

Tonsillar microbial diversity, abundance, and interrelations in atopic and non-atopic individuals

To the Editor,

While the prevalence of allergic diseases has increased worldwide, their underlying mechanisms are still not fully understood.¹ Tonsils, consisting of lymphoid tissue, play an important role in local and systemic immune response to bacterial and viral pathogens and allergens.² Tonsil microbiota is influenced by age and health status.^{3,4}

Allergic diseases, as many other noncommunicable diseases, have been associated with imbalance of the gut, respiratory tract, and skin microbiotas.¹ However, to our knowledge, microbial profiling of tonsils comparing atopic and non-atopic subjects has not been performed as of yet. Therefore, as tonsillar microbiota can closely modulate immune responses and serve as an effective *in vivo* model in investigating their interactions, we wanted to explore whether there are differences in bacterial diversity, abundance, and interrelations in tonsils between atopic and non-atopic elective tonsillectomy patients.

Human tonsil samples were obtained from two consecutive cohorts between 2008 and 2015 from Satakunta Central Hospital, Pori, Turku University Hospital, Turku, and Salo Regional Hospital, Salo, all in Finland.⁵ Study patients completed a standard questionnaire to collect information on health, medication, and respiratory symptoms within 30 days before the operation. Tonsillectomy was performed by routine clinical practice, and the tonsil samples were stored in RNAlater. Blood samples were drawn for allergy tests, and nasopharyngeal aspirate was obtained for respiratory virology analysis. Atopy was defined as positive immunoglobulin E antibody (>0.35 kU/L, Phadiatop Combi®, Phadia, Uppsala, Sweden) against any of the common food or aeroallergens. More details in Online Supplementary.

Intratonillar microbial profiling was performed using RNA sequencing data as previously described.⁶ In total, 53 subjects were included consisting of 24 atopics and 29 non-atopics. After filtering out reads mapped to human, the rest were considered microbial reads (~1 M reads on average, 6% of total). These candidate microbial reads were utilized with Kraken2⁷ for investigating microbial composition with default parameters. Phyloseq⁸ was used to measure diversity of microbial communities and EdgeR to assess differential abundance in atopic and non-atopic subjects. Differences were considered significant based on the corrected *P* value of less than .05. Details in Supplementary methods.

The median age of the study subjects was 11 years (range 2–38) of which 45% were atopic. All the operations were performed during

a “cold phase” of chronic tonsil condition. On the operation day, 36 (68%) had no respiratory tract symptoms, 8 (15%) reported mild respiratory symptoms and 9 (17%) had no data. Main indications for tonsillectomy were recurrent tonsillitis (47%) and tonsillar hypertrophy (45%), see details in Table S1.

Alpha diversity (measures of richness and species distribution in a sample¹) was computed to evaluate the difference in microbial diversity between atopic and non-atopic subjects (Figure 1A). We observed lower alpha diversity in samples from atopic patients compared with non-atopics (*P* = .02, Figure 1A). After adjusting the analysis to age and self-reported allergy, the difference remained significant (*P* = .02). Similar finding was observed when those with clinical allergic disease were compared with non-allergic ones (see Online Supplementary and Figure S2). To investigate the difference in microbial diversity at a group level, we used beta diversity (measure of diversity between samples). However, β -diversity analysis of the tonsillar did not show a significant difference between atopic and non-atopic groups (Supplementary results Figure S3).

The most abundant phylum found in microbial data was Proteobacteria (Figure 1B). No considerable difference (*P* = .3) in the overall bacterial abundance between atopic and non-atopic individuals was observed. However, there was a difference in the expression of single pathogens between the study groups. The differential abundance of two specific taxa, *Clostridium Botulinum* and *Moraxella Osloensis*, was lower (unadjusted *P* < .05, adjusted *P* < .05) in atopic compared with non-atopic study subjects (Figure 1C). The heat maps illustrate correlations between the abundances of pathogenic bacteria in the two study groups (Figure 2A,B; network figures in Figure S4). Positive correlations between pathogenic bacteria were stronger in atopic compared with non-atopic individuals (Figure 2C).

Our finding of low microbial diversity in tonsils of the atopic patients is in line with previous studies on other tissues. For example, exacerbation of atopic dermatitis has been shown to associate with decreased microbial diversity of the skin when compared to healthy skin¹ and asthma has been linked to imbalance of detrimental and beneficial microbes of the respiratory and gastrointestinal epithelium when compared to healthy controls.¹ Therefore, as tonsil tissue contains rich microbiome, it is logical that atopy status is also associated with lower bacterial diversity in tonsils. Collectively, these

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Allergy* published by European Academy of Allergy and Clinical Immunology and John Wiley & Sons Ltd.

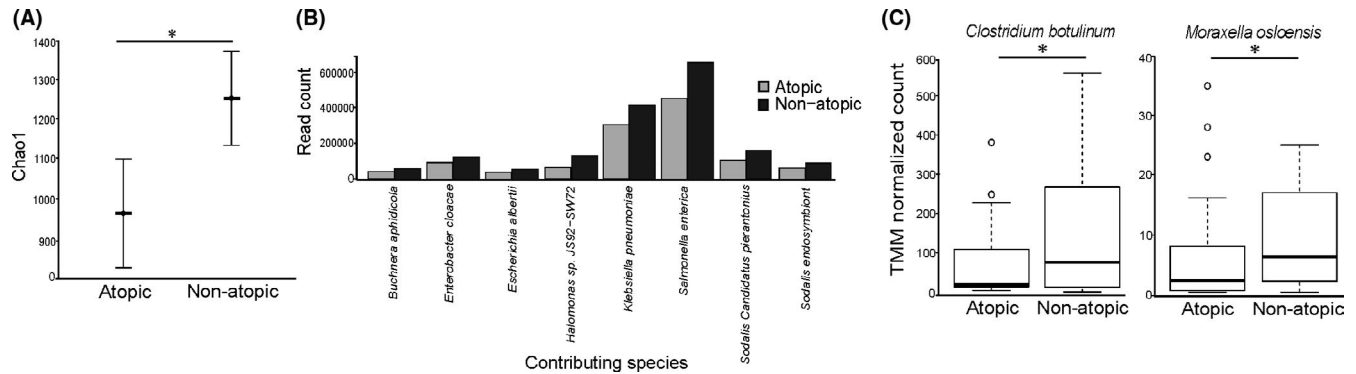


FIGURE 1 Overview of Microbiome data: (A) Atopic subjects ($n = 24$) had lower microbial alpha diversity than non-atopic subjects ($n = 29$) computed with Chao 1 index (estimate of within-group diversity). Data are shown in mean and SEM. B, At phylum level, atopic subjects had a non-significantly lower abundance of Proteobacteria, the largest phylum, than non-atopic subjects ($P = .3$). C, Of all single bacteria, *Clostridium botulinum* and *Moraxella osloensis* showed lower abundance in atopic subjects than in non-atopic subjects. Data are shown in median, quartiles, and range. Please see the statistics in detail in the Online Supplementary. * $P < .05$

