and weighted Unifrac as distance measures ($P < 0.0005$ for all comparisons; Figure 1, Supplementary Figure S1) and were related to shifts in relative abundance of multiple predominant OTUs (Supplementary Table S1).

We observed a significantly higher relative abundance of *Streptococcus (pseudo)pneumoniae,* several *Streptococcus* OTUs, *Rothia* and a lower abundance of Gemellales, *Prevotella melaninogenica*, *Veillonella dispar*, *Parascardovia* and *Leptotrichia* in elderly pneumonia patients compared with elderly controls. Highly similar differences between health and disease were observed for the young adult cohort. Using a multivariate analysis, we confirmed that these differences in microbiota profiles associated with health status were independent of other explanatory variables (comorbidities and smoking; Supplementary Table S2). Additional independent associations between microbiota composition and explanatory variables were only observed for ICS, which was related to increased abundance of *Haemophilus*, *Parascardovia* and *Clostridiaceae* (Supplementary Table S2). The data set lacked the statistical power to assess the associations between disease severity and microbiome composition.

Given the low resolution of 16S-based pyrosequencing regarding discrimination of streptococcal species, we verified the contribution of *Streptococcus pneumoniae* in our study cohorts by qPCR; we observed a positive C_T-value (<40 cycles) for *lytA* in 56% of elderly pneumonia patients respectively, vs 17% in healthy elderly (χ^2-test, $P < 0.0005$).

Microbiota composition in young and elderly controls

Because of the increased susceptibility of elderly to respiratory infections in general, we hypothesized that the trends observed between health and disease would possibly also exist when comparing young vs elderly healthy individuals. We indeed confirmed the negative association between the relative abundance of *Prevotella*, *Veillonella*, *Leptotrichia*, and positive associations between *Rothia* and *Lactobacillus* and elderly age when comparing healthy adults and elderly (Figure 2). The observed age-related dynamics of the above species could be verified within the healthy elderly cohort, confirming moderate–strong correlations between the relative abundance of *Prevotella*, *Leptotrichia* and *Rothia*, and age (Spearman rank $r = -0.19$, -0.35 and 0.25, $P = 0.07$, 0.0008 and 0.02, respectively).

Diversity and bacterial density in relation to health and pneumonia

With regard to ecological diversity measures, species richness was decreased in elderly pneumonia patients compared with healthy elderly (mean 120 (95% confidence interval (CI); 113–127) vs 159 (95% CI; 148–170), respectively, Student's *t*-test $P < 0.0001$), whereas Shannon diversity indices were increased (mean 4.5 (95% CI; 4.4–4.6) vs 4.0 (95% CI; 3.7–4.2), respectively, $P = 0.0002$). These differences were not detected in the young adult cohort.

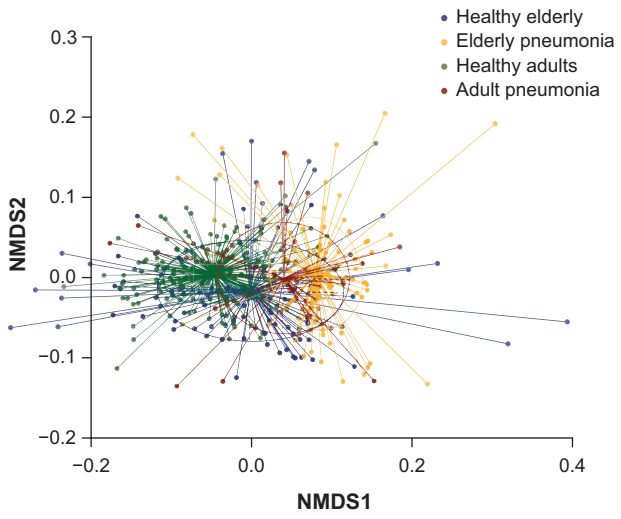

Figure 1 Two-dimensional nonmetric multidimensional scaling (nMDS) plot of the oropharyngeal microbiome composition in adult and elderly pneumonia patients and healthy controls based on the Weighted Unifrac distance measure. Each dot represents the total oropharyngeal microbiome composition of a single individual and was colored according to group; healthy elderly in blue, elderly pneumonia patients in yellow, healthy adults in green and adult pneumonia patients in red. Ellipsoids represent the standard deviation per cohort. Data points are positioned in such a manner that the two-dimensional plot represents the multidimensional data structure the best way possible (that is, (dis)similarities in microbiome structure between individuals are conserved). The microbiome structure was well captured in this nMDS visualization (stress = 0.15).

As an additional measure of dysbiosis, we determined the overall bacterial density of the samples, which was significantly higher in elderly pneumonia patients as opposed to elderly controls (geomean 413 vs 193 pg μl^{-1}, $P = 0.001$), supporting our hypothesis that bacterial replication is less controlled during pneumonia. A similar, but stronger difference in density was observed for the younger adult cohorts, with ultimately the lowest bacterial density being observed in healthy young adults (geomean 38 pg μl^{-1}; Figure 3). In addition, within the pneumonia groups, we observed a significantly higher bacterial density in patients suffering from a moderate/severe pneumonia compared with mild pneumonia cases (geomean 517 vs 246 pg μl^{-1}, $P = 0.02$) supporting bacterial overgrowth being associated with (severity of) disease (Supplementary Figure S2).

