

Supplementary Information

Shotgun metagenomic profiling of bacterial microbiomes, metagenome-assembled genomes and antimicrobial resistance in respiratory and blood samples from Gambian children with pneumonia

Dam Khan^{1,2,*}, Josh L. Espinoza³, Peggy-Estelle Tientcheu¹, Isaac Darko Otchere^{1,4}, Nuredin Ibrahim Mohammed¹, Archibald Worwui¹, Mark P. Nicol⁵, Brenda Kwambana-Adams^{6,7,8}, Martin Antonio^{1,2,9}, and Chris L. Dupont³

Author Affiliations

¹Medical Research Council Unit The Gambia at the London School of Hygiene & Tropical Medicine, Banjul, The Gambia

²Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom

³J. Craig Venter Institute, La Jolla, CA, United States of America

⁴Department of Bacteriology, Noguchi Memorial Institute for Medical Research, Legon, University of Ghana, Accra, Ghana

⁵Marshall Centre, Division of Infection and Immunity, School of Biomedical Sciences, University of Western Australia, Perth, Australia

⁶Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

⁷Malawi Liverpool Wellcome Programme, Blantyre, Malawi

⁸NIHR Global Health Group on Vaccines to Control Respiratory Pathogens and AMR, Division of Infection and Immunity, University College London, London, United Kingdom

⁹Centre for Epidemic Preparedness and Response, London School of Hygiene & Tropical Medicine, London, United Kingdom

Table S1: Specimen types analysed in the cohort

Specimen	Severe case	Very severe case	Total (%)
Nasopharyngeal swab	15	14	29(41)
Lung aspirate	4	3	7(10)
Blood	0	11	11(15)
Induced sputum	0	12	12(17)
Combined nasopharyngeal and oropharyngeal swab	0	12	12(17)
Total	19	52	71 (100)

Table S2: High-quality metagenome-assembled genome (MAG) recovery by specimen type.

Specimen	Number of samples	Raw reads, median (Q1-Q3)	Raw reads, mean (SD)	Final reads, median (Q1-Q3)	Final reads, mean (SD)	Number of high-quality MAGs recovered	MAGs/specimen	Mean MAG completeness (%)	Mean MAG contamination (%)	Mean MAG N50 (bp)
LA	7	17,470,272 (11,975,344–19,625,601)	15,605,543 (7,581,446)	1,659,428 (1,298,046–1,983,546)	1,613,766 (689,447)	8	1.14	98.26	0.97	57,534
Blood	11	21,251,502 (11,189,721–30,924,880)	20,880,923 (14,626,386)	1,409,762 (1,200,388–2,148,311)	1,540,531 (824,932)	11	1	98.53	0.72	91,482
NP-OP	12	3,929,859 (2,207,822–13,578,245)	9,077,555 (10,558,265)	641,092 (392,373–1,761,313)	1,048,808 (864,227)	6	0.5	95.49	0.36	22,821
IS	12	22,246,846 (7,177,610–25,549,118)	18,394,990 (11,316,803)	1,288,607 (487,966–1,691,182)	1,141,267 (708,905)	4	0.33	93.34	0.98	21,603
NPS	29	19,818,820 (16,052,780–21,919,864)	18,468,070 (9,772,912)	1,172,324 (901,630–1,457,200)	1,213,088 (599,067)	6	0.21	95.49	1.44	30,129

“Final reads” refers to read pairs remaining after quality trimming and removal of host-derived sequences. MAGs/specimen = number of high-quality MAGs recovered divided by number of samples analysed for that specimen type.