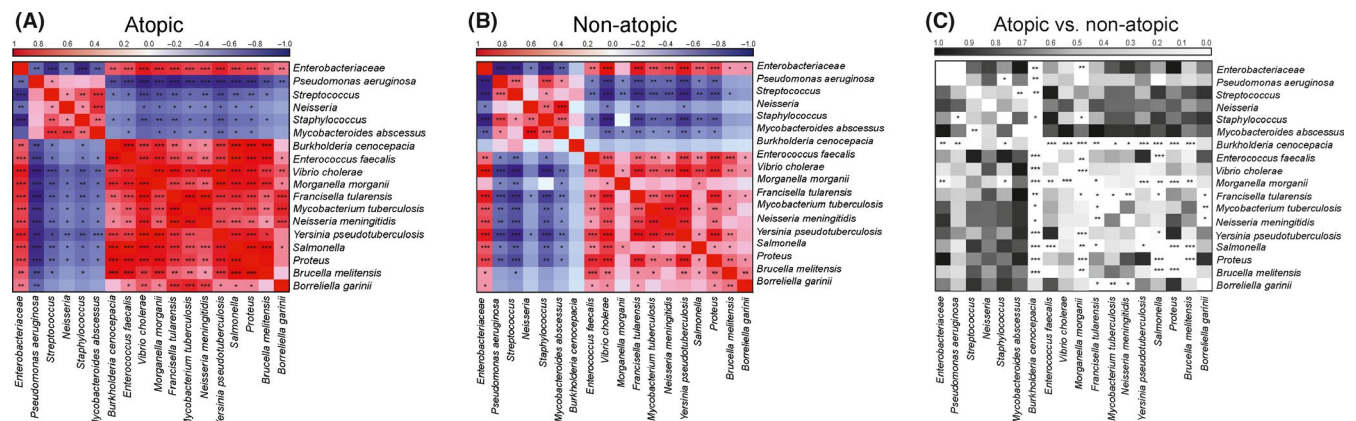


FIGURE 2 Heat maps of intratonsillar pathogenic bacteria-to-bacteria interrelations in atopic subjects (A), non-atopic subjects (B) and comparison of former two (C). Data show stronger positive interrelations between pathogenetic bacteria in atopic than non-atopic subjects. The color scale (blue-red and gray-scale) shows the strength and direction of the correlations. P values $< .001$, $.01$, and $.05$ are indicated with ***, **, and *, respectively. Please see more details in the Online Supplementary

findings suggest that systemic dysbiosis, extending to local lymphoid tissue level, is associated with allergic diseases.¹ It should be noted, however, that our study analyzed gene expression by RNA sequencing whereas many earlier studies¹ have analyzed DNA and genetic information.

In our study, the most abundant phylum was Proteobacteria and the most abundant genera were *Salmonella*, *Klebsiella* and *Sodalis*. This is in line with previous studies, in which Proteobacteria have been highly associated with tonsillar hyperplasia and recurrent tonsillitis in children.^{3,4} Also in earlier studies, the increase in *Moraxella* at genus level has been linked to asthma.¹ However, in our study, the differential abundance of taxa *Moraxella Osloensis* was lower in atopic patients as well as taxa *Clostridium Botulinum*. Comparing the interrelations of pathogenic bacteria, we found more significant correlations in the atopic group. The existence of atopy slightly increased the complexity of bacteria-to-bacteria networks. Moreover, intra-individual variation of microbiome has been demonstrated comparing classical secretory otitis media pathogens in adenoid and tonsil tissues.⁹

This study shows four important findings in elective tonsillectomy patients. First, tonsil tissue contains rich bacterial species in direct contact with immune system. Second, this is a first study to demonstrate that atopic individuals have lower tonsillar bacterial diversity than non-atopics. Third, our data indicate that lower overall bacterial abundance and fourth, stronger positive interrelations between pathogenetic bacteria are associated with atopy. Our study provides interesting insights into the microbiome of human tonsils and suggests differences in microbial balance between atopic and non-atopic subjects. Further studies are warranted to investigate interactions between microbes and lymphoid tissue function.

ACKNOWLEDGEMENTS

Dr Ivaska reports grants from Turku University Hospital Foundation, grants from Turku University Foundation, during the conduct of the study. MSc. Hanif reports grants from The Ella and George Ehrnrooth Foundation, Finland, grants from Ida Montin Foundation, Finland, grants from Finnish-Norwegian Medical Foundation, during the conduct of the study. MSc. Ahmad has

nothing to disclose. Dr Tan has nothing to disclose. Dr Altunbulakli has nothing to disclose. Dr Mikola reports grants from The Finnish Cultural Foundation, during the conduct of the study. Dr Silvoniemi has nothing to disclose. Dr Puhakka has nothing to disclose. Dr Akdis reports grants from Allergopharma, Idorsia, Swiss National Science Foundation, Christine Kühne-Center for Allergy Research and Education, European Commission's Horizon's 2020 Framework Programme, Cure, Novartis Research Institutes, Astra Zeneca, Scibase, advisory role in Sanofi/Regeneron, outside the submitted work. Dr Toppila-Salmi reports personal fees from Sanofi Pharma, grants from GSK, personal fees from Roche, personal fees from Novartis, personal fees from ERT, outside the submitted work. Dr Jartti reports grants from the Sigrid Juselius Foundation, Helsinki, Finland, grants from the Allergy Research Foundation, Helsinki, Finland, grants from the Foundation for Pediatric Research, Helsinki, Finland, during the conduct of the study.