Sample clustering analysis based upon phylogeny

We used weighted Unifrac distance measures to cluster samples from all four cohorts based on phylogenetic (dis)similarity (Figure 4). Only clusters comprising more than four samples were considered for subsequent analyses, resulting in 11 (sub)clusters including 95% (386/405) of samples. We observed three microbiota profiles that were strongly associated with pneumonia (χ^2-test $P < 0.05$); one smaller cluster (IV, $n = 11$) was characterized by a high mean relative abundance of lactobacilli (relative abundance 45% vs 1% in the remaining clusters), with 10/11 (91%) individuals within this cluster suffering from pneumonia (including nine severe pneumonia cases). A second and larger cluster (Va, $n = 51$) was characterized by relative overgrowth of *Rothia* (relative abundance 34% vs 7% in the remaining clusters), with 40/51 (78%) individuals having pneumonia. A third cluster (VII) was typified by predominance of *Streptococcus (pseudo)pneumoniae* (relative abundance 56% vs 13% in the remaining clusters)—4/42 (81%) of the individuals in this cluster were pneumonia patients. Subsequently, three clusters were correlated with health (χ^2-test, $P < 0.05$); two clusters (IIIb, $n = 76$ and IIIa, $n = 7$) showed high relative abundance of *Prevotella melaninogenica* (respectively, 16% and 28%, vs 3% in the remaining clusters). In addition to a high relative abundance of *Prevotella ssp*, cluster IIIb showed a relative increase of *Leptotrichia* (4% vs 1%). Within these two clusters, 80/83 (96%) subjects were healthy individuals. A third large cluster (VIII, $n = 100$) showed a more even profile, although this cluster was also relatively enriched for *Prevotella* and *Leptotrichia* spp.; 98/100 (98%) of the individuals in this cluster were healthy. Five remaining clusters dominated by *Neisseria* (Ia, $n = 6$ and Ib, $n = 24$), Gemellales (VI, $n = 38$), *Actinomyces* (II, $n = 11$) and another *Rothia* spp. (Vb, $n = 20$), respectively, showed no clear association with either health or disease (χ^2-test, $P \geqslant 0.05$).

Association between respiratory viruses and microbiota clusters

Overall, more viruses were present in the young adult and elderly pneumonia cohorts as compared with the elderly controls (15% and 35% vs 6%, χ^2-test, $P < 0.0005$ and $P = 0.075$, respectively; Table 2). Viral presence was more frequently observed in the pneumonia-associated clusters, that is, the *Lactobacillus*-dominated cluster IV (5/10 positive for any virus (50%), one missing), *Rothia*-dominated cluster Va (15/40 positive (38%), 11 missing) and the *Streptococcus (pseudo)pneumoniae*-dominated cluster VII (6/31 positive (19%), 11 missing; Figure 4).

Network analysis

A network was constructed to study the co-occurrence and -exclusion of various OTUs, which yielded a total of 11 distinct OTU clusters, the largest of which consisted of 125 nodes (79% of total number of nodes), interconnected with 777 edges (95% of total number of edges; cluster 1). Within this cluster, we observed strong negative associations between important community members associated with health (*Prevotella* and *Leptotrichia*; subclusters 1A) and species associated with pneumonia (*Streptococcus (pseudo)pneumoniae*, *Rothia* and *Streptococcus*, subclusters 1B–D (Supplementary Figure S3).

