



OPEN

Bacterial composition of nasal discharge in children based on highly accurate 16S rRNA gene sequencing analysis

Kaoru Haro^{1,2✉}, Midori Ogawa¹, Mitsumasa Saito¹, Koichi Kusuhara² & Kazumasa Fukuda¹

Nasopharyngeal colonization by bacteria is a prerequisite for progression to respiratory disease and an important source of horizontal spread within communities. We aimed to perform quantitative analysis of the bacterial cells and reveal the microbiota of the nasal discharge in children at the species level based on highly accurate 16S rRNA gene sequencing. This study enrolled 40 pediatric patients with rhinorrhea. The bacterial cells in the nasal discharge were counted by epifluorescence microscopic analysis. The microbiota was analyzed by using the 16S rRNA gene clone library sequencing method. We demonstrated that a high abundance (median 2.2×10^7 cells/mL) of bacteria was contained in the nasal discharge of children. Of the 40 samples, 37 (92.5%) were dominated by OTUs corresponding to *Haemophilus aegyptius/influenzae*, *Moraxella catarrhalis/nonliquefaciens*, or *Streptococcus pneumoniae*. These samples showed higher cell abundance and lower alpha diversity than the remaining three samples in which the other bacteria coexisted. In addition, 12 sequences with low homology to type strains were considered as previously unknown bacterial lineages. In conclusion, the nasal discharge of most young children contains a large amount of respiratory pathogens and several unknown bacteria, which could not only cause endogenous infection but also be a source of transmission to others.

The nasopharynx of children contains respiratory pathogens, such as *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Moraxella catarrhalis*¹. Nasopharyngeal colonization with *S. pneumoniae* precedes the pathogenesis of pneumococcal diseases, such as acute otitis media, pneumonia, septicemia, and meningitis particularly among children^{2–4}. In the elderly, *S. pneumoniae* is the most common cause of pneumonia and one of the most frequently isolated pathogens in patients with chronic obstructive pulmonary disease (COPD)⁵. *H. influenzae*, especially nontypeable *H. influenzae* (NTHi), is also a common bacterium associated with exacerbation of COPD and acute pneumonia^{6,7}, and colonization of the upper respiratory tract is a prerequisite for developing respiratory disease^{8,9}.

Carriage of these bacteria is also the primary mode of horizontal spread within communities². Recent work has shown that young children play a major role in pneumococcal nasopharynx carriage, acquisition, and transmission¹⁰. Other previous studies have reported that respiratory viral infections increase nasopharyngeal pneumococcal transmission^{11,12}. These findings suggest that nasal discharge can play an important role in the transmission of the bacteria colonizing the nasopharynx. In this regard, analysis of bacteria contained in the nasal discharge of children is important, since these bacteria may be transmitted among communities through nasal discharge and may occasionally cause infectious diseases depending on the condition of the host.

Recent studies conducted next-generation sequencing (NGS) have revealed age-related changes in the upper-airway microbiota¹³ and their association with respiratory tract infections^{14,15} and asthma^{16,17}. However, details about the bacterial community at the species level in nasal discharge have not been well investigated. Therefore, we performed quantitative analysis of the bacterial cells and investigated the microbiota of the nasal discharge in children at the species level based on highly accurate 16S rRNA gene sequencing.

¹Department of Microbiology, School of Medicine, University of Occupational and Environmental Health, Japan, 1-1 Iseigaoka, Yahatanishi-ku, Kitakyushu 807-8555, Japan. ²Department of Pediatrics, School of Medicine, University of Occupational and Environmental Health, Japan, Kitakyushu, Japan. ✉email: kaoru-sato@med.uoeh-u.ac.jp

Characteristics	
Age (year), median, (range)	2.1 (0.2–7.0)
Sex (male), n (%)	22 (55.0)
Antimicrobial drug use ^a , n (%)	13 (32.5)
Day care center ^b , n (%)	23 (67.6)
PCV7/PCV13 and Hib vaccinated (complete 4 doses) ^c , n (%)	24 (75.0)
Symptom	
Fever, n (%)	11 (27.5)
Cough, n (%)	36 (90.0)
Wheeze, n (%)	7 (17.5)
Diagnosis	
Acute upper respiratory tract inflammation, n (%)	22 (55.0)
Acute bronchitis, n (%)	8 (20.0)
Respiratory syncytial virus infection, n (%)	8 (20.0)
Allergic rhinitis, n (%)	1 (2.5)
Nasal congestion, n (%)	1 (2.5)

Table 1. Demographic and clinical characteristics of the subjects. ^aDefined as use of antimicrobial drug within 1 month before sampling. ^bNo information available for 6 subjects. ^cNo information available for 8 subjects.

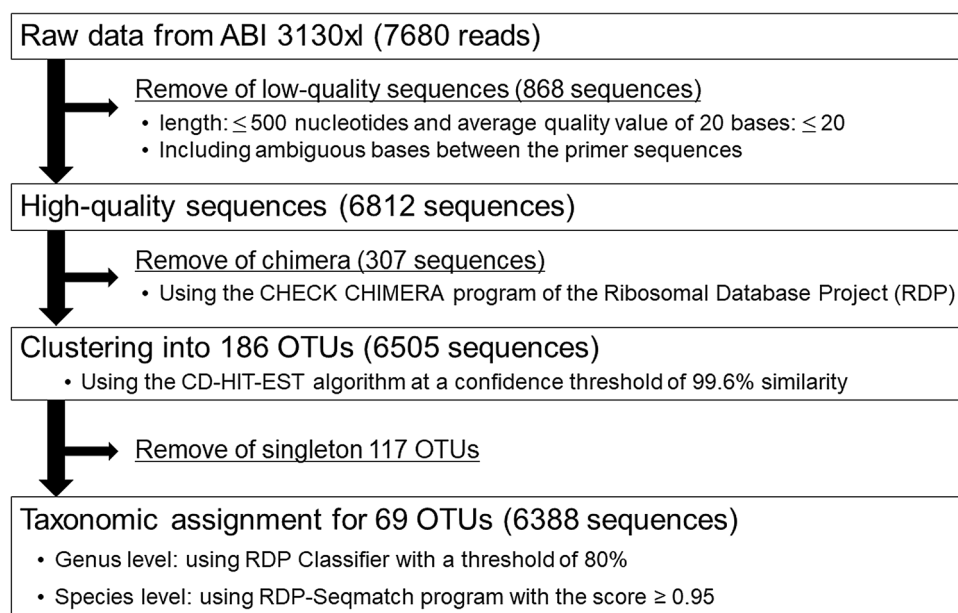


Figure 1. Flow chart of data analysis. *OTU* operational taxonomic unit.