CONFLICTS OF INTEREST

The authors have no conflict of interest in connection with this paper.

FUNDING INFORMATION

TJ and his laboratory are supported by the Sigrid Juselius Foundation, Helsinki; Allergy Research Foundation, Helsinki; and the Foundation for Pediatric Research, Helsinki. LEI and TH are supported by the Turku University Hospital Foundation, Turku; and LEI also by the Turku University Foundation, Turku; EM is supported by the Finnish Cultural Foundation and the Allergy Research Foundation Helsinki, all in Finland.

Lotta E. Ivaska¹ 

Tanzeela Hanif^{2,3}

Freed Ahmad⁴ 

Ge Tan^{5,6,7} 

Can Altunbulakli^{6,7}

Emilia Mikola¹ 

Antti Silvoniemi¹ 

Tuomo Puhakka¹

Cezmi A. Akdis^{6,7}

Sanna Toppila-Salmi^{3,8} 

Tuomas Jartti² 

¹Department of Otorhinolaryngology – Head and Neck Surgery, Turku University Hospital and University of Turku, Turku, Finland

²Department of Paediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland

³Haartman Institute, University of Helsinki, Helsinki, Finland

⁴Department of Biology, University of Turku, Turku, Finland

⁵Functional Genomics Center Zurich, ETH Zurich/University of Zürich, Zürich, Switzerland

⁶Swiss Institute of Allergy and Asthma Research, University of Zürich, Zürich, Switzerland

⁷Christine Kuhne-Center for Allergy Research and Education (CK-CARE), Davos, Switzerland

⁸Skin and Allergy Hospital, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

Correspondence

Lotta E. Ivaska, MD, Department of Otorhinolaryngology – Head and Neck Surgery, Turku University Hospital, Turku, Finland, P.O. Box 52, FIN 20521 Turku, Finland.

Email: lesaar@utu.fi

Ivaska and Hanif equally contributed to this study.

ORCID

Lotta E. Ivaska  <https://orcid.org/0000-0002-0169-2681>

Freed Ahmad  <https://orcid.org/0000-0002-8994-4723>

Ge Tan  <https://orcid.org/0000-0003-0026-8739>

Emilia Mikola  <https://orcid.org/0000-0001-9814-6072>

Antti Silvoniemi  <https://orcid.org/0000-0001-9655-1410>

Sanna Toppila-Salmi  <https://orcid.org/0000-0003-0890-6686>

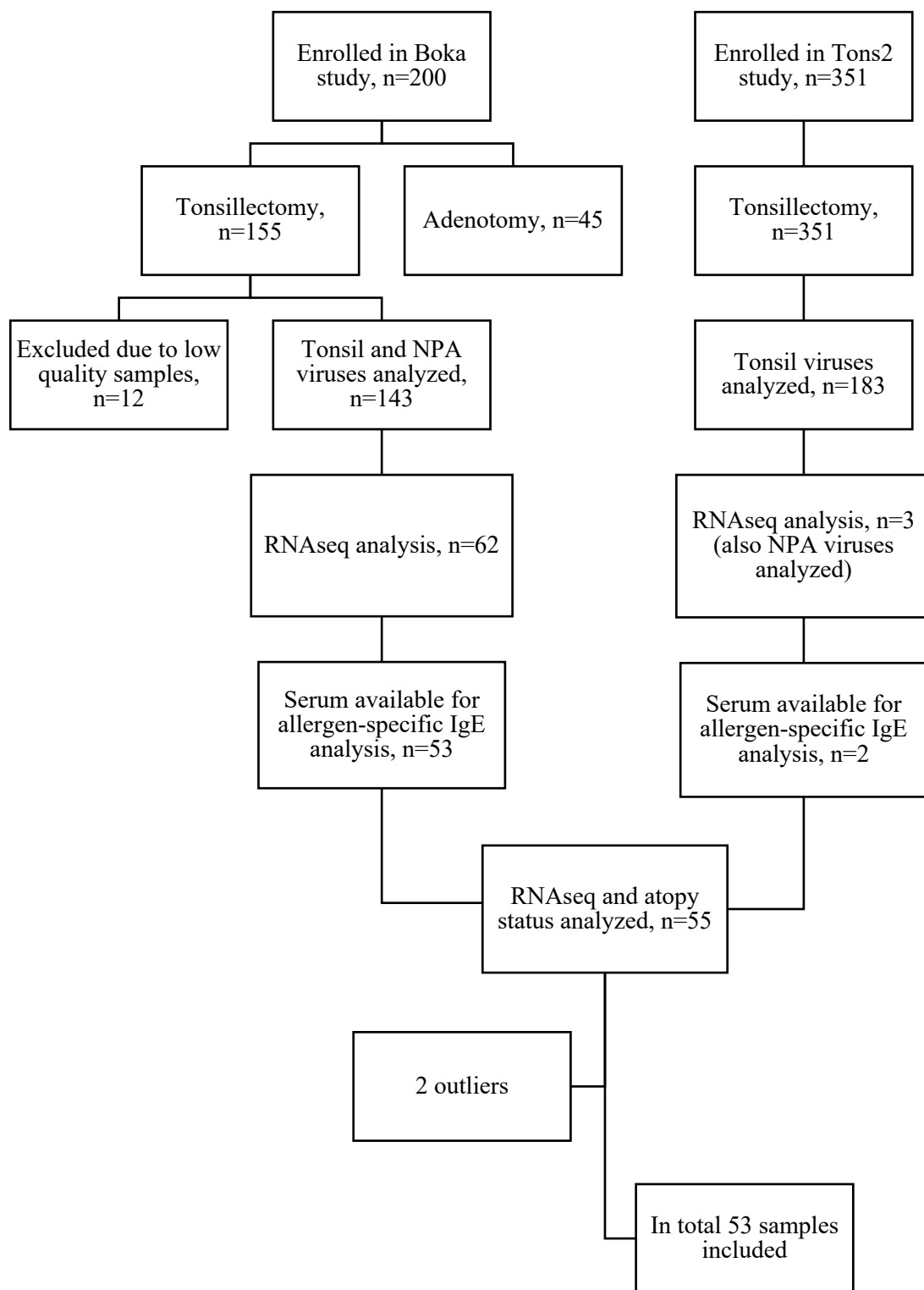
Tuomas Jartti  <https://orcid.org/0000-0003-2748-5362>