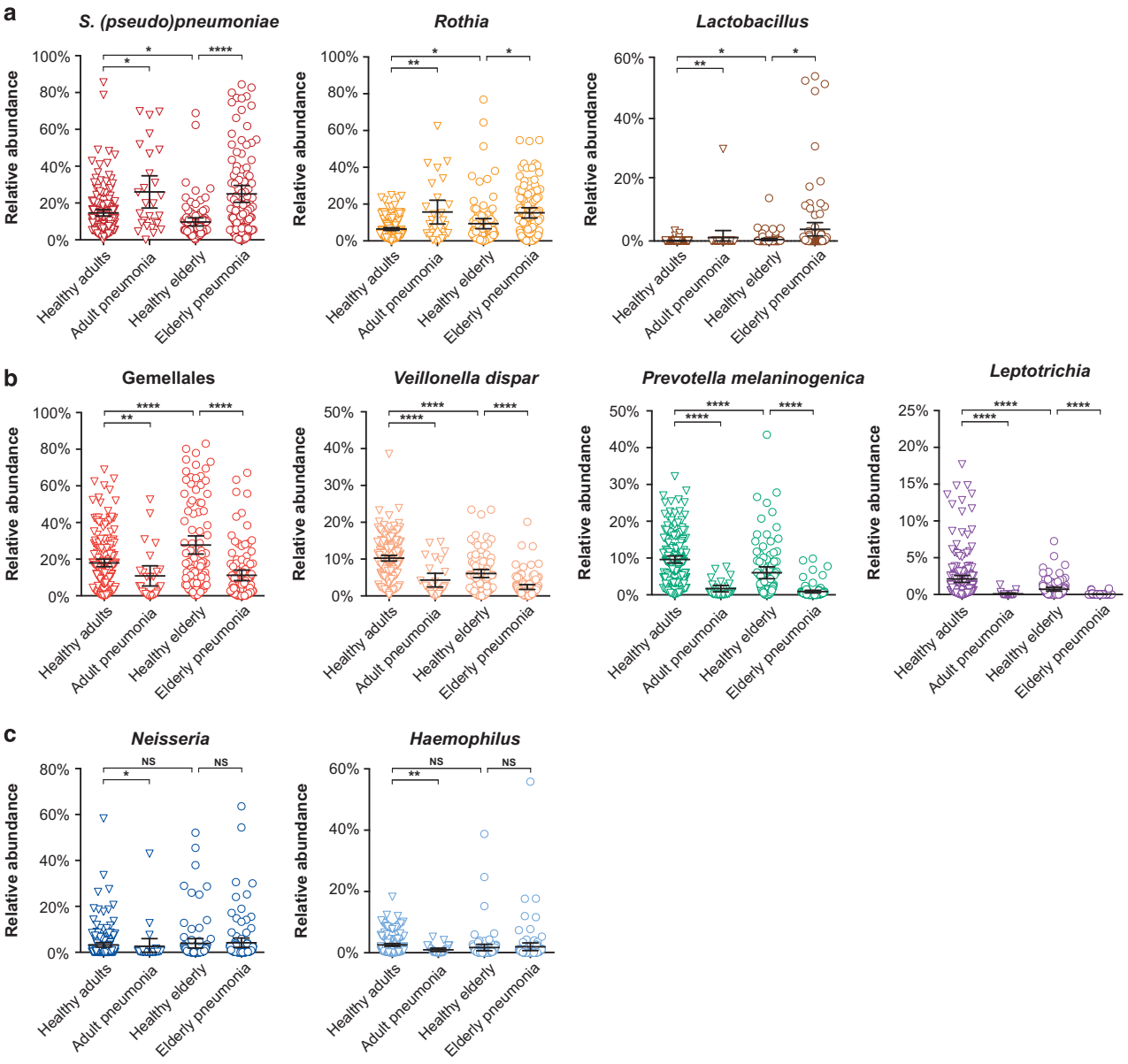

Figure 2 Patterns in relative abundance of frequently occurring OTUs in elderly and adult controls and pneumonia patients. Relative abundance (percentage) is depicted per individual for Gemellales, *Streptococcus (pseudo)pneumoniae*, *Neisseria*, *Rothia*, *Veillonella*, *Prevotella melaninogenica*, *Haemophilus*, *Leptotrichia* and *Lactobacillus*. Statistically significant differences between groups (healthy elderly vs elderly pneumonia (circles), healthy adults vs adult pneumonia (triangles) and healthy adults vs healthy elderly) were calculated using significance analysis of microarrays with false discovery rate correction for multiple tests. $*q \leqslant 0.05$; $**q \leqslant 0.01$; $***q \leqslant 0.001$; $****q \leqslant 0.0001$; NS, not significant. Bars represent geomean and 95% CI of geomean per group. OTUs are color coded based on the phylum level; Firmicutes, red; Actinobacteria, yellow; Bacteroidetes, green; Proteobacteria, blue; and Fusobacteria, purple.

Decision tree model to differentiate healthy controls from pneumonia patients

A random forest analysis including the 25 highest ranking OTUs was performed to identify bacterial community members most potent in partitioning the total study population in healthy individuals and pneumonia patients. Seven OTUs, *Prevotella melaninogenica* OTU rank 5, *Leptotrichia, Streptococcus* OTU rank 8, *Veillonella dispar*, *Parascardovia*, Gemellales and *Prevotella* OTU rank 20, appeared most discriminative, and were therefore included in a decision tree model. The decision tree was trained on the total study cohort and the final model was based on five OTUs; *Prevotella melaninogenica*, *Leptotrichia*, *Streptococcus*, Gemellales and *Parascardovia* (two OTUs were not included; Figure 5). We calculated a specificity of 95% and sensitivity of 84% for the decision tree based on the total population. To rule out skewing of specificity by the relatively high contribution of correctly classified healthy adults, the decision tree was validated separately for the young adult and elderly study

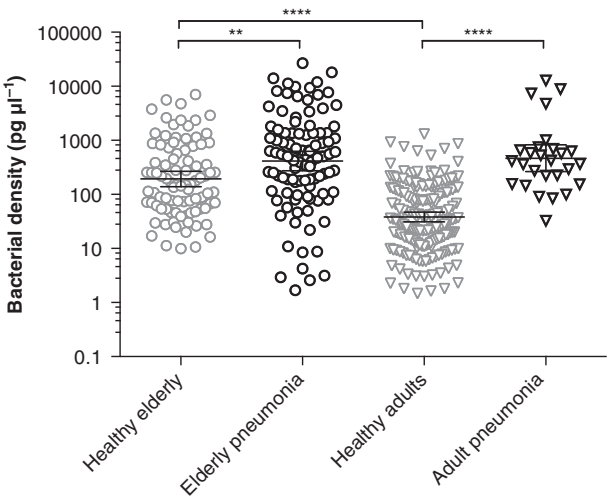

Figure 3 Bacterial density in the oropharynx of elderly and adult controls and pneumonia patients. For each group, the bacterial density per sample is depicted in pg μl^{-1}. Bars represent geomean and 95% CI of geomean per group. Statistically significant differences between groups (healthy elderly vs elderly pneumonia (circles), healthy adults vs adult pneumonia (triangles) and healthy adults vs healthy elderly) were calculated using Mann–Whitney *U*-tests; $*P \leqslant 0.05$; $**P \leqslant 0.01$; $***P \leqslant 0.001$; $****P \leqslant 0.0001$. Healthy subjects are depicted in light gray, patients are shown in dark gray.

population. Also after stratification, we confirmed that for both the within-elderly cohort and the within-young adult cohort analysis, a high specificity was retained (Table 3).

Decision tree models are known for their risk of overfitting (Breiman, 1984) and therefore a fivefold cross-validation was used to more accurately estimate the discriminative power of the model, resulting in a specificity of 89% and sensitivity of 73% (Table 3). A cross-validation estimate for the specificity and sensitivity per age group was not performed due to the low numbers of individuals in especially the adult pneumonia cohort.

Discussion

The URT can be regarded as an ecological niche in which a dynamic interplay takes place between both commensals and potential disease causing bacteria. If balanced, the microbiome is generally advantageous to the host by direct or immune-mediated colonization resistance, which refers to the resilience against invasion and dissemination of pathogens (Buffie and Pamer, 2013). We postulate that the disequilibrium of oropharyngeal microbiota contributes to decreased colonization resistance and reduced containment of potential pathogens, leading to dissemination and subsequent development of pneumonia.