Results

Subject characteristics. The median age of the 40 enrolled children (22 males, 18 females) was 2.1 years (range, 0.2–7.0 years). The demographic and clinical characteristics of the patients are shown in Table 1.

Total bacterial numbers. The bacterial cells counted by epifluorescence microscopic analysis of the nasal discharge are shown in Supplementary Fig. 1. The numbers of bacteria of the samples ranged from 2.5×10^6 to 3.2×10^8 cells/mL (median 2.2×10^7 cells/mL).

Microbiota analysis. We obtained 6812 high-quality sequences from 40 samples of nasal discharge (Fig. 1). Three hundred and seven chimeric sequences were removed, and 6505 sequences were clustered into 186 operational taxonomic units (OTUs), including 117 singletons, using a 99.6% sequence similarity threshold. Rarefaction curves for bacterial OTUs calculated for each sample and Good's coverage are shown in Supplementary Fig. 2. The coverage including and excluding singletons ranged from 0.90 to 1.00 (median, 0.98) and from 0.95 to 1.00 (median, 0.99), respectively. The rarefaction curves were close to saturation in most of the samples with or without singletons. Therefore, singleton OTUs were removed from taxonomic assignment, and 69 OTUs

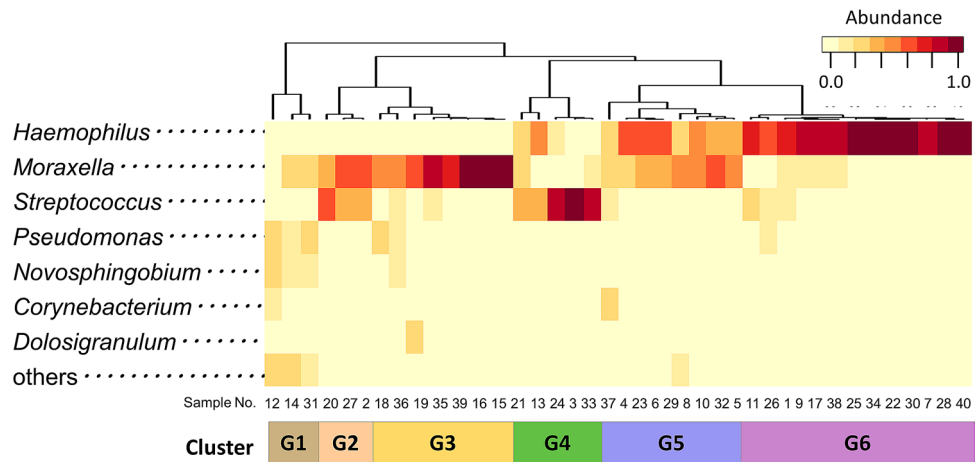


Figure 2. Clustering analysis of the OTUs at the genus level. Clustering of samples based on the relative abundance of the seven major genera and others (relative abundance < 1.0%) revealed six distinct clusters: G1, mixed; G2, *Moraxella* and *Streptococcus*-dominated; G3, *Moraxella*-dominated; G4, *Streptococcus*-dominated; G5, *Haemophilus* and *Moraxella*-dominated; G6, *Haemophilus*-dominated. Bacterial genera in the left row are arranged by frequency.

remained. From the results of RDP-Classifer analysis, 66 of the 69 OTUs were identified at the genus level, of which 57 were identified at the species level based on a Seqmatch score.

The relative abundance of all genera in each sample and the clustering results based on the composition of the genera are shown in Supplementary Fig. 3 and Fig. 2. The most abundant genera in nasal discharge were *Haemophilus* (40.1%), *Moraxella* (31.2%) and *Streptococcus* (15.4%), accounting for 86.7% of all sequences. Other fairly common genera were *Pseudomonas* (4.0%), *Novosphingobium* (3.0%), *Corynebacterium* (1.7%) and *Dolosigranulum* (1.0%). Clustering of samples based on the composition of the genera revealed six distinct clusters: G1) mixed, G2) *Moraxella* and *Streptococcus*-dominated, G3) *Moraxella*-dominated, G4) *Streptococcus*-dominated, G5) *Haemophilus* and *Moraxella*-dominated, and G6) *Haemophilus*-dominated. Most of the samples (92.5%) were classified into clusters dominated by *Haemophilus*, *Moraxella*, *Streptococcus*, or a combination of these genera.

The relative abundance of all OTUs in each sample is shown in Supplementary Fig. 4. The clustering analysis of OTUs at the species level and clinical information for each sample are shown in Fig. 3. Six distinct clusters were revealed by hierarchical clustering of samples based on the composition of the OTUs: S1) mixed, S2) *M. catarrhalis/nonliquefaciens* dominated, S3) *H. aegyptius/influenzae* (OTU 3) dominated, S4) *S. pneumoniae* dominated, S5) *H. aegyptius/influenzae* (OTU 4) dominated and S6) *H. aegyptius/influenzae* (OTU 0) dominated. Thirty-seven (92.5%) of the 40 samples were classified into clusters dominated by OTUs corresponding to *H. aegyptius/influenzae*, *M. catarrhalis/nonliquefaciens*, or *S. pneumoniae*. The genera *Haemophilus*, *Moraxella* and *Streptococcus*, which were the three most abundant genera in the genus-level analysis, were composed of multiple OTUs in the species-level analysis. The genus *Haemophilus* was classified into 5 distinct OTUs representing *H. aegyptius/influenzae* (OTUs 0, 3, 4, 11 and 16) among the top 20 predominant OTUs. Sixteen samples had one of these types of OTUs, while 11 samples had two or more of these types of OTUs. Most of the OTUs classified into the genus *Moraxella* were classified into *M. catarrhalis/nonliquefaciens* in the species-level analysis. The others were classified into *M. lincolnii* or *Moraxella* sp. (unclassified at the species level). The genus *Streptococcus* was classified into two species: *S. pneumoniae* and *S. dentisani*.

Comparisons of the bacterial cell counts and alpha diversity (Shannon and Chao1 indices) between clusters are shown in Fig. 4. The Shannon index was significantly higher in Cluster S1 (mixed) than in the aggregate of the other five clusters (dominated by *H. aegyptius/influenzae*, *M. catarrhalis/nonliquefaciens*, or *S. pneumoniae*) (Fig. 4a, $p < 0.001$) or in cluster S2, cluster S4 and cluster S6 (Fig. 4b). The Chao1 index was significantly higher in Cluster S1 than in the aggregate of the five clusters (Fig. 4c, $p = 0.0124$) or in cluster S2, and cluster S6 (Fig. 4d). In contrast, cluster S2, cluster S4 and the aggregate of the five clusters had significantly higher bacterial cell counts than cluster S1 (Fig. 4e,f). The clinical and demographic characteristics of cluster S1 showed no significant differences compared to those of any other cluster except for wheezing in cluster S1 versus cluster S2 (Table 2).

