Microbial and clinical factors are related to recurrence of symptoms after childhood lower respiratory tract infection

Emma M. de Koff, Wing Ho Man, Marlies A. van Houten, Arine M. Vlieger, Mei Ling J.N. Chu, Elisabeth A.M. Sanders, Debby Bogaert

Online Data Supplement

Contents

SUPPLEMENTARY METHODS

Statistical analysis

Data analysis was performed in R version 3.4.3 within RStudio version 1.1.383. All analyses comparing cases to controls accounted for the matched nature of the samples. A p-value of less than 0.050 or a Benjamini-Hochberg adjusted q-value of less than 0.050 were considered statistically significant.

Chi-square and Wilcoxon tests were used to compare host characteristics between cases with and without recurrence of respiratory symptoms during follow-up. Independent relationships between host characteristics and recurrence of respiratory symptoms during follow-up were assessed in multivariable logistic regression models, with treatment (antibiotics vs. no antibiotics), lower respiratory tract infection (LRTI) phenotype (wheezing illness, bronchiolitis and mixed infection vs. pneumonia) and age in months both as individual explanatory variables and in pairwise interactions, and also correcting for follow-up time in days. The final model was based on backward selection of variables using a p-value of 0.10 as cut-off, until the optimal strength of the model based on the Akaike information criterion was reached.

To compare viral presence between cases and controls, conditional logistic regression was used.

To assess alpha diversity, we calculated the Chao1 index for microbial richness and the Shannon index for diversity (*phyloseq* [1]), and significance of differences between cases and controls was evaluated using linear mixed-effect models. We also compared alpha diversity measures between cases at time of hospital admission with and without subsequent recurrence of respiratory symptoms using linear models adjusted for age, sex and month of hospital admission.

Nonmetric multidimensional scaling (NMDS) biplots based on the Bray-Curtis dissimilarity matrix (*ordinate*-function, *phyloseq* [1], 2 dimensions, maximum 10.000 iterations) were used to visualize differences in the overall microbial community between groups. Statistical

significance of differences in the overall microbial community was assessed by permutational analysis of variance (PERMANOVA) using the *adonis*-function or, when comparing cases with and without recurrence of respiratory symptoms during follow-up, the *adonis2*-function adjusting for age, sex and month of hospital admission (*vegan* [2], 1999 permutations). We considered microbiota recovery to be 'complete' if there was no remaining significant difference in the overall microbial community between cases at recovery and matched controls. Relative abundances of the top 10 most highly abundant operational taxonomic units (OTUs) were visualized in a stacked bar chart.

We used *metagenomeSeq* analysis [3] (*fitZig*-function) to identify differentially abundant OTUs between cases and controls (filtered on OTUs present in >10% of the samples, maximum 100 iterations, mixed model design). Next to that, we identified OTUs with highest discriminative abilities between cases and controls with random forest classifier analysis using a 10-fold cross-validated *VSURF* procedure [4]. In this analysis, OTUs were considered discriminant when they were selected at least twice in the interpretation step. Log2 fold changes of discriminant OTUs as calculated by *metagenomeSeq* were converted to fold changes using the formula fold change = 2^log2 fold change. Combined results from *metagenomeSeq* and *VSURF* were then additionally filtered at a fold change of at least 1.5 or below 0.5 (i.e. a 50% change) to retain only discriminant OTUs with relevant changes.

Above analyses were repeated to assess microbiome recovery in relation to clinical outcome (recurrence vs. no recurrence of respiratory symptoms during follow-up), LRTI phenotype, antibiotic treatment, and presence of respiratory syncytial virus or human rhinovirus. In these sub-analyses, we limited differential abundance testing to the top 100 highest-ranked OTUs, because of limited power, and to avoid false positive results in low abundant OTUs which is a known risk using *metagenomeSeq* analyses in smaller group sizes [5].

References

1. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* 2013;8:e61217.

2. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P, Stevens M, Szoecs E, Wagner H. vegan: Community Ecology Package. 2016; at <http://CRAN.R-project.org/package=vegan>.

3. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nature Methods* 2013;10:1200–1202.

4. Genuer R, Poggi J, Tuleau-malot C. VSURF : An R Package for Variable Selection Using Random Forests. *The R Journal* 2015;7:19–33.

5. Thorsen J, Brejnrod A, Mortensen M, Rasmussen MA, Stokholm J, Al-Soud WA, Sørensen S, Bisgaard H, Waage J. Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. *Microbiome* 2016;4:62.