We chose to type the oropharyngeal microbiota in both elderly and young adult pneumonia patients and asymptomatic community-dwelling controls using high-throughput sequencing as it is generally assumed that the pathogenesis of pneumonia is initiated in the URT due to successive microaspiration (Krone *et al*., 2014). Although *Streptococcus pneumoniae* is the most well-known pathogen responsible for pneumonia, our study suggests that *Rothia* may additionally have a role in the pathogenesis of pneumonia. *Rothia* is unknown as a classical pathogen for CAP, though is capable of causing a diverse array of disease entities among which pneumonia, particularly in immunocompromised hosts (Ramanan *et al*., 2014). Furthermore, the presence and density of *Rothia* in the URT has been linked to an increased risk of otitis media in children (OR 1.35; 1.03–1.77; Laufer *et al*., 2011), also suggesting a potential role for this bacterium in the pathogenesis of respiratory infections in general.

Beside several bacterial OTUs associated with disease, we showed a decreased abundance of several Gram-negative anaerobic bacteria including *Prevotella*, *Veillonella and Leptotrichia* and the Gram-positive genus *Parascardovia* in pneumonia patients. *Prevotella* and *Leptotrichia* commonly reside in the oral cavity (Zaura *et al*., 2009), gut (Wu *et al*., 2011) and vagina (Eribe and Olsen, 2008; Shipitsyna *et al*., 2013). *Prevotella* is associated with a reduced risk of hospital-acquired pneumonia in adult ICU-patients (Bousbia *et al*., 2012), and was reported to be reduced in the oropharynx of adults and children with asthma or COPD (Hilty *et al*., 2010). *Leptotrichia* is a lactic acid-producing bacterium of the *Fusobacteriaceae* and is more commonly observed in healthy nonsmoking individuals (Charlson *et al*., 2010). We observed *Veillonella* was more prevalent in healthy controls, which is in line with observations regarding periodontal health (Kumar *et al*., 2005). *Parascardovia* has been isolated from saliva, gut and human breast milk (Beighton *et al*., 2008; Solís *et al*., 2010) and belongs to the family of *Bifidobacteriaceae*, members of which have been linked to the health-promoting effects of breast milk (Solís *et al*., 2010).

Interestingly, we observed a similar decline in anaerobic colonization with increasing age when only healthy individuals were analyzed, suggesting that the absence of these anaerobic bacteria might be associated with age-related susceptibility to LRTI. Although no human studies are yet available to confirm the age-related changes in (upper) respiratory microbiota, previous research in mice supports these findings; young mice showed a higher relative abundance of Bacteroidetes spp. in the URT compared with elderly mice, and additionally showed a faster and larger Bacteroidetes expansion in response to colonization and clearance of *Streptococcus pneumoniae* (Krone *et al*., 2014).

Although the similar age-related shifts in microbiota composition support the hypothesis that our findings are related to susceptibility rather than consequence of disease, we can not fully rule out the alternative hypothesis, that is, that the observed microbiome profile changes observed both in