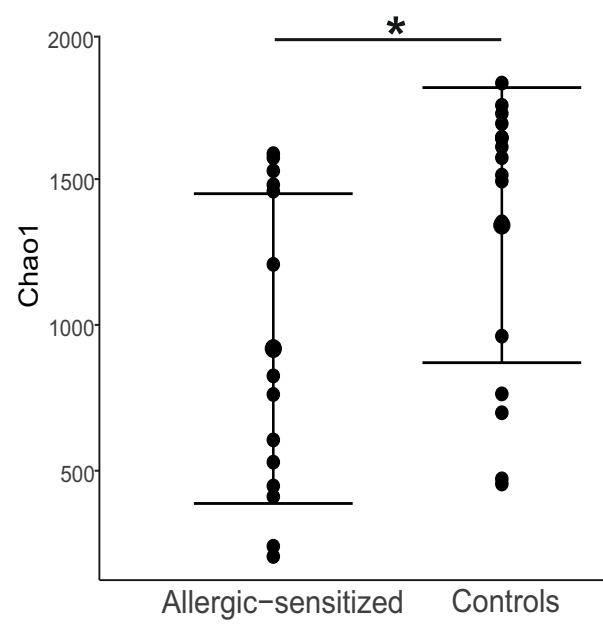
REFERENCES

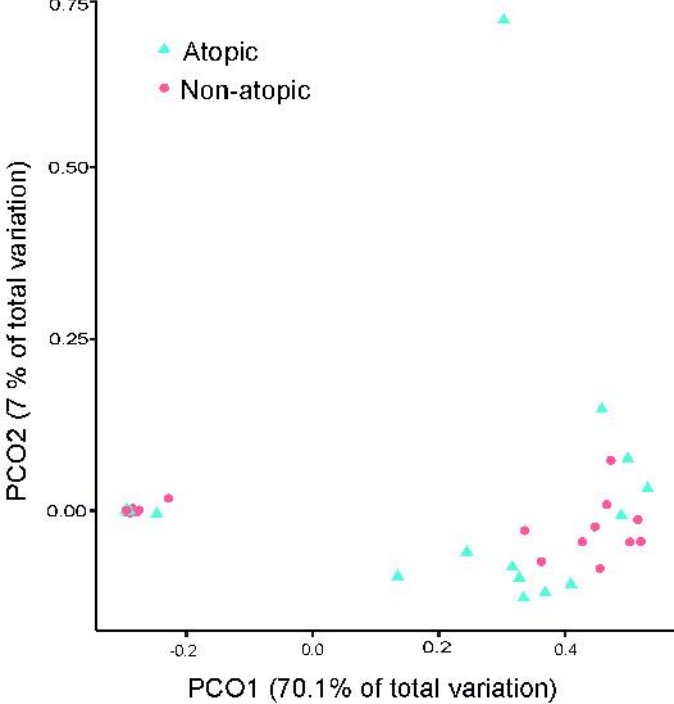
- Huang YJ, Marsland BJ, Bunyavanich S, et al. The microbiome in allergic disease: Current understanding and future opportunities—2017 PRACTALL document of the American Academy of Allergy, Asthma & Immunology and the European Academy of Allergy and Clinical Immunology. *J Allergy Clin Immunol*. 2017;139(4):1099–1110.
- Masters KG, Lasrado S. *Anatomy, Head and Neck, Tonsils*. Treasure Island, FL: StatPearls Publishing; 2019. <http://www.ncbi.nlm.nih.gov/pubmed/30969614>. Accessed May 29, 2019.
- Johnston J, Hoggard M, Biswas K, et al. Paired analysis of the microbiota between surface tissue swabs and biopsies from pediatric patients undergoing adenotonsillectomy. *Int J Pediatr Otorhinolaryngol*. 2018;113:51–57.
- Jensen A, Fagö-Olsen H, Sørensen CH, Kilian M. Molecular mapping to species level of the Tonsillar crypt microbiota associated with health and recurrent tonsillitis. Xu P, ed. *PLoS ONE* 2013;8(2):e56418.
- Jartti T, Palomares O, Waris M, et al. Distinct regulation of tonsillar immune response in virus infection. *Allergy* 2014;69(5):658–667.
- Hanif T, Dhaygude K, Kankainen M, et al. Birch pollen allergen immunotherapy reprograms nasal epithelial transcriptome and recovers microbial diversity. *J Allergy Clin Immunol*. 2019;143(6):2293–2296.
- Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 2014;15(3):R46.
- McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013;8(4):e61217.
- Fagö-Olsen H, Dines LM, Sørensen CH, Jensen A. The adenoids but not the palatine tonsils serve as a reservoir for bacteria associated with secretory otitis media in small children. Cotter PD, ed. *mSystems*. 2019;4(1):00169–218.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