SUPPLEMENTARY TABLES

Table S1. OTUs identified as contaminants

OTU	Method
Tepidimonas (28)	Both
Schlegelella (10)	Both
Acidovorax (66)	Both
Vogesella (69)	Both
Acinetobacter (31)	Both
Acinetobacter seohaensis (64)	Both
Phyllobacteriaceae (52)	Both
Pseudomonas stutzeri (95)	Both
Tardiphaga robiniae (106)	Both
Mesorhizobium (81)	Both
Shewanella (30)	Both
Massilia (88)	Frequency
Pseudomonas aeruginosa (79)	Frequency
Rhizobiales (169)	Prevalence
Xanthomonadales (114)	Prevalence
Cyanobacteria (126)	Prevalence
Hydrotalea (205)	Prevalence
Cyanobacteria (143)	Prevalence
Acinetobacter (139)	Prevalence
Modestobacter (167)	Prevalence
Cupriavidus metallidurans (156)	Prevalence

OTUs were identified as contaminants using their relation with bacterial biomass (frequency method) or their presence in samples compared to negative controls (prevalence method) or both.

Table S2: Respiratory symptoms in cases during 4-8 weeks follow-up.

n	98
Respiratory symptoms	
Rhinorrhea	73 (74.5)
Cough	61 (62.2)
Wheezing	20 (20.4)
Earache	11 (11.2)
Sore throat	7 (7.1)
Hoarseness	4 (4.1)
Severity measures	
Number of respiratory symptoms	2.00 [1.00, 2.00]
>1 respiratory symptoms	57 (58.2)
Fever (>38°C)	47 (48.0)
Physician visit	41 (41.8)
Antibiotic treatment	8 (8.2)

Data are presented as n (%) or median [IQR]. Data were acquired from parent questionnaires.

Table S3: Independent relationships between clinical characteristics during acute lower respiratory tract infection and early recurrence of respiratory symptoms.

	Coefficient	Std. Error	Z-value	P-value
Intercept	-2.588	1.237	-2.093	**0.036**
Age (months)	0.002	0.020	0.079	0.937
Antibiotic treatment	2.473	1.088	2.274	**0.023**
Diagnosis bronchiolitis*	0.751	0.702	1.069	0.285
Diagnosis mixed*	0.479	0.703	0.681	0.496
Diagnosis wheezing illness*	1.190	0.666	1.787	0.074
Follow-up time (days)	0.061	0.024	2.550	**0.011**
Age (months):Antibiotic treatment	-0.078	0.035	-2.202	**0.028**

