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

Wing Ho Man[a,b], Thijs M.A. van Dongen[c], Roderick P. Venekamp[c], Vincent G. Pluimakers[a], Mei Ling J.N. Chu[a,d], Marlies A. van Houten[b], Elisabeth A.M. Sanders[a], Anne G. M. Schilder[e], Debby Bogaert[a,f]

Affiliations:

[a] Department of Paediatric Immunology and Infectious Diseases, Wilhelmina Children's Hospital/University Medical Center Utrecht, Utrecht, The Netherlands;

[b] Spaarne Gasthuis Academy, Hoofddorp and Haarlem, The Netherlands;

[c] Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands;

[d] Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands;

[e] ENT Clinical Trials Programme, Ear Institute, University College London, London, United Kingdom;

[f] Medical Research Council/University of Edinburgh Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom.

Wing Ho Man, MD

Department of Paediatric Immunology and Infectious Diseases

Wilhelmina Children's Hospital/University Medical Center Utrecht

P.O. Box 85090, 3508 AB Utrecht, The Netherlands

winghoman@gmail.com

Thijs M.A. van Dongen, MD, PhD

Department of Epidemiology, Julius Center for Health Sciences and Primary Care

University Medical Center Utrecht

P.O. Box 85500, 3508 GA Utrecht, The Netherlands

T.M.A.vanDongen@umcutrecht.nl

Roderick P. Venekamp, MD, PhD

Department of Epidemiology, Julius Center for Health Sciences and Primary Care

University Medical Center Utrecht

P.O. Box 85500, 3508 GA Utrecht, The Netherlands

R.P.Venekamp@umcutrecht.nl

Vincent G. Pluimakers, MD

Department of Paediatric Immunology and Infectious Diseases

Wilhelmina Children's Hospital/University Medical Center Utrecht

P.O. Box 85090, 3508 AB Utrecht, The Netherlands

vincentpluimakers@gmail.com

Mei Ling J.N. Chu, BSc

Department of Paediatric Immunology and Infectious Diseases

Wilhelmina Children's Hospital/University Medical Center Utrecht

P.O. Box 85090, 3508 AB Utrecht, The Netherlands

M.L.J.N.Chu@umcutrecht.nl

Marlies A. van Houten, MD, PhD

Spaarne Gasthuis Academy

P.O. Box 900, 2000 VB Haarlem, The Netherlands

MvanHouten2@spaarnegasthuis.nl

Elisabeth A.M. Sanders, MD, PhD

Department of Paediatric Immunology and Infectious Diseases

Wilhelmina Children's Hospital/University Medical Center Utrecht

P.O. Box 85090, 3508 AB Utrecht, The Netherlands

L.Sanders@umcutrecht.nl

Anne G. M. Schilder, MD, PhD

ENT Clinical Trials Programme, Ear Institute, University College London

330 Grays Inn Road, London WC1X 8DA, United Kingdom

a.schilder@ucl.ac.uk

Debby Bogaert, MD, PhD (corresponding author)

Medical Research Council/University of Edinburgh Centre for Inflammation Research

Queen's Medical Research Institute, University of Edinburgh

47 Little France Crescent, EH16 4TJ, Edinburgh, United Kingdom

D.Bogaert@ed.ac.uk

Tel: +44 131 2426582

Keywords: respiratory microbiota; nasopharynx; middle ear fluid; acute otitis media; childhood infections

Abbreviated title: Microbiota predicts course of acute otitis media

Runing title: Microbiota predicts course of otitis

Funding

The trial was supported by a grant (170992502) from the Netherlands Organization for Health Research and Development Effects and Costs subprogram.

Conflict of interests

No conflict of interest related to the present study. EAMS declares to have received unrestricted research support from Pfizer, grant support for vaccine studies from Pfizer and GSK and fees paid to the institution for advisory boards or participation in independent data monitoring committees for Pfizer,

GSK. DB declares having received fees paid to the institution for advisory board work for Friesland Campina. No other authors reported financial disclosures.

Role of the funding source

The funding agency of the trial had no role in the study design, in the collection, analysis, and interpretation of data, in writing the report, and the decision to submit the paper for publication. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Correspondence to:

Debby Bogaert, MD, PhD (corresponding author)

Medical Research Council/University of Edinburgh Centre for Inflammation Research

Queen's Medical Research Institute, University of Edinburgh

47 Little France Crescent

EH16 4TJ, Edinburgh, United Kingdom

D.Bogaert@ed.ac.uk

Tel: +44 131 2426582

Word count main text: 3,215/3,000.

ABSTRACT

Background: Acute otitis media (AOM) is one of the most common childhood infections, generally thought to be caused by ascension of bacteria from the nasopharynx (NP) to the middle ear. Using 16S rRNA-based sequencing, we evaluated the relationship between the NP and middle ear fluid (MEF) microbiota in children with acute otitis media with tympanostomy tubes (AOMT) as a proxy for AOM, and explored whether microbiota profiling predicts natural disease course.

Methods: Microbiota profiles of paired NP and MEF samples of 94 children aged below five years with uncomplicated AOMT were determined**.**

Results: Local diversity ($p<0{\cdot}001$) and overall microbiota composition ($p<0{\cdot}001$) of NP and MEF samples differed significantly, though paired NP and MEF samples were much more similar than unpaired samples ($p<0{\cdot}001$). High qualitative agreement between the presence of individual bacteria in both niches was observed. Abundances of *Pseudomonas aeruginosa, Staphylococcus aureus, Streptococcus pyogenes, Turicella otitidis, Klebsiella pneumoniae,* and *Haemophilus* spp. were strongly correlated between the two niches. Additionally, *P. aeruginosa, S. aureus, T. otitidis* and *Streptococcus pneumoniae* abundance in NP were predictive of the presence of a range of oral types of bacteria in MEF. Interestingly, there was no association between *Moraxella catarrhalis* in NP and MEF samples, which was highly present in NP but virtually absent in MEF. Finally, the NP microbiota composition could predict duration of AOMT, even better than MEF microbiota.

Conclusions: We observed substantial correlations between paired NP and MEF microbiota in children with AOMT. Our data also suggest that NP microbiota profiling deserves further exploration as tool for future treatment decisions.

INTRODUCTION

Acute otitis media (AOM) is among the commonest childhood infections; its incidence is highest in children aged 1 to 4 years with 61 new AOM episodes per 100 children per year.[1]

Classically, the three major bacteria involved in AOM are *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Moraxella catarrhalis*,[2] although there is ongoing debate about the role of the latter.[3,4] It is assumed that these bacteria enter the middle ear cavity from the nasopharyngeal (NP) niche by ascending through the Eustachian tube upon a virus-induced inflammatory cascade.[5] Whereas microbial analysis of middle ear fluid (MEF) is regarded as the gold standard to determine AOM etiology,[6] this requires an invasive procedure such as tympanocentesis or myringotomy to obtain a sample; therefore, NP samples are often used as proxy. A recent systematic review of this approach, however, showed only a moderate concordance between conventional cultures of NP and MEF samples.[7] This may reflect limitations of conventional culture techniques, which are less sensitive than molecular methods and do not consider relative abundance of pathogens in the context of the complete microbial ecosystem nor the role of commensals in the pathophysiology.[8]

Episodes of acute ear discharge in children with tympanostomy tubes are thought to be the result of AOM, in which MEF drains trough the tube.[5] The bacteria involved in AOM in children with tympanostomy tubes (AOMT) include the major bacteria found in AOM, as well as *Staphylococcus aureus* and *Pseudomonas aeruginosa*.[9] Because tympanostomy tube otorrhea (TTO) can be easily obtained in children with AOMT, this population is of particular interest when studying the pathogenesis of AOM(T). In addition, characterization of the complete microbial community composition through next-generation sequencing techniques holds a great promise to better understand the relation between NP and TTO microbiota in children with AOM(T). The three studies thus far comparing microbiota compositions of the NP and MEF/TTO in children with otitis media focused either on differences between the NP and MEF/TTO samples or on otitis media with effusion rather than AOM(T), and/or were too small to extensively study the relation between these two niches on the individual patient level.[8,10,11]

Our group recently performed a randomized controlled trial (RCT) on the treatment ofAOMT.[12] As part of this trial, we collected NP and TTO samples of all participants. In the present study, we aim to assess the relevance of the respiratory ecosystem in childhood AOMT by analyzing the relationship between the NP and TTO microbiota in baseline samples of 94 participants. Moreover, we explore whether microbial community profiling predict natural disease course of AOMT.

MATERIALS AND METHODS

Study design

We obtained baseline NP and TTO samples from children under five years of age who participated in our recent RCT of treatment of AOMT. Children were randomly allocated to either antibiotic-corticosteroid (hydrocortisone–bacitracin–colistin) eardrops, oral antibiotics (amoxicillin–clavulanate suspension) or initial observation (no treatment).[9,12] Deep transnasal nasopharyngeal swabs were obtained according to WHO standard procedures,[13] whereas TTO samples were retrieved by swabbing discharge in the ear canal, avoiding skin contact. During follow-up, otoscopy was performed at 2 weeks to assess presence or absence of otorrhea and the parents of participating children kept a daily diary of ear-related symptoms for six months. Further details of the trial entry criteria and methodology are described elsewhere.[12]

Bacterial high-throughput sequencing and bioinformatic processing

Bacterial DNA of the matching TTO and NP sample pairs was isolated, PCR amplicon libraries were generated, 16S ribosomal RNA gene-sequencing was executed and amplicon pools were processed in our bioinformatics pipeline as previously described and detailed in the supplements.[14] All samples fulfilled our quality control standards for reliable analyses, having DNA levels of >0.3 pg/µl over negative controls. The four highest PCR and DNA isolation blanks were also sequenced, and yielded only a median number of 113.5 reads (range 8-667 reads/blank), whereas all samples yielded more than 10.000 sequences. Finally, none of the reagent contaminants published by Salter *et al*.[15] were present in more than half of our negative controls, all indicating that our strict sequencing protocol and

bioinformatics pipeline resulted in no apparent contamination. *Turicella* was not present in any of the negative controls. In addition, culture results of *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, *Staphylococcus aureus* and *Pseudomonas aeruginosa* were used for the *post-hoc* species-level annotations of the corresponding OTU (eFigure 1 in the Supplement). We generated an abundance-filtered dataset by including only those OTUs that were present at or above a confidence level of detection (0.1% relative abundance) in at least two samples, retaining 138 OTUs in total.[16] 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.[17] The OTU-proportions table was used for all other downstream analyses, including hierarchical clustering and random forest modelling. β-Diversity was assessed using the Bray-Curtis similarity metric (calculated by 1 – Bray-Curtis dissimilarity).