Table S3: Serotypes and sequence types of *S. pneumoniae* and *H. influenzae* high-quality MAGs

Specie	MAG	Serotype	ST	Participant	Specimen
<i>Streptococcus pneumoniae</i>	MAG4	19A	10676	3	LA
	MAG6	19A	847	4	NP-OP
	MAG9	Possible non capsular	Unassigned	6	NPS
	MAG11	Possible non capsular	3081	7	LA
	MAG14	6A/6B	5549	10	Blood
	MAG16	19A	847	11	LA
	MAG20	6E	Unassigned	17	Blood
	MAG23	19A	847	20	Blood
	MAG27	19F	925	30	Blood
	MAG32	19A	Unassigned	30	NPS
	MAG34	Possible non capsular	3334	34	NPS
<i>Haemophilus influenzae</i>	MAG7	Nontypeable <i>H. influenzae</i>	2442	5	IS
	MAG10	Nontypeable <i>H. influenzae</i>	503	6	LA
	MAG13	Nontypeable <i>H. influenzae</i>	Unassigned	7	LA
	MAG15	Nontypeable <i>H. influenzae</i>	Unassigned	10	NP-OP
	MAG17	Nontypeable <i>H. influenzae</i>	Unassigned	11	LA
	MAG18	Nontypeable <i>H. influenzae</i>	165	13	LA
	MAG19	Nontypeable <i>H. influenzae</i>	481	15	Blood
	MAG24	Nontypeable <i>H. influenzae</i>	2442	21	Blood
	MAG29	Nontypeable <i>H. influenzae</i>	Unassigned	30	Blood
	MAG31	Nontypeable <i>H. influenzae</i>	Unassigned	30	NPS

Table S4: Comparison of metagenome-assembled genomes with the Respiratory Genome Catalogue based on average nucleotide identity

MAG	Species	Specimen	RGC_bin_file	RGC_Genus	RGC_Species	ANI	Status
MAG1	<i>Neisseria meningitidis</i>	NP/OP	bin.478.fna	<i>Neisseria</i>	<i>s__</i>	85.62	NO RGC HIT
MAG2	<i>Haemophilus parahaemolyticus</i>	IS	bin.699.fna	<i>Haemophilus</i>	<i>s__Haemophilus_D parainfluenzae</i>	82.58	NO RGC HIT
MAG3	<i>Moraxella catarrhalis</i>	IS	bin.172.fna	<i>Moraxella</i>	<i>s__Moraxella catarrhalis</i>	99.18	MATCHED
MAG4	<i>Streptococcus pneumoniae</i>	LA	bin.1084.fna	<i>Streptococcus</i>	<i>s__Streptococcus pneumoniae</i>	97.80	MATCHED
MAG5	<i>Moraxella catarrhalis</i>	NP/OP	bin.172.fna	<i>Moraxella</i>	<i>s__Moraxella catarrhalis</i>	99.21	MATCHED
MAG6	<i>Streptococcus pneumoniae</i>	NP/OP	bin.1084.fna	<i>Streptococcus</i>	<i>s__Streptococcus pneumoniae</i>	97.93	MATCHED
MAG7	<i>Haemophilus influenzae</i>	IS	bin.985.fna	<i>Haemophilus</i>	<i>s__Haemophilus influenzae</i>	97.08	MATCHED
MAG8	<i>Haemophilus parahaemolyticus</i>	IS	bin.699.fna	<i>Haemophilus</i>	<i>s__Haemophilus_D parainfluenzae</i>	82.84	NO RGC HIT
MAG9	<i>Streptococcus pneumoniae</i>	NPS	bin.1084.fna	<i>Streptococcus</i>	<i>s__Streptococcus pneumoniae</i>	97.77	MATCHED
MAG10	<i>Haemophilus influenzae</i>	LA	bin.985.fna	<i>Haemophilus</i>	<i>s__Haemophilus influenzae</i>	97.22	MATCHED