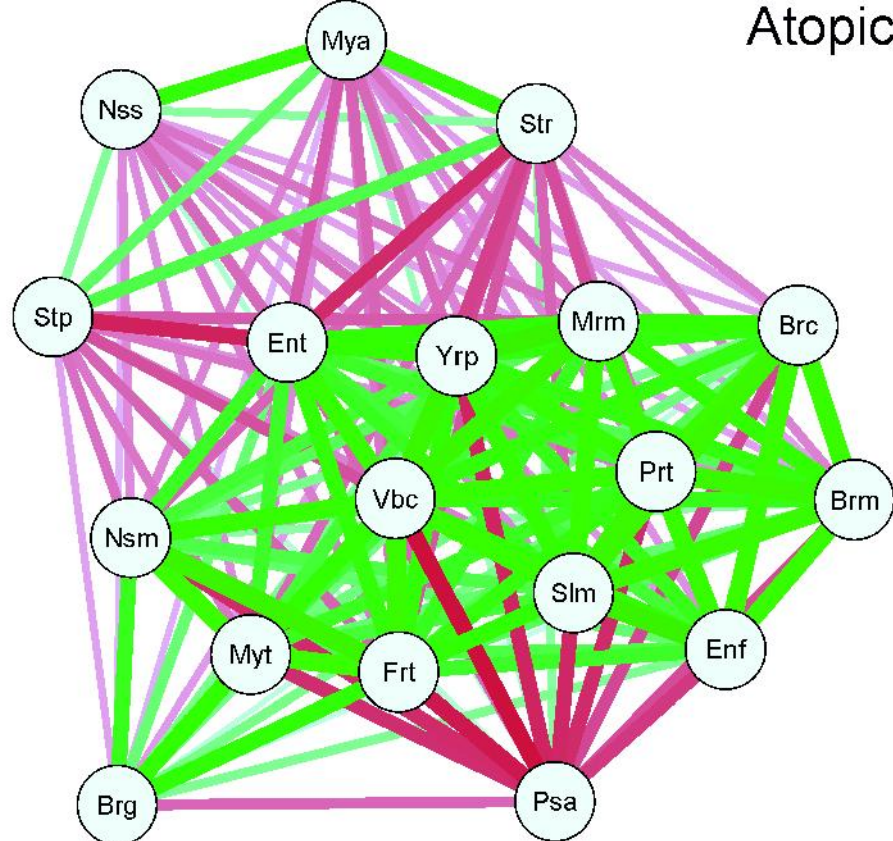






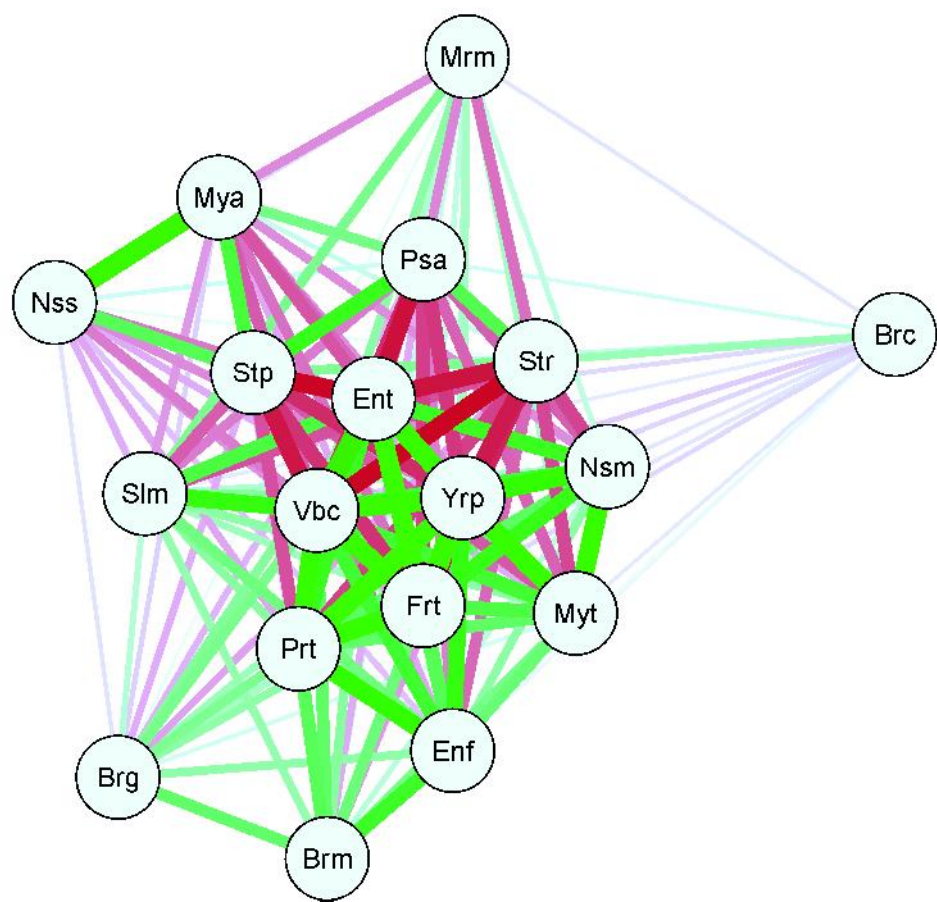
A

Atopic



B

Non-atopic



1 **TABLE S1.** Patient Characteristics.
2

Factor	Atopic n = 24	Non-atopic n = 29	<i>P</i>
Median age (range), years	12.4 (2.0-38.1)	7.5 (3.0-21.1)	.14
Male	12 (50%)	16 (55%)	.70
Indication for adeno- /tonsillectomy			
Recurrent tonsillitis	12 (50%)	13 (45%)	.70
Tonsillar hypertrophy	11 (46%)	13 (45%)	.94
Other indication†	1 (4%)	3 (10%)	.62
Self-Reported Allergy	12 (50%)	8 (28%)	.069
Sensitization			
Food	6 (25%)	0 (0%)	
Aero (i.e. pollen, animal or house dust mite)	14 (58%)	0 (0%)	
Both food and aero	4 (17%)	0 (0%)	
Physician diagnosed asthma	6/22 (27%)	3/26 (12%)	.16
Self-reported allergic rhinitis	11/20 (55%)	3/27 (11%)	.003
Physician-diagnosed atopic dermatitis (AD)	2/20 (10%)	6/29 (21%)	.32
No clinical allergic disease (asthma, rhinitis, AD)	8/22	17/29	.12
Respiratory viruses in tonsillar tissue	12 (50%)	15 (52%)	.90
Respiratory viruses in nasopharynx	18 (75%)	24 (83%)	.49
Smoking or exposure to smoking	11/23 (48%)	15/28 (54%)	.683
Use of antibiotics (ab) within 1 year	16/20 (80%)	18/28 (64%)	.238
0 ab courses	4/20	10/28	
1 ab courses	3/20	5/28	
2 ab courses	3/20	4/28	
3 ab courses	3/20	1/28	
4 ab courses	2/20	2/20	

5 ab courses	3/20	4/28	
6 ab courses	0	1/28	
8 ab courses	0	1/28	
11 ab courses	1/20	0	
12 ab courses	1/20	0	
Season of sampling			
April-August	10 (42%)	11 (38%)	.78
September-March	14 (58%)	18 (62%)	

3 Data are expressed as number of subjects (%) except age.

4 Differences between atopic vs. nonatopic subjects were calculated with Mann Whitney U test, Chi
5 square test or Fischer exact test (when counts <5).

