Outbreak and Postnatal Antibiotic Exposures Drive the Development Trajectory of the Nasopharyngeal Microbiota in the First Year of Life

Polona Rajar[1,2]†, Achal Dhariwal[2]†, Gabriela Salvadori[2], Heidi Aarø Åmdal[2], Dag Berild[3,4], Ulf R. Dahle[5], Drude Fugelseth[1,4], Gorm Greisen[6], Ulrik Lausten-Thomsen[6], Ola Didrik Saugstad[7], Fernanda Cristina Petersen[2]*, Kirsti Haaland[1]*

[1] Department of Neonatal Intensive Care, Division of Paediatric and Adolescent Medicine, Oslo University Hospital, Oslo, Norway.
[2] Institute of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway.
[3] Department of Infectious Diseases, Oslo University Hospital, Oslo, Norway.
[4] Institute of Clinical Medicine, Faculty of Medicine, Oslo University, Oslo, Norway.
[5] Centre for Antimicrobial Resistance, Norwegian Institute of Public Health, Oslo, Norway
[6] Department of Neonatology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark.
[7] Department of Pediatric Research, University of Oslo, Oslo, Norway.

† These authors contributed equally: Polona Rajar, Achal Dhariwal.

* Correspondence:
Kirsti Haaland
Department of Neonatal Intensive Care Unit Ullevål,
Oslo University Hospital, Oslo, Norway
Tel: +47-22117622
Email: uxkila@ous-hf.no

Fernanda Cristina Petersen
Institute of Oral Biology
Faculty of Dentistry,
University of Oslo, Oslo, Norway
Tel: +47-22840312
E-mail: f.c.petersen@odont.uio.no

ABSTRACT

Early exposure to antibiotics and prolonged hospitalization in preterm infants may perturb microbiome development and contribute to adverse health outcomes. Although nasopharyngeal microbiomes are linked to respiratory infections, their early development is underexplored and often assessed with 16S rRNA sequencing, which lacks species resolution. Here, we investigated nasopharyngeal microbiota dynamics in 66 preterm infants by performing deep shotgun metagenomics on 369 nasopharyngeal aspirates collected from birth until 6 months corrected age (~7-10 months chronological age). The nasopharyngeal microbiota evolved dynamically, exhibiting age-structured and individualized patterns shaped by postnatal antibiotic exposure and hospitalization. Early-life antibiotic exposure (ampicillin + gentamicin) had transient but significant effects on microbial diversity, composition, stability, and community dynamics. Conversely, an unexpected NICU outbreak left a lasting signature, with persistent *S. marcescens* carriage observed at 6 months corrected age, which our machine learning algorithm accurately predicted based on microbiome composition in the first days of life. A *S. marcescens*-dominated community type was present in 45% of total samples and exhibited remarkable stability over time, with minimal transitions to other types regardless of antibiotic exposure. These findings underscore the need for deeper insight into how antibiotics and potential outbreaks can lead to alterations in microbiome trajectories, with possible long-term health implications.

Keywords: whole metagenomic sequencing, nasopharyngeal, respiratory, microbiome, preterm birth, hospital outbreak, antibiotics

INTRODUCTION

Microbial colonization after birth is a dynamic process, influenced by environmental and host factors, which holds significant implications for long-term health [1-3]. Most of the current knowledge originates from studies on the intestinal microbiota, leaving a knowledge gap regarding the microbiota development in other body niches, particularly in the respiratory tract, where infections are notably common and sometimes critical [4-6].

Immediately after birth, the upper respiratory tract mucosa is rapidly colonized by microbes originating from mothers' genito-rectal, skin and oral flora, as well as from the surrounding environment. This initial colonization is followed by an increase in bacterial abundance and a shift towards distinct bacterial communities within weeks [3, 7, 8]. For term infants, the nasopharyngeal microbiota further differentiates towards stable microbial profiles by six months of age, influenced by factors such as the mode of delivery, nutrition regime, and antibiotic therapy [3, 6, 7, 9]. Timed signals from the respiratory microbiota are necessary for the infant's immune system maturation [10-13]. Alterations in airway microbiota during this critical developmental phase has been linked to severity and susceptibility to respiratory diseases later in life, such as asthma, otitis media, and respiratory syncytial virus infections [14-19].

Preterm infants are particularly vulnerable due to the development of inflammation and infection-related pathologies in the respiratory tract, even without confirmed neonatal lung injury [20-23]. For preterm infants, disruption in microbiota development may pose additional risks, exacerbated by prolonged hospitalization and early antibiotic treatment [2, 24]. Both factors dramatically influence the microbial ecological succession in preterm infants, though their effects on non-intestinal sites remain less understood. In Norway, 77% of infants born at gestational age (GA) <32 weeks are exposed to antibiotics during their first week of life, and

24 75% are treated with antibiotics within 72 hours after birth due to suspected early onset neonatal
25 sepsis (EONS) [25].

26 To date, research on the respiratory microbiome has primarily utilized 16S rRNA amplicon
27 sequencing. This method, however, lacks the species-level resolution necessary to distinguish
28 between commensals and potential pathogens [20]. This gap in knowledge is largely attributed to
29 the challenges associated with low microbial biomass and the high content of human DNA in
30 nasopharyngeal samples, which complicates the retrieval of sufficient microbial DNA for non-
31 targeted sequencing using shotgun metagenomics [26]. To address these challenges, we recently
32 developed and optimized a methodology that enabled us to perform an in-depth analysis of the
33 nasopharyngeal resistome in preterm infants [27, 28]. Here, we focus on the developmental
34 dynamics of the nasopharyngeal microbiome in preterm infants and explore the impact of
35 different clinical variables and exposures on its trajectories and maturation.

36

37 **MATERIALS & METHODS**

38 **Study design and sample collection**

39 Between July 2019 and January 2021, we approached parents within 48 hours after birth of all
40 infants with a gestational age from 28 weeks 0 days to 31 weeks 6 days who were born at or
41 transferred to the Ullevål Neonatal Intensive Care Unit (NICU), a tertiary NICU at Oslo
42 University Hospital. Patients' metadata were collected from the electronic records. We divided
43 infants into three groups (Table 1): "Only Early Antibiotics" (n = 24), "Other Antibiotics" (n =
44 21) and "Antibiotic Naive" (n = 21). Infants in the "Only Early Antibiotics" group were treated
45 with intravenous ampicillin and gentamicin due to increased risk for early-onset neonatal sepsis
46 (EONS), initiated within 24 hours after birth and discontinued after a mean (SD) duration of 4
47 (2) days. The "Other Antibiotics" group received different combinations of either early
48 antibiotics (initiated within 72 hours after birth); late antibiotics (initiated later than 72 hours

49 after birth); and/or antibiotics after discharge (intravenous or per oral route). Five infants
50 received antibiotics in the period between discharge from the NICU and six months corrected
51 age. The "Antibiotic Naive" group included infants who did not receive any antibiotics from
52 birth until 6 months corrected age.

53 Nasopharyngeal aspirates (NPAs) were obtained within 48 hours at six time points: day of life
54 (DOL) 0 (= day of birth), DOL 7, DOL 14, DOL 28, DOL 56 or discharge and at six months
55 corrected age [28]. Samples were stored at -80°C in 20% glycerol (2 ml) solution until further
56 processing in the laboratory. Throughout all further steps, samples belonging to the same infant
57 were processed concomitantly.

58 **Metagenomic DNA extraction**

59 Metagenomic DNA extraction was performed using standard operating procedures published
60 previously [28] with some minor modifications (Method Supplement). After host DNA removal
61 (MolYsis™ Basic5, Molzym), all samples were spiked with 20 µl of Spike-in Control II Low
62 Microbial Load (Catalog D6321 & D6321-10, ZymoBIOMICS™, Irvine, CA), immediately
63 followed by DNA extraction (MasterPure™ Gram Positive DNA Purification Kit, Epicentre).
64 Zymo Femto™ Quantification kits (ZymoBIOMICS™) were used to quantify human (Zymo
65 E2005) and microbial DNA (Zymo E2006) in all extracted samples with real time qPCR.
66 Further, Nextera DNA Flex kit (Illumina Inc., San Diego, CA) was used for metagenomic
67 library preparation following manufacturer's instructions. Extracted metagenomic DNA was
68 eluted in 35 µl, quantified using Qubit™ dsDNA HS kit on a Qubit 4.0 Fluorometer (Invitrogen,
69 Thermo Fisher Scientific, Waltham, MA) and stored at -80°C for up to four weeks before
70 further processing.

71 **Bacterial and host DNA quantification with real time qPCR**

72 Zymo Femto™ Quantification kits (ZymoBIOMICS™) were used to quantify human (Zymo
73 E2005) and microbial DNA (Zymo E2006) in all extracted samples, according to the
74 manufacturer's instructions. Used sample volume was 1 µl.

75 **Controls for low biomass samples**

76 Six aliquots of 20 µl Spike-in Control II (ZymoBIOMICS™) were extracted with MasterPure™
77 Gram Positive DNA Purification Kit (Epicentre) in two independent experiments and used as a
78 positive control (expected yield 0.4 ng). One glycerol saline solution sample from each batch
79 prepared for the cryopreservation of patient samples (n = 5) as well as reagent blanks (n = 5)
80 were extracted and used as negative controls. The DNA yield from negative controls was
81 extremely low or undetected (n = 10; mean = 0.00043 ng/µL), as measured by RT-qPCR.

82 **Metagenomic Library Preparation and Sequencing**

83 Nextera DNA Flex kit (Illumina Inc., San Diego, CA) was used for library preparation
84 following manufacturer's instructions. Input yield was set to 10 ng DNA according to Qubit
85 results, and all available extracted DNA was used in samples that yielded less than 10 ng
86 according to both Qubit and PCR measurements. Twenty-four samples from 16 infants were
87 excluded from library preparation due to negative 16S rRNA qPCR values, determined by
88 subtracting their bacterial DNA yield from the average yield of positive spike-in controls
89 (Supplementary Table E1).

90 Twelve PCR amplification cycles were used for metagenomic library preparation as per the
91 producer's recommendations. Library concentration and purity were measured using Qubit™
92 dsDNA HS kit (Invitrogen) and Bioanalyzer 2100 (Agilent, Santa Clara, CA). Libraries of
93 remaining 345 NPA samples proceeded to WMS in one pool with 5 nM concentration (Figure
94 1B). Three samples with insufficient sequencing reads after contaminant removal were also
95 excluded (Supplementary Table E5). Furthermore, for four infants with multiple samples

96 available at the same time points, only one sample per time point was retained, leading to the
97 exclusion of six additional samples before downstream analysis. Twenty-two of the 25 samples
98 excluded before WMS analysis belonged to infants exposed to postnatal antibiotics ($p < 0.05$)
99 (Supplementary Table E1). Libraries of positive ($n = 6$) and negative controls ($n = 10$)
100 proceeded to WMS in one pool with 9.4 nM concentration. Details on sample inclusion and
101 exclusion at various stages, from sampling to analysis, are provided in Supplementary Table
102 E1.

103 Whole metagenome sequencing (WMS) was conducted at the Norwegian Sequencing Centre
104 (Oslo, Norway), on a NovaSeq S4 platform (Illumina Inc.) using a paired-end sequencing
105 approach with a targeted read length of 150 bp in high-output mode.