Unclassified clusters. Nine OTUs of all 69 OTUs analyzed in this study were unclassified at the species level based on a Seqmatch score ≥ 0.95 (Table 3). Two OTUs (OTU 42, OTU 59) were unclassified at the genus level but were identified at the family level, whereas 1 OTU (OTU 30) was unclassified at even the phylum level based on assignment using the RDP Classifier with a confidence threshold of 0.8. The analysis using the RDP Classifier showed that the representative sequences of OTUs 42 and 59 showed 75% and 77% similarity, respectively, with the genus *Snodgrassella*, and 100% similarity with the family *Neisseriaceae*. According to the results, OTUs 42 and 59 were possibly previously unknown bacteria at the genus level within the family *Neisseriaceae*. NCBI BLAST search (<https://blast.ncbi.nlm.nih.gov>) results revealed that the representative sequence of OTU

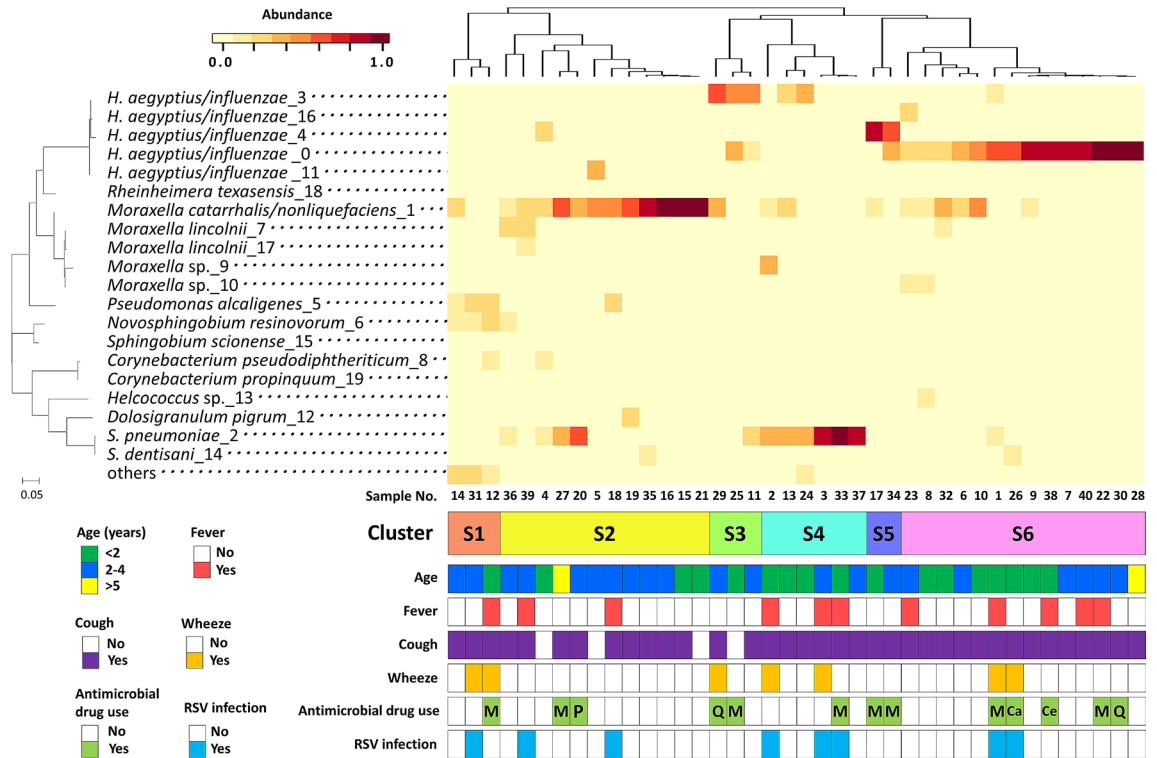


Figure 3. Clustering analysis of the OTUs at the species level and clinical information. The annotated bacterial species and the number of OTUs are shown on the left side of the heatmap. OTUs with relative abundance < 0.4% in the total sequences were grouped as “others”. Six distinct clusters were revealed by hierarchical clustering of samples based on the composition of the OTUs: S1, mixed; S2, *M. catarrhalis/nonliquefaciens* dominated; S3, *H. aegyptius/influenzae* (OTU 3) dominated; S4, *S. pneumoniae* dominated; S5, *H. aegyptius/influenzae* (OTU 4) dominated; S6, *H. aegyptius/influenzae* (OTU 0) dominated. Clinical information such as age, symptoms, antimicrobial drug use and diagnosis of respiratory syncytial virus infection are shown under each sample. RSV respiratory syncytial virus, M macrolide, P penicillin, Q new quinolone, Ca carbapenem, Ce cephem.

30 shared 99.8% similarity with an unclassified bacterium. Recently, the unclassified bacterium was identified in bronchoalveolar lavage fluids (BALFs) in adults with respiratory disease and was temporally named IOLA (Infectious Organism Lurking in Airways)¹⁸. IOLA was also detected in specimens from patients with respiratory tract disease and in pig lungs^{19–21}. To detect IOLA, a nested PCR performed on all 40 samples revealed that 4 samples (10%) were positive for IOLA (Supplementary Fig. 5).

Phylogenetic trees based on 16S rRNA gene sequences, including unclassified clones in this study and type strains belonging to the same genus or family, are shown in Supplementary Figs. 6–12. The results of the phylogenetic analysis showed that the unclassified clones in this study were located in a separate branch from previously known type strains.

Discussion

In this study, we found a high abundance (median 2.2×10^7 cells/mL) of bacteria in the nasal discharge of children, and most of the samples were dominated by OTUs corresponding to *H. aegyptius/influenzae*, *M. catarrhalis/nonliquefaciens*, or *S. pneumoniae*, which are potentially pathogenic bacteria associated with respiratory diseases in both children and adults^{22,23}.