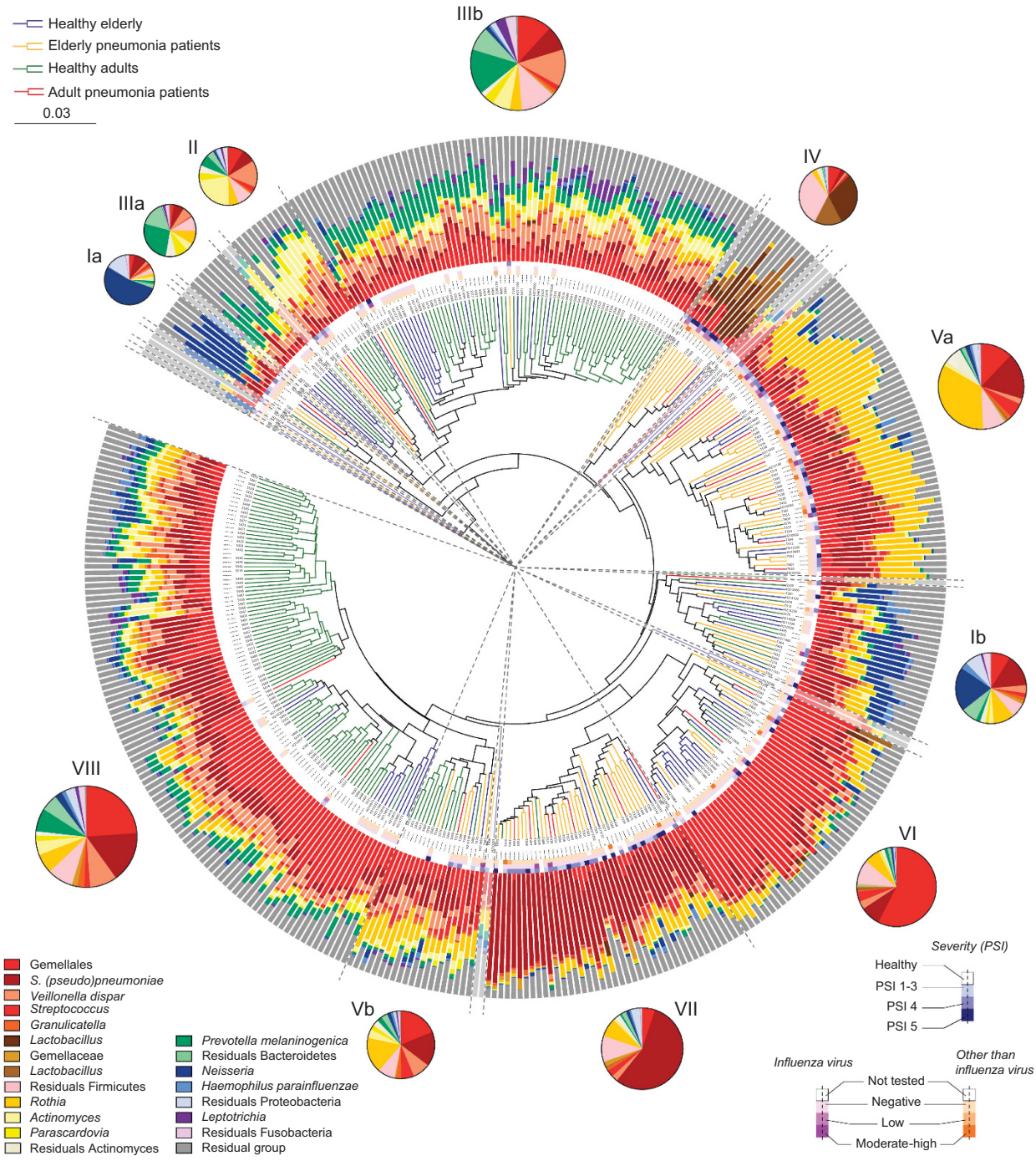

Figure 4 Weighted Unifrac (phylogenetic) average linkage hierarchical clustering analysis in elderly and adult healthy controls and pneumonia patients. The weighted Unifrac distance matrix was used to construct a circle dendrogram including all 405 individuals; healthy elderly (blue branches), elderly pneumonia patients (yellow), healthy adults (green) and adult pneumonia patients (red). Adjacent to the dendrogram branch ends, viral data and information on disease severity are visualized and stacked bar charts show the relative abundance of the 15 highest ranked OTUs. Color coding: presence of influenza virus and non-influenza virus is shown in purple and orange respectively; dark purple/orange, virus present in moderate-high amount (C_t-value <35 cycles); moderate light purple/orange, virus present in low amount (C_t-value 35–45 cycles); light purple/orange, viral qPCR performed and tested negative (C_t-value >45 cycles); white, no viral qPCR performed. Clusters (defined as more than four samples; bright) are designated by dotted gray lines originating from the center of the dendrogram. Groups of samples comprising less than four samples are shown opaque. Three clusters were significantly associated with health; these clusters showed predominance of *Prevotella melaninogenica* (IIIa and IIIb) and *Leptotrichia* (IIIb) or showed a balanced microbial community (VIII). Clusters that were correlated with disease showed high relative abundance of either lactobacilli (IV), *Rothia* (Va) or *Streptococcus (pseudo)pneumoniae* (VII). Cluster IV was most strongly associated with the presence of influenza virus (50%). Five clusters showed no association with either health or disease, these clusters showed high relative abundance of *Neisseria* (Ia and Ib), *Actinomyces* (II), Gemellales (VI) or *Rothia* (Vb). Per cluster, the mean relative abundance of the 15 highest ranked oropharyngeal OTUs and the residual bacteria separated by phylum is given in circle graphs that were size-scaled according to the number of individuals within each cluster. PSI, pneumonia severity score.

Table 2 Viral detection rates in oropharyngeal samples taken from elderly and adult controls and pneumonia patients

	Healthy elderly	*Elderly pneumonia*	*Healthy adults*	*Adult pneumonia*
qPCR resp. viruses (no. available (%))	88 (96.7%)	100 (100.0%)	NA	27 (100.0%)
Influenza virus (A/B)	**0 (0.0%)**[a]	**19 (19.0%)**[a]	NA	2 (7.4%)
RSV	2 (2.2%)	5 (5.0%)[b]	NA	0 (0.0%)
Rhinovirus	0 (0.0%)	4 (4.0%)	NA	0 (0.0%)
Coronavirus	2 (2.2%)	7 (7.0%)[c]	NA	1 (3.7%)[c]
PIV	0 (0.0%)	4 (4.0%)	NA	1 (3.7%)
Other viruses[d]	1 (1.1%)	5 (5.0%)	NA	1 (3.7%)
Any virus	**5 (5.5%)**[a]	**34 (35.1%)**[a,b]	NA	4 (14.8%)
Multiple viruses (⩾2)	0 (0.0%)	6 (6.0%)	NA	1 (3.7%)

Abbreviations: PIV, parainfluenza virus; q-PCR, quantitative-PCR; RSV, respiratory syncytial virus; NA, not available. [a]Viral presence was determined by qPCR on oropharyngeal samples; a χ^2-test was used to determine statistically significant differences between groups (bold values). [b]Three missing. [c]One missing. [d]Detailed results on other respiratory viruses detected by qPCR; healthy elderly, bocavirus ($n=1$); Ki polyomavirus ($n=1$); Wu polyomavirus ($n=1$); human metapneumovirus ($n=2$); adult pneumonia patients: bocavirus ($n=1$). Number (percentage) was reported.