106 **Bioinformatics processing and microbiome profiling**

107 FASTQC (v.0.11.9) was used to assess the quality of raw and clean reads [29]. Quality and adapter
108 trimming, as well as quality control were performed with Trim galore (v.0.6.1) [30]. All reads
109 were aligned with human reference genome (GRCh38) and human DNA sequences were
110 removed using Bowtie2 (v.2.3.4.2) [31]. MetaPhlAn 3.0 [32] was used for taxonomic microbiome
111 profiling of the remaining clean, high-quality reads, using a customized database by adding
112 clade-specific marker gene from the mock species (*q=0.05*).

113 **Statistical analysis and data visualization**

114 Details regarding statistical analysis of cohort demographics, 16S rRNA qPCR results and
115 downstream analysis of WMS data are described in the Method Supplement. Downstream
116 analysis was conducted in R (v.4.2.1) within RStudio (v.2022.07.2+576) [33, 34]. Figures were
117 created mainly created using ggplot2 (v.3.4.0) package [35] and GraphPad Prism 9.4.1(458)©
118 GraphPad Software, LLC (Boston, MA, USA).

119

RESULTS

Study cohort characteristics and overview of sequencing results

We obtained 369 NPA samples from 66 infants at six sampling time points between birth and six months corrected age (Figure 1), including 108 samples from 21 infants not exposed to any postnatal antibiotics (Figure S1A). Infants exposed to postnatal antibiotic (n = 45) had lower GA and birth weight (BW), and their mothers were more often treated with antibiotics during pregnancy (Table 1). The average (SD) early antibiotic (ampicillin and gentamicin) treatment duration was 4 (2) days. None of the infants had a positive blood culture obtained prior to initiation of early antibiotics, and 13 had a positive blood culture prior to initiation of late antibiotics.

In total, 12.13 billion raw sequencing reads were generated from NPA samples, with an average of 35 million (M) reads per sample (range: 2.55–93 M) (Supplementary Table E2). Of the total reads, 99.09–99.99% survived quality filtering and trimming across all samples. In contrast, the average number of raw sequencing reads detected in negative control samples was ~2.5 logs lower (mean: 0.15 M; range: 24,113–355,861) compared to nasopharyngeal samples (Supplementary Table E2). On average, 55% of quality-filtered reads (range: 0.86%–98.6%) were associated with the human genome. After human contamination removal, an average of 6 M high-quality clean reads per sample (range: 439–24 M) were assigned to bacteria (Supplementary Table E2). A total of 28 microbial species, including three ZymoBIOMICS spike-in species, were identified as microbial contaminants and excluded from the raw microbial abundance tables (Supplementary Table E5). Overall, 336 nasopharyngeal samples collected from 66 preterm infants were included in downstream analysis, containing 537 species distributed across 13 bacterial phyla.

143 **Characterization of nasopharyngeal microbial load, composition, and dynamics in**
144 **preterm infants**

145 Firstly, we quantified total bacterial load in nasopharyngeal samples from preterm infants using
146 16S rRNA gene-targeted qPCR. Using a generalized linear mixed-effects model (GLMM)
147 corrected for individual variation and age, we observed a significant increase in absolute
148 bacterial abundance up to time point 4 compared to the initial sample (time point 1) (Tukey
149 HSD post hoc test: $p.adj < 0.05$; Figure 2A). This increase was followed by stabilization and a
150 slight declining trend observed at 6 months corrected age. The mean absolute bacterial
151 abundance across infant groups was lowest at time point 1 (mean log 16S qPCR: -1.15 ng),
152 indicating low bacterial loads in the nasopharynx of preterm infants at birth (Figure 2B).
153 Further, infants exposed to "Other antibiotics" had significantly lower bacterial loads at time
154 point 3 compared to those in the "Only Early antibiotics" group (pairwise Wilcoxon test: $p.adj$
155 $= 0.012$; Figure 2B).

156 We next characterized the microbial community composition in the developing nasopharyngeal
157 microbiome using shotgun metagenomic sequencing. Absolute bacterial abundance (measured
158 by qPCR) showed no significant correlation with the number of bacterial reads (Pearson: $R =$
159 0.09, $p = 0.10$; Figure S1C). *Proteobacteria* (mean: 53.3%; SD: 13.5%) and *Firmicutes* (mean:
160 38.1%; SD: 12.1%) were the dominant phyla, followed by *Actinobacteria* (mean: 6.8%; SD:
161 2.46%), *Fusobacteria* (mean: 1.2%; SD: 1.69%), and *Bacteroidetes* (mean: 0.6%; SD: 0.59%)
162 across all groups and time points (Figure S2).

163 At the genus level, the five most abundant bacteria were *Serratia* (mean: 26.73%; SD: 10.93%),
164 *Streptococcus* (mean: 23.45%; SD: 9.50%), *Staphylococcus* (mean: 17.08%; SD: 9.18%),
165 *Gemella* (mean: 7.16%; SD: 4.09%), and *Corynebacterium* (mean: 3.50%; SD: 2.49%) (Figure
166 2C). *Serratia* predominated in the naive group (mean: 38.18%; SD: 16.8%), *Streptococcus* in

167 the "Only Early Antibiotics" group (mean: 25.36%; SD: 7.19%), and *Staphylococcus* in the
168 "Other Antibiotics" group (mean: 24.69%; SD: 14.60%) (Figure 2D). At time point 1, *Serratia*
169 had the highest mean relative abundance (45.72%), followed by *Streptococcus* (14.85%) and
170 *Staphylococcus* (12.68%). *Staphylococcus* and *Streptococcus* were the most prevalent genera,
171 detected in over 90% of samples.

172 *Streptococcus mitis* and *Staphylococcus epidermidis* were the most prevalent species, identified
173 in over 89% of the samples. The most abundant species were *Serratia marcescens* (mean:
174 23.76%; SD: 9.66%), followed by *S. epidermidis* (mean: 18.48%; SD: 5.51%), *S. mitis* (mean:
175 14.13%; SD: 7.04%), *Gemella haemolysans* (mean: 6.56%; SD: 7.04%), and *Staphylococcus*
176 *aureus* (mean: 6.44%; SD: 4.21%) (Figure 2E). However, no species exhibited dominance
177 (>50% relative abundance) in more than 10% of samples, irrespective of antibiotic exposure.
178 *S. epidermidis* increased in relative abundance until time point 3, after which *S. aureus* became
179 the dominant *Staphylococcus* species. *S. mitis* remained the most abundant *Streptococcus*
180 species across all time points (Figure 2E). At six months corrected age (time point 6), the
181 microbiome composition was discernible from earlier time points. In the "Only Early
182 Antibiotics" (mean: 18.67%) and "Other Antibiotics" (mean: 15.88%) groups, *Dolosigranulum*
183 *pigrum*, often associated with health, emerged as one of the dominant species, while *Moraxella*
184 *catarrhalis*, linked to respiratory health risks, predominated in the naive group (mean: 25.28%)
185 (Figure 2F). Both *D. pigrum* and *M. catarrhalis* had the highest mean relative abundances at
186 time point 6 across all groups. Additionally, *S. aureus* accounted for less than 0.5% to the total
187 mean relative abundance in groups dominated by *M. catarrhalis* at 6 months corrected age.

188 **Postnatal antibiotic exposure affects the nasopharyngeal microbiome diversity,**
189 **composition and stability**

190 Metagenomic analysis revealed a gradual increase in species-level α-diversity (measured using
191 Shannon index) over time (time point: 1-4) in the nasopharyngeal microbiome before
192 plateauing across infants (time point: 4-6) (Figure 3A). To determine which clinical variables
193 (in Supplementary Table E6) were associated with microbiota diversity (Shannon), we used a
194 GLMM approach, including subject as a random effect to account for individual variation. No
195 clinical variable, other than age (Tukey HSD post hoc test: $p.adj < 0.05$), was significantly
196 associated with infant nasopharyngeal microbiome diversity. When stratified by postnatal
197 antibiotic exposure, we observed a similar trend of increasing α-diversity over time, stabilizing
198 at later time points in the "Other Antibiotics" and "Naive" groups. However, in the "Only Early
199 Antibiotics" group, we noticed significantly lower α-diversity at time points (time point 2 and
200 3) mostly coinciding with the discontinuation of ampicillin and gentamicin (Tukey HSD post
201 hoc test: $p.adj < 0.05$; Figure 3B). These changes remained significant even after correcting for
202 age, suggesting a potential effect of early antibiotic administration on the nasopharyngeal
203 microbiome.

204 Next, we performed β-diversity analysis to assess the impact of postnatal antibiotic use on the
205 overall nasopharyngeal microbial community composition at each time point. Here, we
206 explored the effect of early, late and after discharge antibiotics separately. At time point 1, no
207 significant differences were detected in the composition between treatment groups
208 (PERMANOVA-test: $R^2 = 4.33\%$, $p.adj = 0.189$; Figure S3A). The effect of early antibiotics
209 was most pronounced in the first samples collected after the cessation of early antibiotic
210 treatment *i.e.,* time point 2 ($R^2 = 2.74\%$, $p.adj = 0.016$; Figure 3C), with significant differences
211 observed in the composition of infants exposed to "Only Early antibiotics" compared to
212 antibiotic-naive infants ($R^2 = 4.42\%$, $p.adj = 0.018$; Figure S3B). Further, our results from
213 differential abundance analysis revealed a significant reduction in *Staphylococcus argenteus*
214 and *Staphylococcus aureus* in the "Only Early Antibiotics" group compared to the "Naive"

group (ANCOM-BC2: *p.adj* < 0.05), supporting the observed compositional differences. However, this effect of early antibiotics was transient and was not evident at any subsequent time points up to 6 months corrected age (PERMANOVA: all *p.adj* > 0.05; Figure S3C-F).

When examining other postnatal antibiotic regimens, we found a minor but significant effect of late antibiotics following treatment at time point 3 (R^2 = 2.91%, *p.adj* = 0.014; Figure 3D). On the other hand, antibiotics administered after discharge from the NICU did not significantly affect the microbiome composition at six months corrected age (R^2 = 2%, *p.adj* = 0.262).

We also investigated whether microbial community stability differed among treatment groups over time. Community stability, measured by Aitchison dissimilarity between consecutive time points within each infant, was highest in the "Naive" group (mean dissimilarity = 39.4), which showed significantly greater temporal stability compared to the "Other Antibiotics" group (mean dissimilarity = 43.5) over the study period (pairwise Wilcoxon test: *p.adj* = 0.003; Figure 3E, F). Across all groups, the lowest stability (highest dissimilarity) was observed between time points 5–6, potentially driven by a combination of changing environmental, infant-related, perinatal, and healthcare-associated factors. During the antibiotic discontinuation period (time points 2–3), infants in the "Only Early Antibiotics" group showed a tendency toward greater stability compared to the "Naïve" group (mean dissimilarity = 36.6 vs. 38.7), though the difference was not statistically significant (pairwise Wilcoxon test: *p.adj* = 1; Figure 2F). Notably, significant differences in stability between time points 5–6 and earlier post-antibiotic periods were observed only in this group (pairwise Wilcoxon test, *p.adj* < 0.05; Figure S4). This suggests that early antibiotic exposure drives microbial communities to become more similar across individuals, reflecting a homogenizing effect on nasopharyngeal microbiota development.