6 †Chronic white patches in tonsils (n=1), teeth braces (n=1), periodic fever (n=2)

7

Online Supplementary

Tonsillar microbial diversity, abundance and interrelations in atopic and non-atopic individuals

Lotta E. Ivaska^{a*}, MD, Tanzeela Hanif^{b,c*}, MSc, Freed Ahmad^d, MSc, Ge Tan^{e,f,g}, PhD, Can Altunbulakli^{f,g}, PhD, Emilia Mikola^a, MD, Antti Silvoniemi^a, MD, Tuomo Puhakka^a, MD, Cezmi A. Akdis^{f,g}, MD, Sanna Toppila-Salmi^{c,h}, MD, Tuomas Jartti^b, MD

^aDepartment of Otorhinolaryngology – Head and Neck Surgery, Turku University Hospital and University of Turku, Turku, Finland

^bDepartment of Paediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland

^cHaartman Institute, University of Helsinki, Helsinki, Finland

^dDepartment of Biology University of Turku, Turku, Finland

^eFunctional Genomics Center Zurich, ETH Zurich/University of Zürich, Zürich, Switzerland

^fSwiss Institute of Allergy and Asthma Research (SIAF), University of Zürich, Zürich, Switzerland

^gChristine Kuhne-Center for Allergy Research and Education (CK-CARE) Davos, Switzerland

^hSkin and Allergy Hospital, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

*Equal contribution

Correspondence to: Lotta E. Ivaska, MD, Department of Otorhinolaryngology – Head and Neck Surgery, Turku University Hospital, Turku, Finland, P.O. Box 52, FIN 20521 Turku, Finland. Phone: +358 2 313 0000 Email: lesaar@utu.fi Fax: +358 2 313 3525.

Abbreviations

NPA: Nasopharyngeal aspirate

IgE: Immunoglobulin E

PCR: Polymerase chain reaction

PBS: Phosphate-buffered saline

TMM: Trimmed Mean of M-values

CPM: Counts per millions

PcoA: Principal coordinates analysis

Methods

Study design and subjects

Atopic and non-atopic patients of all ages and samples were collected from two consecutive cohorts, see Fig E1. Boka study was carried out in the Satakunta Central Hospital, Pori, Finland between April 2008 and March 2009. Tons2 study was carried out in the Turku University Hospital, Turku, Finland and Salo Regional Hospital, Salo, Finland between October 2013 and December 2015. The inclusion criteria were elective adeno-/tonsillectomy according to clinical indication and written informed consent from the study subject or his/her parents. The study protocols were approved by the Ethics Committee of Satakunta Central Hospital and by the Ethics Committee of the Hospital District of Southwest Finland, respectively.

The internal tonsillar tissue removed from the patient was instantly cut into 3-4 mm cubes, stored in RNAlater, an RNA stabilization reagent (Qiagen, Hilden, Germany), incubated at +2-8 °C until the next working day and finally stored at -80 °C¹. Nasopharyngeal aspirate samples were collected through a nostril using a standardized procedure². If the aspirate yield was small, the collection was repeated after administering 2 ml of physiologic saline into a nostril. For viral analyses, a piece of the removed tonsils and the nasopharyngeal aspirate were stored in dry tubes at -80 °C¹.

Self-reported allergies and other clinical data were asked with standard questionnaire¹. Allergic sensitization was tested with allergen specific immunoglobulin IgE antibodies to any of the common food (codfish, cow's milk, egg, peanut, wheat, soybean) or aeroallergens (cat, dog, horse, birch, mugwort, timothy, *Cladosporium herbarum* and *Dermatophagoides pteronyssinus*) with cut-off level 0.35 kU/L (Phadiatop Combi®, Phadia, Uppsala, Sweden). Animal sensitization was defined as positive IgE antibodies to cat, dog, horse or *Dermatophagoides pteronyssinus*. Birch, mugwort, timothy and *Cladosporium herbarum* were considered as pollen aeroallergens³.

Viral diagnostics

The detection of respiratory viruses in nasopharyngeal aspirates and intratonsillar samples was performed using PCRs (including reverse transcription step when applicable) on nucleic acid extracts². A nasopharyngeal aspirate sample was suspended into 1 ml of PBS and nucleic acid was isolated from 550 µl of the suspension using NucliSense easyMag automated nucleic acid extractor

(BioMerieux, Boxtel, The Netherlands) with on-board lysis. An intratonsillar sample (approximately 300 µg) was homogenized in a 2 ml tube with 400 µl of PBS, 400 µl of protease buffer AL and 100 µl of QIAGEN protease using Tissue Lyzer LT (Qiagen, Hilden, Germany) and ø7 mm stainless steel bead for 3 min at 50 beats/s and incubated for 1 h at 56°C. The beating was repeated with an additional 500 µl of PBS and the tube was centrifuged for 30 s at 10000 g. A 220-µl aliquot of the supernatant was incubated overnight in the NucliSense easyMag lysis buffer at room temperature and processed in the easyMag extractor. An elution volume of 55 µl was used for both sample types¹.

In-house real-time PCRs were used as described for bocavirus-1, enterovirus, rhinovirus and respiratory syncytial virus⁴⁻⁶. Seeplex RV12 ACE Detection (Seegene, Seoul, Korea) multiplex PCR assay was used for adenovirus, coronaviruses (229E/NL63 and OC43/HKU1), influenza A and B viruses, metapneumovirus, parainfluenza virus types 1-3, respiratory syncytial virus group A and B and rhinovirus. Virus diagnostics were carried out in the Department of Virology, University of Turku, Turku, Finland and in the Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden¹.

To isolate total RNA directly from palatine tonsils, tissues (previously stabilized in RNAlater) were homogenized in grinding tubes containing CK28 ceramic beads by using a Precellys 24 homogenizer (Bertin Technologies, Montigny le Bretonneux, France) two times at 6000 rpm for 50 seconds. Total RNA from cell samples was isolated using the RNeasy mini kit (Qiagen, Hilden, Germany). The tonsil samples were handled carefully as described previously¹. No sequencing controls were used.