Statistical analysis

All analyses were performed in R version 3.3.2. Good's estimator of coverage was calculated using the formula: (1-(singletons/total number of sequences)) x 100.[18] α-Diversity was estimated by the Chao 1 estimate of richness and the Shannon's diversity index, which takes into account both richness and evenness of the samples. Statistical significance of the differences in α-diversity was calculated using linear mixed models with the participant as random factor. Nonmetric multidimensional scaling (NMDS) plots were used to visualize differences of total microbiota communities between groups and statistical significance was calculated by *adonis* and Multi-Response Permutation Procedures (MRPP) (both 9,999 permutations) with samples from the same participant grouped in the analysis (as random factor). The overall qualitative concordance between NP and TTO microbiota was evaluated according to previously described methods.[7] In short, we calculated the prevalence in both niches, the positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity using the TTO sample as the reference. The quantitative correlations were calculated with Spearman's rank correlation coefficient. Average linkage hierarchical clustering including the determination of biomarker species

was performed as described previously.[19] We used *metagenomeSeq* to identify the microbial taxa associated between groups (i.e. NP vs. TTO).[17]

To confirm with an unsupervised quantitative method whether the abundances of NP biomarker species were related to their respective abundances of the paired TTO samples, we used a random forest approach. This also allowed us to determine the relation of biomarker species in the NP with other species in the paired TTO samples. We performed 100-times repeated, 10-times cross-validated sparse random forest models generating 10,000 trees (*train* function, *randomForest* package) for each of the biomarker species. Variables for this sparse model were selected using the bacterial species determined by the interpretation step of a 20-times cross-validated VSURF procedure, generating 10,000 trees each iteration, with 100 iterations for the thresholding step and 50 iterations for the interpretation step.[20] The direction of the associations was estimated post-hoc using the partial Spearman's correlations. The importance of each bacterial species is determined by evaluating the increase in the mean square error (MSE; i.e. the decrease in prediction accuracy) between observations and model when the data for that bacterial species is randomly permuted. The increase in MSE averaged over all trees produces the final measure of importance.[21]

To assess whether respiratory microbiota composition predicts AOMT natural disease course, we studied the association between NP and TTO microbiota of the 27 children who were not treated (initial observation group). We used the trial's prespecified clinical outcome measures, i.e. otoscopically confirmed otorrhea two weeks after randomization (binary outcome), the duration of the initial otorrhea episode, total number of days with otorrhea and number of recurrent otorrhea episodes during six months of follow-up (numerical outcomes). To this purpose, we built separate cross-validated sparse random forest classification and prediction models as described above for the clinical outcomes, respectively. The performance of the classification models was evaluated by calculating the area under the ROC curve (AUC) using the out-of-bag predictions for classification (*pROC* package[22]). The performance of the prediction models was assessed by calculating the Spearman's rank correlations between the model predicted and the observed outcome values.

A p-value of less than 0.05 for single parameter outcome or Benjamini-Hochberg (BH) adjusted q-value less than 0.05 when multiple variables were tested was considered statistically significant.

RESULTS

Participants

In 98 out of 107 (92%) children under 5 years of age from whom paired NP and TTO samples were available, a sufficient amount of DNA was isolated for reliable 16S rRNA-based sequencing analyses.[14] MiSeq PCR followed by MiSeq sequencing was successful in 94 of 98 children (96%). Fifteen of these children had bilateral AOMT, resulting in 94 NP samples and 109 paired TTO samples (eFigure 2 in the Supplement). Characteristics of the study population are shown in eTable 1 in the Supplement.

Characterization of sequencing results and diversity

A total of 8,758,772 reads were used for analysis (mean 43,147 ± 16,199 reads per sample). These were binned into 138 97%-identity OTUs, representing 66 taxonomic genera from eight phyla. Good's coverage of >99.9% was reached for all samples and rarefaction curves on raw count data approached plateau in all samples (eFigure 3 in the Supplement), suggesting that the sequence results of each sample represented the majority of bacteria present in the NP and TTO samples under study.

The estimated number of species and Shannon diversity was higher in NP samples than in TTO samples (Chao mean 37.8 and 25.6 species for NP and TTO, respectively; Shannon mean 0.97 and 0.73 for NP and TTO, respectively, both $p<0.001$; eFigure 4 in the Supplement).

The total microbiota composition differed significantly between NP and TTO (adonis, $R^2=0.054$, $p<0.001$; MRPP, $A=0.031$, $p<0.001$; Figure 1A). However, paired NP and TTO samples were considerably more similar than unpaired samples underlining the same biological source (median Bray-Curtis similarity 0.26 and 0.04, respectively, $p<0.001$, Figure 1B). The similarity of paired NP and TTO samples did vary slightly with age (median Bray-Curtis similarity; <2 years, 0.27; >2 years, 0.11; $p=0.093$), but not with number of previous tympanostomy tubes (1 tube, 0.25; >1 tube, 0.15; $p=0.446$), duration of tube presence (0-5 days, 0.20; >5 days, 0.16; $p=0.849$), history of prior adenoidectomy (yes, 0.14; no 0.26; $p=0.595$), nor with season of sampling ($p=0.899$; eFigure 5 in the Supplement). TTO samples from both ears of the same child (n=15 with bilateral AOMT) were substantially more similar than TTO samples of different children (Bray-Curtis similarity 0.50 and 0.02, respectively, $p<0.001$).

Microbiota profiles and biomarker species

Hierarchical clustering showed the presence of 10 distinct microbiota profiles, which were mainly driven by the abundance of 12 biomarker species (Figure 2A). Most biomarker species were differentially abundant in NP and TTO samples, except for *Streptococcus*(7), *Klebsiella* and *Haemophilus* (91), which showed high concordance for presence as well as abundance between niches. In contrast, *Moraxella spp., S. pneumoniae* (6), *H. influenzae* (1), *Corynebacterium*, and *Dolosigranulum* were stronger associated with the NP, whereas *Turicella*, *P. aeruginosa* (5), and *S. aureus* (2) abundances were more associated with TTO (metagenomeSeq absolute log2 fold change, all >2; q<0.01; Figure 2B). A posteriori plotting of all biomarker species in the NMDS ordination supported the niche-preferential abundance as described above (Figure 1A).

On the individual level, 30% of the paired NP and TTO samples, however, shared the exact same microbiota profile (Figure 2C-D). This one-to-one association was most obvious for the *Haemophilus*-, *S. aureus* (2)-, *Streptococcus* (7) - and *Klebsiella*-dominated profiles. The *Streptococcus* (7) NP-profile was additionally associated with the same profile in TTO, also associated with a *S. pneumoniae*-dominated TTO-profile. The *M. catarrhalis* NP-profile was rarely found in TTO. However, a strong association was observed between Moraxella-dominated NP and *P. aeruginosa*-dominated TTO (Figure 2C-D).

Agreement in microbiota composition

In contrast to the relatively low correlation between paired NP and TTO samples on total microbiota profile level (Figure 2), the concordance on the single bacterial species level (OTU level) was considerably higher with a substantial agreement of 79% for the presence/absence of individual species (95% CI 78-80%; eTable 2 in the Supplement). The high NPV underlines that the NP might be the common biological source of TTO bacteria (91%, 95% CI 91-92%).

The quantitative correlation between the bacterial abundances of individual species in the paired NP and TTO samples was in line with the qualitative results, with 12 out of the 15 most abundant bacterial species showing a significant correlation between NP and TTO (p<0.05; Spearman's rho range, 0.193-0.548; Figure 3); *H. influenzae* (1) (Spearman's rho 0.548, p<0.001), *P. aeruginosa* (5) (Spearman's rho

0.489, p<0.001) and *S. aureus* (2) abundance (Spearman's rho 0.439, p<0.001) showed the strongest correlations, whereas *Moraxella* spp. (including *M. catarrhalis* [3]) and *Streptococcus* spp. (including *S. pneumoniae* [6], Spearman's rho 0.180, p=0.061) abundances were clearly not correlated between NP and TTO. When analysing also the lower abundant bacterial species, only 46 of the 138 species showed a significant correlation (p<0.05; median Spearman's rho 0.337; IQR, 0.247–0.436; combined relative abundance of 81.5%), suggesting low abundant species are less likely seeded from NP to middle ear.

Random forest associations

All results together confirmed our hypothesis that the NP microbiota composition does not fully reflect TTO microbiota in a simple one-to-one fashion. Despite this, we found that microbial profiles of NP samples still predicted the microbial community in the paired TTO samples fairly well, with an almost one-to-one association when dominated by *H. influenzae* (1) and *Haemophilus* (91), *Klebsiella*, *Corynebacterium*, and *Streptococcus* (7) (Figure 4). Moreover, *S. aureus* (2) abundance in the NP was predictive for either *S. aureus* (2) or *Neisseria* overgrowth in TTO as well as absence of other species. Similarly, *P. aeruginosa* (5) abundance in the NP swab was predictive for either *Pseudomonas* or *Staphylococcus* abundance in TTO. *Dolosigranulum* abundance in NP demonstrated a less specific association with TTO bacterial abundances. *M. catarrhalis* (3) was highly predictive of other species but itself, especially *Pseudomonas*. *S. pneumoniae* (6) abundance in the NP was mostly associated with presence of a diverse group of (oral) anaerobes, though not itself.