Factors influencing the nasopharyngeal microbiota development in preterm infants

239 To identify drivers of infant nasopharyngeal microbiota composition, we used PERMANOVA
240 tests (all *p.adj* ≤ 0.001) based on Aitchison distance and visualized the results using Principal
241 component analysis (PCA) of CLR-transformed data from samples collected from preterm birth
242 to 6 months corrected age (~7-10 months chronological age). Inter-individual variation had a
243 pronounced effect on nasopharyngeal microbiota composition, explaining 33.08% of the
244 community variation (PERMANOVA: R^2 = 33.08%, *p.adj* = 0.001, Figure S5A), highlighting
245 the highly individualized nature of longitudinal microbiota changes and underscoring the
246 importance of accounting for inter-individual variation in microbiome studies. Among clinical
247 variables, postnatal age was significantly associated with overall nasopharyngeal microbial
248 community composition across the infant cohort. Postmenstrual age (PMA) exerted a greater
249 influence on overall microbiota composition (R^2 = 7.7%, *p.adj* = 0.001; Figure S5B) compared
250 to chronological age, represented by sampling time points (R^2 = 6.33%, *p.adj* = 0.001; Figure
251 4A), and days of life (DOL) (R^2 = 2.75%, *p.adj* = 0.001; Figure S5C), while adjusting for inter-
252 individual variation.

253 Mode of birth (R^2 = 0.36%, *p.adj* = 1) and antibiotic treatment (during pregnancy, birth, and
254 postnatally), including its duration, did not significantly influence microbiota composition
255 (Supplementary Table E3). Additionally, baseline variables that differed between the three
256 antibiotic exposure groups, such as gestational age (R^2 = 1.97%, *p.adj* = 1), birth weight (R^2 =
257 1.31%, *p.adj* = 1), and duration of ruptured membranes (R^2 = 0.30%, *p.adj* = 1), were not
258 significantly associated with microbial community composition during the first ~7-10 months
259 of life (Supplementary Table E3).

260 To further explore whether infant nasopharyngeal microbiota maturation is ubiquitously
261 associated with age, we employed the Random Forests machine learning algorithm to regress
262 microbial species relative abundances against the chronological age of preterm infants, as

263 previously described [36]. A total of 24 taxa were identified as the minimum set required for
264 accurate age prediction (Figure 4B). Using these 24 most informative age-discriminatory taxa,
265 we trained, refined, and validated a model to characterize nasopharyngeal microbiota
266 development in preterm infants. The taxa most strongly associated with age in the
267 nasopharyngeal microbiota were *Dolosigranulum pigrum*, *Granulicatella elegans* and
268 *Moraxella catarrhalis* (Figure 4C). This finalized sparse model was then applied to predict
269 infant chronological age based on the relative abundances of these 24 taxa, producing a
270 microbiota-derived predicted age that reflects microbiota maturity. A strong linear relationship
271 was observed between microbiota age and chronological age of preterm infants (Pearson: R =
272 0.97, $p < 0.0001$, Figure 4D), confirming the model's accuracy in predicting preterm infant age.
273 However, in infants exposed to postnatal antibiotics, predicted microbiota age showed a non-
274 significant tendency to lag behind chronological age at 6 months corrected age (Figure 4D).

275 **Developmental trajectory of the infant nasopharynx microbiota in early life based on**
276 **community types**

277 Using the Dirichlet multinomial mixture (DMM) method, we identified four optimal
278 community type clusters (CT1-4) in the nasopharyngeal microbiota of all preterm infants based
279 on its composition (Figure S6A). These clusters accounted for 45.2% (CT1), 38.7% (CT2),
280 8.6% (CT3), and 7.5% (CT4) of all the samples (Figure 5D, E). These community types were
281 mainly characterized by variations in the relative abundances of four bacterial genera:
282 *Streptococcus*, *Serratia*, *Staphylococcus*, and *Gemella* (Figure 5B, Figure S6B). CT1 and CT4
283 were predominated by *Serratia* and *Staphylococcus*, respectively, while CT2 and CT3 exhibited
284 higher abundances of both *Streptococcus* and *Gemella* (Figure 5B). Notably, CT2 had a
285 significantly higher relative abundance of *Staphylococcus* compared to CT3. These genera
286 dominated their respective community types with mean relative abundances of 51.8% for

287 *Serratia* (CT1), 40.16% (CT2) and 35.27% (CT3) for *Streptococcus*, 14.4% (CT2) and 11.5%
288 (CT3) for *Gemella*, and 21.92% for *Staphylococcus* (CT4).

289 We also found significant differences in microbial diversity (α-diversity, pairwise Wilcoxon
290 test: all *p.adj* < 0.05; Figure 5C) and overall composition (β-diversity, PERMANOVA: R^2 =
291 12.61%, *p.adj* = 0.001; Figure 5A) between all the identified community types. Moreover,
292 sequencing depth did not differ significantly between community types (pairwise Wilcoxon
293 test: *p.adj* > 0.05; Figure S6C), suggesting that it was unlikely to influence their separation.
294 Sampling time point was significantly associated with community type clustering in the "Only
295 Early Antibiotics" and Naive groups (Fisher's exact test: *p.adj* < 0.05; Figure 5D), while no
296 such association was observed in the "Other Antibiotics" group (Fisher's exact test: *p.adj* =
297 0.15; Figure 5D). Additionally, there was no significant difference in community type
298 clustering when stratified according to antibiotic exposure group (Fisher's exact test: *p.adj* =
299 0.44) (Figure 5E).

300 Next, we performed a Markov chain-based analysis to quantify the likelihood of transitions
301 between community types in the early-life nasopharyngeal microbiota of preterm infants. The
302 modeling revealed distinct transition patterns across treatment groups, with CT1 (*Serratia*-
303 driven) emerging as one of the most stable (*i.e.,* higher self-transition probability) states (mean
304 prob. = 0.53), followed by CT2 (0.47) (Figure 5F). Community types were most stable in the
305 naive group (mean prob. = 0.5) compared to the other groups. However, CT3 and CT4 in the
306 naive group transitioned more frequently or equally toward CT2 (0.6 and 0.5, respectively)
307 compared to their self-transition probabilities (Figure 5F). In contrast, in the "Only Early
308 Antibiotics" group, CT3 (0.6) and CT4 (0.67) transitioned predominantly to CT1, a community
309 type with the lowest species diversity, underlining our earlier observation that early antibiotic
310 exposure reduces microbial diversity. This may explain the disappearance of CT3 and CT4

311 following treatment at time points 2 and 3 (Figure 5D). On the other hand, the "Other
312 Antibiotics" group exhibited lower stability in dominant communities, such as CT1, and more
313 frequent transitions between states (Figure 5F), resulting in greater dynamism. This variability
314 may reflect the influence of different antibiotic regimens within this group. Overall, these
315 findings highlight the pronounced effects of postnatal antibiotic exposures on early-life
316 microbial community dynamics, reducing overall community stability by making transient
317 states less stable and promoting more frequent transitions across communities.

318 We further leveraged metagenomics sequencing data to investigate functional differences
319 among community types by characterizing their antibiotic resistance (resistome) profiles [27].
320 Striking differences were observed between the community types. The *Streptococcus*-
321 dominated community type (CT2) exhibited a significantly higher AMR load, reflected in both
322 total ARG abundance and ARG diversity (Shannon index) (pairwise Wilcoxon test: $p.adj <$
323 0.05; Figure 5G). These findings suggest that *Streptococcus* plays a key role in driving
324 antibiotic resistance within the infant nasopharyngeal microbiome.

325 **Prolonged signature of an outbreak on nasopharyngeal microbiota in preterm infants**

326 A few months into our study, a clinical outbreak with *S. marcescens* was detected in the NICU.
327 Six infants, including three from our study cohort, tested positive for *S. marcescens* in various
328 clinical samples. After two screenings conducted seven days apart on all hospitalized infants,
329 the second screening found no positive nasopharyngeal cultures, and the outbreak was declared
330 over (Figure S7A). Environmental samples were negative, and no source of the outbreak was
331 identified. However, metagenomic analysis of nasopharyngeal aspirate samples revealed
332 persistent carriage of *S. marcescens* in infants hospitalized after the outbreak (Figure 6A).
333 Notably, *S. marcescens*, which was detected early after birth (time point 1), persisted in over
334 90% of preterm infants across all treatment groups and remained detectable at ~7–10 months

335 of chronological age (time point 6), extending well beyond hospital discharge (Figure S7B).
336 Moreover, no significant differences in the abundance of *S. marcescens* were detected among
337 treatment groups at any time point (Figure 6B).

338 Further, our analysis revealed that *S. marcescens* emerged as the dominant species, accounting
339 for the most variation in overall microbiota composition (PERMANOVA: $R^2 = 6.76\%$, *p.adj* =
340 0.001; Figure 6C). *S. marcescens* showed a significantly higher abundance in samples collected
341 from infants hospitalized after the outbreak compared to those discharged or transferred to a
342 different NICU prior to it (pairwise Wilcoxon test: *p.adj* < 0.05; Figure 6D). Although
343 hospitalization status did not significantly influence the overall nasopharyngeal microbiota
344 composition in preterm infants (PERMANOVA: $R^2 = 1.46\%$, p.adj = 0.824; Figure 6E), our
345 cross-sectional analysis revealed significant differences in composition at the first two sampling
346 time points (time points 1 and 2), with these differences becoming most pronounced by six
347 months corrected age (time point 6, $R^2 = 4.91\%$, p.adj = 0.001; Figure S7C).

348 To further analyze the persistent effects of the outbreak, we used a supervised machine learning
349 algorithm, Random Forest, to predict the presence or absence of *S. marcescens* at six months
350 corrected age (time point 6) based on microbiome composition in early life (time points 1 and
351 2). Our model demonstrated high performance, achieving 94.74% and 100% accuracy at time
352 point 1 and 2, respectively, in classifying *S. marcescens*-positive and *S. marcescens*-negative
353 groups (Figure S8). These results indicate that the nasopharyngeal microbiome composition in
354 early life can predict the persistence of *S. marcescens* colonization at six months corrected age.
355 Overall, these findings suggest that *S. marcescens* establishes a prolonged microbial signature,
356 likely reflective of its acquisition during the NICU outbreak.

357 **DISCUSSION**

358 The nasopharyngeal microbiota is a critical reservoir for microbes associated with respiratory
359 tract infections, influencing immune development and disease susceptibility in infants [10, 14, 17,
360 37]. This study provides an unprecedented high-resolution analysis of the longitudinal
361 development of the preterm nasopharyngeal microbiota using deep whole metagenome
362 sequencing. We demonstrated that the nasopharyngeal microbiota development from birth to 6
363 months corrected age was characterized by dynamic and age-structured communities, with
364 highly individualized patterns influenced by postnatal antibiotic exposure and hospitalization.
365 Early-life antibiotic exposure (ampicillin + gentamicin) had transient but pronounced
366 ecological effects on microbial diversity, composition, stability, and community type transition
367 dynamics. An unexpected *Serratia marcescens* outbreak during the study gave us a unique
368 opportunity to assess its impact on microbiota development by comparing the microbiome
369 before and after the outbreak. We found a persistent outbreak signature that, at ~7–10 months
370 chronological age (6 months corrected age), could be accurately predicted from the
371 nasopharyngeal microbiota composition in the first days of life.