MAG11	<i>Streptococcus pneumoniae</i>	LA	bin.1084.fna	<i>Streptococcus</i>	<i>s__Streptococcus pneumoniae</i>	98.01	MATCHED
MAG12	<i>Moraxella catarrhalis</i>	LA	bin.172.fna	<i>Moraxella</i>	<i>s__Moraxella catarrhalis</i>	96.09	MATCHED
MAG13	<i>Haemophilus influenzae</i>	LA	bin.985.fna	<i>Haemophilus</i>	<i>s__Haemophilus influenzae</i>	97.37	MATCHED
MAG15	<i>Haemophilus influenzae</i>	NP/OP	bin.985.fna	<i>Haemophilus</i>	<i>s__Haemophilus influenzae</i>	97.31	MATCHED
MAG16	<i>Streptococcus pneumoniae</i>	LA	bin.1084.fna	<i>Streptococcus</i>	<i>s__Streptococcus pneumoniae</i>	97.88	MATCHED
MAG17	<i>Haemophilus influenzae</i>	LA	bin.141.fna	<i>Haemophilus</i>	<i>s__Haemophilus influenzae</i>	97.17	MATCHED
MAG18	<i>Haemophilus influenzae</i>	LA	bin.985.fna	<i>Haemophilus</i>	<i>s__Haemophilus influenzae</i>	97.27	MATCHED
MAG25	<i>Limosilactobacillus Sp.</i>	NP/OP				NA	NO RGC HIT
MAG30	<i>Moraxella catarrhalis</i>	NPS	bin.172.fna	<i>Moraxella</i>	<i>s__Moraxella catarrhalis</i>	99.29	MATCHED
MAG31	<i>Haemophilus influenzae</i>	NPS	bin.985.fna	<i>Haemophilus</i>	<i>s__Haemophilus influenzae</i>	97.39	MATCHED
MAG32	<i>Streptococcus pneumoniae</i>	NPS	bin.1084.fna	<i>Streptococcus</i>	<i>s__Streptococcus pneumoniae</i>	97.94	MATCHED
MAG33	<i>Moraxella lincolnii</i>	NP/OP	bin.78.fna	<i>Moraxella</i>	<i>s__Moraxella_C lincolnii</i>	97.63	MATCHED
MAG34	<i>Streptococcus pneumoniae</i>	NPS	bin.1084.fna	<i>Streptococcus</i>	<i>s__Streptococcus pneumoniae</i>	98.09	MATCHED
MAG35	<i>Moraxella nonliquefaciens</i>	NPS	bin.210.fna	<i>Moraxella</i>	<i>s__Moraxella nonliquefaciens</i>	98.07	MATCHED