Relation between microbiota and clinical outcome

Although the baseline respiratory microbiota community profiles of the children allocated to the initial observation group could not predict the otoscopically confirmed presence or absence of otorrhea two weeks after onset of symptoms very accurately (AUC 0.71 and 0.62 for the sparse RF models using NP and TTO microbiota, respectively), the microbiota composition of NP samples could predict the duration of symptoms and recurrence of otorrhea as reported by the parents fairly well (Pearson's r between predicted and observed outcome 0.40-0.54, all p<0.05, random forest R^2 0.69-0.70; Figure 5A), whereas the models using TTO microbiota did not demonstrate a significant correlation between predicted and

observed outcome values (all p>0.10). Within this untreated group, especially the NP abundance of *Acinetobacter*, followed by *Klebsiella*, *Neisseria*, and *H. influenzae* (1) (positive partial Spearman's correlation) were associated with longer duration of otorrhea, whereas abundance of *Corynebacterium*, followed by *Dolosigranulum* and *Haemophilus* (91) were associated with shorter duration of otorrhea (negative partial Spearman's correlation; Figure 5B).

DISCUSSION

This study, comparing paired NP and TTO samples of 94 children with AOMT, shows a substantive qualitative and moderate quantitative correlation between NP and TTO thereby supporting the hypothesis that the microbiota in the middle ear originates from the NP. Moreover, NP microbiota composition predicts presence and absence of other microbiota in the TTO well, with for example *S. aureus* abundance in the NP predicting either the presence of *S. aureus* or *Pseudomonas* in the middle ear. Second, our study indicates that the TTO microbiota of children with AOMT is a rich community comprising of on average 26 species, suggesting the existence of a complex middle ear microbiome in those children rather than the presence of a single pathogen.

In accordance with previous small studies, NP samples show a higher α-diversity compared to TTO and the total microbiota composition differed significantly between both niches.[10,23] Although high qualitative concordance was found, our analyses also showed that some biomarker species are overrepresented in NP samples, whereas other biomarker species are more abundant in TTO samples, suggesting niche preference. Especially the association of bacteria like *M. catarrhalis*, other *Moraxella spp.*, *S. pneumoniae*, and *Corynebacterium* /*Dolosigranulum* with NP rather than TTO presence, confirms previous findings that these microbes are key commensals of the NP niche ecosystem.[24–27] Moreover, the association of *Turicella*, *P. aeruginosa*, and *S. aureus* within TTO samples corroborates reports that describe these species as otopathogens.[10,27–29]

The difference in niche preference between bacteria is presumably driven by the niche specific growth condition of both sites such as oxygen tension, temperature, humidity, presence of nutrients or immune cells.[26] Moreover, seeding of microorganisms from the NP to the middle ear through the Eustachian tube

and local outgrowth might not solely depend on the presence and/or abundance of these microorganisms in the ascending community, but also on their relative abundance, as well as on the presence of other microbial community members that may either support or prevent their dissemination. By analysing the association of the microbial profiles on the individual level as well as using quantitative correlations associating NP and TTO microbiota, a significant one-to-one relationship between the NP and TTO abundances was found for the majority of the microbiota (81.5%). The strength of the correlations was generally modest but was highest for potential (AOMT) pathogens such as *P. aeruginosa*, *S. aureus*, *Streptococcus*, *Turicella*, and *Haemophilus* spp. This was confirmed using unsupervised random forest analysis. Moreover, random forest analysis also demonstrated that the abundances of *P. aeruginosa*, *S. aureus*, and *Turicella* in NP were additionally associated with a range of gram-negative oral type of bacteria in the TTO, including *Neisseria*, *Bradyrhizobium*, *Bergeyella* and *Actinomyces* spp. A possible explanation for this symbiotic behavior might be the ability of *P. aeruginosa*[30] to rapidly reduce these species' toxic oxygen levels, and vice versa the known facilitation of *P. aeruginosa* growth by the metabolites of these oral bacteria.[31] Interestingly, *S. pneumoniae* abundance in the NP predicted mostly the presence of a diverse group of anaerobes in TTO, whereas in the few occasions *S. pneumoniae* occurred in TTO this was mostly predicted by the abundance of *Streptococcus* (7). This suggests that collaboration between both species (which is a well-known phenomenon for streptococcal species[26]) is needed for the currently circulating serotypes to colonize the middle ear niche, and render pathogenic behavior.

Over the last years, with the advent of microbial community profiling, evidence is accumulating that *M. catarrhalis* is associated with a stable bacterial community composition and a state of respiratory health.[26] In our study, only three TTO samples had a *M. catarrhalis* dominated profile, suggesting a limited role of this species in AOM(T) pathogensis. While our study population consisted of children with previous otitis media episodes requiring tympanostomy tubes, NP abundance of *M. catarrhalis* showed no association with its presence in TTO samples across all our analyses. These data are corroborated by a recent study from Australia [27], therefore suggest that this bacterium is rather a NP commensal than a pathogen. Other studies however have also reported that *M. catarrhalis* could play a in acute otitis media in children,[32,33] although in those cases, it generally reflects a mild infection.[3] In

addition, the historical common detection of *M. catarrhalis* in conventional culture-based studies might mirror the easy identification of this species by culture, rather then a high abundance in middle ear fluid. In all, this warrants a careful consideration of future vaccination strategies against this microorganism.[34]

Some limitations deserve further attention. First, we did not include children with a body temperature higher than 38.5°C, who might have different microbiota profiles. Second, TTO was sampled from the ear canal after a median otorrhea duration of two days (IQR, 1-4); the contamination by external ear canal microbiota might have led to an overestimation of *P. aeruginosa*, *Turicella*, and/or *S. aureus* detection as these species are common constituent of the microbiota in the ear canal.[11,35] However, we have previously compared bacterial presence in otorrhea samples swabbed from the ear canal with those taken from the lumen of the tympanostomy tube in a subset of 20 children participating in the trial and did observe a high concordance, suggesting limited outer ear canal contamination.[9] Also, the high correlation between the abundances of *P. aeruginosa*, *Turicella*, and *S. aureus* in NP and TTO samples might further indicate that their abundances in TTO are not merely the result of contamination from the outer ear canal, but that these species rather originate from the nasopharyngeal niche. Our results are in line with other recent studies that detect *Pseudomonas* and *Turicella* in low abundance in the nasopharynx of the majority of children without tympanostomy tubes.[27,36,37] We cannot exclude, however, that the presence of *P. aeruginosa* and *T. otitidis* in the nasopharynx may be the result of its reversed transition from the TTO to the nasopharynx.

Although NP microbiota did aptly predict the natural disease course of AOMT as defined by three different clinical outcome measures as reported by parents, we did not observe a significant relation between NP microbiota and otorrhea two weeks after randomization as confirmed by a physician through otoscopic examination. This ambiguity may well be due to sample size constraints, as only 27 of the 94 children included in the current study were allocated to the initial observation group. Although our prediction algorithms may not be accurate enough for their direct implementation in the clinical setting, they might open-up new avenues to refrain from treatment in children with AOMT whom NP microbiota profiles indicate a favorable natural disease course and to initiate treatment in those with a less favorable predicted outcome. Further testing and validation in prospective cohorts is, however, warranted. With AOM being the single most important cause of childhood antibiotic prescribing, it does

seem worthwhile to further study the potential usefulness of microbiota analysis to predict clinical outcome and its impact on antimicrobial use and subsequent development of antimicrobial resistance. Of particular note is that we could not predict the natural disease course using TTO microbiota, although TTO reflects the site of infection. This may suggest that the NP microbiota not only seed the middle ear with potential pathogens that initiate disease but also determines recovery to health. This is strengthened by the finding of a strong association between *Dolosigranulum* and *Corynebacterium* abundance and a better clinical outcome, supporting evidence that these bacteria are associated with respiratory health.[26,27] We hypothesize that children colonized with these beneficial microbes have diminished mucosal inflammation, leading to more rapid restoration of Eustachian tube function, and subsequently clinical recovery.

In conclusion, this study offers valuable insights into the association between NP and TTO microbiota compositions in children with AOMT. Our findings of substantial niche-niche relationships endorse the hypothesis that the middle ear microbiota is seeded by the NP microbiota through ascending the Eustachian tube. Moreover, our results suggest that *M. catarrhalis* could be a NP commensal rather than a pathogen, which is a relevant finding with regard to future vaccine strategies and that warrants further investigation. Finally, our data indicate that NP microbiota profiles may be useful for clinical decision-making in the future, but for this more research is needed.

ACKNOWLEDGMENT

The trial was supported by a grant (170992502) from the Netherlands Organization for Health Research and Development Effects and Costs subprogram. We thank the children and their parents who participated in the study; Pauline Winkler, Nelly van Eden, Lidian Izeboud, Dicky Mooiweer, and our team of medical students for administrative and practical support; the participating family physicians and the ear, nose, and throat surgeons at the participating hospitals.

AUTHOR CONTRIBUTIONS

WHM, MAvH, EAMS, and DB designed the experiments in this study. TMAvD, RPV, and AGMS were investigators of the primary randomized controlled trial, contributed to the study design, and were responsible for patient recruitment and clinical data collection. MLJNC and VGP were responsible for sample preparation and 16S-rRNA gene amplicon sequencing. WHM, VGP, and DB were responsible for bioinformatic processing and statistical analyses. WHM and DB wrote the paper. All authors significantly contributed to interpreting the results, critically revised the manuscript, and approved the final manuscript.

AVAILABILITY OF DATA AND MATERIALS

The 16S rRNA sequence reads were submitted to the National Center for Biotechnology and Information Sequence Read Archive (accession number SRP128433).