372 After birth, preterm infants exhibited low bacterial abundance and diversity in the
373 nasopharyngeal microbiota, followed by a sharp increase in the first few weeks and a more
374 gradual rise for up to two months before stabilization. Similar trends have been described for
375 term as well as preterm infants [3, 7, 38, 39], though our metagenomic approach offers enhanced
376 taxonomic resolution at species-level. In addition, samples from postnatal antibiotic-exposed
377 infants were more likely to have low biomass from birth up to 6 months corrected age. Postnatal
378 age was found to be associated with infant nasopharyngeal microbiome diversity, with a
379 significant reduction observed directly upon treatment in the "Only Early antibiotics" group.
380 Infants exposed to "Other Antibiotics" had significantly lower absolute bacterial abundance at

381 two weeks after birth compared with "Only Early antibiotics" group. Overall, these trends in
382 our results suggest a general suppression of bacterial colonization and diversity, associated with
383 the discontinuation of early and/or late antibiotics in the nasopharyngeal microbiota of infants.

384 Microbiome development in early life is a dynamic process, and transformations in the
385 nasopharyngeal microbiota are expected, even in antibiotic-naive infants. In our cohort, the
386 nasopharyngeal microbiota of infants was predominantly colonized by four genera: *Serratia*,
387 *Streptococcus*, *Gemella*, and *Staphylococcus*, forming different community types. At the
388 species level, *S. marcescens*, *S. epidermidis*, *S. aureus*, *Gemella haemolysans*, and *S. mitis* were
389 commonly identified during the first two months of life. Carriage of specific bacterial species
390 in the upper respiratory tract, such as methicillin-resistant *Staphylococcus aureus* (MRSA), has
391 been closely linked to increased infection risks for newborns in the NICU [40]. Of particular
392 concern are pathogens identified by the World Health Organization (WHO) as critical priorities
393 due to their potential to cause severe infections and resistance to multiple antibiotics [41]. In our
394 study, among the 20 most abundant species, we identified several on the WHO priority list
395 pathogens, including *Serratia marcescens* that became prominent during the outbreak,
396 alongside *Klebsiella pneumoniae, Staphylococcus aureus, Streptococcus pneumoniae,* and
397 additional species from the Enterobacteriaceae family.

398 Because shotgun metagenomics enables higher taxonomic resolution than 16S rRNA studies,
399 we were able to distinguish highly relevant pathogens from species generally considered
400 commensals in immunocompetent individuals. For example, most streptococci identified
401 belonged to *Streptococcus mitis* across all ages. *S. mitis* is a close relative of *S. pneumoniae* and
402 is considered a commensal that may provide protection against its pathogenic counterpart [42]. In
403 contrast, *S. pneumoniae*, a human colonizer with high pathogenic potential, appeared at lower
404 abundances, reaching its highest proportion at 6 months corrected age. This timing coincides

405 with when *S. pneumoniae* typically becomes more common in the nasopharynx of infants [43].
406 Among the staphylococci, we could distinctly identify *S. epidermidis*, initially more prevalent,
407 and *S. aureus*, a pathogen of greater concern, which became the dominant species in this genus
408 after approximately 3 weeks of life. Expanding our understanding of such species-specific
409 dynamics is relevant as it may help guiding future preventative strategies.

410 By six months corrected age, the nasopharyngeal microbiome shifted significantly,
411 characterized by the dominance of *Dolosigranulum pigrum* and *Moraxella catarrhalis*,
412 alongside the presence of some oral bacteria such as *Neisseria flavescens* and *Veillonella spp*.
413 While *Corynebacterium* and *Dolosigranulum* are often associated with respiratory health
414 benefits [3, 6, 44], their scarcity in our cohort may reflect the unique conditions of preterm birth
415 and the impact of an outbreak. Instead, we observed a high prevalence of *Moraxella* across all
416 time points, with 16% of samples at 6 months corrected age dominated by this genus, and no
417 significant differences between antibiotic exposure groups. Although early *Moraxella*
418 colonization and low abundances of *Corynebacterium* and *Dolosigranulum* may predispose
419 infants to chronic inflammation [3, 45], some studies suggest that Moraxella-dominant profiles are
420 associated with community stability and fewer respiratory infections [6]. A longer follow-up of
421 our cohort may further clarify these dynamics and their impact on future respiratory health
422 outcomes [16, 46, 47].

423 Using machine learning, we identified *Dolosigranulum pigrum* and *Moraxella catarrhalis* as
424 key age-discriminatory species in the nasopharyngeal microbiota of preterm infants, aligning
425 with 16S rRNA studies that highlight *Dolosigranulum* and *Moraxella* as significant age-
426 associated taxa [3]. Moreover, for infants who were exposed to postnatal antibiotics, the
427 development of their nasopharyngeal microbiota seemed to lag slightly behind their
428 chronological ages at 6 months corrected age, indicating a potential delay in microbiota

429 maturation due to early antibiotic exposure; however, this observation was not statistically
430 significant. This trend aligns with prior research on the gut microbiota suggesting that early-
431 life antibiotic use disrupts microbial succession, with potential immune consequences [48]. In the
432 respiratory tract, changes in maturation of the nasopharyngeal microbiota have been correlated
433 with increased risk of respiratory tract infections during the first year of life [3]. Such relationships
434 indicate that insights into the developmental dynamics of the nasopharyngeal microbiome could
435 be key in identifying factors contributing to infection susceptibility and shaping strategies to
436 reduce respiratory health risks in infancy.

437 Inter-individual variability was the most significant factor influencing nasopharyngeal
438 microbiome composition, consistent with previous findings in both term and preterm infants [3,
439 15, 49]. In addition to individual differences, postnatal age also significantly impacted microbiome
440 composition in our cohort, aligning with earlier studies [3, 7, 50]. However, in contrast to Bosch et
441 al. [7], we found no significant association between mode of delivery and nasopharyngeal
442 microbiome composition in the first year of life. Similar findings have been published for the
443 gut microbiome, with postnatal age having a greater influence than mode of delivery-both
444 recognized as known factors shaping the microbiome composition of preterm infants [36, 51].

445 Early-life antibiotic exposure has been linked to several diseases [16, 52, 53]. However, conflicting
446 results have been reported in the literature regarding the impact of antibiotic exposure on
447 nasopharyngeal microbiome composition in infants [54]. In our cohort, early antibiotic treatment
448 (ampicillin + gentamicin) had a pronounced homogenizing effect on the nasopharyngeal
449 microbiome directly following treatment, characterized by reduced bacterial diversity and
450 absolute abundance, as well as marked shifts in overall composition and microbiota stability.
451 Moreover, our statistical modeling of random ecological processes offers a fresh perspective
452 on how early antibiotic exposure can disrupt microbial colonization dynamics by promoting a

453 transition from *Staphylococcus* to a *Serratia*-dominated community. This shift may result in the
454 selection of nosocomial strains (mainly *S. marcescens*), whose resistome contains genes
455 capable of conferring resistance to the administered antibiotic classes [27]. Our differential
456 abundance analysis corroborates these findings, showing a significant decrease in
457 *Staphylococcus argenteus* and *Staphylococcus aureus* in the "Only Early Antibiotics" group
458 compared to the "Naive" group directly after treatment at 4 to 9 DOL (time point 2; ANCOM-
459 BC2: *p.adj* <0.05). In addition to early antibiotic effects, we observed minor yet significant
460 shifts in taxonomic composition following late antibiotic exposure at 12 to 19 DOL (time point
461 3). Infants in the "Other antibiotics" group, exposed to multiple classes of antibiotics (Beta-
462 lactam, Aminoglycoside, Glycopeptides, Nitroimidazole, Lincosamide), exhibited significantly
463 lower temporal microbiome stability and a higher frequency of transitions between community
464 types compared to the naive group. This suggests that exposure to multiple antibiotic classes
465 may further disrupt the natural course of microbiota development typically observed in
466 antibiotic-naive infants.

467 Although no discernible effects of postnatal antibiotics on the nasopharyngeal microbiota were
468 detected by NICU discharge or at 6 months corrected age, even transient alterations to the
469 infant's microbiome may have lasting consequences. Such changes can increase the risk of
470 diseases later in life [17], influence immune system development, and promote the selection and
471 spread of antibiotic resistance genes in both pathogenic and commensal bacteria [36]. This is
472 particularly concerning, given that leading pathogens commonly found in the nasopharyngeal
473 microbiota account for the highest global burden of deaths attributable to antimicrobial
474 resistance [55]. Implementation of stewardship actions focused on optimizing or reducing the
475 initiation of early antibiotic therapy appears promising and has been adopted in some NICUs,
476 though not yet for the smallest preterm infants [56, 57].

477 Besides antibiotics, intensive care procedures such as intubation, feeding tubes and intravenous
478 lines, may also influence colonization patterns [58]. Additionally, prolonged hospital stays expose
479 preterm infants to potentially pathogenic bacteria while their immune systems are still
480 developing [59]. The local NICU environment has been shown to influence the gut, skin
481 microbiome, and resistome of infants [60]. *S. marcescens*, an opportunistic pathogen prevalent in
482 NICUs, can cause outbreaks and nosocomial infections in preterm infants [61]. Its persistence
483 post-hospitalization and outbreak in our study underscores its role as a marker of hospital-
484 acquired colonization, potentially influencing early-life microbiome development in preterm
485 infants. This real-time outbreak likely coincided with its NICU introduction, illustrating the
486 dynamic nature of hospital-acquired infections. The detection of *S. marcescens* was strongly
487 linked to an outbreak that occurred during the study period. Although the outbreak was
488 officially declared over after a short interval, our data indicate that *S. marcescens* persisted to
489 the end of the study. Notably, previous studies do not report *Serratia* in the nasopharyngeal
490 microbiota of term infants, and in preterm infants, it has been described only in very low
491 abundance[38]. Understanding its prolonged presence is vital for assessing the long-term effects
492 of hospital-acquired dysbiosis on microbiota development.

493 A significant strength of our study lies in the serendipitous timing of the outbreak in conjunction
494 with our longitudinal study, which gave us a unique chance to analyze its effects on the
495 nasopharyngeal microbiome. In addition, the study's inclusion of a naive group adds a valuable
496 control, as comprehensive studies on preterm infants involve not seldom cohorts where all have
497 been administered antibiotics [62, 63], which is common practice in some hospitals for very and
498 extremely low-birth-weight newborns [64]. Also, the consistent administration of early
499 antibiotics, specifically ampicillin and gentamicin, ensures that comparisons are not diluted in
500 significance due to variability in treatment, as different antibiotics often have distinct impacts
501 on the microbiota [65]. Moreover, we followed recommended low-biomass research practices for

502 our nasopharyngeal samples [66], including sequencing negative and positive controls and
503 rigorous *in silico* contaminant removal to minimize contamination and bias, ensuring reliable
504 and accurate findings. Infants remained in a controlled NICU environment for most sampling
505 periods, had no daycare exposure before the final sampling, and followed a consistent
506 vaccination schedule, minimizing external influences [14, 17]. Additionally, we used
507 nasopharyngeal aspirates instead of nasal swabs to reduce contamination from the anterior
508 nostril microbiome [67]. Notably, this is the first study to track the development of the infant
509 nasopharyngeal microbiome with species-level resolution, offering detailed insights into
510 microbial dynamics.