* with pneumonia as a reference

SUPPLEMENTARY FIGURES

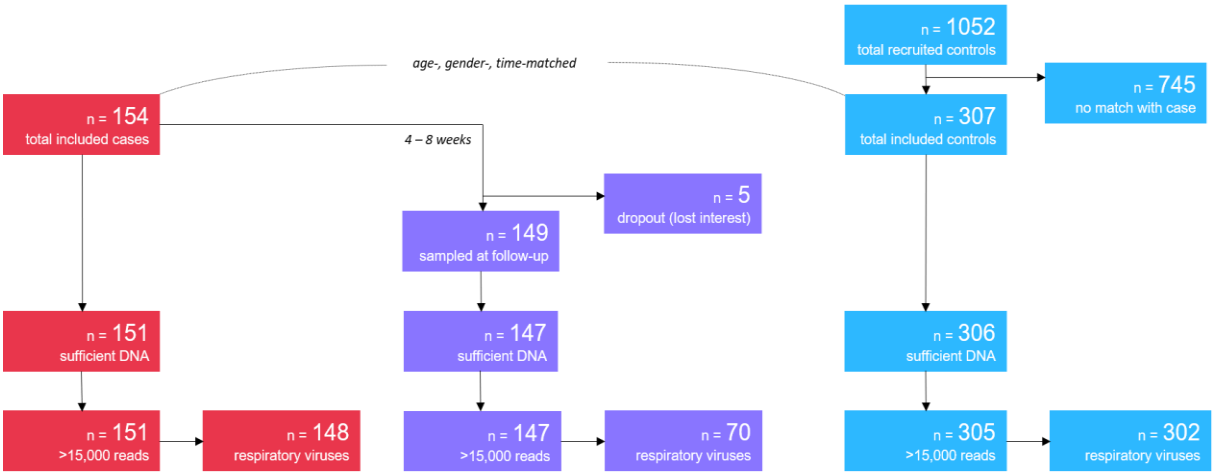


Figure S1. Study design

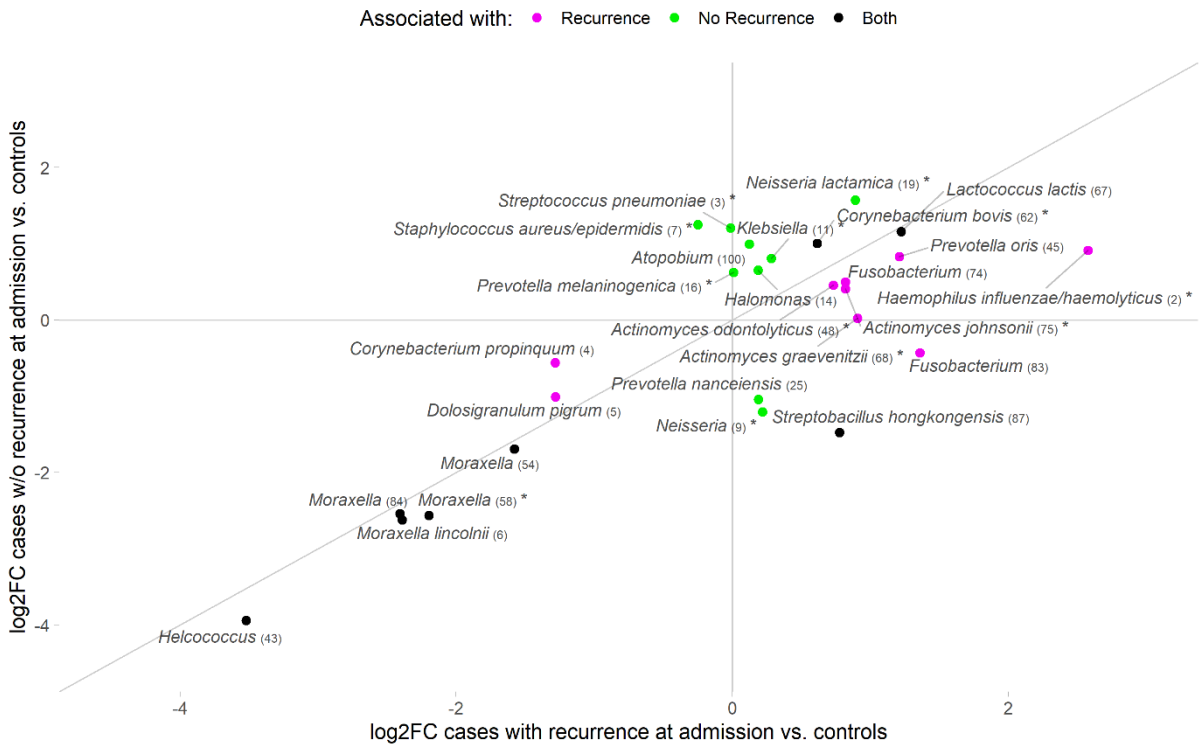


Figure S2. Discriminant OTUs between cases at admission and controls, stratified by recurrence of respiratory symptoms during follow-up.

Log2 fold changes (log2FC) of differentially abundant OTUs between cases at time of admission with subsequent recurrence of respiratory symptoms during follow-up and controls (purple, x-axis), or between cases at time of admission without subsequent recurrence of respiratory symptoms during follow-up and controls (green, y-axis), or both (black). Significance was assessed by *metagenomeSeq* analysis and cross-validated *VSURF* analysis limited to the top 100 most highly ranked OTUs, and results were combined and filtered at a fold change of at least 1.5 or below 0.5. OTUs marked by an asterisk were identified by cross-validated *VSURF* analysis.