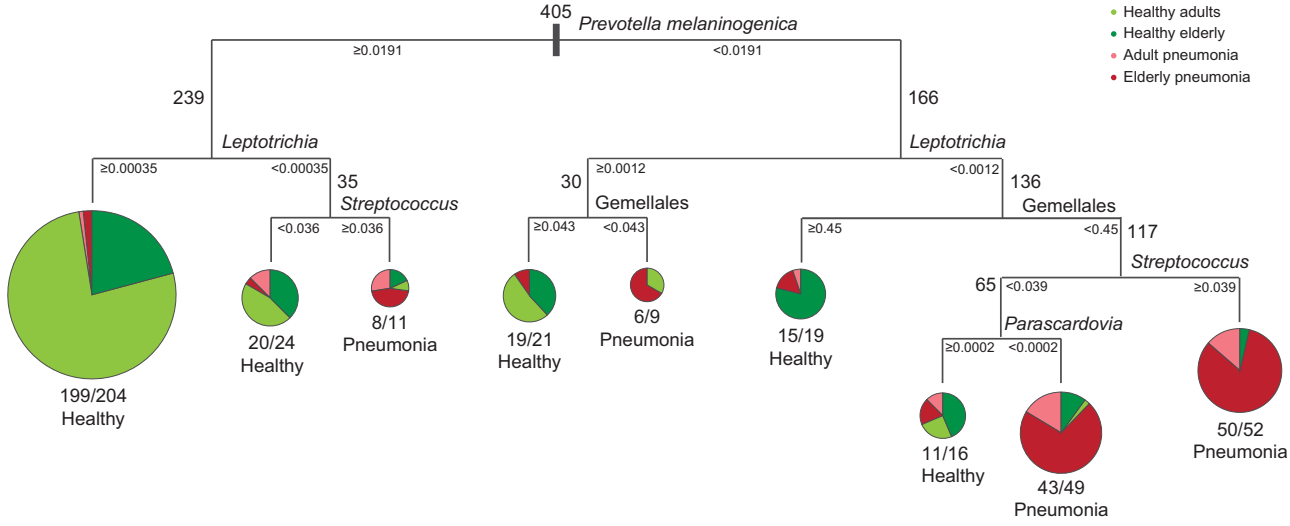

Figure 5 Decision tree model to distinguish healthy individuals from pneumonia patients based on URT microbiota composition. On the basis of random forest analysis, seven variables (*Prevotella melaninogenica* OTU rank 5, *Leptotrichia*, *Streptococcus* OTU rank 8, Gemellales, *Parascardovia*, *Prevotella* OTU rank 20 and *Veillonella dispar*) were found to be most important in the discrimination between healthy controls and pneumonia patients. Measures of impurity (Gini- and information indices) were used to identify the most discriminative OTUs based on the random forest analysis. Subsequently, these OTUs were used to construct a decision tree model, which was based on the relative abundance of *Prevotella melaninogenica*, *Leptotrichia*, *Streptococcus*, Gemellales and *Parascardovia*. From the root of the tree, each branch represents a division of the initial group based on the relative abundance (shown in small font) of an OTU. The numbers (regular font size) neighboring the branches depict the number of individuals before each split. Adjacent to the branch ends are circle graphs illustrating the number of individuals (represented by the surface of the circles, colored by the study group they originate from) that are stratified by the prior split. Light green, healthy adults; dark green, healthy elderly; light red, adult pneumonia and dark red, elderly pneumonia. The numbers below the graphs represent the exact proportions. The combined presence and abundance of *Prevotella* and *Leptotrichia* appears to be the strongest determinant for respiratory health.

pneumonia patients and with increasing age might be the consequence rather than the cause of infection. This could, for example, be explained by inflammation-induced acidification of the local environment of the oropharyngeal niche (Lardner, 2001), which has been related to a reduction of members of the Bacteroidetes phylum and outgrowth of acidophilic species like lactobacilli (Duncan *et al.*, 2009). Therefore, new, preferably longitudinal studies should be executed to elucidate a possible role of the observed species in relation to increased or decreased risk of developing respiratory infections.

Strengths of our study include the large cohort size including pneumonia patients of different age groups with season- and age-matched asymptomatic controls, and sampling executed according to identical protocols. Also our data did not seem to be skewed owing to antibiotic usage before admission, probably as a consequence of the restrictive antimicrobial prescription policy in the Netherlands. Finally, we were able to execute age-dependent analyses within the subgroups allowing for studies in both the general population as well as a major risk group for pneumonia: elderly age.

Table 3 Results decision tree model

	Healthy elderly	*Elderly pneumonia*	*Healthy adults*	*Adult pneumonia*
Correctly classified (*n*, (%))	82 (90.1%)	89 (89.0%)	182 (97.3%)	18 (66.7%)
Incorrectly classified (*n*, (%))	9 (9.9%)	11 (11.0%)	5 (2.7%)	9 (33.3%)
	Elderly cohort[a]		**Adult cohort**[a]	
Specificity (95% CI)	90% (82–95%)		97% (94–99%)	
Sensitivity (95% CI)	89% (81–94%)		67% (46–83%)	
	Combined[b]			
Specificity (95% CI)	95% (92–97%)			
Sensitivity (95% CI)	84% (76–90%)			
	Fivefold cross-validation combined[c]			
Specificity (95% CI)	89% (88–91%)			
Sensitivity (95% CI)	73% (70–76%)			

Abbreviation: CI, confidence interval. [a]The specificity and sensitivity of the model were calculated for the elderly and adult cohorts separately. [b]The specificity and sensitivity of the model were calculated for the total data set. [c]Since these measures of diagnostic accuracy are likely overestimating the true discriminative power of the model, a fivefold cross-validation was performed. A decision tree model (Figure 5) was trained on the total study population and included *Prevotella melaninogenica*, *Leptotrichia*, *Streptococcus*, Gemellales and *Parascardovia*. Fivefold cross-validation was not performed per study group due to the small number of individuals, specifically in the adult pneumonia cohort, which potentially can hamper a proper estimation of diagnostic potency.

Potential limitations of our study include the relatively high number of 'healthy' non-pneumonia elderly suffering from one or more comorbidities, although this seems to reflect the normal population distribution of comorbidities in elderly, and, moreover, we thoroughly assessed the independent effect of pneumonia, comorbidities and smoking on the oropharyngeal microbial composition using a multivariate linear regression model. Our healthy adult cohort consisted solely of parents of young children, potentially limiting generalizability to the general population. Nevertheless, oropharyngeal microbiota profiles in our young adult cohort largely resembled those observed in the Human Metagenome Project (Segata *et al*., 2012). Owing to the 24/7 setting, sampling was performed by multiple different caregivers in different hospitals and home-based settings, which could have introduced a sampling bias; however, we tried to prevent this type of bias by confirming our observations using within-group analyses. Last, as stated, the cross-sectional study design precludes solid statements on the causal relationship between microbiome composition and development of pneumonia.