511 However, our study has some limitations. Firstly, we focused on known factors influencing the
512 microbiome and may have overlooked other relevant variables. Notably, we did not assess
513 potential microbial sources such as hospital surfaces, healthcare personnel, or maternal
514 microbiota, which could contribute to infant nasopharyngeal microbiome development. Since
515 this is a human microbiome study, a causal relationship between antibiotic treatment and the
516 observed ecological effects cannot be fully established. Confirming causality would necessitate
517 complementary mechanistic studies, such as those involving animal models. However, our
518 findings strongly suggest a direct biological impact of antibiotics, as the microbiota
519 composition in the treatment group showed significant alterations only immediately following
520 exposure, with effects that gradually diminished over time, even after adjusting for age-related
521 changes. While there were significant differences in gestational age between infants exposed to
522 antibiotics and those who were not, our analysis did not reveal a significant association between
523 gestational age and microbiota composition. This suggests that, although gestational age may
524 be a relevant factor, it did not impart a notable effect on the microbiome outcomes observed in
525 our study. The study's single-center design may also limit the broader applicability of our
526 findings. Additionally, concurrent processing of each infant's samples (all samples in the same

527 batch) may have reinforced individual microbiome signatures. Despite equimolar library
528 pooling prior to WMS, variation in raw and bacterial reads was observed. Nonetheless, we
529 rigorously accounted for inter-individual variation and employed analytical methods that
530 mitigate sensitivity to sequencing depth while addressing the compositional nature of
531 microbiome data [68, 69], ensuring the reliability and robustness of our findings.

532 Overall, this study highlights the significant influences of age, outbreak events, and postnatal
533 antibiotic exposure on the developmental dynamics of the nasopharyngeal microbiota in
534 preterm infants. The transient alterations in microbial diversity, composition and stability
535 observed in response to antibiotic exposure, along with the enduring impact of microbial
536 signatures from outbreak events, underscore the importance of understanding these factors in
537 shaping the microbiome in early life. Preterm infants are exposed to multiple factors known to
538 shape the microbiota across various body niches [70 36 71 72]. There is also growing evidence that
539 respiratory health outcomes may be influenced by the intestinal and oral microbiota [9, 15, 73].
540 Future studies are warranted to investigate the interconnected colonization dynamics of these
541 body niches during infancy. Additionally, ongoing efforts to develop strategies for restoring
542 microbial balance and minimizing unnecessary antibiotic use during this critical phase of
543 development represent a valuable pathway for preventing disturbances in microbiota-host
544 interactions [74, 75], which may have potential implications for health.

545 **Ethical statement**

546 The study was performed in accordance with the Declaration of Helsinki and approved by the
547 Hospital's Data Protection Officer and the Regional Committee for Medical and Health
548 Research Ethics-South East, Norway (2018/1381 REKD).

549 **Consent statement**

550 Written informed consent was obtained from the infant's parents. The participants received no
551 compensation.

552 **Data availability**

553 The metagenomic sequencing data (clean reads *i.e.,* after quality filtering and removal of human
554 DNA) generated from all nasopharyngeal samples are publicly available on the NCBI Sequence
555 Read Archive database under BioProject ID: PRJNA1009231.

556 **Acknowledgments**

557 We would like to thank the parents of participating infants, and the clinical staff at the Neonatal
558 Intensive Care units at Ullevål hospital for their assistance with sample collection. The
559 sequencing service for this work was provided by the Norwegian Sequencing Centre
560 (www.sequencing.uio.no) and the computations were performed on resources provided by
561 Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in
562 Norway.

563 **Competing interests**

564 The authors declare that the research was conducted in the absence of any commercial or
565 financial relationships that could be construed as a potential conflict of interest.

566 **Author contributions**

FCP and KH conceived the study. All authors contributed to the design of the study and interpretation of data. PR and KH included the infants, collected samples and patient metadata. PR, GS and HAA processed the samples under FCP supervision. AD analysed shotgun metagenomic data. AD and PR carried out statistical data analysis and created the figures. PR and AD wrote the manuscript. All authors discussed the results, critically revised the manuscript, and agreed to the published version of the manuscript.

Funding

This work was primarily funded by the Research Council of Norway (RCN) project numbers 273833 and 322375, and by the Olav Thon Foundation, with additional support from the Faculty of Dentistry at the University of Oslo and by Oslo University Hospital.

578 **References**

579 1. Gallacher, D.J. & Kotecha, S. Respiratory Microbiome of New-Born Infants. *Front Pediatr* **4**
580 (2016).
581 2. Staude, B. et al. The Microbiome and Preterm Birth: A Change in Paradigm with Profound
582 Implications for Pathophysiologic Concepts and Novel Therapeutic Strategies. *BioMed*
583 *research international* **2018**, 7218187 (2018).
584 3. Bosch, A. et al. Maturation of the Infant Respiratory Microbiota, Environmental Drivers, and
585 Health Consequences. A Prospective Cohort Study. *American journal of respiratory and*
586 *critical care medicine* **196**, 1582-1590 (2017).
587 4. Groer, M.W. et al. Development of the preterm infant gut microbiome: a research priority.
588 *Microbiome* **2**, 38 (2014).
589 5. Selway, C.A., Collins, C.T., Makrides, M., Sullivan, T.R. & Weyrich, L.S. Variable preterm oral
590 microbiome stabilizes and reflects a full-term infant profile within three months. *Pediatric*
591 *research* (2023).
592 6. Biesbroek, G. et al. Early respiratory microbiota composition determines bacterial succession
593 patterns and respiratory health in children. *Am J Respir Crit Care Med* **190**, 1283-1292 (2014).
594 7. Bosch, A. et al. Development of Upper Respiratory Tract Microbiota in Infancy is Affected by
595 Mode of Delivery. *EBioMedicine* **9**, 336-345 (2016).
596 8. Dominguez-Bello, M.G. et al. Delivery mode shapes the acquisition and structure of the initial
597 microbiota across multiple body habitats in newborns. *Proceedings of the National Academy*
598 *of Sciences of the United States of America* **107**, 11971-11975 (2010).
599 9. Biesbroek, G. et al. The impact of breastfeeding on nasopharyngeal microbial communities in
600 infants. *American journal of respiratory and critical care medicine* **190**, 298-308 (2014).
601 10. Flynn, M. & Dooley, J. The microbiome of the nasopharynx. *J Med Microbiol* **70** (2021).
602 11. Gollwitzer, E.S. et al. Lung microbiota promotes tolerance to allergens in neonates via PD-L1.
603 *Nat Med* **20**, 642-647 (2014).
604 12. Sharma, A.A., Jen, R., Butler, A. & Lavoie, P.M. The developing human preterm neonatal
605 immune system: a case for more research in this area. *Clin Immunol* **145**, 61-68 (2012).
606 13. Liu, C. et al. Microbial dysbiosis and childhood asthma development: Integrated role of the
607 airway and gut microbiome, environmental exposures, and host metabolic and immune
608 response. *Frontiers in immunology* **13**, 1028209 (2022).
609 14. Teo, S.M. et al. The infant nasopharyngeal microbiome impacts severity of lower respiratory
610 infection and risk of asthma development. *Cell Host Microbe* **17**, 704-715 (2015).
611 15. Binia, A. et al. The Influence of FUT2 and FUT3 Polymorphisms and Nasopharyngeal
612 Microbiome on Respiratory Infections in Breastfed Bangladeshi Infants from the Microbiota
613 and Health Study. *mSphere* **6**, e0068621 (2021).
614 16. Toivonen, L. et al. Antibiotic Treatments During Infancy, Changes in Nasal Microbiota, and
615 Asthma Development: Population-based Cohort Study. *Clinical Infectious Diseases* **72**, 1546-
616 1554 (2021).
617 17. Man, W.H., de Steenhuijsen Piters, W.A. & Bogaert, D. The microbiota of the respiratory
618 tract: gatekeeper to respiratory health. *Nature reviews. Microbiology* **15**, 259-270 (2017).
619 18. Kristensen, M. et al. The respiratory microbiome is linked to the severity of RSV infections
620 and the persistence of symptoms in children. *Cell Reports Medicine* **5** (2024).
621 19. Zelasko, S. et al. Early-life upper airway microbiota are associated with decreased lower
622 respiratory tract infections. *Journal of Allergy and Clinical Immunology* **155**, 436-450 (2025).
623 20. Perin, J. et al. Global, regional, and national causes of under-5 mortality in 2000-19: an
624 updated systematic analysis with implications for the Sustainable Development Goals. *Lancet*
625 *Child Adolesc Health* **6**, 106-115 (2022).
626 21. Tracy, M.C. & Cornfield, D.N. Bronchopulmonary Dysplasia: Then, Now, and Next. *Pediatr*
627 *Allergy Immunol Pulmonol* **33**, 99-109 (2020).

628 22. Harju, M. et al. The burden of childhood asthma and late preterm and early term births. *J*
629 *Pediatr* **164**, 295-299.e291 (2014).
630 23. Priante, E. et al. Respiratory Outcome after Preterm Birth: A Long and Difficult Journey. *Am J*
631 *Perinatol* **33**, 1040-1042 (2016).
632 24. Costello, E.K., Carlisle, E.M., Bik, E.M., Morowitz, M.J. & Relman, D.A. Microbiome assembly
633 across multiple body sites in low-birthweight infants. *mBio* **4**, e00782-00713 (2013).
634 25. Vatne, A. et al. Early Empirical Antibiotics and Adverse Clinical Outcomes in Infants Born Very
635 Preterm: A Population-Based Cohort. *J Pediatr* **253**, 107-114.e105 (2023).
636 26. O'Connor, L. & Heyderman, R. The challenges of defining the human nasopharyngeal
637 resistome. *Trends in Microbiology* **31**, 816-831 (2023).
638 27. Dhariwal, A. et al. Prolonged hospitalization signature and early antibiotic effects on the
639 nasopharyngeal resistome in preterm infants. *Nature Communications* **15**, 6024 (2024).
640 28. Rajar, P. et al. Microbial DNA extraction of high-host content and low biomass samples:
641 Optimized protocol for nasopharynx metagenomic studies. *Frontiers in microbiology* **13**,
642 1038120 (2022).
643 29. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data.
644 Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
645 30. Krueger, F. "Trim Galore: a wrapper tool around Cutadapt and FastQC." Trim Galore (2012).
646 31. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
647 357-359 (2012).
648 32. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse
649 microbial communities with bioBakery 3. *Elife* **10**, e65088 (2021).
650 33. Allaire, J. RStudio: integrated development environment for R. *Boston, MA* **770**, 165-171
651 (2012).
652 34. R Core Team, R. R: A language and environment for statistical computing. (2013).
653 35. Wickham, H. ggplot2. *Wiley interdisciplinary reviews: computational statistics* **3**, 180-185
654 (2011).
655 36. Gasparrini, A.J. et al. Persistent metagenomic signatures of early-life hospitalization and
656 antibiotic treatment in the infant gut microbiota and resistome. *Nat Microbiol* **4**, 2285-2297
657 (2019).
658 37. Organization, W.H. in Executive summary: ending preventable child deaths from pneumonia
659 and diarrhoea by 2025: the integrated global action plan for pneumonia and diarrhoea
660 (GAPPD) (2013).
661 38. Gallacher, D. et al. Dissimilarity of the gut-lung axis and dysbiosis of the lower airways in
662 ventilated preterm infants. *The European respiratory journal* **55**, 1901909 (2020).
663 39. Lauer, T. et al. Bacterial Colonization within the First Six Weeks of Life and Pulmonary
664 Outcome in Preterm Infants <1000 g. *J Clin Med* **9** (2020).
665 40. Nurjadi, D. et al. Surveillance for colonization, transmission, and infection with methicillin-
666 susceptible Staphylococcus aureus in a neonatal intensive care unit. *JAMA network open* **4**,
667 e2124938-e2124938 (2021).
668 41. Organization, W.H. WHO bacterial priority pathogens list, 2024: bacterial pathogens of public
669 health importance, to guide research, development, and strategies to prevent and control
670 antimicrobial resistance. (World Health Organization, 2024).
671 42. Shekhar, S. et al. Antibodies reactive to commensal Streptococcus mitis show cross-reactivity
672 with virulent Streptococcus pneumoniae serotypes. *Frontiers in Immunology* **9**, 747 (2018).
673 43. Lebon, A. et al. The inverse correlation between Staphylococcus aureus and Streptococcus
674 pneumoniae colonization in infants is not explained by differences in serum antibody levels
675 in the Generation R Study. *Clinical and Vaccine Immunology* **18**, 180-183 (2011).
676 44. de Steenhuijsen Piters, W.A. et al. Early-life viral infections are associated with
677 disadvantageous immune and microbiota profiles and recurrent respiratory infections.
678 *Nature Microbiology* **7**, 224-237 (2022).