REFERENCES

1. Monasta L, Ronfani L, Marchetti F, et al. Burden of Disease Caused by Otitis Media: Systematic Review and Global Estimates. *PLoS One*. 2012;7(4):e36226.
2. Coker TR, Chan LS, Newberry SJ, et al. Diagnosis, Microbial Epidemiology, and Antibiotic Treatment of Acute Otitis Media in Children. *JAMA*. 2010;304(19):2161.
3. Aebi C. *Moraxella Catarrhalis – Pathogen or Commensal?* Vol 697. (Curtis N, Finn A, Pollard AJ, eds.). New York, NY: Springer New York; 2011.
4. Broides A, Dagan R, Greenberg D, Givon-Lavi N, Leibovitz E. Acute Otitis Media Caused by Moraxella catarrhalis: Epidemiologic and Clinical Characteristics. *Clin Infect Dis*. 2009;49(11):1641-1647.
5. Schilder AGM, Chonmaitree T, Cripps AW, et al. Otitis media. *Nat Rev Dis Prim*. 2016;2:16063.
6. Preciado D. *Otitis Media: State of the Art Concepts and Treatment*. (Preciado D, ed.). Cham: Springer International Publishing; 2015.
7. van Dongen TM, van der Heijden GJ, van Zon A, Bogaert D, Sanders EAM, Schilder AGM. Evaluation of concordance between the microorganisms detected in the nasopharynx and middle ear of children with otitis media. *Pediatr Infect Dis J*. 2013;32(5):549-552.
8. Liu CM, Cosetti MK, Aziz M, et al. The otologic microbiome: a study of the bacterial microbiota in a pediatric patient with chronic serous otitis media using 16SrRNA gene-based pyrosequencing. *Arch Otolaryngol Head Neck Surg*. 2011;137(7):664-668.
9. van Dongen TMA, Venekamp RP, Wensing AMJ, Bogaert D, Sanders EAM, Schilder AGM. Acute otorrhea in children with tympanostomy tubes: prevalence of bacteria and viruses in the post-pneumococcal conjugate vaccine era. *Pediatr Infect Dis J*. 2015;34(4):355-360.
10. Jervis-Bardy J, Rogers GB, Morris PS, et al. The microbiome of otitis media with effusion in Indigenous Australian children. *Int J Pediatr Otorhinolaryngol*. 2015;79(9):1548-1555.
11. Chan CL, Wabnitz D, Bassiouni A, Wormald P-J, Vreugde S, Psaltis AJ. Identification of the Bacterial Reservoirs for the Middle Ear Using Phylogenic Analysis. *JAMA Otolaryngol Neck Surg*. 2016;13(2):Doc 11.
12. van Dongen TMA, van der Heijden GJMG, Venekamp RP, Rovers MM, Schilder AGM. A trial of treatment for acute otorrhea in children with tympanostomy tubes. *N Engl J Med*. 2014;370(8):723-733.
13. O'Brien K, Nohynek H, World Health Organization Pneumococcal Vaccine Trials Carriage Working Group. Report from a WHO Working Group: standard method for detecting upper respiratory carriage of Streptococcus pneumoniae. *Pediatr Infect Dis J*. 2003;22(2):e1-11.
14. Bosch AATM, de Steenhuijsen Piters WAA, van Houten MA, et al. Maturation of the Infant Respiratory Microbiota, Environmental Drivers, and Health Consequences. A Prospective Cohort Study. *Am J Respir Crit Care Med*. 2017;196(12):1582-1590.
15. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12(1):87.
16. Subramanian S, Huq S, Yatsunenko T, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*. 2014;510(7505):417.
17. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods*. 2013;10(12):1200-1202.
18. Good IJ. The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika*. 1953;40(3-4):237-264.
19. De Steenhuijsen Piters WAA, Heinonen S, Hasrat R, et al. Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. *Am J Respir Crit Care Med*. 2016;194(9):1104-1115.
20. Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests. *Pattern Recognit Lett*. 2010;31(14):2225-2236.
21. Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5-32.
22. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
23. Chan CL, Wabnitz D, Bardy JJ, et al. The microbiome of otitis media with effusion.

Laryngoscope. 2016;126(12):2844-2851.
24. Biesbroek G, Tsivtsivadze E, Sanders EAM, et al. Early Respiratory Microbiota Composition Determines Bacterial Succession Patterns and Respiratory Health in Children. *Am J Respir Crit Care Med*. 2014;190(11):1283-1292.
25. Pettigrew MM, Laufer AS, Gent JF, Kong Y, Fennie KP, Metlay JP. Upper respiratory tract microbial communities, acute otitis media pathogens, and antibiotic use in healthy and sick children. *Appl Environ Microbiol*. 2012;78(17):6262-6270.
26. Man WH, de Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol*. 2017;15(5):259-270.
27. Lappan R, Imbrogno K, Sikazwe C, et al. A microbiome case-control study of recurrent acute otitis media identified potentially protective bacterial genera. *BMC Microbiol*. 2018;18(1):13.
28. Krueger A, Val S, Pérez-Losada M, et al. Relationship of the Middle Ear Effusion Microbiome to Secretory Mucin Production in Pediatric Patients With Chronic Otitis Media. *Pediatr Infect Dis J*. 2017;36(7):635-640.
29. Daniel M, Imtiaz-Umer S, Fergie N, Birchall JP, Bayston R. Bacterial involvement in otitis media with effusion. *Int J Pediatr Otorhinolaryngol*. 2012;76(10):1416-1422.
30. Kim E-J, Wang W, Deckwer W-D, Zeng A-P. Expression of the quorum-sensing regulatory protein LasR is strongly affected by iron and oxygen concentrations in cultures of Pseudomonas aeruginosa irrespective of cell density. *Microbiology*. 2005;151(4):1127-1138.
31. Tunney MM, Field TR, Moriarty TF, et al. Detection of Anaerobic Bacteria in High Numbers in Sputum from Patients with Cystic Fibrosis. *Am J Respir Crit Care Med*. 2008;177(9):995-1001.
32. Sillanpää S, Oikarinen S, Sipilä M, et al. Moraxella catarrhalis Might Be More Common than Expected in Acute Otitis Media in Young Finnish Children. *J Clin Microbiol*. 2016;54(9):2373-2379.
33. Hilty M, Qi W, Brugger SD, et al. Nasopharyngeal microbiota in infants with acute otitis media. *J Infect Dis*. 2012;205(7):1048-1055.
34. Murphy TF, Parameswaran GI. Moraxella catarrhalis, a Human Respiratory Tract Pathogen. *Clin Infect Dis*. 2009;49(1):124-131.
35. Brook I, Yocum P, Shah K. Aerobic and Anaerobic Bacteriology of Otorrhea Associated with Tympanostomy Tubes in Children. *Acta Otolaryngol*. 1998;118(2):206-210.
36. Chonmaitree T, Jennings K, Golovko G, et al. Nasopharyngeal microbiota in infants and changes during viral upper respiratory tract infection and acute otitis media. Miyaji EN, ed. *PLoS One*. 2017;12(7):e0180630.
37. Teo SM, Mok D, Pham K, et al. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe*. 2015;17(5):704-715.

FIGURES

Figure 1: **Total microbiota composition differs between NP and TTO samples, although paired NP and TTO samples of the same participants were more similar compared to unpaired samples.** **(A)** NMDS biplot depicting the individual NP (red) and TTO (blue) microbiota compositions. Ellipses represent the standard deviation of all points within a niche. In addition, the biplot depicts the interrelation with the 12 biomarker species (determined by random forest analysis on hierarchical clustering results). The shape and color of the biomarker species represents its phylum: Proteobacteria green square; Firmicutes orange triangle; Actinobacteria purple circle.

(B)

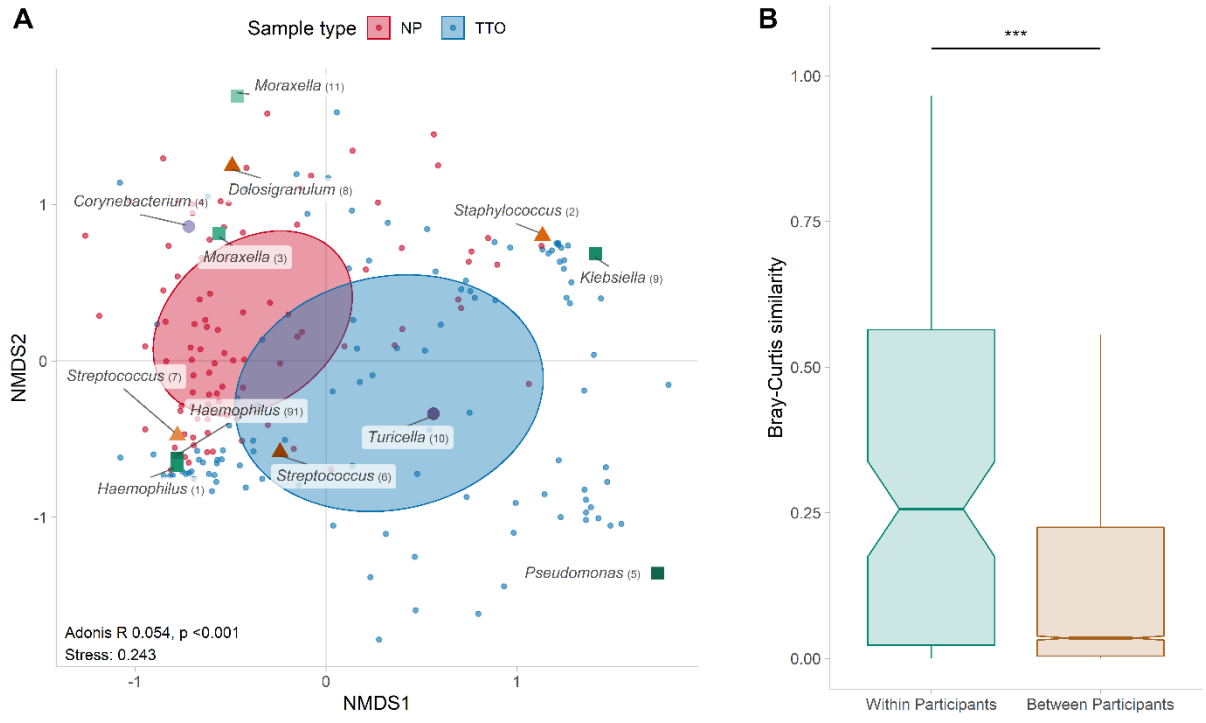


Figure 2: **Hierarchical clustering reveals 10 profiles.**

(A) Hierarchical clustering of all study samples identified 13 clusters, of which 10 represented 5 or more samples each. Biomarker species for these 10 profiles were determined using random forest analysis: *Pseudomonas aeruginosa* (5) (PSEUD, mint); *Klebsiella* (9) (KLEBS, yellow); *Staphylococcus aureus* (2) (STAPH, purple); *Streptococcus pneumoniae* (6) (STREP1, red); *Turicella* (10) (TURIC, blue); *Streptococcus* (7) (STREP2, orange); *Haemophilus* spp. (HAEMO, green); *Moraxella catarrhalis* (3) (MORA1, pink); *Moraxella spp.* (MORA2, grey); *Corynebacterium* (4) and *Dolosigranulum* (8) (CO_DO, dark purple). The figure visualizes the clustering dendrogram, including information on the distribution of the NP samples (red) and TTO samples (blue), and the relative abundances of the 12 biomarker species.

