Supplementary Online Content

Respiratory microbiota predicts clinical disease course of acute otorrhea in children with tympanostomy tubes

Wing Ho Man, Thijs M.A. van Dongen, Roderick P. Venekamp, Vincent G. Pluimakers, Mei Ling J.N. Chu, Marlies A. van Houten, Elisabeth A.M. Sanders, Anne G.M. Schilder, Debby Bogaert

Supplemental methods

16S rRNA Gene Amplification and Sequencing

Bacterial DNA was isolated from samples and quantified as previously described.[1,2] In short, an aliquot of 200µl of each sample was added to 650µl lysis buffer with 0·1 mm zirconium beads and 550µl phenol. All samples were mechanically lysed with a bead beater procedure. Amplification of the V4 hypervariable region of the 16S rRNA gene was performed using barcoded universal primer pair 533F/806R. Amplicons were quantified by PicoGreen (Thermofisher) and pooled in equimolar amounts. Amplicon pools of samples and controls were sequenced using the Illumina MiSeq platform (San Diego, CA, USA).

Bioinformatics Analysis

Raw sequences were trimmed using an adaptive, window-based trimming algorithm (Sickle, Q>20, length threshold of 150 nucleotides).[3] We aimed to further reduce the number of sequence errors in the reads by applying an error correction algorithm (BayesHammer, SPAdes genome assembler toolkit).[4] Forward and reverse reads were then assembled into contigs using PANDAseq.[5] Merged reads were demultiplexed using QIIME v1·9.[6] After removal of singleton sequences, we removed chimeras using both *de novo* and reference (against Gold database) chimera identification (UCHIME algorithm in VSEARCH).[7,8] VSEARCH abundance-based greedy clustering was used to pick OTUs at a 97% identity threshold.[9] Taxonomic annotation was executed using the RDP-II naïve Bayesian classifier on SILVA v119 training set.[10] After aligning the node representative sequences to the Silva v119 core alignment database using the PyNAST method,[11] a rooted phylogenetic tree was calculated using FastTree.[12] We generated an abundance-filtered dataset by including only those OTUs that were present at or above a confident level of detection (0·1% relative abundance) in at least 2 samples, retaining 138 OTUs in total.[13] To avoid OTUs with identical annotations, we refer to OTUs using their taxonomical annotations combined with a rank number based on the abundance of each given OTU. The raw OTU-counts table was used for calculations of α-diversity and analyses using the *metagenomeSeq* package.[14] The OTU-proportions table was used for all other downstream analyses, including hierarchical clustering and random forest modelling. Moreover, the Bray-Curtis

(dis)similarity metric was consistently used to express ecological distance (β-diversity) in all analyses because it includes proportional abundance information and excludes joint-absence information, and thereby yields useful insights into the specific structure of our data.[15]

eTable 1. Characteristics of the study population at baseline.

	Overall (n=94)
Boys, n (%)	57 (60.6)
Mean age, yrs (SD)	3.38 (1.41)
Indication for tube insertion, n (%)	
Otitis media with effusion	43 (45.7)
Acute otitis media	32 (34)
Both	19 (20.2)
Mean duration of otorrhea in days before enrolment (SD)	2.70 (1.78)
Vaccinated, n (%)	
Received PCV7	73 (78.5)
Antibiotics in previous 14 days, n (%)	
Eardrops	0 (0)
Oral	0 (0)
Mean number of tympanostomy tube insertions (SD)	1.24 (0.54)
Mean number of siblings (SD)	1.27 (0.59)
Day care or school, n (%)	
Yes, day care	54 (57.4)
Yes, school	31 (33)
No	9 (9.6)
Breastfed, n (%)	68 (72.3)
Household smoking, n (%)	12 (12.9)