Microbiome analysis

We assessed the quality of raw reads using FASTQC⁷ and filtered out low quality reads (Phred quality score <30, length <36 bp). Trimmomatic⁸ was used for removing low quality reads, Illumina adapters and short read-length. We aligned the quality filtered reads on human genome GRCh38-v93 using Hisat2⁹. Human and non-human reads were separated and the latter subset was used as microbial reads for taxonomic identification of microbial communities. Filtered out non-human candidate microbial reads were utilized with Kraken2¹⁰ for investigating the microbial composition with default parameters. Kraken2¹⁰ is a read based metagenomic classification approach that classifies the overlapping read k-mers and allows to make custom database. In order to get more

functional units and increase the accuracy of the results, we mapped the reads into nucleotide databases using Kraken2¹⁰. The outputs from Kraken2¹⁰ were then converted into biom format using kraken-biom tool. The biom file contained the read counts for every microbial species found in any of the samples. The biom file was read in R and the replicates were merged at this point for downstream analysis.

We transformed raw read counts into weighted trimmed mean of M-values algorithm (TMM) using edgeR¹¹ package version 3.2.1 in R version 3.5.2 (www.r-project.org) to determine the expression estimates. For assessing microbial diversity and abundance, we filtered out the lowly abundant taxa. After initial filtering, we kept only taxa with an average expression level of CPM > 1 in at least one sample. Microbial richness was measured on the normalized taxa table using alpha diversity measures. Alpha diversity in each sample was estimated with phyloseq¹² using Chao1 index at taxa-level. Principal Coordinate Analysis (PCoA) was performed on the selected samples and filtered out taxa, using the state-of-art Bray-Curtis beta diversity distance method. PCoA plot was visualized in R using the Bioconductor package vegan. We identified differentially abundant taxa in R, version 3.5.2, with the edgeR¹¹ version 3.2.1. For measuring the differential abundance, we did an additional filtering and kept only those taxa that were present in half of the either group. The model matrix was designed as follows: ~ batch+condition, where “condition” refers to atopic and non-atopic, “batch” refers to the 3 different sets of sequencing batches. Fit of the generalized linear model and tests for differences in expression were performed with the “glmFit” and “glmLRT” functions, respectively. The resulting P values were corrected with Benjamini-Hochberg method and the threshold for statistical significance was set at a false discovery rate (FDR) < .05. Conventional statistics in comparing patient groups was carried out using SPSS software Statistics for Windows, version 25.0 (IBM SPSS Statistics for Macintosh, Armonk, NY). Correlations between the abundances of pathogenic bacteria in atopic and non-atopic samples were measured with Pearson correlation in R, version 3.5.2 (www.r-project.org). Difference in correlation coefficients between atopic and non-atopic samples was measured using the methods described previously¹.

Results

Patient characteristics

Allergic diseases were enquired from the study patients (see Table S1). The division into groups of atopic and non-atopic study subjects was made by IgE sensitization result. Only self-reported allergic rhinitis was more common in the atopic group ($P = .003$). There was no difference in the intra-tonsillar or nasopharyngeal respiratory virus rates between atopics and non-atopics. Also, active and passive smoking prevalence were similar in both study groups, as well as the use of antibiotics. Eighty percent of the atopic patients and 64% of the non-atopic patients had had at least one antibiotic course within a year (see Table S1). There were no data of five patients. There was no difference in the microbiome analysis between the two groups when adjusting it for the use antibiotics.

Studying microbial RNA in tonsils

Using our filtering pipeline, an average of ~1M unmapped reads were used as candidate microbial reads in our analyses. From these, Kraken2¹⁰ estimated 28.5M reads (0.5M mean, 0.74M standard deviation) to the bacterial, viral and archaeal gene families. A total of 4137 taxa were assigned, with 45 taxa at the phylum level. Most of the taxa we observed were derived from bacteria and a smaller portion derived from archaea. Based on initial data filtering, 3456 unique taxa were used for further analysis.

The most abundant phylum found in microbial data was Proteobacteria (with sub-categories alpha, beta and gamma), Actinobacteria and Firmicutes. We did an additional filtering (described in methods) and used 709 taxa for differential abundance analysis. In this analysis *Clostridium botulinum* (taxa id: 1491) and *Moraxella osloensis* (taxa id: 34062) showed lower abundance in the atopic group compared to the non-atopic group. (corrected $P < .05$, Fig 1C). In general, the microbiome was not dependent on the sex or age group of the study subjects. PCoA was done to see the real clustering of samples. The tonsillar microbiome of the subjects did not cluster completely by disease status in a PCoA plot based on the Bray-curtis distance (Fig S3). So, there could be batch effect behind the lack of clear clustering.

We also investigated the correlation between the abundances of 20 pathogenic bacteria in atopic and non-atopic individuals. Two sequences imitating *Yersinia pestis* and *Bacillus anthracis* were not included due to false positive result. In the atopic group, the 18 pathogenic bacteria seemed to organize themselves into two clusters which showed relatively high positive intracluster correlations and negative correlation between the clusters (Fig 2A-B, the P values are shown in the

figure). In the non-atopic group, these correlations were less prominent (Fig 2B). The intracluster correlations of *Burkholderia cenocepacia* and *Morganella morganii* were positive in the atopic group but negative or indifferent in the non-atopic group (Fig 2A-C). The network figures depicted slightly stronger interrelations between pathogenic bacteria in the atopy group compared to the non-atopy group (Fig S4).

Additional analysis of the microbial diversity depending on the clinical allergic illness and sensitization status was made by computing the Chao index (Fig S2). Alfa diversity was lower in the allergic and sensitized than the non-allergic and non-sensitized study patients ($P < .05$).

Discussion

As a methodological consideration we want to point out that most of the earlier studies concerning tonsillar microbiome have performed 16s rRNA sequencing for the identification of bacteria at genus level or even species level^{13–16}. This method is targeted to search microbes and is based on PCR amplification of the 16s rRNA genes, which are present in all bacteria. RNA sequencing on the other hand is based on analyzing wide scale transcriptomes so the detection of bacterial species and diversity is enhanced compared to the 16s rRNA sequencing method. RNA sequencing lacks the risk of biases concerning primers used on 16s rRNA sequencing¹⁷. RNA sequencing has previously been used successfully for the detection of microbiomes¹⁸.