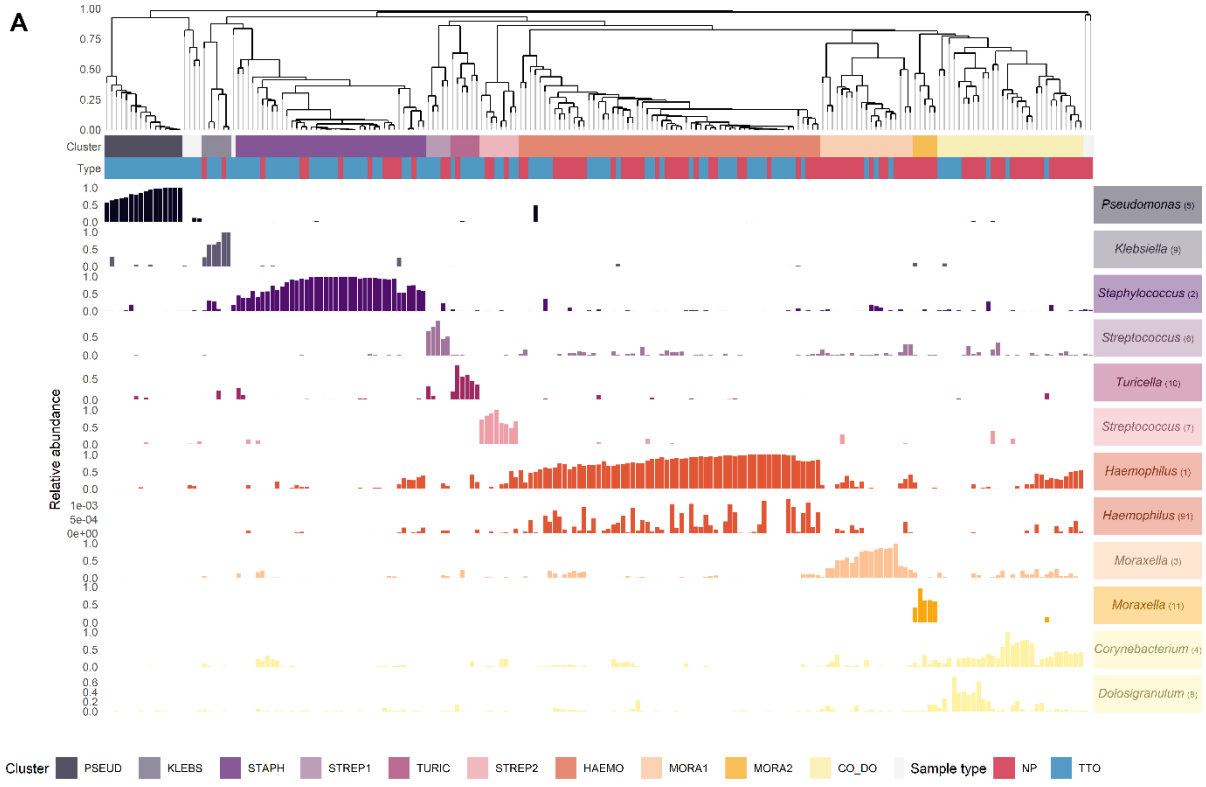
(B) Associations between the biomarker species abundance and either the NP niche or the TTO niche (determined by *metagenomeSeq*). *Turicella*, *P. aeruginosa* (5) and *S. aureus* (2) were positively associated with the TTO, whereas multiple *Moraxella* spp., *H. influenzae* (1), *S. pneumoniae* (6), *Corynebacterium* and *Dolosigranulum* were associated with the NP.

(C) Visualization of the relation between the overall microbiota profile of the NP samples and that of the paired TTO samples of the same participant as a parallel alluvial diagram. The alluvial diagram depicts the direct links between the microbiota profile of the NP samples (left) and that of the paired TTO samples (right). Green lines represent participants that have the same profile in both niches (n = 33) and brown lines represent participants that have different profiles in both niches (n = 76).

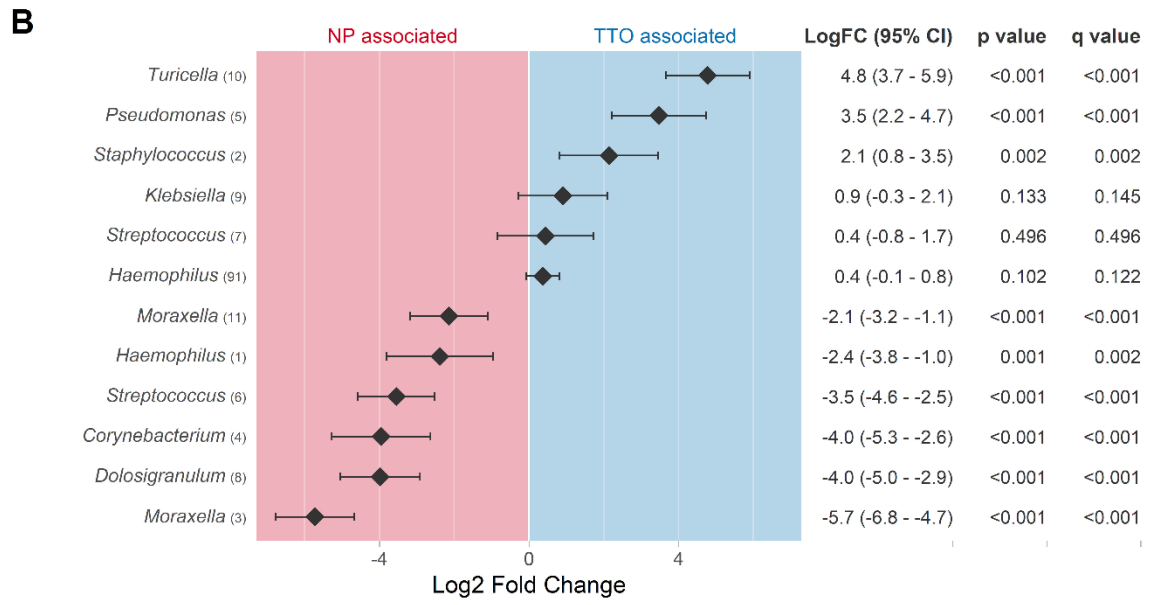
(D) Visualization of the relation between the overall microbiota profile of the NP samples and that of the paired TTO samples of the same participant as a heatmap. The circles show the absolute number of paired samples per profile, whereas the size and color of the circles represent the proportion of these sample in relation to the total NP microbiota profile. The emphasis of the main diagonal for some microbiota profiles (i.e. *Klebsiella* [KLEBS], *Staphylococcus aureus* [STAPH], *Streptococcus* (7) [STREP2] and *Haemophilus* [HAEMO]) indicates that these profiles have a strong one-to-one relation in both niches. Especially the *Staphylococcus*, *Streptococcus* and *Haemophilus*-dominated profiles in the NP are highly predictive for the correlating profile in the TTO fluid or an alternative profile (e.g.

549 *Haemophilus*-dominated profiles predicts for either *Haemophilus* or *Staphylococcus*-dominated TTO