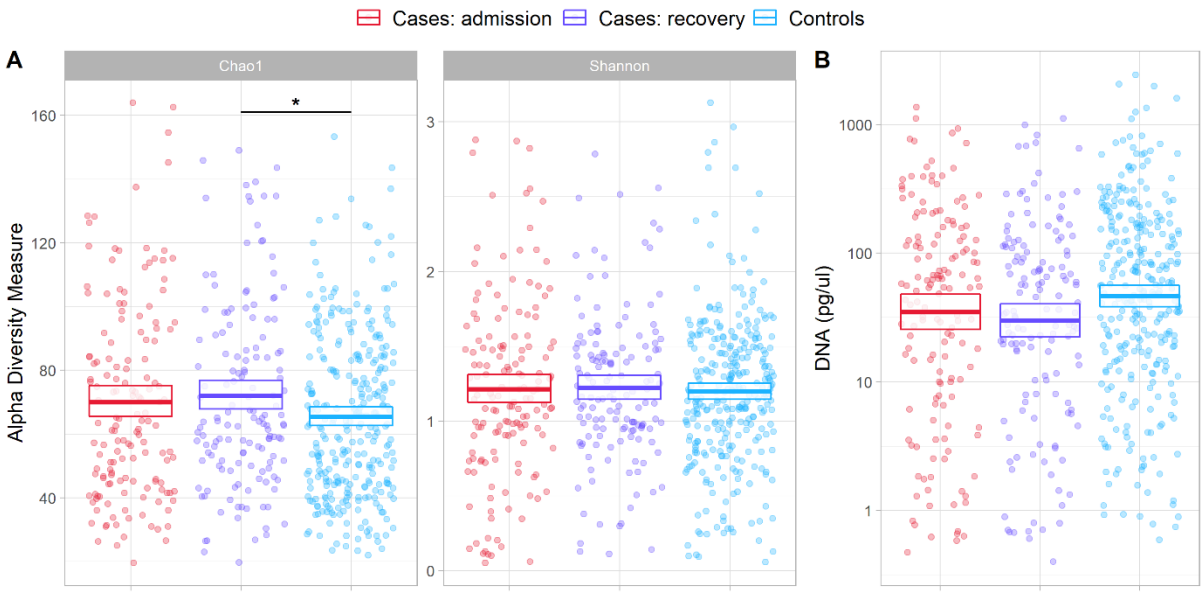


Figure S3. Microbial richness, diversity and biomass

(**A**) Alpha diversity measures Chao1 index and Shannon diversity index and (**B**) microbial biomass estimated by quantitative PCR of the 16S rRNA gene are shown for cases at admission, at recovery and controls. Boxes denote means with 95% confidence intervals. Significance was tested by linear mixed-effect models and indicated by *: $p<0.05$.

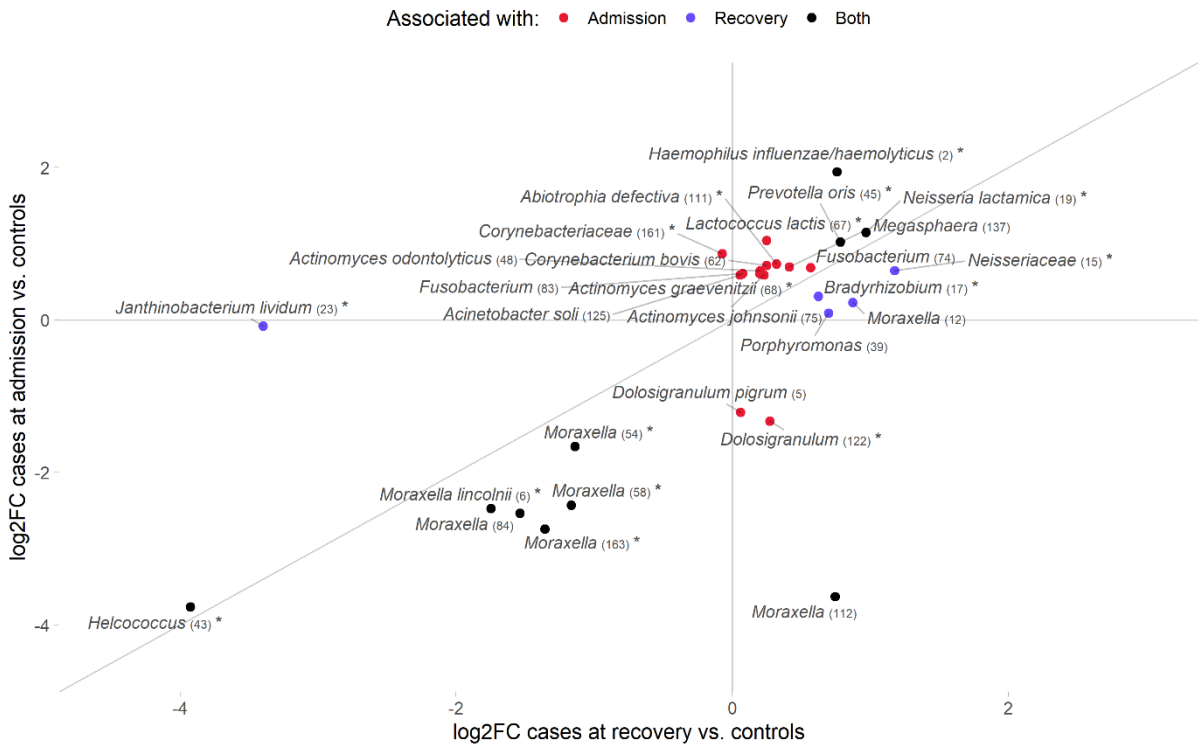


Figure S4. Discriminant OTUs between cases and controls.