45. Følsgaard, N.V. et al. Pathogenic bacteria colonizing the airways in asymptomatic neonates stimulates topical inflammatory mediator release. *Am J Respir Crit Care Med* **187**, 589-595 (2013).
46. Man, W.H. et al. Infant respiratory syncytial virus prophylaxis and nasopharyngeal microbiota until 6 years of life: a subanalysis of the MAKI randomised controlled trial. *Lancet Respir Med* **8**, 1022-1031 (2020).
47. Zeng, Y. & Liang, J.Q. Nasal Microbiome and Its Interaction with the Host in Childhood Asthma. *Cells* **11** (2022).
48. Bokulich, N.A. et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Science translational medicine* **8**, 343ra382-343ra382 (2016).
49. Perez, G.F. et al. Nasopharyngeal microbiome in premature infants and stability during rhinovirus infection. *J Investig Med* **65**, 984-990 (2017).
50. Reyman, M. et al. Microbial community networks across body sites are associated with susceptibility to respiratory infections in infants. *Communications biology* **4**, 1233 (2021).
51. Aguilar-Lopez, M., Dinsmoor, A.M., Ho, T.T.B. & Donovan, S.M. A systematic review of the factors influencing microbial colonization of the preterm infant gut. *Gut Microbes* **13**, 1-33 (2021).
52. Kelderer, F. et al. Associations between pre- and postnatal antibiotic exposures and early allergic outcomes: A population-based birth cohort study. *Pediatr Allergy Immunol* **33**, e13848 (2022).
53. Lu, Y. et al. Early-Life Antibiotic Exposure and Childhood Asthma Trajectories: A National Population-Based Birth Cohort. *Antibiotics* **12**, 314 (2023).
54. Zimmermann, P. Exploring the microbial landscape of the nasopharynx in children: a systematic review of studies using next generation sequencing. *Frontiers in Microbiomes* **2**, 1231271 (2023).
55. Murray, C.J. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The lancet* **399**, 629-655 (2022).
56. Salas Garcia, S. et al. Early-onset sepsis: a cohort study comparing serial clinical observation with laboratory testing. *J Perinatol* **41**, 2482-2487 (2021).
57. Vatne, A. et al. Reduced Antibiotic Exposure by Serial Physical Examinations in Term Neonates at Risk of Early-onset Sepsis. *Pediatr Infect Dis J* **39**, 438-443 (2020).
58. Parker, L.A. et al. Neonatal Feeding Tube Colonization and the Potential Effect on Infant Health: A Review. *Frontiers in Nutrition* **9** (2022).
59. Hartz, L.E., Bradshaw, W. & Brandon, D.H. Potential NICU Environmental Influences on the Neonate's Microbiome: A Systematic Review. *Adv Neonatal Care* **15**, 324-335 (2015).
60. Hourigan, S.K. et al. Comparison of Infant Gut and Skin Microbiota, Resistome and Virulome Between Neonatal Intensive Care Unit (NICU) Environments. *Frontiers in Microbiology* **9** (2018).
61. Cristina, M.L., Sartini, M. & Spagnolo, A.M. Serratia marcescens infections in neonatal intensive care units (NICUs). *International journal of environmental research and public health* **16**, 610 (2019).
62. Gibson, M.K. et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nature microbiology* **1**, 1-10 (2016).
63. Thänert, R., Sawhney, S.S., Schwartz, D.J. & Dantas, G. The resistance within: antibiotic disruption of the gut microbiome and resistome dynamics in infancy. *Cell host & microbe* **30**, 675-683 (2022).
64. Flannery, D.D. et al. Temporal trends and center variation in early antibiotic use among premature infants. *JAMA network open* **1**, e180164-e180164 (2018).
65. Schwartz, D.J., Langdon, A.E. & Dantas, G. Understanding the impact of antibiotic perturbation on the human microbiome. *Genome medicine* **12**, 1-12 (2020).
66. Kennedy, K.M. et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature* **613**, 639-649 (2023).

731 67. Luna, P.N. et al. The association between anterior nares and nasopharyngeal microbiota in
732 infants hospitalized for bronchiolitis. *Microbiome* **6**, 2 (2018).
733 68. Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V. & Egozcue, J.J. Microbiome Datasets Are
734 Compositional: And This Is Not Optional. *Frontiers in Microbiology* **8** (2017).
735 69. Knight, R. et al. Best practices for analysing microbiomes. *Nature Reviews Microbiology* **16**,
736 410-422 (2018).
737 70. Brooks, B. et al. Microbes in the neonatal intensive care unit resemble those found in the gut
738 of premature infants. *Microbiome* **2**, 1 (2014).
739 71. Ferretti, P. et al. Mother-to-infant microbial transmission from different body sites shapes
740 the developing infant gut microbiome. *Cell host & microbe* **24**, 133-145. e135 (2018).
741 72. Sampaio-Maia, B. & Monteiro-Silva, F. Acquisition and maturation of oral microbiome
742 throughout childhood: An update. *Dental research journal* **11**, 291 (2014).
743 73. Luoto, R. et al. Prebiotic and probiotic supplementation prevents rhinovirus infections in
744 preterm infants: a randomized, placebo-controlled trial. *The Journal of allergy and clinical*
745 *immunology* **133**, 405-413 (2014).
746 74. Kang, H.M. & Kang, J.H. Effects of nasopharyngeal microbiota in respiratory infections and
747 allergies. *Clinical and Experimental Pediatrics* **64**, 543 (2021).
748 75. Shekhar, S., Schwarzer, M., Dhariwal, A. & Petersen, F.C. Nasal microbiota transplantation: a
749 gateway to novel treatments. *Trends in Microbiology* (2025).

750

751

752 **Figure Legends**

753 **Figure 1. Patient inclusion and sample processing.** (**A**) Flowchart describing patient
754 inclusion process and availability of infants from birth until last follow up at six months
755 corrected age. (**B**) Overview of obtained nasopharyngeal aspirate samples and excluded
756 samples prior to main microbiome analysis. WMS: whole metagenomic sequencing.

757 **Figure 2. Taxonomic composition of the nasopharyngeal microbiome. (A)** Boxplot showing
758 the total bacterial DNA (16S rRNA, ng), as measured by qPCR, for all infants across sampling
759 time points spanning from birth to six months corrected age. Significant differences between
760 time points were determined using a GLMM model that accounted for individual variation and
761 age, followed by two-sided Tukey's HSD post hoc tests. **(B)** Comparison of bacterial load at
762 each sampling time point among different antibiotic (AB) exposure groups, with differences
763 tested using pairwise Wilcoxon rank-sum tests. In panels **(A)** and **(B)**, the central horizontal
764 line in each boxplot represents the median, the box indicates the interquartile range (IQR), and
765 the whiskers extend to 1.5× IQR. P-values were corrected using the Benjamini–Hochberg (BH)
766 method (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$). **(C–F)** Stacked bar plots
767 displaying the mean relative bacterial abundance of the 20 most abundant taxa across sampling
768 time points and AB exposure groups at the genus **(C–D)** and species **(E–F)** levels. Panels **(C)**
769 and **(E)** illustrate the mean relative bacterial abundance by sampling time point (x-axis), while
770 panels **(D)** and **(F)** present the data stratified by both sampling time point and postnatal AB
771 exposure group. Low abundant taxa (genus and species) are grouped into the "Others" category.
772 GLMM: generalized linear mixed-effects model.

773 **Figure 3: Effect of postnatal antibiotics on the nasopharyngeal microbial diversity,**
774 **composition and stability. (A, B)** Boxplots comparing species α-diversity (measured using the
775 Shannon index) across sampling time points for all nasopharyngeal samples **(A)** and stratified

776 by postnatal antibiotic exposure group **(B)**. In these boxplots, the lower and upper horizontal
777 line indicate the 25th and 75th percentiles, respectively, with the central line showing the
778 median. Whiskers extend to the most extreme values within 1.5 times the interquartile range
779 (IQR). Significant differences between time points were determined using a GLMM approach
780 that corrected for individual variation and age, followed by two-sided Tukey's HSD post hoc
781 tests. **(C, D)** Principal component analysis (PCA) plots based on Aitchison dissimilarities of
782 CLR-transformed species abundances, visualizing the overall nasopharyngeal microbiome
783 composition at time point 2 **(C)** and time point 3 **(D)**. Each data point represents one sample
784 and is colored according to early **(C)** or late **(D)** antibiotic exposure, with ellipses representing
785 the 95% confidence interval around the centroid of the microbiome composition for each group.
786 Statistical significance, indicated by adjusted P-values (calculated using the Benjamini–
787 Hochberg method) and effect sizes (R^2), was assessed using a PERMANOVA test and is
788 displayed at the bottom right of each plot. **(E)** Boxplot comparing the changes in
789 nasopharyngeal microbiota stability, calculated as the Aitchison distance between consecutive
790 sampling time points within the same subjects, among the three antibiotic exposure groups.
791 Pairwise Wilcoxon tests with Benjamini–Hochberg (BH) correction for multiple comparisons
792 were performed. **(F)** Dot plot showing the stability of nasopharyngeal microbiota profiles over
793 sampling time points stratified by postnatal antibiotic exposure. Stability was measured by
794 calculating the Aitchison dissimilarity between each pair of consecutive time points for each
795 infant. Loess lines were fitted to the data for each antibiotic exposure group, with shaded areas
796 representing the 95% confidence intervals. In **A**, **B** and **E**, statistical significance is denoted by
797 asterisks as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$. ns: not
798 significant.

799 **Figure 4. Nasopharyngeal microbiome composition of preterm infants evolves with age.**
800 **(A)** Principal Component Analysis (PCA) plot of the overall nasopharyngeal microbiota

801 composition, based on CLR-transformed species abundances. Each point represents one sample
802 (n = 336), colored by age group (or sampling time point). Boxplots on the bottom and right
803 margins illustrate the distribution of samples along the first two principal components. In the
804 boxplots, the center line denotes the median, box edges mark the 25th and 75th percentiles
805 (interquartile range, IQR), and whiskers extend to 1.5 × IQR. Adjusted p-values were calculated
806 using two-sided pairwise Wilcoxon tests with Benjamini–Hochberg (BH) correction (*$p < 0.05$;
807 **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **(B)** Tenfold cross-validation error as a function of
808 the number of input species used to regress against infants' chronological age in the training
809 set. The results indicate that 24 species are sufficient for random forest–based predictions of
810 chronological age from the microbiota composition. Data are shown as the mean ± SEM from
811 100 iterations. **(C)** The 24 most informative, age-discriminatory bacterial species identified by
812 the random forest model, listed in descending order of their importance for model accuracy.
813 These species were included in a sparse model. Data are represented as the mean ± SEM from
814 100 iterations. **(D)** Scatterplot showing the relationship between chronological age and the
815 microbiome-predicted age (in months), with each dot representing a sample (n = 336). Dots are
816 colored by antibiotic exposure group. The line of best fit and its 95% confidence interval
817 (shaded area) were estimated by linear regression (Pearson's R = 0.97, $p < 0.0001$).