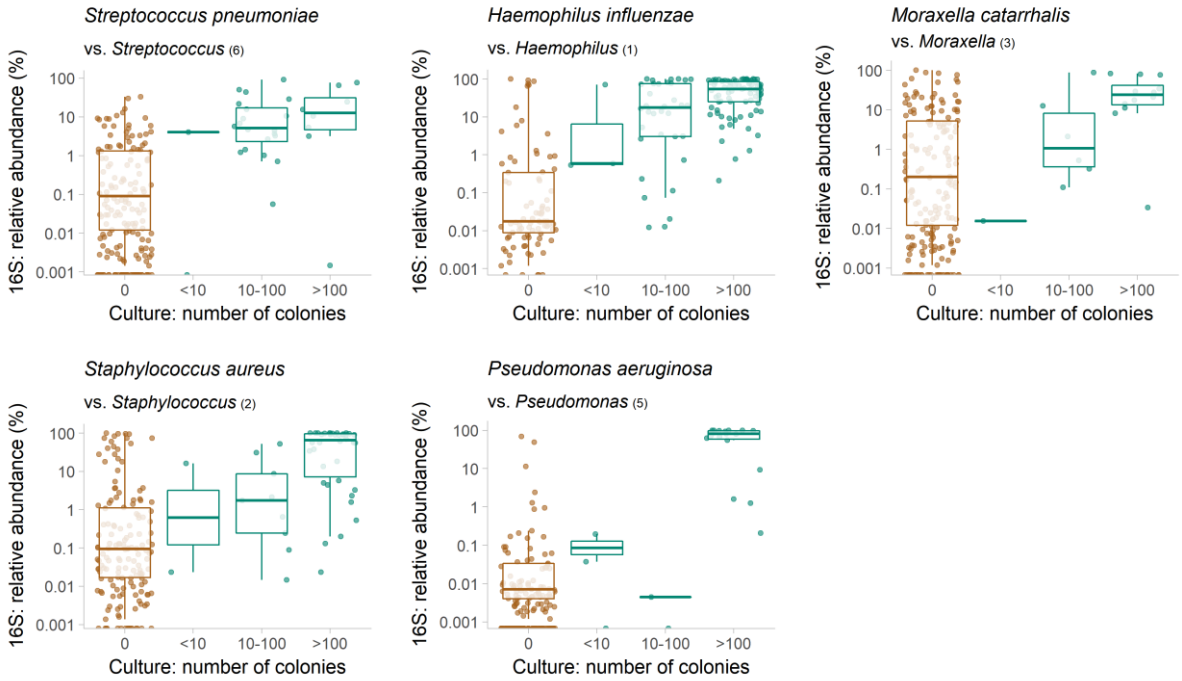
eTable 2. Qualitative agreement between matched pairs of NP and TTO samples on OTU level.

We calculated the overall positive predictive value, negative predictive value, sensitivity and specificity using the TTO sample as the reference. Also, we calculated the prevalence of OTU's in both niches and the concordance expressed as the proportion of overall agreement.

	Point Estimate	95% CI
Targets	15042	-
Sensitivity	0.59	0.57 - 0.61
Specificity	0.83	0.82 - 0.84
Positive predictive value	0.40	0.39 - 0.42
Negative predictive value	0.91	0.91 - 0.92
Prevalence NP	0.24	0.23 - 0.25
Prevalence TTO	0.16	0.16 - 0.17
Agreement	0.79	0.78 - 0.8

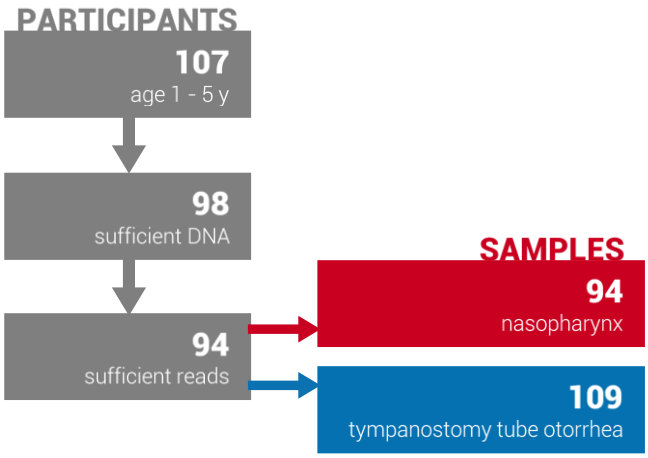
eFigure 1. Culture results confirm the taxonomic annotation of the corresponding OTU's.