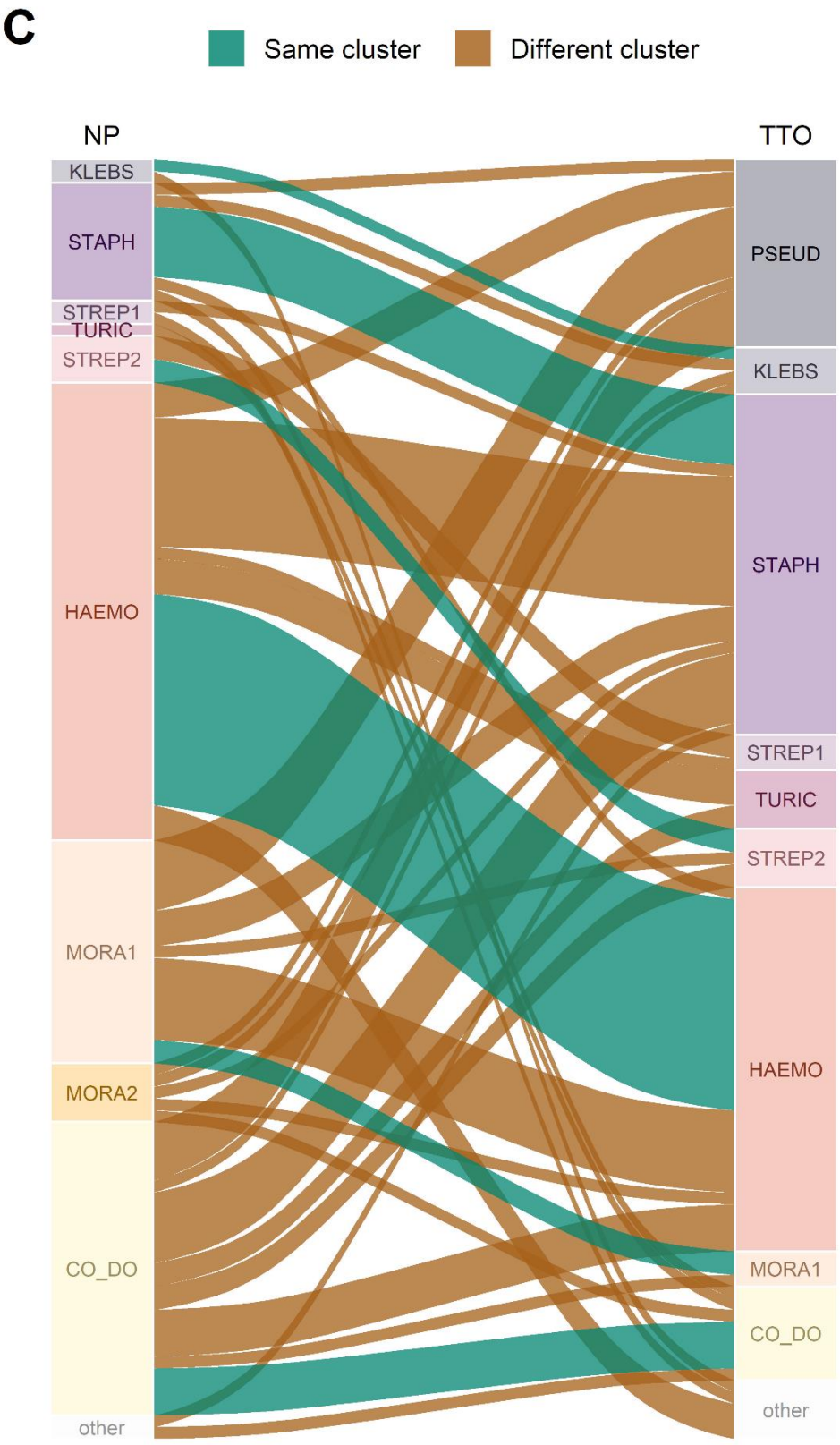
550 profile). Significance symbols (Fischer's exact tests): * = p<0.05; . = p<0.10.



551



552


C
Same cluster
Different cluster
NP
TTO
KLEBS
STAPH
STREP1
TURIC
STREP2
HAEMO
MORA1
MORA2
CO_DO
other
PSEUD
KLEBS
STAPH
STREP1
TURIC
STREP2
HAEMO
MORA1
CO_DO
other

553

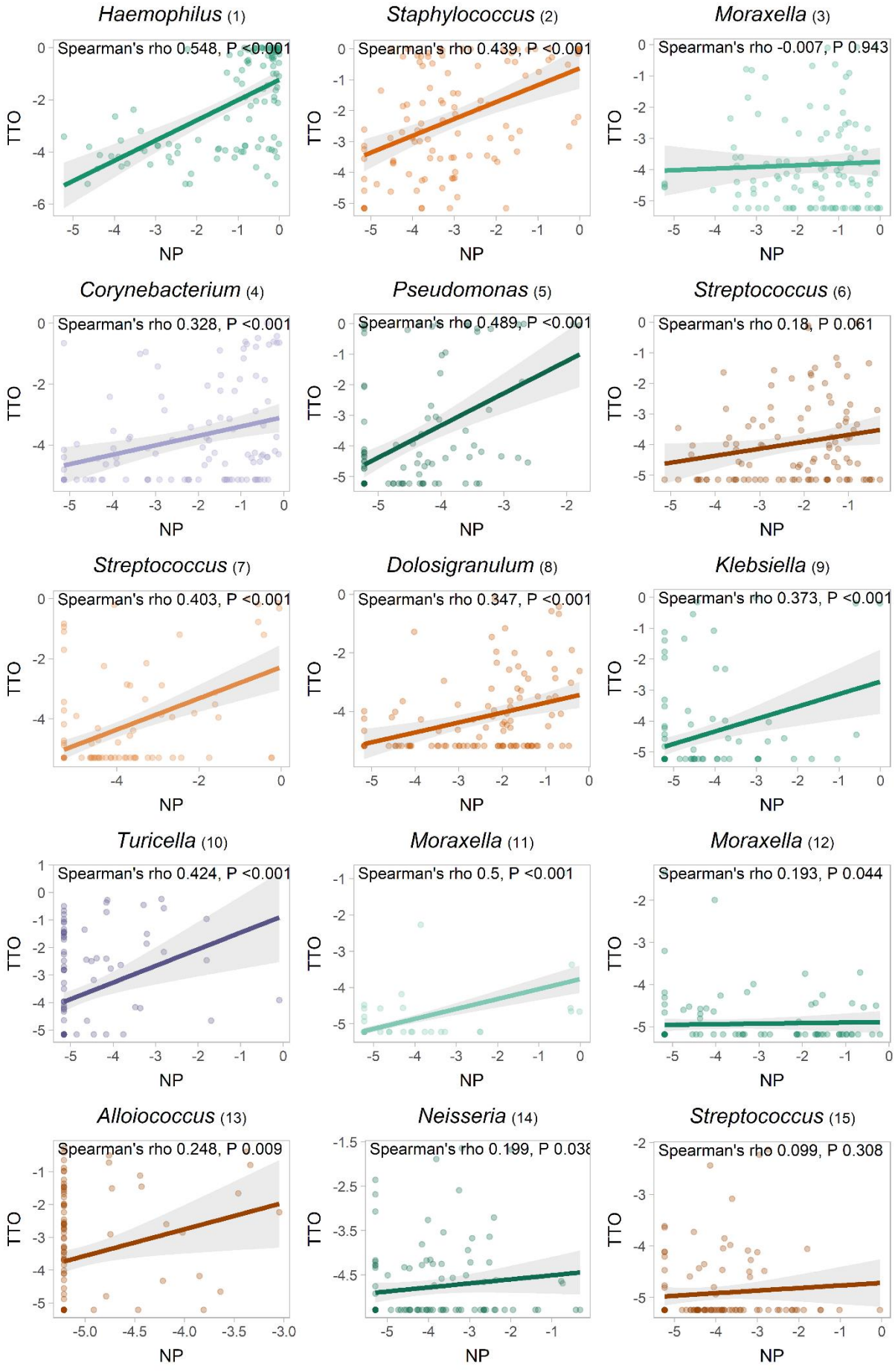
D


% within NP cluster
0%
6%
12%
25%
50%
100%
NP
PSEUD
KLEBS
STAPH
STREP1
TURIC
STREP2
HAEMO
MORA1
MORA2
CO_DO
other
TTO
PSEUD KLEBS STAPH STREP1 TURIC STREP2 HAEMO MORA1 MORA2 CO_DO other

554

Figure 3: **Twelve of the 15 most abundant bacterial species have a strong positive quantitative correlation of their abundances in the paired NP and TTO samples.**

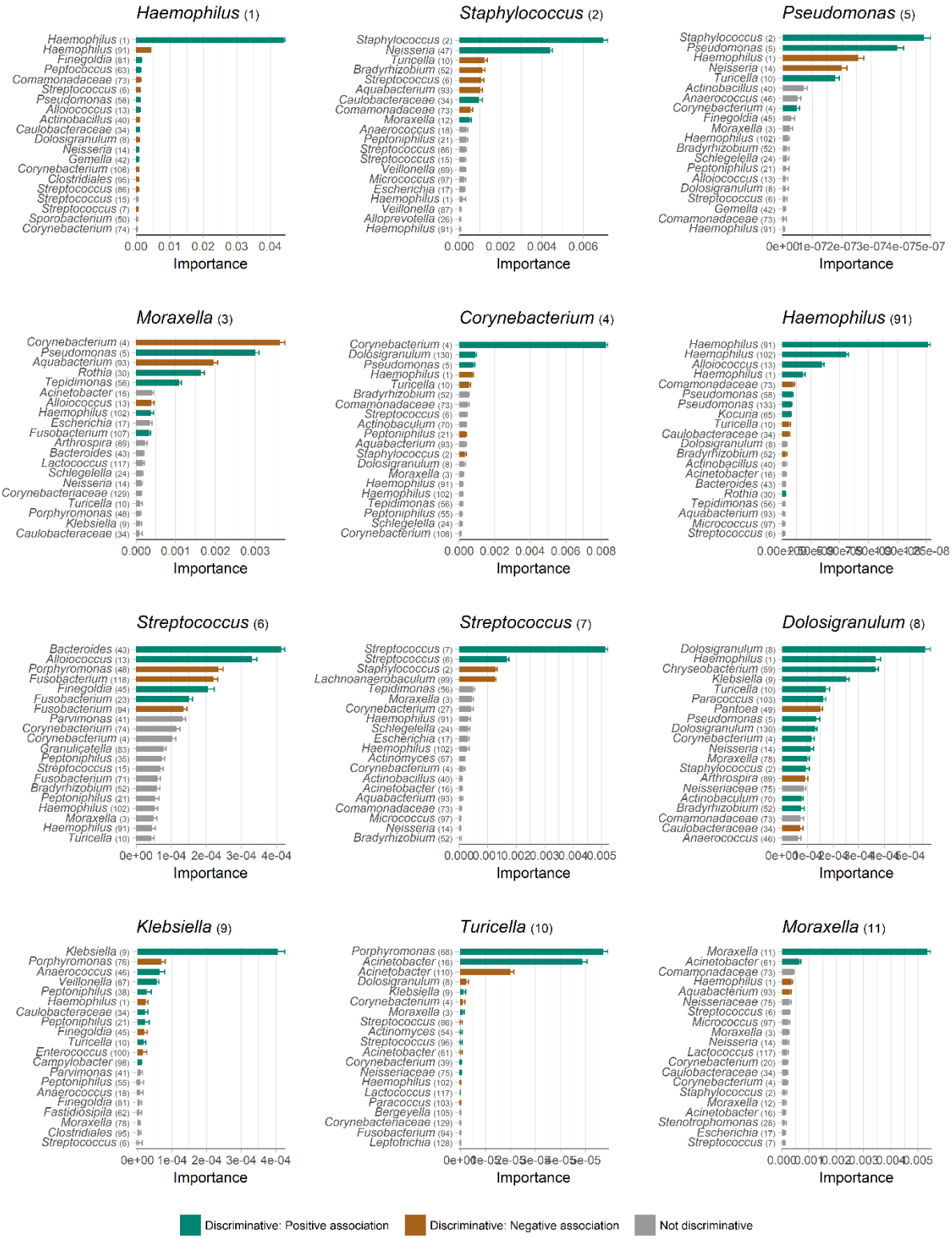
Linear dependency and Spearman's correlation coefficient between the relative abundance of the NP samples and that of the paired TTO samples (logarithmic scales).