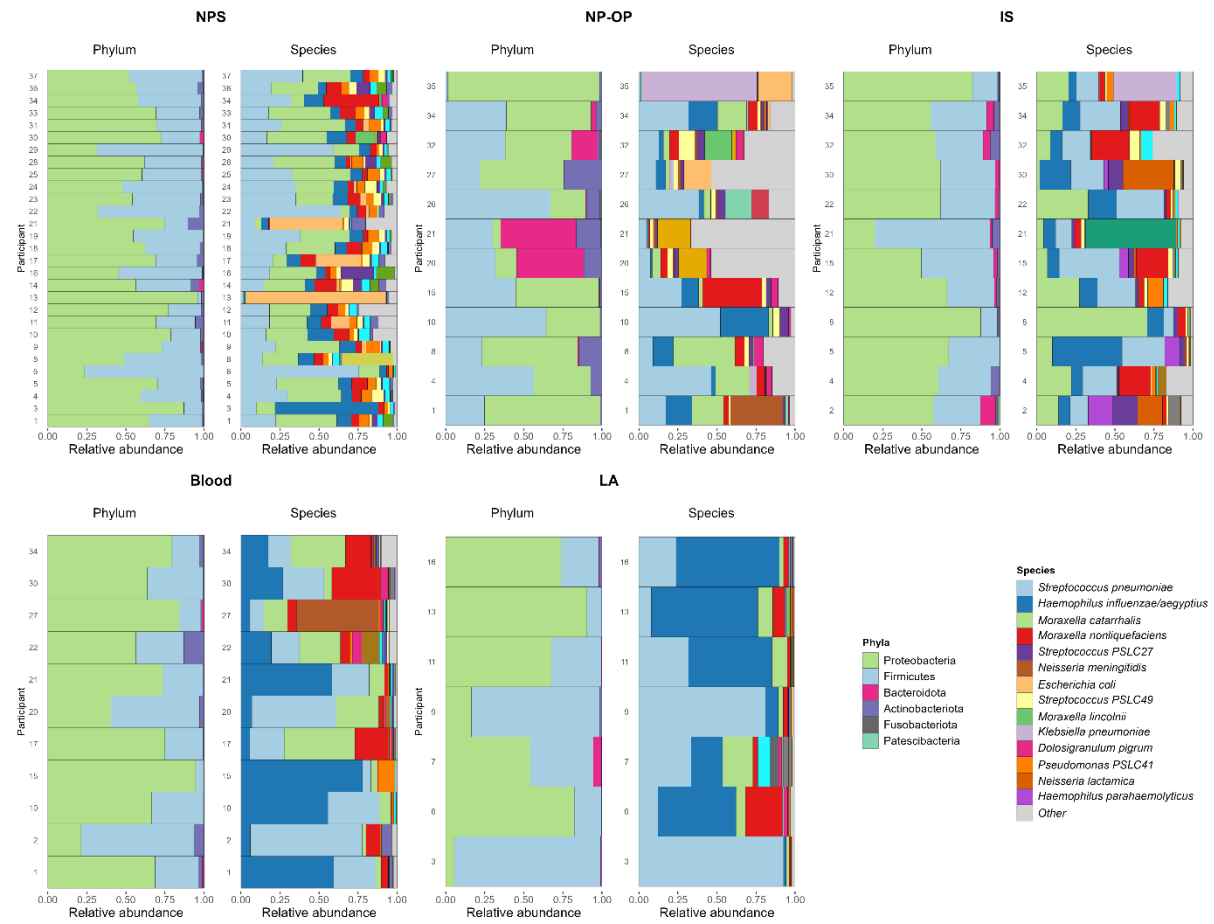


Figure S1: Barplots of mean relative abundances of bacteria DNA at the phylum and species level in individual samples grouped by specimen type. The bars which represent individual samples are arranged on the Y axis based on similarity in profiles. DNA extraction methods for each specimen type: DNeasy PowerSoil Kit (NPS, NP-OP, IS), QIAamp MinElute Virus Spin Kit (LA), and QIAamp DNA Blood Mini Kit (blood). NPS: nasopharyngeal swab, NP-OP: nasopharyngeal/oropharyngeal swab, IS: induced sputum, LA: lung aspirate