818 **Figure 5: Identification of four community types in the nasopharyngeal microbiome of**
819 **preterm infants.** (**A**) Dirichlet Multinomial Mixture (DMM) analysis clustered the 336
820 samples into four distinct clusters, which were visualized using Principal Component Analysis
821 (PCA) based on Aitchison distances of species composition. Each point represents a single
822 sample and is colored according to its community type. (**B**) Box plots show the relative
823 abundance of the main contributing taxa at the genus level within each community type. (**C**)
824 Boxplot of species diversity (Shannon index) in the nasopharyngeal microbiome samples of
825 preterm infants, grouped by the four identified community types. (**D**) Bar chart shows the

826 proportions of community types identified at different sampling time points across the whole
827 cohort (left) and in three groups stratified according to postnatal antibiotic exposure (right) (n
828 = 336). (**E)** The distribution of community types across the infant postnatal antibiotic exposure
829 groups. (**F**) Stability and transition probabilities among the four microbial community types,
830 evaluated *via* Markov Chain modelling across different antibiotic exposure groups. Each node
831 represents a community type identified by the DMM model, and arrow width is proportional to
832 the maximum likelihood estimate of transition probabilities among the community types. **(G)**
833 Boxplots comparing the antibiotic resistance gene (ARG) diversity (Shannon index, left) and
834 total ARG abundance (sum of RPKM, right) in nasopharyngeal samples (n = 336) from preterm
835 infants, stratified by community types. In **B, C**, and **G**, the horizontal lines in each boxplot
836 represent the first quartile, the median, and the third quartile. Pairwise Wilcoxon rank-sum tests
837 with Benjamini–Hochberg (BH) correction were used to assess statistical significance.
838 *$p < 0.05$, **$p < 0.01$, ***$p<0.001$ and ****$p < 0.0001$. ns: not significant; CT: community
839 type.

840 **Figure 6: *S. marcescens* outbreak at the NICU. (A)** Samples are shown in chronological order
841 by sampling date. Each bar represents one sample, grouped by whether the infant was
842 discharged or transferred to a different NICU before the outbreak, or hospitalized and
843 discharged from the NICU after the outbreak. The relative abundance of *S. marcescens* is shown
844 in blue, while the abundance of all other bacterial species is combined into the "Others"
845 category (gray). **(B)** Boxplots illustrating the CLR-transformed abundance of *S. marcescens*
846 across different postnatal antibiotic exposure groups at each time point (1–6). The line within
847 each box denotes the median, the box edges represent the 25th and 75th percentiles, and
848 whiskers extend to 1.5× the interquartile range. **(C)** PCA plot of the nasopharyngeal microbiota,
849 colored according to the CLR-transformed abundance of *S. marcescens*. Points shaded in blue
850 indicate higher *S. marcescens* abundance; points in gray indicate lower abundance. Gradient

values are shown in the legend on the right. **(D)** Boxplots of the CLR-transformed and relative abundance (%) of *S. marcescens* in samples collected from birth to six months corrected age. Samples are grouped by hospitalization status (discharged/moved to a different NICU before the outbreak vs. hospitalized and discharged after the outbreak). Statistical significance of differences was determined using two-sided pairwise Wilcoxon rank-sum tests with Benjamini–Hochberg (BH) correction. Statistically significant differences are denoted with asterisks: $*p < .05$, $**p < .01$, $***p < .001$ and $****p < .0001$; ns: not statistically significant. **(E)** PCA plot based on Aitchison dissimilarities (species level) of all nasopharyngeal samples collected from birth to six months corrected age. Each point represents a sample, colored red if collected from infants discharged or moved before the outbreak and blue if from those hospitalized and discharged afterward. Effect sizes (R^2) and corresponding adjusted p-values (p.adj, BH-corrected) were calculated using PERMANOVA, as shown in the bottom left of the plot. Percentages on the axes indicate the variance explained by each principal component. PCA: principal component analysis; PERMANOVA: permutational multivariate analysis of variance; *p.adj*: adjusted *p*-values.

Figure S1: Overview of obtained samples and sampling time points in our study cohort. **(A)** Nasopharyngeal aspirate (NPA) samples collected from each infant (left y-axis) are connected by horizontal lines. Infants are organized by postnatal antibiotic (AB) exposure (right y-axis). The x-axis represents the day of life (DOL). In total, 369 NPA samples were obtained from 66 infants within the first six months of corrected age. **(B)** NPA samples were collected at six designated sampling time points (y-axis). If a sample could not be obtained according to the protocol at the designated time (e.g., if the infant was unstable), a new sample was collected as soon as possible, leading to greater variation in DOL for each time point. The horizontal bars indicate the DOL range for each sampling time, and the vertical line within each bar denotes the mean. Mean (SD) values are listed on the right side of the figure. **(C)** Scatter plot showing

876 the correlation between bacterial abundance measured by qPCR (log 16S qPCR [ng]) on the x-
877 axis and bacterial read counts (in millions) from metagenomic sequencing on the y-axis. Each
878 dot represents a single sample. R = Pearson's correlation coefficient.

879 **Figure S2:** Stacked bar plots showing the mean relative abundances of the five most abundant
880 bacterial phyla by sampling time point (left) and by antibiotic (AB) exposure group (right)
881 across the sampling period. All other phyla with mean relative abundances <1% are grouped
882 into the "Others" category.

883 **Figure S3:** Principal Component Analysis (PCA) plots, based on Aitchison dissimilarity,
884 visualizing the overall nasopharyngeal microbial community composition of preterm infants at
885 each sampling time point **(A–F)**. Each point represents a sample and is colored according to the
886 postnatal antibiotic (AB) exposure group. Ellipses represent the 95% confidence interval for
887 each AB exposure group. Effect sizes (R^2) from permutational multivariate analysis of variance
888 (PERMANOVA), along with corresponding adjusted P-values (*p.adj*, calculated using the
889 Benjamini–Hochberg method), are shown at the bottom right of each plot.

890 **Figure S4:** Boxplots illustrating the temporal stability of the nasopharyngeal microbiota,
891 stratified by postnatal antibiotic exposure. Aitchison dissimilarities between consecutive time
892 points (1–2, 2–3, 3–4, 4–5, and 5–6) within each subject were used as a measure of stability.
893 The horizontal line in each box indicates the median, while the lower and upper box boundaries
894 correspond to the 25th and 75th percentiles, respectively. Whiskers extend to the most extreme
895 data points within 1.5 × IQR of the box boundaries. Brackets highlight statistically significant
896 differences between time points within each group (Adjusted P-values (*p.adj*) < 0.05; pairwise
897 Wilcox test), with **p.adj* < 0.05, ***p.adj* < 0.01, and ****p.adj*< 0.001.
898

Figure S5: Individual variability and postnatal age significantly influence the nasopharyngeal microbiome composition in preterm infants. Principal Component Analysis (PCA) plots (based on CLR-transformed species abundance) illustrate β-diversity of the nasopharyngeal microbiota. Panels **(A)**, **(B)**, and **(C)** show the effects of individual variability, postmenstrual age (PMA), and days of life (DOL), respectively. Each point represents one patient sample (n = 336), color-coded according to the relevant grouping (individual, PMA, or DOL). Effect sizes (R^2) and corresponding adjusted P-values (*p.adj*, Benjamini–Hochberg correction) were calculated using PERMANOVA and are displayed in the bottom left of each plot.

Figure S6: Dirichlet multinomial mixture (DMM) analysis of nasopharyngeal samples reveals four distinct community types. **(A)** The Laplace approximation from the DMM fit indicates four optimal nasopharyngeal microbiota structures in our preterm infant cohort, and (B) shows the top 10 driver species for each DMM-based community type, ordered by their assignment strength on the y-axis.

Figure S7. (A) Overview of the 13 NICU infants (y-axis) screened at least once during the two main screening days (x-axis). Within a few days of receiving a negative screening result, nine infants were found to have reads assigned to *S. marcescens*, including four with a lower relative abundance (<1.3%) and five with a higher relative abundance (16.5–85.0%). **(B)** Timeline of *S. marcescens* detection in nasopharyngeal samples from each preterm infant across multiple time points, where each point represents a sample, colored by whether it was collected before or after the outbreak detection, and point shapes indicate whether *S. marcescens* was identified in whole metagenomic analyses. **(C)** Principal Components Analysis (PCA) plots of β-diversity (Aitchison distances) at each time point (1–6), with each point representing a sample and colored by whether it was collected before or after the outbreak. Adjusted p-values (Benjamini–

922 Hochberg correction) and effect sizes (R^2) for all time points were calculated using
923 PERMANOVA and summarized in the panel on the right.

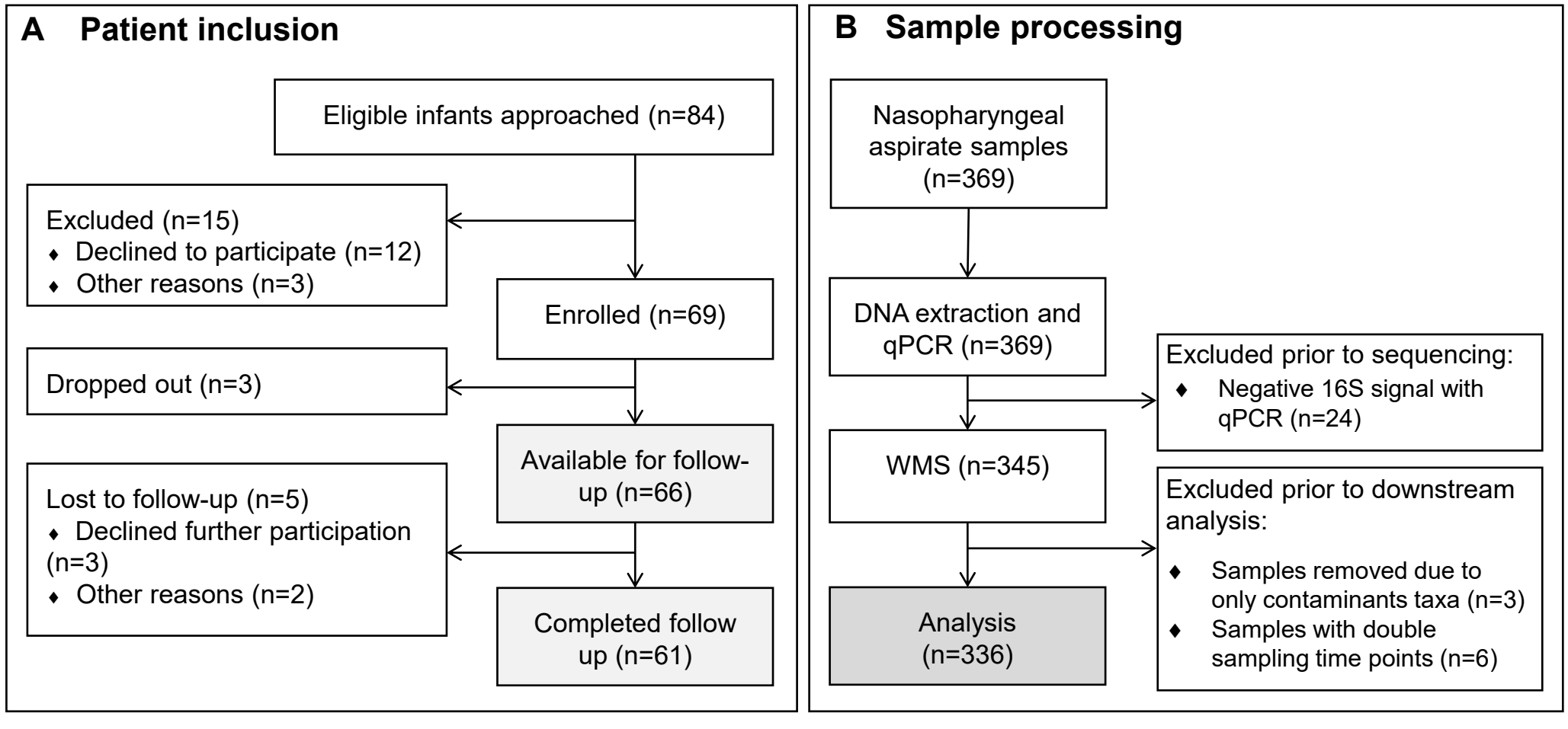
924 **Figure S8.** Random Forest confusion matrices for classifying preterm infants at six months
925 corrected age as *S. marcescens* persistence or non-persistence, based on nasopharyngeal
926 microbiota composition at time points 1 **(A)** and 2 **(B)**. Panels **(C)** and **(D)** show the top 10
927 species that most strongly contributed to the classification at time points 1 and 2, respectively.