In conclusion, this is the first explorative study on correlations between oropharyngeal microbiota composition and health and pneumonia in large cohorts of elderly and young adult individuals. We showed that pneumonia in elderly patients is not only associated with increased abundance of well-known pathogens such as *Streptococcus pneumoniae*, but also with *Rothia* and *Lactobacillus*. Furthermore, we observed an even more profound association between the absence of anaerobic commensals and pneumonia in both young and elderly adults. In addition, we observed an increased bacterial density in oropharyngeal samples of pneumonia patients in general, and decreased bacterial richness in elderly pneumonia patients, suggesting a dysbiotic ecological niche with lack of pathogen containment. Whether the lack of 'protective' bacteria predisposes to the observed imbalance and whether the resulting dysbiosis is part of the causative mechanism as posed by Dickson *et al.* (2014), a consequence of disease, or an epiphenomenon, has yet to be determined.

Conflict of Interest

EAMS declares to have received unrestricted research support from Pfizer, grant support for vaccine studies from Pfizer and GlaxoSmithKline and fees paid to the institution for advisory boards or participation in independent data monitoring committees for Pfizer and GSK. RHV reported receiving grant support from GlaxoSmithKline and Wyeth/Pfizer for vaccine studies and consulting fees from GlaxoSmithKline. KT received grant support and consulting fees from Pfizer. DB received consulting fees from Pfizer. These grants and fees were not received for the research described in this paper. The remaining authors declare no conflict of interest.

Acknowledgements

This research has received funding from the Wilhelmina Children's Hospital Fund, ZonMW (grant 91209010) and NWO-VENI (grant 91610121). We thank Mei Ling Chu for her technical assistance and Dr Marinus JC Eijkemans for his input on the statistical analysis. This work is dedicated to the memory of Dr Reinier H Veenhoven who contributed to the planning of the design and execution of the clinical studies in healthy controls and Dr Marcel F Peeters, who was investigator in the adult pneumonia study.

Author contributions

RHV, MJB, EAMS, JWAR and DB designed the study. EGWH, GB and MRvdB wrote the study protocols. EGWH, GB, MRvdB and RHV were responsible for the recruitment of patients and collection of the samples. KT, EGWH and ALW were responsible for the conventional culturing/

qPCR on the samples. ALW was responsible for sample preparation for pyrosequencing. WAAdSP, DB and XW were responsible for the post-processing of sequences and data analysis. All authors were involved in data interpretation and drafting of the manuscript.

References

Beighton D, Gilbert SC, Clark D, Mantzourani M, Al-Haboubi M, Ali F *et al.* (2008). Isolation and identification of bifidobacteriaceae from human saliva. *Appl Environ Microbiol* **74**: 6457–6460.

Biesbroek G, Sanders EAM, Roeselers G, Wang X, Caspers MPM, Trzciński K *et al.* (2012). Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PLoS One* **7**: e32942.

Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E *et al.* (2011). Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS One* **6**: e17035.

Bousbia S, Papazian L, Saux P, Forel JM, Auffray J-P, Martin C *et al.* (2012). Repertoire of intensive care unit pneumonia microbiota *PLoS One* **7**: e32486.

Breiman L. (1984). *Classification and regression trees*. Chapman & Hall/CRC: New York, NY.

Buffie CG, Pamer EG. (2013). Microbiota-mediated colonization resistance against intestinal pathogens. *Nat Rev Immunol* **13**: 790–801.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Carvalho MDGS, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW *et al.* (2007). Evaluation and improvement of real-time PCR assays targeting lytA, ply, and psaA genes for detection of pneumococcal DNA. *J Clin Microbiol* **45**: 2460–2466.

Chaban B, Albert A, Links MG, Gardy J, Tang P, Hill JE. (2013). Characterization of the upper respiratory tract microbiomes of patients with pandemic H1N1 influenza. *PLoS One* **8**: e69559.

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R *et al.* (2010). Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One* **5**: e15216.

Cho I, Blaser MJ. (2012). The human microbiome: at the interface of health and disease. *Nat Rev Genet* **13**: 260–270.

Dethlefsen L, McFall-Ngai M, Relman DA. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* **449**: 811–818.

Dickson RP, Erb-Downward JR, Huffnagle GB. (2014). Towards an ecology of the lung: new conceptual models of pulmonary microbiology and pneumonia pathogenesis. *Lancet Resp Med* **2**: 238–246.

Duncan SH, Louis P, Thomson JM, Flint HJ. (2009). The role of pH in determining the species composition of the human colonic microbiota. *Environ Microbiol* **11**: 2112–2122.

Eribe ERK, Olsen I. (2008). Leptotrichia species in human infections. *Anaerobe* **14**: 131–137.

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J *et al.* (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* **8**: e1002606.

File TM. (2003). Community-acquired pneumonia. *Lancet* **362**: 1991–2001.

Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE *et al.* (1997). A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* **336**: 243–250.

Hansel TT, Johnston SL, Openshaw PJ. (2013). Microbes and mucosal immune responses in asthma. *Lancet* **381**: 861–873.

Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C *et al.* (2010). Disordered microbial communities in asthmatic airways. *PLoS One* **5**: e8578.

Huijskens EGW, van Erkel AJM, Palmen FMH, Buiting AGM, Kluytmans JAJW, Rossen JWA. (2012). Viral and bacterial aetiology of community-acquired pneumonia in adults. *Influenza Other Respir Viruses* **7**: 567–573.

Iwai S, Fei M, Huang D, Fong S, Subramanian A, Grieco K *et al.* (2012). Oral and airway microbiota in HIV-infected pneumonia patients. *J Clin Microbiol* **50**: 2995–3002.

Jackson ML, Neuzil KM, Thompson WW, Shay DK, Yu O, Hanson CA *et al.* (2004). The burden of community-acquired pneumonia in seniors: results of a population-based study. *Clin Infect Dis* **39**: 1642–1650.