Log2 fold changes (log2FC) of differentially abundant OTUs between cases at admission and controls (red, y-axis), or cases at recovery and controls (blue, x-axis), or both (black). Significance was assessed by *metagenomeSeq* analysis and cross-validated *VSURF* analysis, and results were combined and filtered at a fold change of at least 1.5 or below 0.5. OTUs marked by an asterisk were identified by cross-validated *VSURF* analysis.

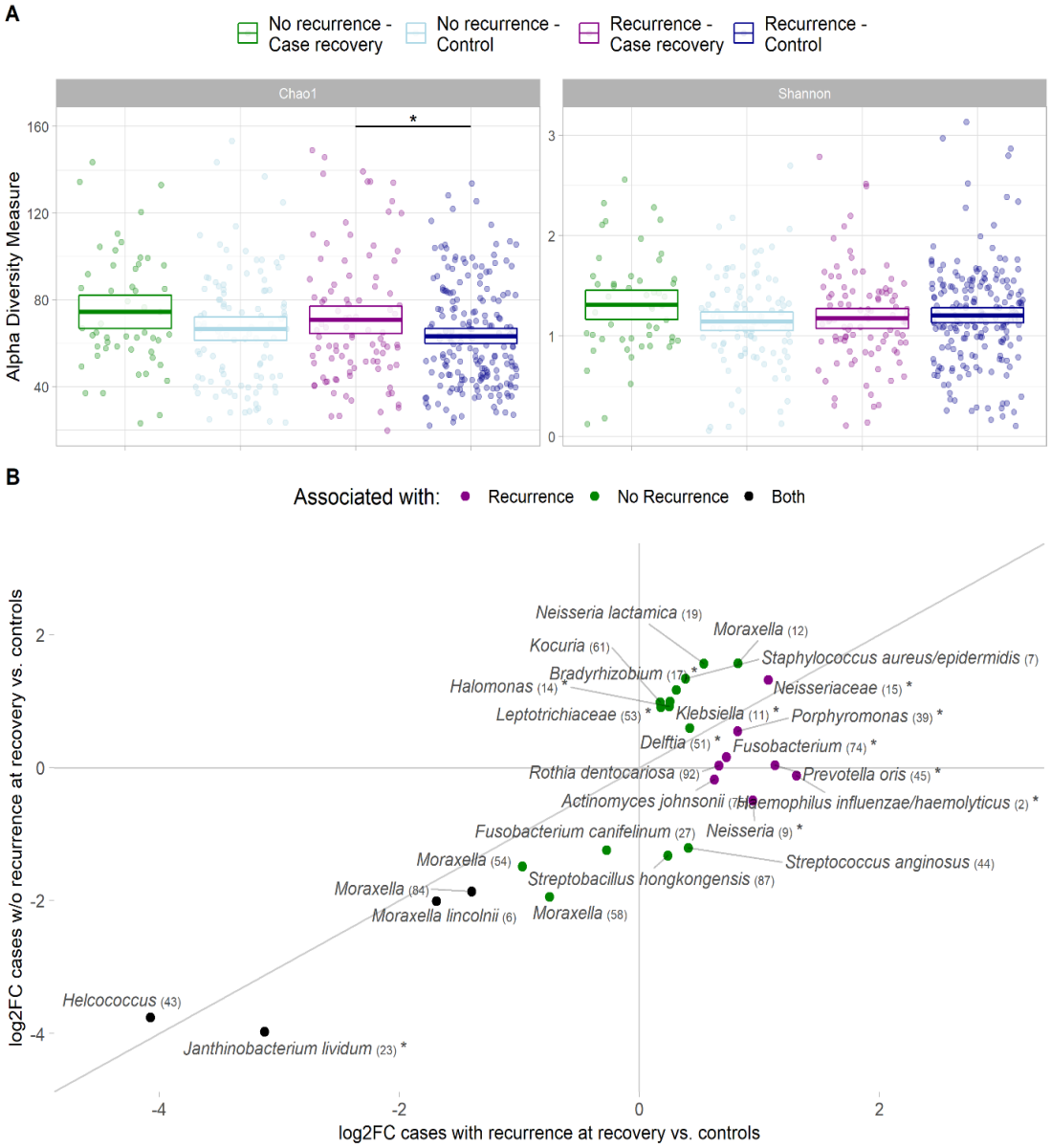


Figure S5. Microbial recovery in cases with and without recurrence of respiratory symptoms during follow-up.