In addition, the samples dominated by these pathogenic bacteria showed higher cell abundance and lower alpha diversity than those in which the other bacteria coexisted. These results suggest that most young children have a large amount of pathogens in their nasal discharge, which may be not only a cause of infections in those children but also a source of pathogens that can be transmitted to others. Our results are consistent with previous studies on the nasopharyngeal microbiota in children, that revealed a relatively high abundance of the genera *Moraxella*, *Streptococcus*, *Haemophilus*, *Staphylococcus*, *Corynebacterium* and *Dolosigranulum*²⁴. However, most of these studies used nasal swabs for specimen collection and performed no quantitative evaluation of bacteria in nasal discharge. Furthermore, these previous studies on the nasal microbiota were analyzed by NGS. Although NGS has been widely applied in human microbiota studies, NGS analyses are often insufficient at the species level because respiratory samples contain many kinds of *Streptococcus* spp. which show high similarity in their 16S rRNA gene sequences²⁵. In our analysis, the sequences determined by the Sanger method were clustered into OTUs at a threshold of 99.6% sequence similarity, indicating that sequences with more than two base differences

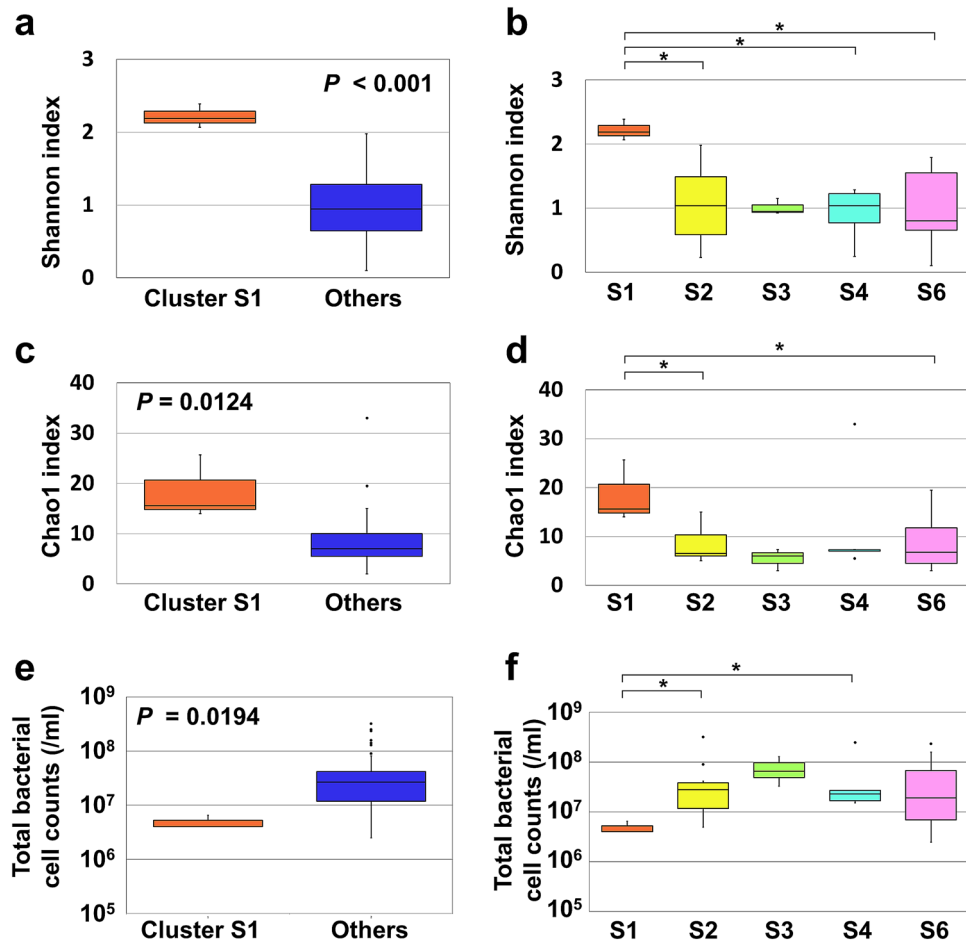


Figure 4. Comparison of the Shannon index, Chao1 index and bacterial cell counts between clusters based on the species-level analysis. The clusters were determined with clustering analysis at the species level (see Fig. 3). The Shannon index (a), Chao1 index (c) and total bacterial cell counts (e) of cluster S1 were compared with those of the other five clusters combined. The Shannon index (b), Chao1 index (d) and total bacterial cell counts (f) of cluster S1 were compared with those of the other four clusters. Cluster S5 was excluded from the analyses due to low subject numbers in cluster S5 (n = 2). The box plots show the 25th and 75th percentiles (bottom and top edges of the boxes, respectively), the median (middle horizontal line), and the smallest and largest values (whiskers), with outlier defined as 1.5 times the interquartile range (points). Statistical significance between cluster S1 and the aggregate of the five clusters was tested using the Mann–Whitney U test. Comparing cluster S1 to each four clusters, the Steel test was applied (bars; *p < 0.05).

Cluster	S1 (n = 3)	S2 (n = 12)	S3 (n = 3)	S4 (n = 6)	S5 (n = 2)	S6 (n = 14)
Age (year), median	2.3	2.1	2.0	1.9	2.0	1.9
Sex (male), n (%)	1 (33)	6 (50)	1 (33)	5 (83)	1 (50)	8 (57)
Fever, n (%)	1 (33)	2 (17)	0 (0)	3 (50)	0 (0)	5 (36)
Cough, n (%)	3 (100)	9 (75)	2 (67)	6 (100)	2 (100)	14 (100)
Wheeze, n (%)	2 (67)	0 (0)*	1 (33)	2 (33)	0 (0)	2 (14)
Antimicrobial drug use ^a	1 (33)	2 (17)	2 (67)	1 (33)	2 (100)	5 (36)
Day care center ^b , n (%)	1 (50)	4 (40)	2 (100)	4 (67)	2 (100)	10 (83)
Hib vaccination completed ^c , n (%)	3 (100)	7 (70)	1 (100)	5 (83)	2 (100)	6 (60)
PCV vaccination completed ^c , n (%)	3 (100)	7 (70)	1 (100)	5 (83)	2 (100)	6 (60)
RSV infection, n (%)	1 (33)	2 (17)	0 (0)	3 (50)	0 (0)	2 (14)

Table 2. Comparison of clinical and demographic characteristics between clusters based on the species-level analysis. RSV Respiratory syncytial virus, Hib Haemophilus influenzae type b, PCV pneumococcal conjugate vaccine. ^aDefined as use of antimicrobial drug within 1 month before sampling ^bNo information available for 6 samples. ^cNo information available for 8 samples. *p < 0.05 compared with cluster S1 by Fisher’s exact test.