Boxplots visualizing the relation between the culture results for *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, *Staphylococcus aureus* and *Pseudomonas aeruginosa* and the relative abundance of the corresponding OTU as determined by 16S rRNA sequencing.



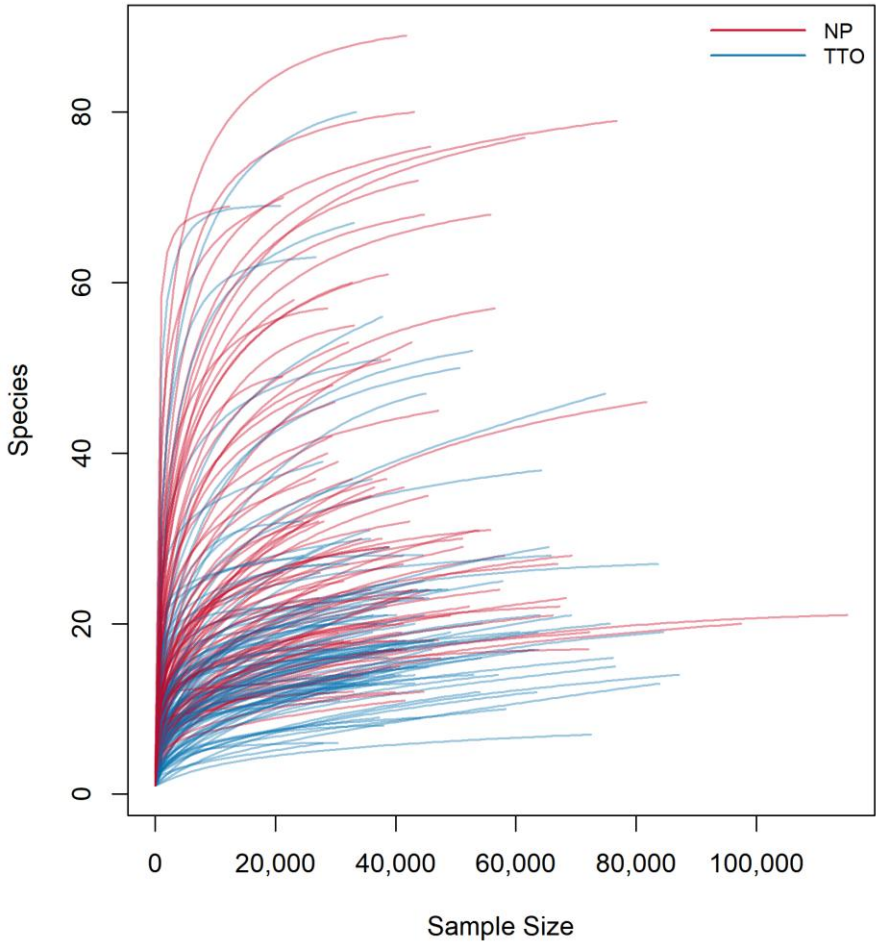
eFigure 2. Flow chart participants and samples.

Flow chart describing the number of participants and samples analyzed in this study. Only participants that had both a high-quality nasopharynx sample and a high-quality TTO samples were used for downstream analysis.



eFigure 3. Rarefaction curves on raw count data.