(**A**) Alpha diversity measures Chao1 index and Shannon diversity index are shown for cases after recovery with and without recurrence of respiratory symptoms and matched controls. Boxes denote means with 95% confidence intervals. Significance was tested by linear mixed-effect models and indicated by *: p<0.05. (**B**) Log2 fold changes (log2FC) of differentially abundant OTUs between cases at time of recovery who had had recurrence of respiratory symptoms during follow-up and controls (purple, x-axis), or between cases at time of recovery who had not had recurrence of respiratory symptoms during follow-up and controls

(green, y-axis), or both (black). Significance was assessed by *metagenomeSeq* analysis and cross-validated *VSURF* analysis limited to the top 100 most highly ranked OTUs, and results were combined and filtered at a fold change of at least 1.5 or below 0.5. OTUs marked by an asterisk were identified by cross-validated *VSURF* analysis.

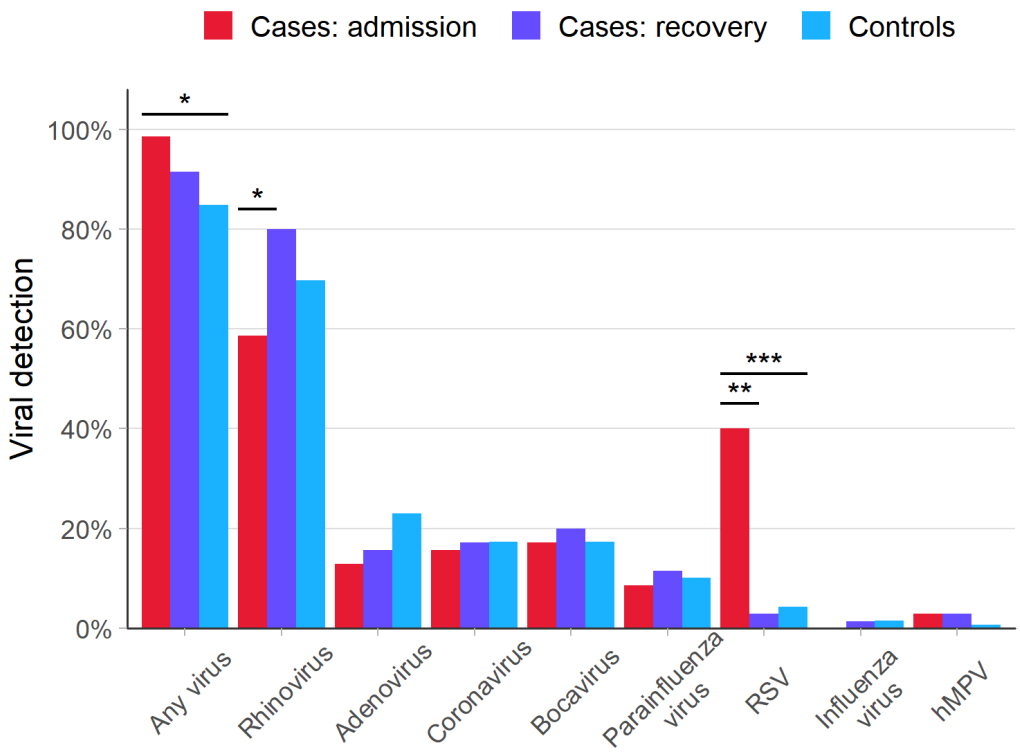
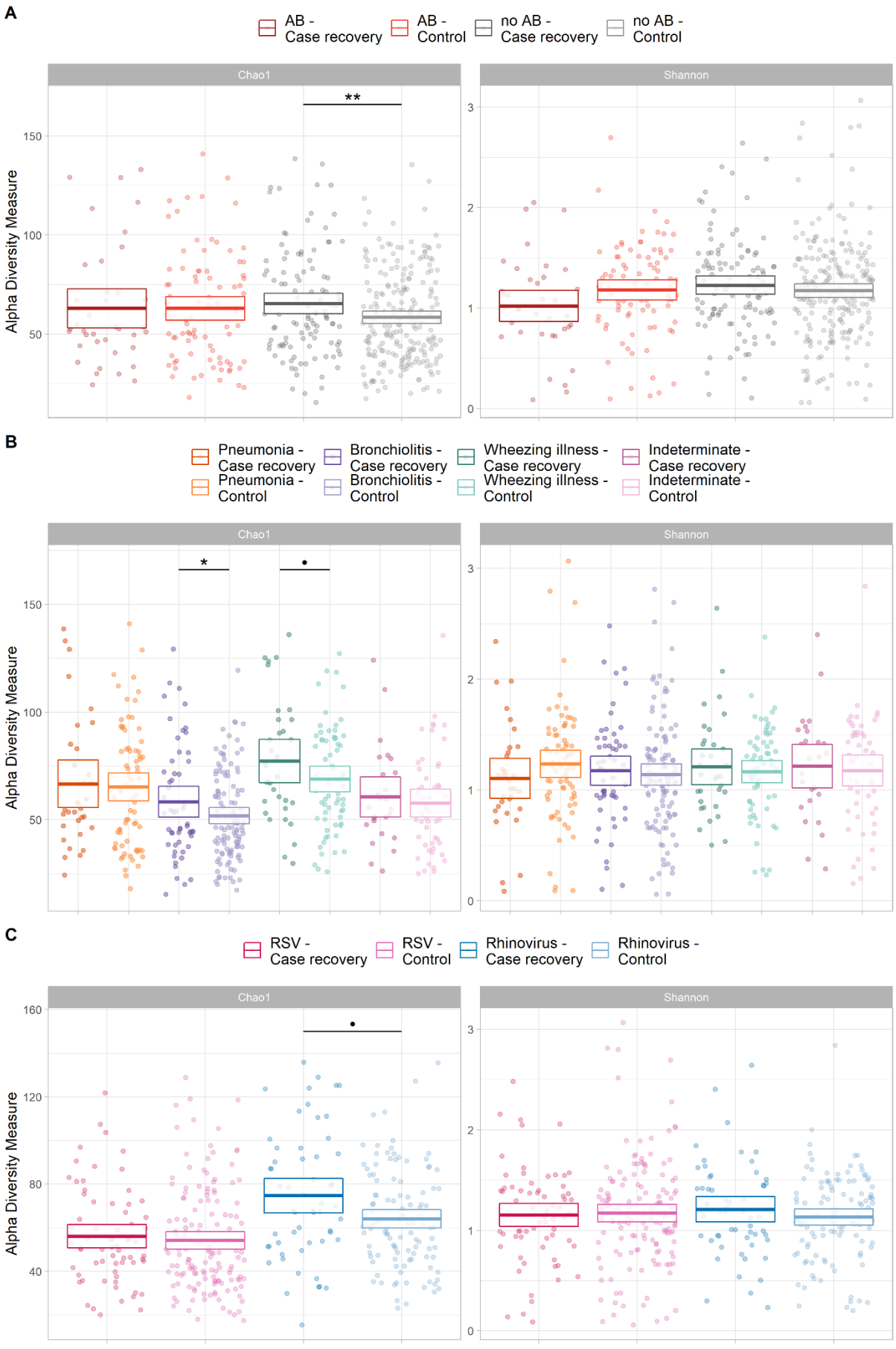


Figure S6. Viral carriage in cases and controls as detected by quantitative PCR

Bars denote percentages of samples positive for any virus and for each individual respiratory virus for cases at admission, cases at recovery, and controls. Significance was assessed by conditional logistic regression and is indicated by ***: p<0.001, **: p<0.005, or *: p<0.05.


A
AB - Case recovery
AB - Control
no AB - Case recovery
no AB - Control
Chao1
Shannon
Alpha Diversity Measure
**
B
Pneumonia - Case recovery
Pneumonia - Control
Bronchiolitis - Case recovery
Bronchiolitis - Control
Wheezing illness - Case recovery
Wheezing illness - Control
Indeterminate - Case recovery
Indeterminate - Control
*
•
C
RSV - Case recovery
RSV - Control
Rhinovirus - Case recovery
Rhinovirus - Control

1

Figure S7. Microbial recovery according to antibiotic treatment, LRTI phenotype and viral presence

Alpha diversity measures Chao1 index and Shannon diversity index are shown for cases at recovery and matched controls, stratified by (**A**) antibiotic treatment or no antibiotic treatment, (**B**) LRTI phenotype and (**C**) type of virus present at admission. Boxes denote means with 95% confidence intervals. Significance was tested by linear mixed-effect models and indicated by **: $p<0.005$; *: $p<0.05$; •: $p<0.10$.