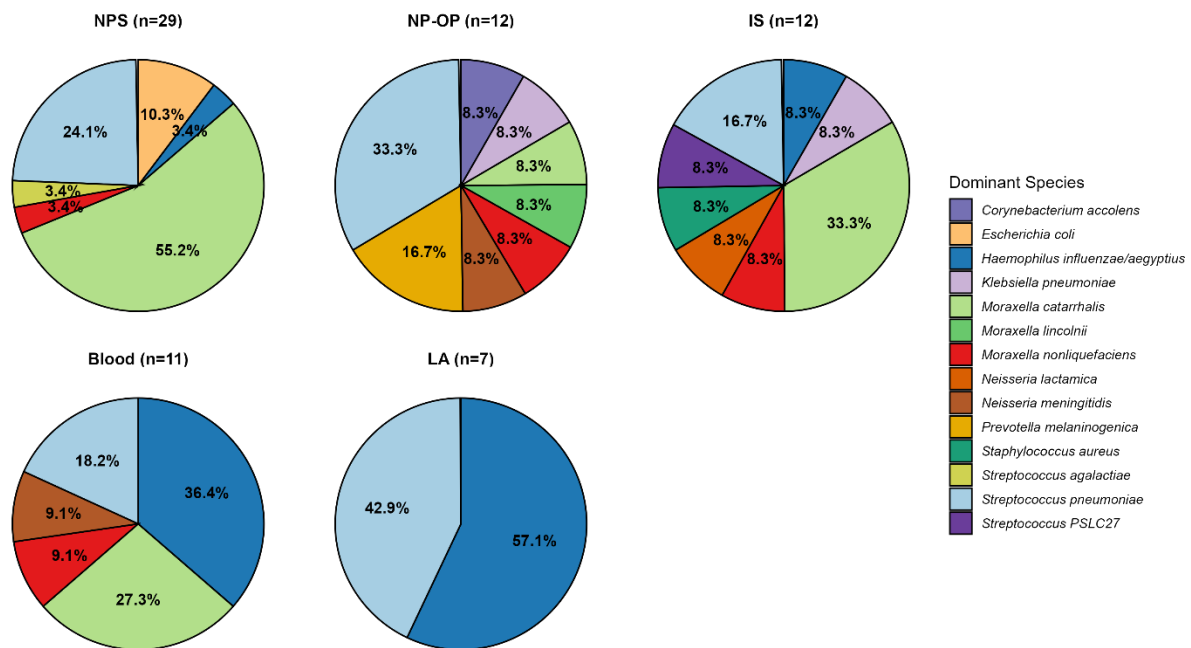


Figure S2: Distribution of dominant bacterial species across specimen type. A species was defined as dominant if it had the highest relative abundance in a sample. Each sector in the pie represents the percentage of the number of samples in which that species was dominant relative to the number of samples analysed for that specimen type. The number of samples (equal to the number of participating individuals contributing each specimen type) is indicated in parentheses. DNA extraction methods: DNeasy PowerSoil Kit (NPS, NP-OP, IS), QIAamp MinElute Virus Spin Kit (LA), and QIAamp DNA Blood Mini Kit (blood). NPS: nasopharyngeal swab, NP-OP:

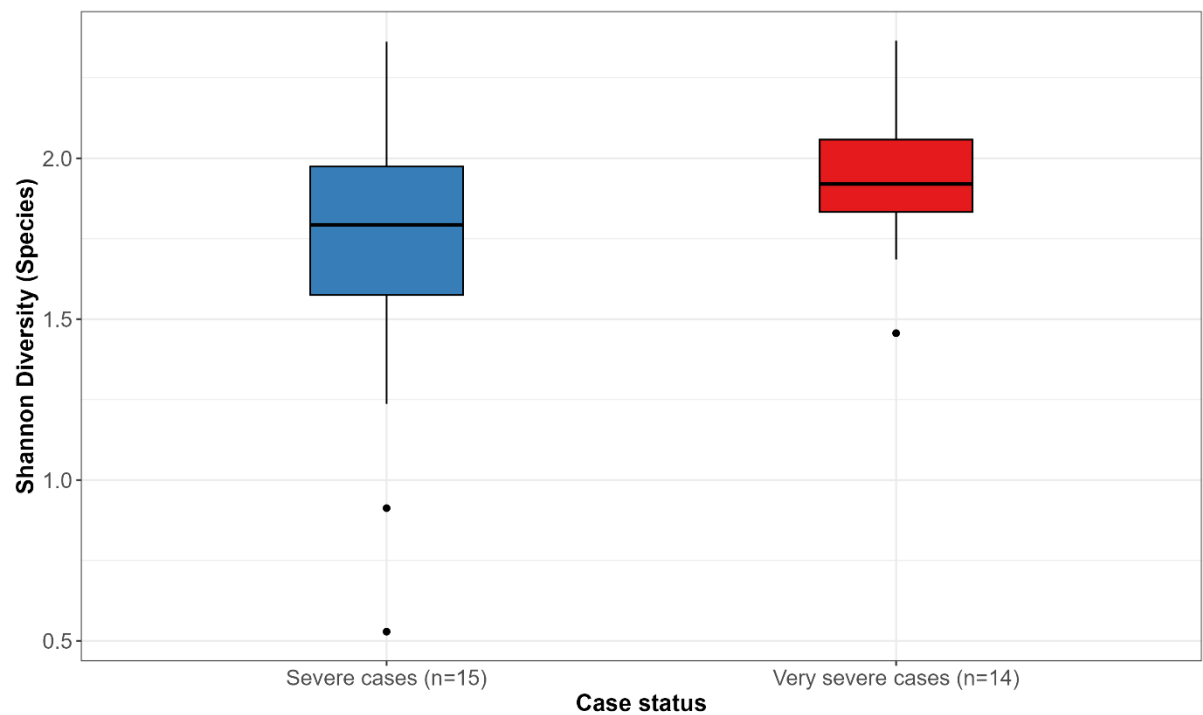


Figure S3: Shannon diversity of the nasopharynx microbial community by case type. Median Shannon diversity values are represented by horizontal lines within each boxplot while the upper and lower ranges of each boxplot represent the 75% and 25% quartiles. The number of samples in each category is indicated in parentheses.

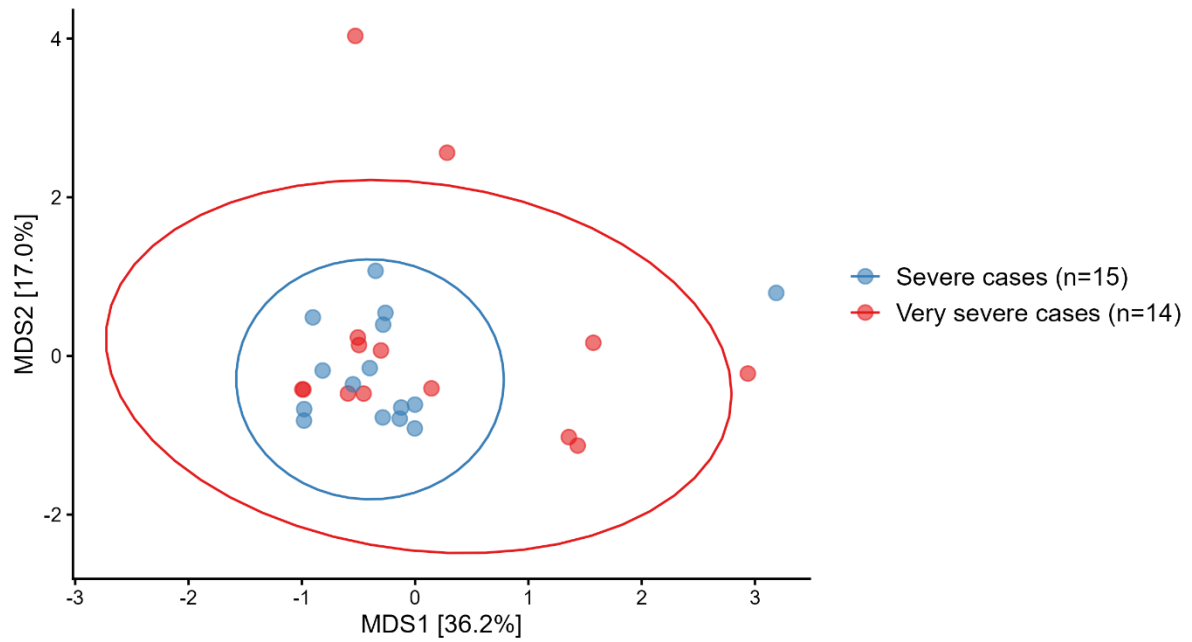


Figure S4: Principal Coordinates Analysis (PCoA) plot based on Aitchison distance of severe cases and very severe cases. The number of samples in each category is indicated in parentheses.