Phylum level (less than 80% RDP classifier score to known bacterial type strains)				
OTU No	Domain (RDP-classifier score)	Phylum (RDP-classifier score)	Number of samples	Number of clones
30	Bacteria (96%)	Bacteroidetes (61%)	1	8
Genus level (less than 80% RDP-classifier score to known bacterial type strains)				
OTU No	Family (RDP-classifier score)	Genus (RDP classifier score)	Number of samples	Number of clones
42	Neisseriaceae (100%)	Snodgrassella (75%)	1	3
59	Neisseriaceae (100%)	Snodgrassella (77%)	1	2
Species level (less than 0.95 Seqmatch score to known bacterial type strains)				
OTU No	Genus (RDP-classifier score)	Species (Seqmatch score)	Number of samples	Number of clones
9	Moraxella sp. (99%)	<i>M. lincolni</i> (0.868)	4	72
10	Moraxella sp. (100%)	<i>M. lincolni</i> (0.940)	5	69
13	Helcococcus sp. (100%)	<i>Helcococcus ovis</i> (0.818)	7	53
27	Flavobacterium sp. (100%)	<i>Flavobacterium suzhouense</i> (0.825)	4	12
28	Rheinheimera sp. (100%)	<i>Rheinheimera aquatica</i> (0.924)	6	10
29	Haemophilus sp. (100%)	<i>H. aegyptius</i> (0.942)	1	8
37	Haemophilus sp. (89%)	<i>Actinobacillus minor</i> (0.885)	2	4
41	Pseudomonas sp. (99%)	<i>Pseudomonas alcaligenes</i> (0.921)	2	3
43	Haemophilus sp. (91%)	<i>H. sputorum</i> (0.895)	2	3

Table 3. Unclassified OTUs (possibly previously unknown bacteria).

were clustered into different OTUs. These methods make it possible to accurately identify the bacterial flora in nasal discharge at the species level.

Our findings are consistent with a previous report that indicated that the genera *Moraxella*, *Streptococcus* and *Haemophilus* colonized the nasopharynx of children significantly more frequently in acute respiratory infection than in noninfection states¹⁴. Many of our study subjects who visited a pediatric clinic with rhinorrhea did not have fever (72.5% nonfever). Such children often participate in normal life activities, such as attending a daycare center. However, from the viewpoint of transmission of pathogenic bacteria within a community, we consider that even rhinorrhea, which is considered a mild symptom, plays an important role as a source of infection. Transmission of nasopharyngeal bacteria has been most commonly reported for *S. pneumoniae*². Mosser et al. showed that children play an important role in bringing pneumococci into and maintaining carriage and transmission within households¹⁰. In this study, not only *S. pneumoniae* but also several bacteria related to respiratory diseases in adults, such as the genera *Haemophilus*, *Moraxella*, and *Novosphingobium*, were observed in the nasal discharge of children. *H. influenzae* is a cause of community acquired pneumonia in adults as well as children²⁶, and the most common cause of exacerbation of COPD²⁷. *Moraxella* and *Novosphingobium* have also been reported to associate with COPD^{28,29}. Colonization by these bacteria in the upper respiratory tract is a prerequisite for the development of invasive infections^{2,3,8,9}. Therefore, our results indicate that contact with children with rhinorrhea in closed communities (families, children attending schools or daycare centers etc.) may be a risk factor for community-acquired pneumonia and COPD exacerbation.

In our analysis, there were five different OTUs with the same taxonomy, namely, “*H. aegyptius/influenzae*”, among the top 20 predominant OTUs (OTUs 0, 3, 4, 11 and 16). Several samples had two or more of these OTUs. Simultaneous colonization with several different *H. influenzae* strains is often observed^{30,31}. Therefore, our results indicate that there were multiple strains of *H. influenzae* in the nasal discharge of children. In addition, with the widespread use of Hib vaccination, the incidence of invasive infection caused by NTHi has recently increased³². Several studies also reported nosocomial outbreaks of NTHi infection^{33,34}. Although analysis targeting the 16S rRNA gene sequence cannot identify serotypes among strains, pathogenicity, and antibacterial susceptibility, high-precision analysis based on this gene would be useful for understanding disease transmission and tracking the spread of strains³⁵.

In this study, using the Sanger method targeting 550 bp of the 16S rRNA gene including three variable regions (V3, V4 and V5), most of the sequences were classified into known species and only twelve OTUs (composed of 239 sequences) of 69 OTUs (6388 sequences) were assigned to unclassified bacteria at the species level or higher. One of the twelve OTUs unclassified at the phylum level using the RDP Classifier showed high similarity with IOLA according to the NCBI BLAST search. We have previously reported that IOLA is a novel microorganism found in BALFs in adults with respiratory disease¹⁸. This is the first report of IOLA detected in nasal discharge of children. In this study, IOLA was detected by nested-PCR in four samples (10%), which suggested that children are colonized or infected with IOLA in the nasal cavity at a certain rate. IOLA is unlikely to have an obvious influence on the health of children but might have some effects because IOLA has been detected in the human respiratory tract only in patients with lung disease^{18–20}. In this study, we also detected some sequences assigned as previously unknown species in the genera *Moraxella*, *Helcococcus*, *Flavobacterium*, *Rheinheimera*, *Haemophilus*, and *Pseudomonas* and in the family *Neisseriaceae*. Thus, many previously unknown bacteria may be contained in nasal discharge. Further investigation is needed to clarify the influence of these bacteria on human health and disease.

This study has some limitations. First, our small sample size may have been insufficient to detect associations between nasal discharge microbiota patterns and the clinical and demographic characteristics of patients. Further studies are needed to clarify the role of the microbiota of nasal discharge in development of disease and symptom expression, and to evaluate the influence of antimicrobial drug use, vaccination, and environmental factors on the microbiota. Second, the samples for DNA analysis were stored at 4 °C until time of analysis (within 7 days: average 2.4 days). In preliminary experiments, we confirmed that major bacterial constitutions in a sample were not significantly changed during store at 4 °C for 5 days by using a clone library method. However, the possibility cannot be denied that storage at 4 °C may have some effect on the bacterial composition. Finally, the number of clones identified in this study was a maximum of 192 per sample, suggesting that this method cannot evaluate bacterial species in minor populations (approximately < 0.5%). In fact, IOLA was detected in only one sample by the clone library but was detected in four samples by specific PCR in this study. However, the rarefaction analysis in this study revealed that the clone library was sufficient to evaluate the bacterial diversity of nasal discharge in the major population.

In conclusion, we found that the potentially pathogenic bacteria associated with respiratory disease were predominantly detected at high abundance in most samples of nasal discharge from children. Furthermore, several previously unknown bacteria were discovered by highly accurate 16S rRNA gene sequencing. Close contact with children with rhinorrhea may be a risk factor for developing or exacerbating respiratory diseases especially for the elderly and patients with underlying diseases.

Methods

Subject enrollment. Forty patients who visited the Sato Children's Clinic, Kitakyushu, Japan, with rhinorrhea in January 2017 were enrolled in this study. Clinical and demographic information was collected using a standardized questionnaire. The questionnaire completed by nurses or physicians at the time of enrollment contained information about gender, date of birth, symptoms, clinical diagnosis, and antimicrobial drug use within 1 month before sampling. Rapid antigen detection tests for adenovirus (ImunoAce Adeno, TAUNS Laboratories Inc., Shizuoka, Japan), influenza virus (ImunoAce Flu, TAUNS Laboratories Inc.), and respiratory syncytial virus (ImunoAce RSV Neo, TAUNS Laboratories Inc.) were performed when clinically indicated. Additional clinical information about daycare centers and vaccination status with a 7-valent or, 13-valent pneumococcal conjugate vaccine (PCV) and *H. influenzae* type b (Hib) vaccine was collected using a chart review.