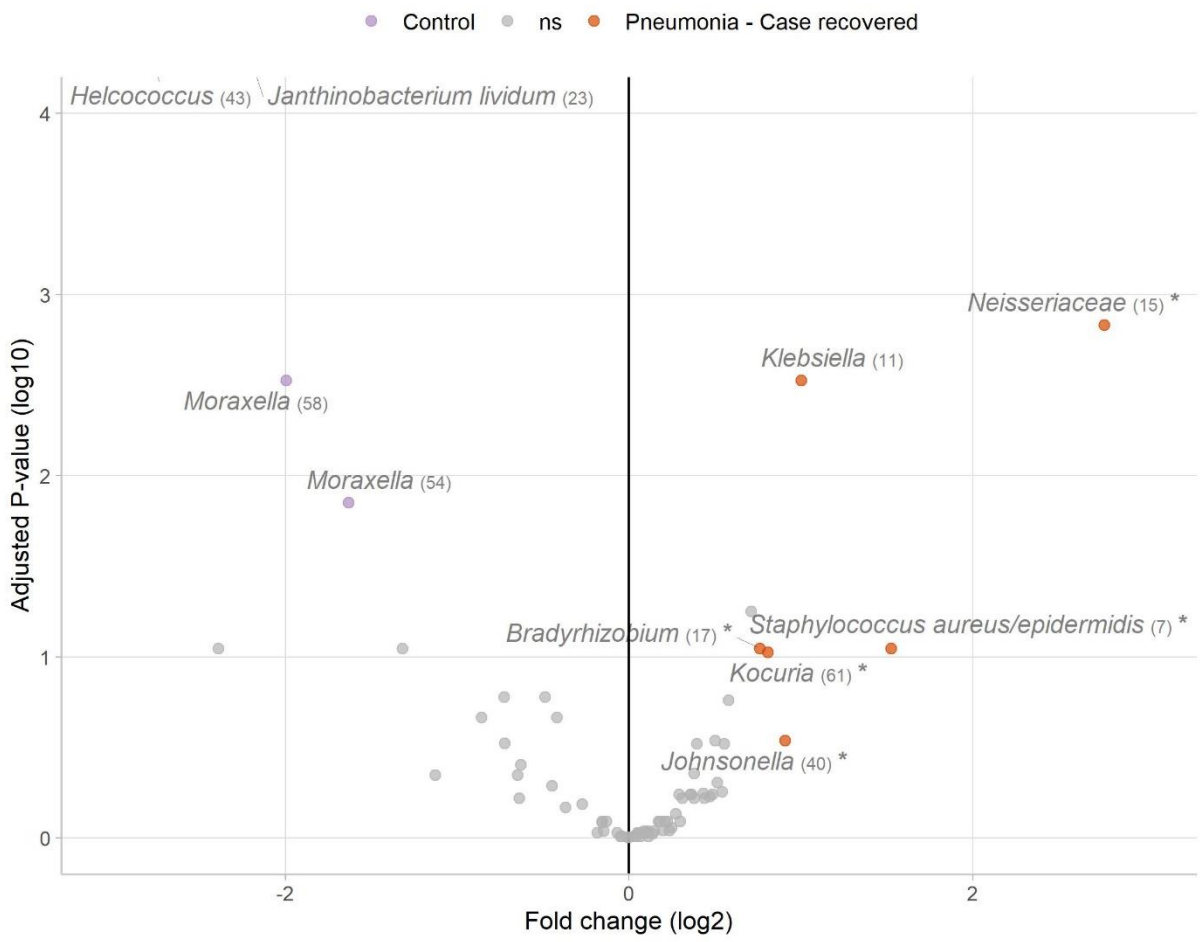


Figure S8. Discriminant OTUs between cases after recovery from pneumonia and controls.

Volcano plot of differentially abundant OTUs between cases with pneumonia at recovery and controls. Significance was assessed by *metagenomeSeq* analysis and cross-validated *VSURF* analysis limited to the top 100 most highly ranked OTUs, and combined results were filtered at a fold change of at least 1.5 or below 0.5. OTUs marked by an asterisk were identified by cross-validated *VSURF* analysis. Results of data points falling beyond the limits of the plot: *Helcococcus* (43) log2FC -4.21, adjusted p-value (log10) 8.30; *Janthinobacterium lividum* (23) log2FC -3.08, adjusted p-value (log10) 6.82.