Haemophilus (1)
Spearman's rho 0.548, P <0.001
Staphylococcus (2)
Spearman's rho 0.439, P <0.001
Moraxella (3)
Spearman's rho -0.007, P 0.943
Corynebacterium (4)
Spearman's rho 0.328, P <0.001
Pseudomonas (5)
Spearman's rho 0.489, P <0.001
Streptococcus (6)
Spearman's rho 0.18, P 0.061
Streptococcus (7)
Spearman's rho 0.403, P <0.001
Dolosigranulum (8)
Spearman's rho 0.347, P <0.001
Klebsiella (9)
Spearman's rho 0.373, P <0.001
Turicella (10)
Spearman's rho 0.424, P <0.001
Moraxella (11)
Spearman's rho 0.5, P <0.001
Moraxella (12)
Spearman's rho 0.193, P 0.044
Alloiococcus (13)
Spearman's rho 0.248, P 0.009
Neisseria (14)
Spearman's rho 0.199, P 0.038
Streptococcus (15)
Spearman's rho 0.099, P 0.308
TTO
NP

559

Figure 4: **Predictive value of NP biomarker abundances for TTO abundances.**

Sparse random forest analyses determining the predictive value of biomarker species abundance in NP for abundance of corresponding and other species in TTO (cross validated *VSURF* selection). The importance (increase of MSE) reflects how important a certain TTO species for the accuracy of the random forest model, i.e. how strong the association is with the NP biomarker species tested. The direction of the associations was estimated post-hoc using the partial Spearman's correlations. An almost one-to-one relationship was only observed for *H. influenzae* (1) and *Haemophilus* (91), *Klebsiella*, *Corynebacterium*, and *Streptococcus* (7). *Staphylococcus aureus* (2) dominance in the NP was associated with *S. aureus* (2) and *Neisseria* abundance in TTO, whereas *P. aeruginosa* (5) dominance was associated with both *P. aeruginosa* (5) and *S. aureus* (2) dominance in TTO. *M. catarrhalis* (3) was highly predictive of other species but itself, underlining its limited role in otitis pathogenesis.

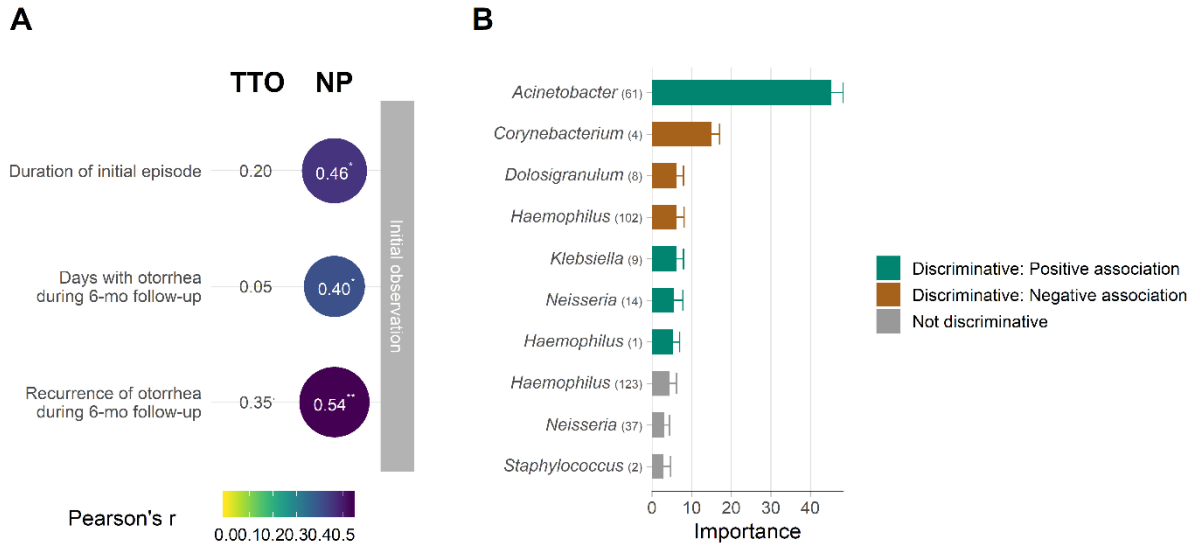

Haemophilus (1)
Staphylococcus (2)
Pseudomonas (5)
Moraxella (3)
Corynebacterium (4)
Haemophilus (91)
Streptococcus (6)
Streptococcus (7)
Dolosigranulum (8)
Klebsiella (9)
Turicella (10)
Moraxella (11)
Importance
Discriminative: Positive association
Discriminative: Negative association
Not discriminative

571

Figure 5: **NP microbiota is superior to TTO microbiota in predicting short- and long-term outcome.**

(A) Sparse random forest regression analyses using cross validated *VSURF* selected bacterial species were performed to predict within the initial observation group the duration of the initial otorrhea episode, the total number of days with otorrhea during 6 months of follow-up and the recurrence of otorrhea during 6 months of follow-up. The duration of the initial otorrhea episode was defined as the interval from the day of study-group assignment up to the first day of otorrhea that was followed by 7 or more days without otorrhea. Recurrence of otorrhea during 6 months of follow-up was defined as an episode of otorrhea lasting 1 or more days after an otorrhea-free period of 7 or more days. The correlation between the predicted and observed values of the outcome is visualized in a heatmap. Circles are only depicted for significant correlations. The size and color of the circles correspond to the Pearson correlation coefficients.

(B) The duration of the natural course of recovery from the otorrhea episode could best be predicted (random forest R2 0.70) using the NP abundances of *Acinetobacter*, *Klebsiella*, *Neisseria*, *H. influenzae* (1) (positive partial Spearman's correlation), as well as *Corynebacterium*, *Dolosigranulum* and *Haemophilus* (91) (negative partial Spearman's correlation). Significance symbols: *** = p<0.001; ** = p<0.01; * = p<0.05.



SUPPLEMENTAL DIGITAL CONTENT

- **eTable 1.** Characteristics of the study population at baseline.
- **eTable 2.** Qualitative agreement between matched pairs of NP and TTO samples on OTU level.
- **eFigure 1.** Culture results confirm the taxonomic annotation of the corresponding OTU's.
- **eFigure 2.** Flow chart participants and samples.
- **eFigure 3.** Rarefaction curves on raw count data.
- **eFigure 4.** α-Diversity.

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 *32*

16S rRNA Gene Amplification and Sequencing 32

Bioinformatics Analysis 32

eTable 1. Characteristics of the study population at baseline. *34*
eTable 2. Qualitative agreement between matched pairs of NP and TTO samples on OTU level. *35*
eFigure 1. Culture results confirm the taxonomic annotation of the corresponding OTU's. *36*
eFigure 2. Flow chart participants and samples. *37*
eFigure 3. Rarefaction curves on raw count data. *38*
eFigure 4. α-Diversity. *39*
eFigure 5. Similarity of paired NP and TTO samples does not vary with clinical variables. *40*
References *42*

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

641 random forest modelling. Moreover, the Bray-Curtis (dis)similarity metric was consistently used to
642 express ecological distance (β-diversity) in all analyses because it includes proportional abundance
643 information and excludes joint-absence information, and thereby yields useful insights into the specific
644 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)

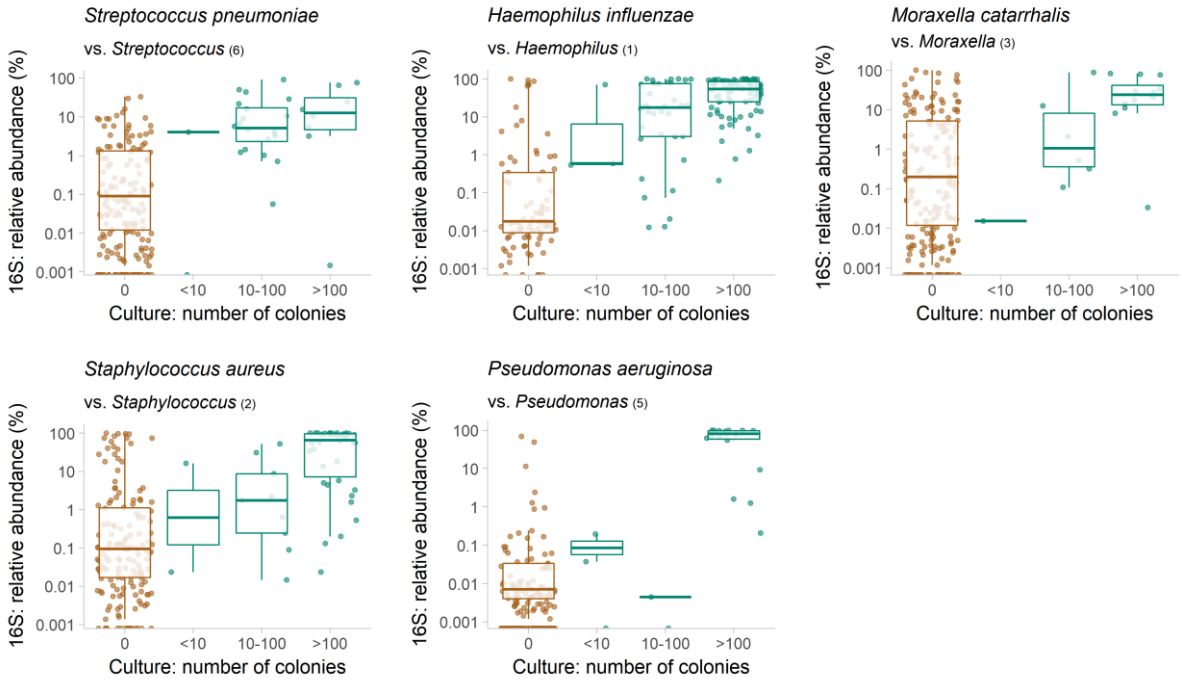
eTable 2. Qualitative agreement between matched pairs of NP and TTO samples on OTU level.