Ethics statement. This study was approved by Ethics Committee of Medical Research, University of Occupational and Environmental Health, Japan (No. H28-184). All experiments and methods in this study were performed in accordance with relevant guidelines and regulations. Informed consent was obtained from all parents/legally authorized representatives as children are involved in the study.

Sample collection. The nasal discharge was aspirated through a nasal olive tip (AS ONE Corporation, Osaka, Japan) attached to a Specimen Collection Container (Nippon Covidien Ltd., Tokyo, Japan) from both nostrils by a trained nurse. The samples were stored immediately at 4 °C until analysis (within 7 days: average 2.4 days). Each nasal discharge sample was diluted 10 times with phosphate-buffered saline (PBS) and vigorously shaken by a Micro Smash MS-100 apparatus (Tomy Seiko Co., Ltd., Tokyo, Japan) for 15 s at 4,500 rpm. The diluted sample was used for cell counting and DNA extraction.

Epifluorescence staining method. The total bacterial cells in the sample were counted by an epifluorescence staining method³⁶. One-hundred micro-liter of the diluted nasal discharge sample was added to 900 µL of 4',6-diamidino-2-phenylindole (DAPI) solution (1 µg/mL in 0.1 M phosphate [pH 8.5] buffer, 5% NaCl, 0.5 mM sodium EDTA) and then, incubated at room temperature for 20 min. After washing with 2 mL of water filtrated with a filter pore-size 0.2 µm, objects shaped like bacteria were counted under an Olympus BX40 epifluorescence microscope (Olympus Optical, Tokyo, Japan), and the number of bacteria per milliliter of nasal discharge specimen was then calculated³⁷. All samples were counted 3 times by the author and coauthor.

DNA extraction method. A 630-µL aliquot of the nasal discharge diluted 10 times with PBS as mentioned above was mixed with 70 µL of 30% sodium dodecyl sulfate (final concentration, 3.0%) and approximately 0.3 g of a mixture of glass beads that consisted of equal weights of 0.1 mm- and 1 mm-diameter beads. The mixture was vigorously shaken by a Micro Smash MS-100 apparatus (Tomy Seiko Co., Ltd.) for 30 s at 5500 rpm. Then, the mixture was treated with 500 µL of TE-saturated phenol (Nacalai Tesque, Kyoto, Japan) (vortex-mixed for 30 s) and centrifuged at 15,000 rpm for 5 min. DNA in the aqueous phase was washed with PBS buffer and TE buffer (10 mM Tris-HCl, 1 mM EDTA-2Na, [pH8.0]) two times and concentrated to a final volume of approximately 50 µL in TE buffer using Amicon Ultra-100K filters (Merck Millipore Ltd., Billerica, MA, USA). DNA extracted from 630-µL of PBS without nasal discharge was used as negative control template.

Polymerase chain reaction conditions. The 16S rRNA genes were amplified as described previously³⁷. Briefly, 25- µL PCR mixtures containing the universal primer set³⁸ (E341F; 5'-CCTACGGGAGGCAGCAG-3' and E907R; 5'-CCGTCGAATTCMTTTRAGTTT-3') and AmpliTaq Gold DNA Polymerase LD (Applied Biosystems, Foster City, CA) were prepared. Then the reaction mixtures were incubated at 96 °C for 5 min, followed by 30 cycles of 96 °C for 30 s, 53 °C for 30 s and 72 °C for 1 min and a final elongation step of 72 °C for 7 min. The PCR products were evaluated by 2% agarose gel electrophoresis. No DNA band was observed in the negative control samples.

Clone library construction and nucleotide sequencing analysis. All PCR products except for a negative control were confirmed as positive using 2% agarose gel electrophoresis and cloned using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, California). A total of 192 colonies from each sample were randomly selected for sequencing analysis. Sequencing analysis was performed as described previously³⁶. Briefly, the partial fragments of the cloning vectors (pCR II) containing inserted PCR products were amplified with AmpliTaq Gold 360 DNA Polymerase and a primer set (M13Forward: 5'-GTAAAACGACGGCCAG-3' and M13Reverse: 5'-CAGGAAACAGCTATGAC-3'). After the primers and deoxyribonucleotide triphosphate were eliminated from the PCR mixture with an ExoSAP-IT (GE Health care UK Ltd., England, UK) according to the manufacturer's instructions, an aliquot (1 µL) was used for the sequencing reaction. The sequencing reactions were accomplished with the primer "M13Forward" and the BigDye Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems). The nucleic acid sequences were determined on a 3130xl Genetic Analyzer (Applied Biosystems).

Data analysis was performed using 'Sequencing Analysis V.5.2' software (Applied Biosystems) as described previously³⁹. The sequences remaining after the quality check (read length: > 500 nucleotides and average quality value of 20 bases: ≥ 20) and only the reads containing precise sequences of the universal primers E341F and E907R were selected. In addition, the raw data were checked, and sequences including ambiguous bases between the primer sequences were excluded. The remaining highly accurate sequences were trimmed to remove primer and vector sequences. Approximately 550 base pairs were screened for chimeras with the Check Chimera program of the Ribosomal Database Project (RDP)⁴⁰. Then, the sequences were clustered into OTUs⁴¹ using the cd-hit-est algorithm⁴² at a threshold of 99.6% sequence similarity. Representative OTU sequences were classified at the genus using the RDP Classifier⁴³ with a confidence threshold of 80%. Sequences with less than 80% were deemed unclassified and classified into the next higher taxonomic rank with RDP Classifier confidences greater than 80%. Species level taxonomies were also assigned by using the RDP-Seqmatch program according to the highest score. When no species matched the sequence with a Seqmatch score ≥ 0.95 , the sequences were deemed unclassified at the species level. *Haemophilus aegyptius* and *H. influenzae* were described as "*H. aegyptius/influenzae*" because they cannot be distinguished by their 16S rRNA genes⁴⁴. In the same way, *M. catarrhalis* and *M. nonliquefaciens* were described as "*Moraxella catarrhalis/nonliquefaciens*" because they cannot be distinguished by analysis of the 550 bp of the 16S rRNA gene targeted in this study. For the species-level analysis, different OTUs with the same taxonomy were distinguished by OTU number following the taxonomy name.

Detection of IOLA. Nested PCR was performed on all samples for detection of IOLA. The 16S rRNA genes were amplified with a universal primer set as mentioned above (25 cycles). Subsequent PCR was performed using IOLA specific primers (30 cycles) as described previously¹⁸.