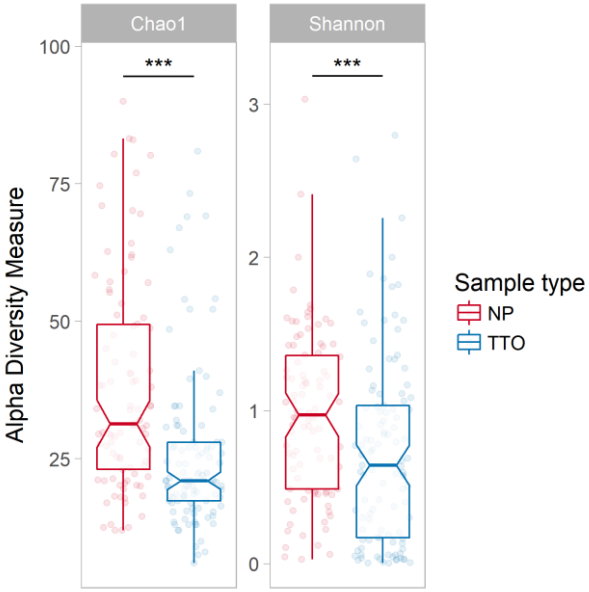
Rarefaction curves on raw count data approached plateau for both NP samples (red) and TTO samples (blue).



eFigure 4. α-Diversity.

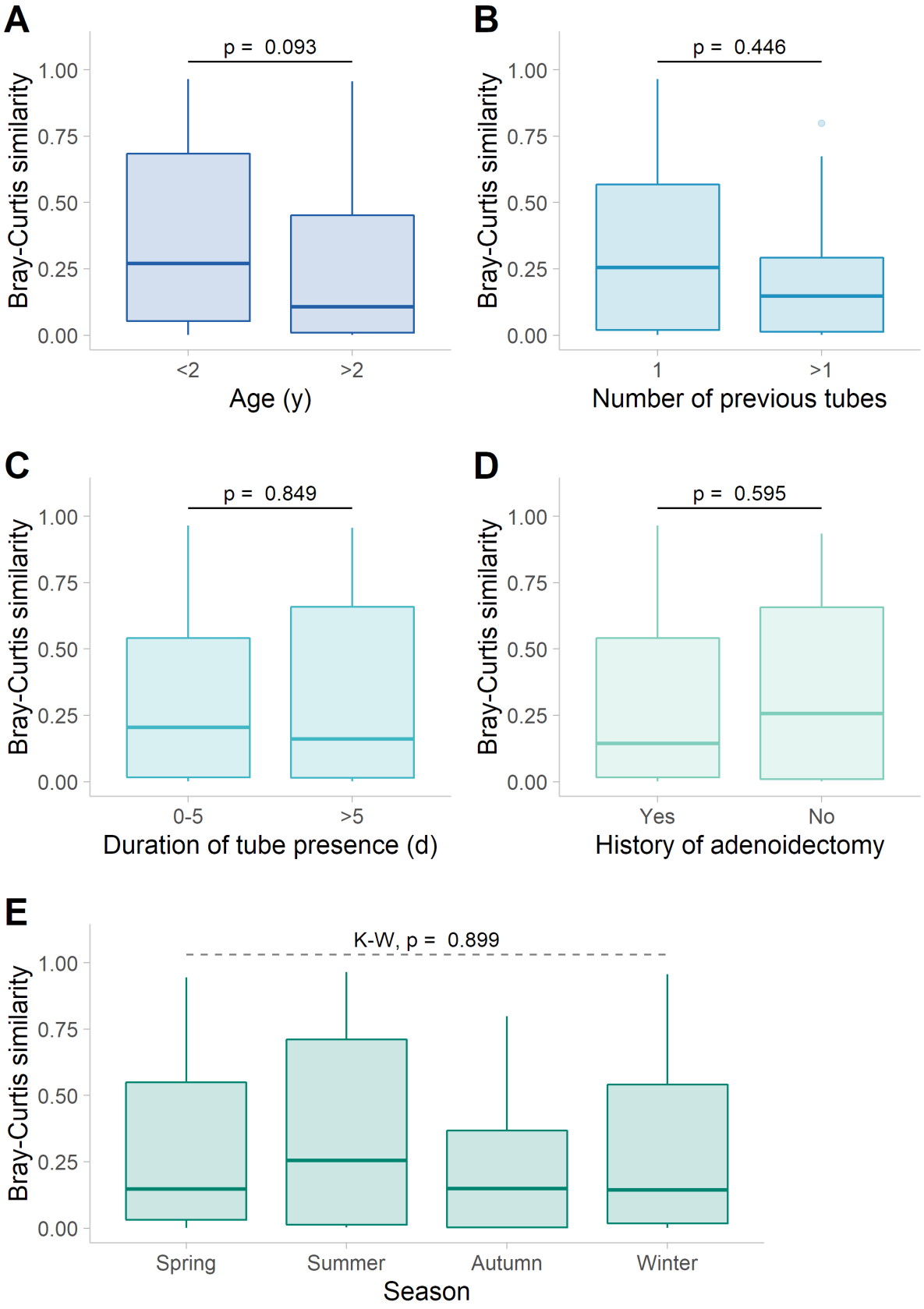
The ecological diversity was significantly higher in NP samples (red) compared to the TTO samples (blue), according to the Chao 1 estimate and Shannon's diversity index.