Jansen AGSC, Rodenburg GD, van der Ende A, van Alphen L, Veenhoven RH, Spanjaard L *et al.* (2009). Invasive pneumococcal disease among adults: associations among serotypes, disease characteristics, and outcome. *Clin Infect Dis* **49**: e23–e29.

Kaplan V, Clermont G, Griffin MF, Kasal J, Watson RS, Linde-Zwirble WT *et al.* (2003). Pneumonia. *Arch Intern Med* **163**: 317–323.

Krone CL, Biesbroek G, Trzciński K, Sanders EAM, Bogaert D. (2014). Respiratory microbiota dynamics following Streptococcus pneumoniae acquisition in young and elderly mice. *Infect Immun* **82**: 1725–1731.

Krone CL, van de Groep K, Trzciński K, Sanders EA, Bogaert D. (2014). Immunosenescence and pneumococcal disease: an imbalance in host–pathogen interactions. *Lancet Resp Med* **2**: 141–153.

Kumar PS, Griffen AL, Moeschberger ML, Leys EJ. (2005). Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *J Clin Microbiol* **43**: 3944–3955.

Lardner A. (2001). The effects of extracellular pH on immune function. *J Leukoc Biol* **69**: 522–530.

Laufer AS, Metlay JP, Gent JF, Fennie KP, Kong Y, Pettigrew MM. (2011). Microbial communities of the upper respiratory tract and otitis media in children. *mBio* **2**: e00245–10.

Letunic I, Bork P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475–W478.

Leung RKK, Zhou JW, Guan W, Li SK, Yang ZF, Tsui SKW. (2013). Modulation of potential respiratory pathogens by pH1N1 viral infection. *Clin Microbiol Infect* **19**: 930–935.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. (2014). Cluster: cluster analysis basics and extensions.

Magurran AE. (2004). *Measuring Biological Diversity*. Blackwell Science: Oxford.

Mathers CD, Boerma T, Fat DM. (2008). The Global Burden of Disease: 2004 Update. World Health Organization: Geneva , Switzerland, 1–60.

Millett ERC, Quint JK, Smeeth L, Daniel RM, Thomas SL. (2013). Incidence of community-acquired lower respiratory tract infections and pneumonia among older adults in the United Kingdom: a population-based study. *PLoS One* **8**: e75131.

Molyneaux PL, Mallia P, Cox MJ, Footitt J, Willis-Owen SAG, Homola D *et al.* (2013). Outgrowth of the bacterial airway microbiome after rhinovirus exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **188**: 1224–1231.

Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL *et al.* (2013). Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med* **187**: 1067–1075.

Ramanan P, Barreto JN, Osmon DR, Tosh PK. (2014). Rothia bacteremia - A 10 year experience at Mayo Clinic, Rochester, Minnesota. *J Clin Microbiol* **52**: 3184–3189.

Sakwinska O, Bastic Schmid V, Berger B, Bruttin A, Keitel K, Lepage M *et al.* (2014). Nasopharyngeal microbiota in healthy children and pneumonia patients. *J Clin Microbiol* **52**: 1590–1594.

Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D *et al.* (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol* **13**: R42.

Shipitsyna E, Roos A, Datcu R, Hallén A, Fredlund H, Jensen JS *et al.* (2013). Composition of the vaginal microbiota in women of reproductive age–sensitive and specific molecular diagnosis of bacterial vaginosis is possible? *PLoS One* **8**: e60670.

Solís G, de los Reyes-Gavilan CG, Fernández N, Margolles A, Gueimonde M. (2010). Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe* **16**: 307–310.

Tichopad A, Roberts C, Gembula I, Hajek P, Skoczynska A, Hryniewicz W *et al.* (2013). Clinical and economic burden of community-acquired pneumonia among adults in the Czech Republic, Hungary, Poland and Slovakia Borrow. *PLoS One* **8**: e71375.

Tusher VG, Tibshirani R, Chu G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98**: 5116–5121.

van de Pol AC, van Loon AM, Wolfs TFW, Jansen NJG, Nijhuis M, Breteler EK *et al.* (2007). Increased detection of respiratory syncytial virus, influenza viruses, parainfluenza viruses, and adenoviruses with real-time PCR in samples from patients with respiratory symptoms. *J Clin Microbiol* **45**: 2260–2262.

van de Pol AC, Wolfs TFW, Jansen NJG, Kimpen JLL, van Loon AM, Rossen JWA. (2009). Human bocavirus and KI/WU polyomaviruses in pediatric intensive care patients. *Emerging Infect Dis* **15**: 454–457.

van Gils EJM, Veenhoven RH, Hak E, Rodenburg GD, Bogaert D, IJzerman EPF *et al.* (2009). Effect of reduced-dose schedules with 7-valent pneumococcal conjugate vaccine on nasopharyngeal pneumococcal carriage in children. *JAMA* **302**: 159–167.

Wang X, Eijkemans MJC, Wallinga J, Biesbroek G, Trzciński K, Sanders EAM *et al.* (2012). Multivariate approach for studying interactions between environmental variables and microbial communities. *PLoS One* **7**: e50267.

Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA *et al.* (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**: 105–108.

Wyllie AL, MLJN Chu, Schellens MHB, van Engelsdorp Gastelaars J, Jansen MD, van der Ende A *et al.* (2014). *Streptococcus pneumoniae* in saliva of Dutch primary school children. *PLoS One* **9**: e102045.

Zaura E, Keijser BJ, Huse SM, Crielaard W. (2009). Defining the healthy 'core microbiome' of oral microbial communities. *BMC Microbiol* **9**: 259.

Zhou Y, Lin P, Li Q, Han L, Zheng H, Wei Y *et al.* (2010). Analysis of the microbiota of sputum samples from patients with lower respiratory tract infections. *Acta Biochim Biophys Sin* **42**: 754–761.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)