Data analysis. All data analyses were performed using R software V3.6.1. Heatmaps were generated using the heatmap3 package. To assess the diversity coverage of the clone library, rarefaction curves were obtained using the vegan package. Coverage was calculated according to the equation proposed by Good⁴⁵, where $C = 1 - (n/N)$, n = the number of OTUs represented by only one clone and N = the total number of clones. Alpha-diversity assessed by the Shannon and Chao1 indices was calculated using the estimateR function in the vegan package.

Phylogenetic analysis. In all phylogenetic analyses to classify the twelve unclassified sequences in this study, the data sets of nucleic acid sequences were obtained from the RDP databases. The multiple alignments were built using MAFFT⁴⁶ with default settings. IQ-TREE software v1.6.10⁴⁷ was used to construct maximum likelihood (ML) trees, and the appropriate models for each phylogenetic analysis were selected by Model Finder. The trees were visualized with FigTree v1.4.4. All sequences were submitted to a public database (DNA Data Bank of Japan, accession number: LC558357-LC564861).

Statistical analysis. Comparisons of the proportions of sex, symptoms and antimicrobial drug use were analyzed by Fisher's exact test. The Kruskal–Wallis test and Mann–Whitney U test were used to compare the median age, the Shannon index, the Chao1 index and the total bacterial counts among groups of clusters. The Steel test was used to compare cluster S1 to each four clusters. P-values less than 0.05 were considered to be statistically significant. STATA 14.2 software (Stata Corporation, College Station, TX), BellCurve for Excel, version 3.20 software (Social Survey Research Information, Tokyo, Japan), and R software V3.6.1. were used for the analysis.

Received: 25 August 2020; Accepted: 9 November 2020

Published online: 19 November 2020

References

1. DeMuri, G. P., Gern, J. E., Eickhoff, J. C., Lynch, S. V. & Wald, E. R. Dynamics of Bacterial Colonization With Streptococcus pneumoniae, Haemophilus influenzae, and Moraxella catarrhalis during symptomatic and asymptomatic viral upper respiratory tract infection. *Clin. Infect. Dis.* **66**, 1045–1053. <https://doi.org/10.1093/cid/cix941> (2018).
2. Bogaert, D., De Groot, R. & Hermans, P. W. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect. Dis.* **4**, 144–154. [https://doi.org/10.1016/S1473-3099\(04\)00938-7](https://doi.org/10.1016/S1473-3099(04)00938-7) (2004).
3. Simell, B. *et al.* The fundamental link between pneumococcal carriage and disease. *Expert Rev Vaccines* **11**, 841–855. <https://doi.org/10.1586/erv.12.53> (2012).