The strenghts of this study are the direct analyses of tonsil tissue samples and the detailed characterization of the study subjects. The limitations of the study include a relatively small number of study subjects and potential uncertainties concerning the measuring of RNA expression for bacterial detection. Due to ethical reasons, we did not study healthy control group or an atopic group without chronic tonsillar disease. Atopy was defined only by serum immunoglobulin E antibodies (IgE) and we did not use skin prick tests even though a small proportion of atopy patients are IgE negative but skin prick test positive.

This study shows the complexity of tonsil tissue microbiome and supports the results of earlier 16s rRNA sequencing studies, in which it has been depicted that the microbiome of tonsil tissues differs depending on the site of the sample¹⁴, between obstructive sleep apnea patients and recurrent tonsillitis patients¹³, between children with secretory otitis media and children with adenotonsillar hyperplasia¹⁹ and between adults and children¹⁵. Our study adds to the knowledge of the previous

studies in that alterations in microbiota can be identified between atopic and non-atopic patients undergoing tonsillectomy. Yet, follow-up studies are needed to evaluate whether the detected low microbiome diversity in atopics is a cause or a consequence of atopy.

References

1. Jartti T, Palomares O, Waris M, et al. Distinct regulation of tonsillar immune response in virus infection. *Allergy*. 2014;69(5):658-667.
2. Jartti T, Lehtinen P, Vuorinen T, et al. Respiratory Picornaviruses and Respiratory Syncytial Virus as Causative Agents of Acute Expiratory Wheezing in Children. *Emerg Infect Dis*. 2004;10(6):1095-1101.
3. Elenius V, Palomares O, Waris M, et al. The relationship of serum vitamins A, D, E and LL-37 levels with allergic status, tonsillar virus detection and immune response. *PLoS One*. 2017;12(2):119-124.
4. Peltola V, Waris M, Österback R, Susi P, Ruuskanen O, Hyypiä T. Rhinovirus Transmission within Families with Children: Incidence of Symptomatic and Asymptomatic Infections. *J Infect Dis*. 2008;197(3):382-389.
5. McLeish NJ, Wittevelde J, Clasper L, et al. Development and Assay of RNA Transcripts of Enterovirus Species A to D, Rhinovirus Species A to C, and Human Parechovirus: Assessment of Assay Sensitivity and Specificity of Real-Time Screening and Typing Methods. *J Clin Microbiol*. 2012;50(9):2910-2917.
6. Koskenvuo M, Möttönen M, Waris M, Allander T, Salmi TT, Ruuskanen O. Human bocavirus in children with acute lymphoblastic leukemia. *Eur J Pediatr*. 2008;167(9):1011-1015.
7. Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data : <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed May 15, 2019.
8. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
9. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-360.
10. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 2014;15(3):R46.
11. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential

- expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140.
12. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217.
 13. Johnston J, Hoggard M, Biswas K, et al. The bacterial community and local lymphocyte response are markedly different in patients with recurrent tonsillitis compared to obstructive sleep apnoea. *Int J Pediatr Otorhinolaryngol*. 2018;113:281-288.
 14. Johnston J, Hoggard M, Biswas K, et al. Paired analysis of the microbiota between surface tissue swabs and biopsies from pediatric patients undergoing adenotonsillectomy. *Int J Pediatr Otorhinolaryngol*. 2018;113:51-57.
 15. Jensen A, Fagö-Olsen H, Sørensen CH, Kilian M. Molecular Mapping to Species Level of the Tonsillar Crypt Microbiota Associated with Health and Recurrent Tonsillitis. Xu P, ed. *PLoS One*. 2013;8(2):e56418.
 16. Johnston JJ, Douglas R. Adenotonsillar microbiome: An update. *Postgrad Med J*. 2018:398-403.
 17. Olde Loohuis LM, Mangul S, Ori APS, et al. Transcriptome analysis in whole blood reveals increased microbial diversity in schizophrenia. *Transl Psychiatry*. 2018;8(1):96.
 18. Hanif T, Dhaygude K, Kankainen M, et al. Birch pollen allergen immunotherapy reprograms nasal epithelial transcriptome and recovers microbial diversity. *J Allergy Clin Immunol*. 2019;143(6):2293-2296.
 19. Fagö-Olsen H, Dines LM, Sørensen CH, Jensen A. The Adenoids but Not the Palatine Tonsils Serve as a Reservoir for Bacteria Associated with Secretory Otitis Media in Small Children. Cotter PD, ed. *mSystems*. 2019;4(1): e00169-18.

Legends to the Online Supplementary figures

FIG S1. Flowchart of the study. Two cohorts (Boka- and Tons2 study) were combined to find suitable cases for RNAseq analysis. Plain adenotomy patients were excluded. The cases included for RNAseq analysis were men and women preferably under 21 years, with different combinations of tonsillar and NPA viruses.

FIG S2. Alpha diversity compared in samples that were sensitized and allergic (N= 14) with the non-allergic-non-sensitized samples (N= 17) computed with Chao 1 index (estimate of diversity in microbiome). Shown in mean and quartiles. The comparisons are made by Anova test. *P < .05.

FIG S3. Principal coordinate analysis plot of microbial community based on Bray-Curtis distance shows no difference ($P = .2$) in atopic and non-atopic subject groups.

FIG S4. Nodes illustrate the bacteria and edges represent the correlation between the bacteria. Edge width indicates the strength of correlation (pearson correlation > 0.1, $P < .05$). Nodes are colored by Phylums and the edges' colors demonstrate positive and negative correlations (green and red, respectively). Bacterial nodes are labeled as 3 letter abbreviations and opened in the text below. The heatmap shows the difference in correlations between atopic and non-atopic patients. P values < .001, .01 and .05 are indicated with ***, ** and *, respectively.

Brg: *Borrelia garinii*, Brm: *Brucella melitensis*, Brc: *Burkholderia cenocepacia*, Enf: *Enterococcus faecalis*, Ent: Enterobacteriaceae, Frt: *Francisella tularensis*, Mrm: *Morganella morganii*, Mya: *Mycobacteroides abscessus*, Myt: *Mycobacterium tuberculosis*, Nsm: *Neisseria meningitidis*, Nss: *Neisseria*, Prt: *Proteus*, Psa: *Pseudomonas aeruginosa*, Slm: *Salmonella*, Str: *Streptococcus*, Stp: *Staphylococcus*, Vbc: *Vibrio cholerae*, Yrd: *Yersinia pseudotuberculosis*