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

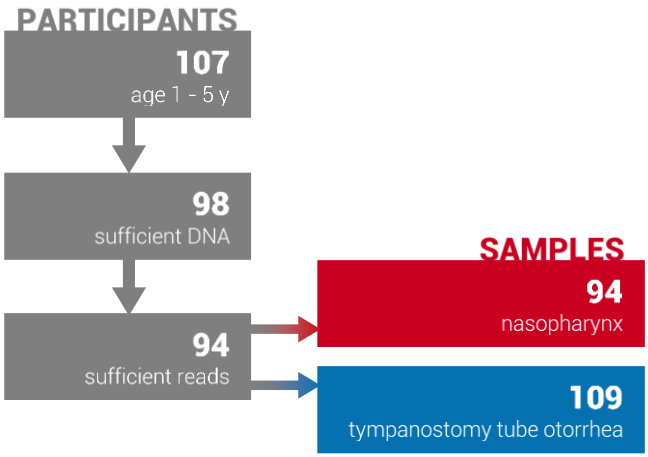
eFigure 1. Culture results confirm the taxonomic annotation of the corresponding OTU's.

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.



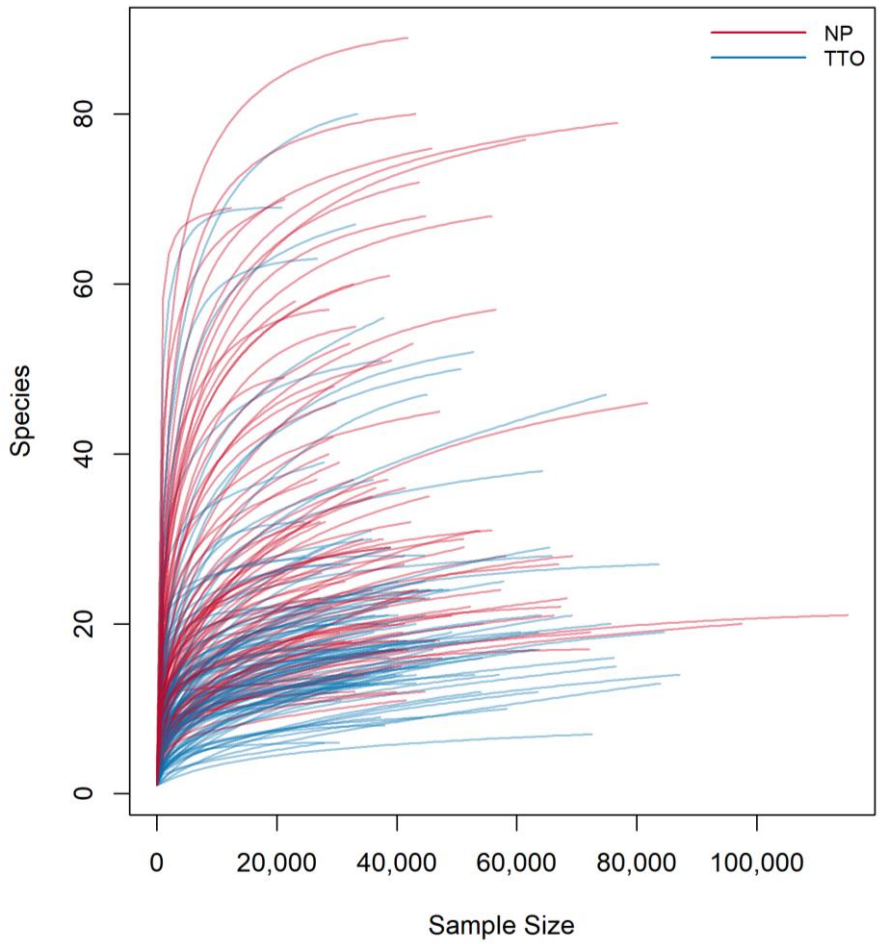
eFigure 2. Flow chart participants and samples.

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.

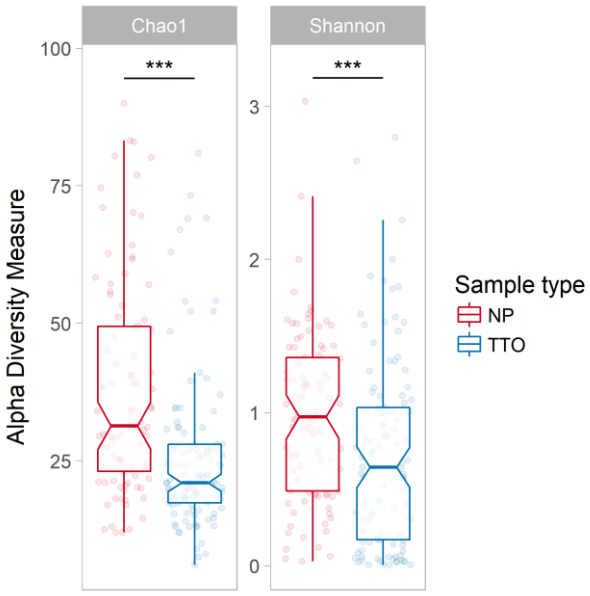
Rarefaction curves on raw count data approached plateau for both NP samples (red) and TTO samples (blue).



eFigure 4. α-Diversity.

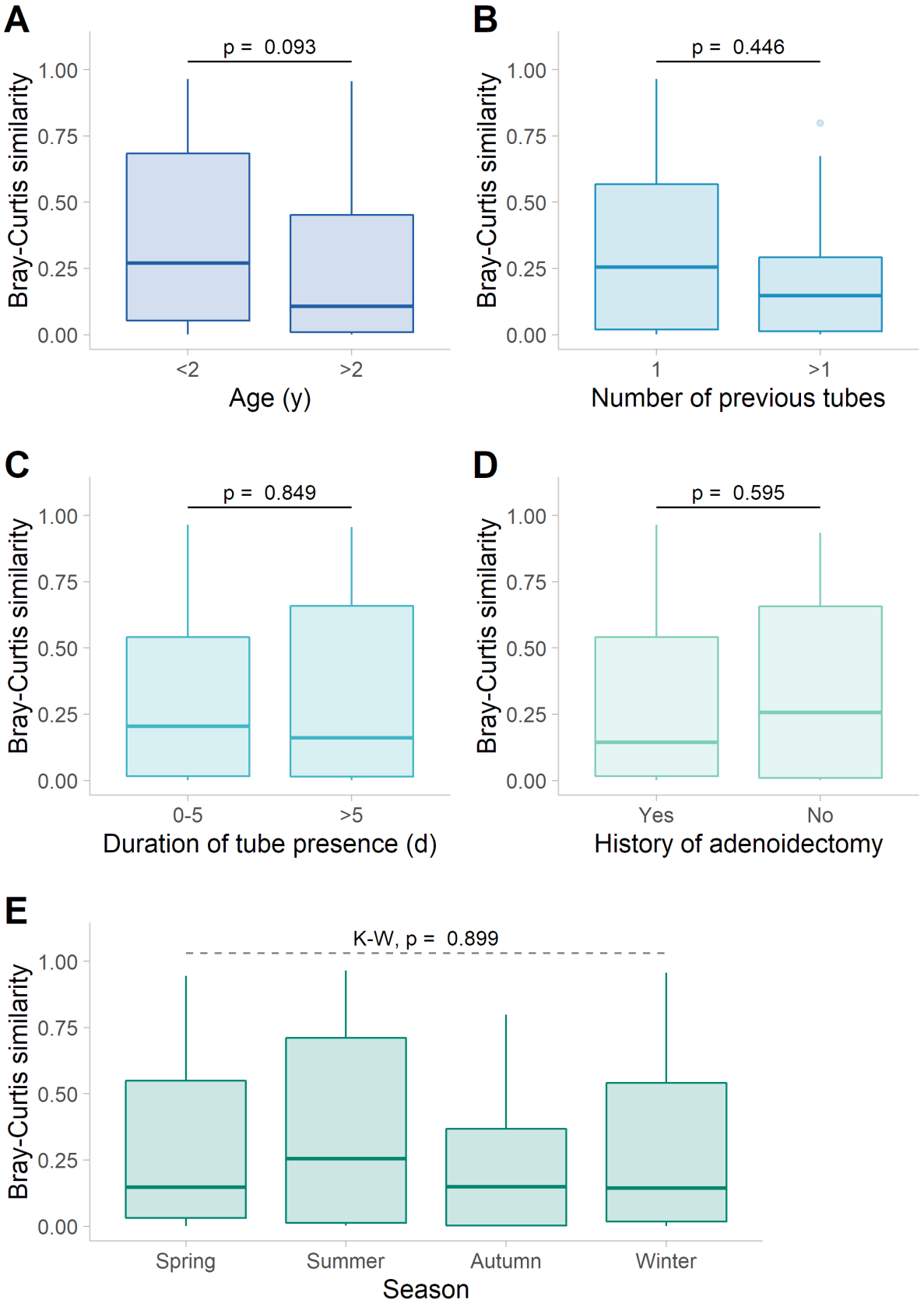
The ecological diversity was significantly higher in NP samples (red) compared to the TTO samples (blue), according to the Chao 1 estimate and Shannon's diversity index.

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

683

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.