4. Faden, H. *et al.* Relationship between nasopharyngeal colonization and the development of otitis media in children, Tonawanda/Williamsville Pediatrics. *J. Infect. Dis.* **175**, 1440–1445. <https://doi.org/10.1086/516477> (1997).
5. Perez-Trallero, E., Marimon, J. M., Larruskain, J., Alonso, M. & Ercibengoa, M. Antimicrobial susceptibilities and serotypes of *Streptococcus pneumoniae* isolates from elderly patients with pneumonia and acute exacerbation of chronic obstructive pulmonary disease. *Antimicrob. Agents Chemother.* **55**, 2729–2734. <https://doi.org/10.1128/AAC.01546-10> (2011).
6. Bandi, V. *et al.* Nontypeable *Haemophilus influenzae* in the lower respiratory tract of patients with chronic bronchitis. *Am. J. Respir. Crit. Care Med.* **164**, 2114–2119. <https://doi.org/10.1164/ajrcm.164.11.2104093> (2001).
7. Bandi, V. *et al.* Infectious exacerbations of chronic obstructive pulmonary disease associated with respiratory viruses and nontypeable *Haemophilus influenzae*. *FEMS Immunol. Med. Microbiol.* **37**, 69–75. [https://doi.org/10.1016/S0928-8244\(03\)00100-7](https://doi.org/10.1016/S0928-8244(03)00100-7) (2003).
8. Cerquetti, M. & Giufre, M. Why we need a vaccine for non-typeable *Haemophilus influenzae*. *Hum. Vaccin. Immunother.* **12**, 2357–2361. <https://doi.org/10.1080/21645515.2016.1174354> (2016).
9. Park, D. E. *et al.* Colonization density of the upper respiratory tract as a predictor of pneumonia-haemophilus influenzae, *Moraxella catarrhalis*, *Staphylococcus aureus*, and *Pneumocystis jirovecii*. *Clin. Infect. Dis.* **64**, S328–S336. <https://doi.org/10.1093/cid/cix104> (2017).
10. Mosser, J. F. *et al.* Nasopharyngeal carriage and transmission of *Streptococcus pneumoniae* in American Indian households after a decade of pneumococcal conjugate vaccine use. *PLoS ONE* **9**, e79578. <https://doi.org/10.1371/journal.pone.0079578> (2014).
11. Gwaltney, J. M. Jr., Sande, M. A., Austrian, R. & Hendley, J. O. Spread of *Streptococcus pneumoniae* in families. II. Relation of transfer of *S. pneumoniae* to incidence of colds and serum antibody. *J. Infect. Dis.* **132**, 62–68. <https://doi.org/10.1093/infdis/132.1.62> (1975).
12. Diavatopoulos, D. A. *et al.* Influenza A virus facilitates *Streptococcus pneumoniae* transmission and disease. *FASEB J.* **24**, 1789–1798. <https://doi.org/10.1096/fj.09-146779> (2010).
13. Bosch, A. *et al.* Development of upper respiratory tract microbiota in infancy is affected by mode of delivery. *EBioMedicine* **9**, 336–345. <https://doi.org/10.1016/j.ebiom.2016.05.031> (2016).
14. Teo, S. M. *et al.* The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe* **17**, 704–715. <https://doi.org/10.1016/j.chom.2015.03.008> (2015).
15. Bosch, A. *et al.* Maturation of the infant respiratory microbiota, environmental drivers, and health consequences: a prospective cohort study. *Am. J. Respir. Crit. Care Med.* **196**, 1582–1590. <https://doi.org/10.1164/rccm.201703-0554OC> (2017).
16. McCauley, K. *et al.* Distinct nasal airway bacterial microbiotas differentially relate to exacerbation in pediatric patients with asthma. *J. Allergy Clin. Immunol.* **144**, 1187–1197. <https://doi.org/10.1016/j.jaci.2019.05.035> (2019).
17. Zhou, Y. *et al.* The upper-airway microbiota and loss of asthma control among asthmatic children. *Nat. Commun.* **10**, 5714. <https://doi.org/10.1038/s41467-019-13698-x> (2019).
18. Fukuda, K. *et al.* An unclassified microorganism: novel pathogen candidate lurking in human airways. *PLoS ONE* **9**, e103646. <https://doi.org/10.1371/journal.pone.0103646> (2014).
19. Hilty, M. *et al.* Disordered microbial communities in asthmatic airways. *PLoS ONE* **5**, e8578. <https://doi.org/10.1371/journal.pone.0008578> (2010).
20. Harris, J. K. *et al.* Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. *Proc. Natl. Acad. Sci. USA.* **104**, 20529–20533. <https://doi.org/10.1073/pnas.0709804104> (2007).
21. Nichols, J. E. *et al.* Production and transplantation of bioengineered lung into a large-animal model. *Sci. Transl. Med.* <https://doi.org/10.1126/scitranslmed.aao3926> (2018).
22. Jiang, W. *et al.* Etiologic spectrum and occurrence of coinfections in children hospitalized with community-acquired pneumonia. *BMC Infect. Dis.* **17**, 787. <https://doi.org/10.1186/s12879-017-2891-x> (2017).
23. Torres, A., Blasi, F., Peetermans, W. E., Viegi, G. & Welte, T. The aetiology and antibiotic management of community-acquired pneumonia in adults in Europe: a literature review. *Eur. J. Clin. Microbiol. Infect. Dis.* **33**, 1065–1079. <https://doi.org/10.1007/s10096-014-2067-1> (2014).
24. Bomar, L., Brugger, S. D. & Lemon, K. P. Bacterial microbiota of the nasal passages across the span of human life. *Curr. Opin. Microbiol.* **41**, 8–14. <https://doi.org/10.1016/j.mib.2017.10.023> (2018).
25. Yatera, K., Noguchi, S. & Mukae, H. The microbiome in the lower respiratory tract. *Respir. Investig.* **56**, 432–439. <https://doi.org/10.1016/j.resinv.2018.08.003> (2018).
26. Murphy, T. F. Respiratory infections caused by non-typeable *Haemophilus influenzae*. *Curr. Opin. Infect. Dis.* **16**, 129–134. <https://doi.org/10.1097/00001432-200304000-00009> (2003).
27. Sethi, S. & Murphy, T. F. Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N. Engl. J. Med.* **359**, 2355–2365. <https://doi.org/10.1056/NEJMra0800353> (2008).
28. Mayhew, D. *et al.* Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. *Thorax* **73**, 422–430. <https://doi.org/10.1136/thoraxjnl-2017-210408> (2018).
29. Rutebemberwa, A. *et al.* Novosphingobium and its potential role in chronic obstructive pulmonary diseases: insights from microbiome studies. *PLoS ONE* **9**, e111150. <https://doi.org/10.1371/journal.pone.0111150> (2014).
30. Farjo, R. S. *et al.* Diversity and sharing of *Haemophilus influenzae* strains colonizing healthy children attending day-care centers. *Pediatr. Infect. Dis. J.* **23**, 41–46. <https://doi.org/10.1097/01.inf.0000106981.89572.d1> (2004).
31. Saito, M., Okada, K., Takemori, K. & Yoshida, S. I. Clonal spread of an invasive strain of *Haemophilus influenzae* type b among nursery contacts accompanied by a high carriage rate of non-disease-associated strains. *J. Med. Microbiol.* **49**, 845–847. <https://doi.org/10.1099/0022-1317-49-9-845> (2000).
32. Resman, F. *et al.* Invasive disease caused by *Haemophilus influenzae* in Sweden 1997–2009; evidence of increasing incidence and clinical burden of non-type b strains. *Clin. Microbiol. Infect.* **17**, 1638–1645. <https://doi.org/10.1111/j.1469-0691.2010.03417.x> (2011).
33. Van Dort, M. *et al.* An outbreak of infections caused by non-typeable *Haemophilus influenzae* in an extended care facility. *J. Hosp. Infect.* **66**, 59–64. <https://doi.org/10.1016/j.jhin.2007.02.001> (2007).
34. Andersson, M. *et al.* Outbreak of a beta-lactam resistant non-typeable *Haemophilus influenzae* sequence type 14 associated with severe clinical outcomes. *BMC Infect. Dis.* **15**, 581. <https://doi.org/10.1186/s12879-015-1319-8> (2015).
35. Sacchi, C. T. *et al.* High level of sequence diversity in the 16S rRNA genes of *Haemophilus influenzae* isolates is useful for molecular subtyping. *J. Clin. Microbiol.* **43**, 3734–3742. <https://doi.org/10.1128/JCM.43.8.3734-3742.2005> (2005).
36. Yamaguchi, N. *et al.* Rapid monitoring of bacteria in dialysis by fluorescent vital staining and microcolony methods. *Nephrol. Dial. Transplant.* **22**, 612–616. <https://doi.org/10.1093/ndt/gfl675> (2007).
37. Kawanami, T. *et al.* A higher significance of anaerobes: the clone library analysis of bacterial pleurisy. *Chest* **139**, 600–608. <https://doi.org/10.1378/chest.10-0460> (2011).
38. Lane, D. J. 16S/23S rRNA sequencing. *Nucleic acid techniques in bacterial systematics.* (1991).
39. Otsuji, K. *et al.* Dynamics of microbiota during mechanical ventilation in aspiration pneumonia. *BMC Pulm. Med.* **19**, 260. <https://doi.org/10.1186/s12890-019-1021-5> (2019).
40. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200. <https://doi.org/10.1093/bioinformatics/btr381> (2011).

41. Blaxter, M. *et al.* Defining operational taxonomic units using DNA barcode data. *Philos. Trans. R. Soc. Lond. B* **360**, 1935–1943. <https://doi.org/10.1098/rstb.2005.1725> (2005).
42. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682. <https://doi.org/10.1093/bioinformatics/btq003> (2010).
43. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267. <https://doi.org/10.1128/AEM.00062-07> (2007).
44. Nørskov-Lauritsen, N. Classification, identification, and clinical significance of *Haemophilus* and *Aggregatibacter* species with host specificity for humans. *Clin. Microbiol. Rev.* **27**, 214–240. <https://doi.org/10.1128/cmr.00103-13> (2014).
45. Good, I. J. The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–264 (1953).
46. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166. <https://doi.org/10.1093/bib/bbx108> (2019).
47. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. <https://doi.org/10.1093/molbev/msu300> (2015).

Acknowledgements

We are grateful to Dr. Katsuko Sato and the staffs of Sato Children's Clinic for collecting nasal discharge and providing patients' information.

Author contributions

K.H. conceptualized the study, drafted the initial manuscript, and carried out the initial analysis. M.O., M.S. and K.K. considered clinical significance of the results and revised the manuscript. K.F. designed the study, supervised data collection, and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-77271-z>.

Correspondence and requests for materials should be addressed to K.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020