Significance symbols: *** = p<0.001; ** = p<0.01; * = p<0.05.



eFigure 5. Similarity of paired NP and TTO samples does not vary with clinical variables.

Bray-Curtis similarity (1 – Bray-Curtis dissimilarity) of the paired NP and TTO samples of the same participant stratified by age (<2 years, n=32; >2 years n=47; A), number of previous tympanostomy tubes (including the insertion of the current tympanostomy tube; 1 tube, n=63; >1 tube, n=16; B), duration of tube presence (0-5 days, n=40; >5 days, n=39; C), history of prior adenoidectomy (yes, n=47; no, n=32; D), and season of sampling (Spring, March-May, n=18; Summer, June-August, n=17; Autumn, September-November, n=20; Winter, December-February, n=21). The Bray-Curtis similarity is bounded between 0 and 1, where 0 means that two samples are completely dissimilar, and 1 means the two sites are completely similar. P-values are based on Wilcoxon rank-sum tests (A-D) and a Kruskal-Wallis test (E).


A
p = 0.093
Bray-Curtis similarity
1.00
0.75
0.50
0.25
0.00
<2
>2
Age (y)
B
p = 0.446
Bray-Curtis similarity
1.00
0.75
0.50
0.25
0.00
1
>1
Number of previous tubes
C
p = 0.849
Bray-Curtis similarity
1.00
0.75
0.50
0.25
0.00
0-5
>5
Duration of tube presence (d)
D
p = 0.595
Bray-Curtis similarity
1.00
0.75
0.50
0.25
0.00
Yes
No
History of adenoidectomy
E
K-W, p = 0.899
Bray-Curtis similarity
1.00
0.75
0.50
0.25
0.00
Spring
Summer
Autumn
Winter
Season

86

References

1. Prevaes SMPJ, de Winter-de Groot KM, Janssens HM, et al. Development of the Nasopharyngeal Microbiota in Infants with Cystic Fibrosis. *Am J Respir Crit Care Med*. 2016;193(5):504-515.
2. Wyllie AL, Chu MLJN, Schellens MHB, et al. Streptococcus pneumoniae in saliva of Dutch primary school children. *PLoS One*. 2014;9(7):e102045.
3. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software].
4. Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*. 2013;14(Suppl 1):S7.
5. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics*. 2012;13(1):31.
6. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335-336.
7. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27(16):2194-2200.
8. Rognes T, Mahé F, Flouri T, Quince C, Nichols B. VSEARCH.
9. Westcott SL, Schloss PD. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ*. 2015;3:e1487.
10. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2012;41(D1):D590-D596.
11. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*. 2010;26(2):266-267.
12. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5(3):e9490.
13. Subramanian S, Huq S, Yatsunenko T, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*. 2014;510(7505):417.
14. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods*. 2013;10(12):1200-1202.
15. Anderson MJ, Crist TO, Chase JM, et al. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol Lett*. 2011;14(1):19-28.