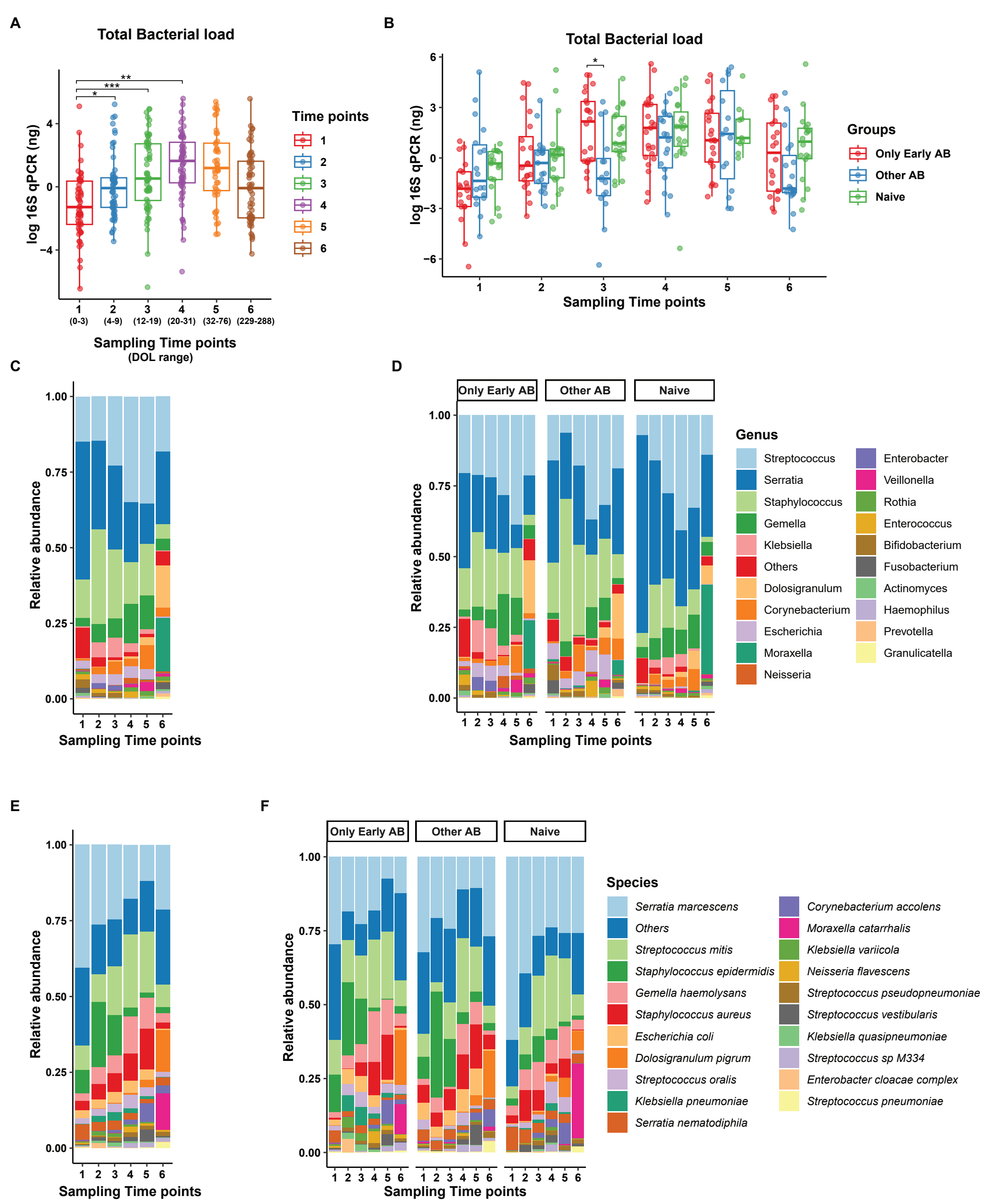
928 **Tables**

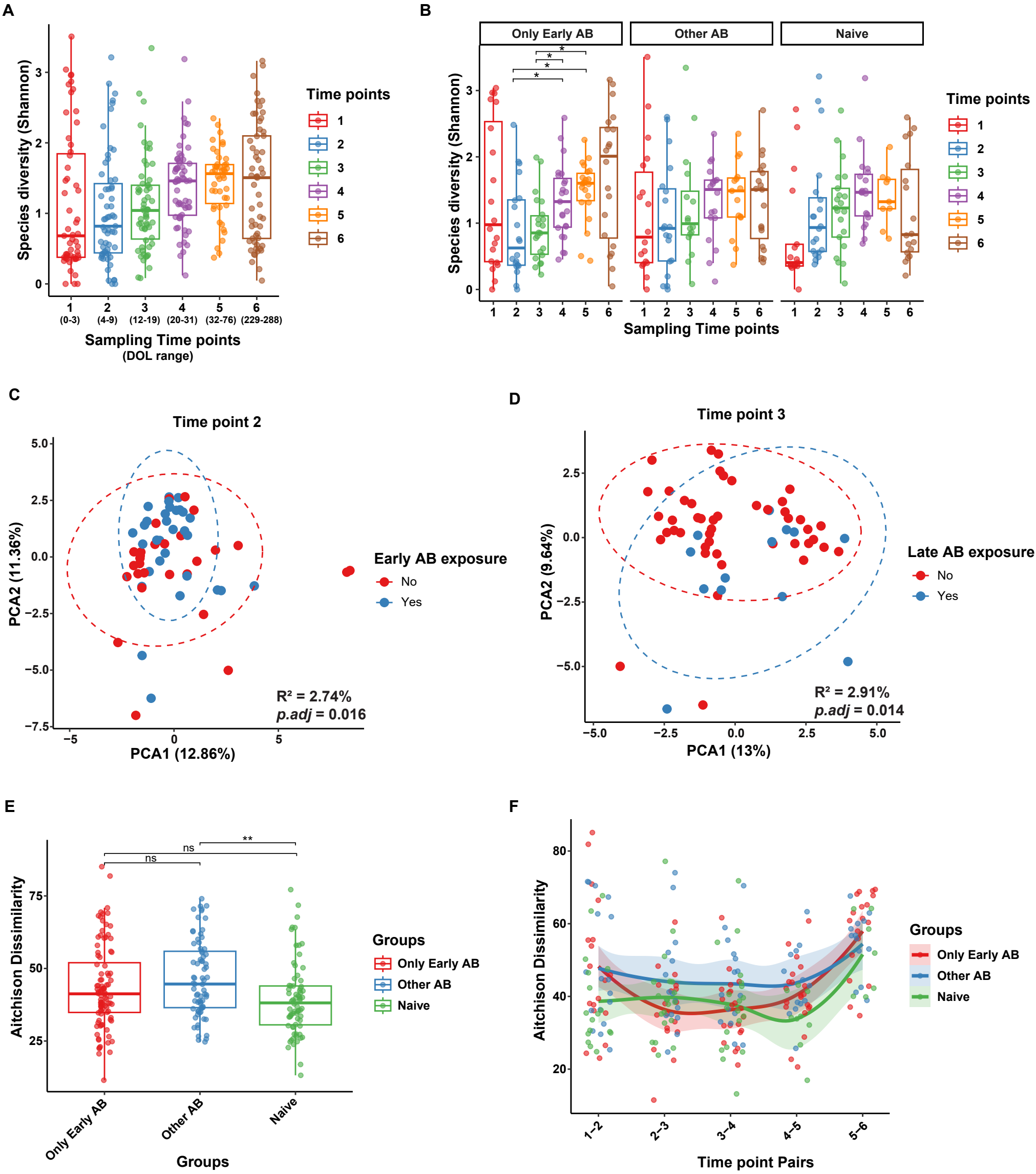
Whole cohort (n=66)	**AB naive (n=21)**	**AB exposed (n=45)**	**Only early ABs (n=24)**	**Other ABs (n=21)**
Male, n (%)	12 (57%)	33 (73%)	18 (75%)	15 (71%)
GA, weeks (mean, SD)	31, 6/7	**29 4/7, 1 ****	**29 4/7, 1 ****	**29 3/7, 1 ****
BW, grams (mean, SD)	1562, 188	**1280, 262 ****	**1278, 255 ****	**1282, 275 ****
BW, group, n (%)		*	*	*
ELBW	0 (0%)	**6 (13%)**	**3 (12%)**	**3 (14%)**
VLBW	8 (38%)	**31 (69%)**	**17 (71%)**	**14 (67%)**
LBW	13 (62%)	**8 (18%)**	**4 (17%)**	**4 (19%)**
Vaginal delivery n (%)	4 (19%)	**16 (35.5%)**	**9 (37.5%)**	**8 (38%)**
Caesarian section, n (%)	17 (81%)	29 (64.5%)	15 (62.5%)	13 (62%)
ABs during pregnancy, n (%)				
None	21 (100%)	**32 (71%) ***	**13 (54%) ****	19 (90%)
< 10 days	0 (0%)	10 (22%)	10 (42%)	0 (0%)
≥10days	0 (0%)	3 (7%)	1 (4%)	2 (10%)
Intrapartum AB prophylaxis, n (%)				
None documented	2 (10%)	3 (7%)	1 (4%)	2 (10%)
Given	19 (90%)	42 (93%)	23 (96%)	19 (90%)
ROM, hours (means, SD)	3, 12	201, 491	**258, 422 ***	136, 563

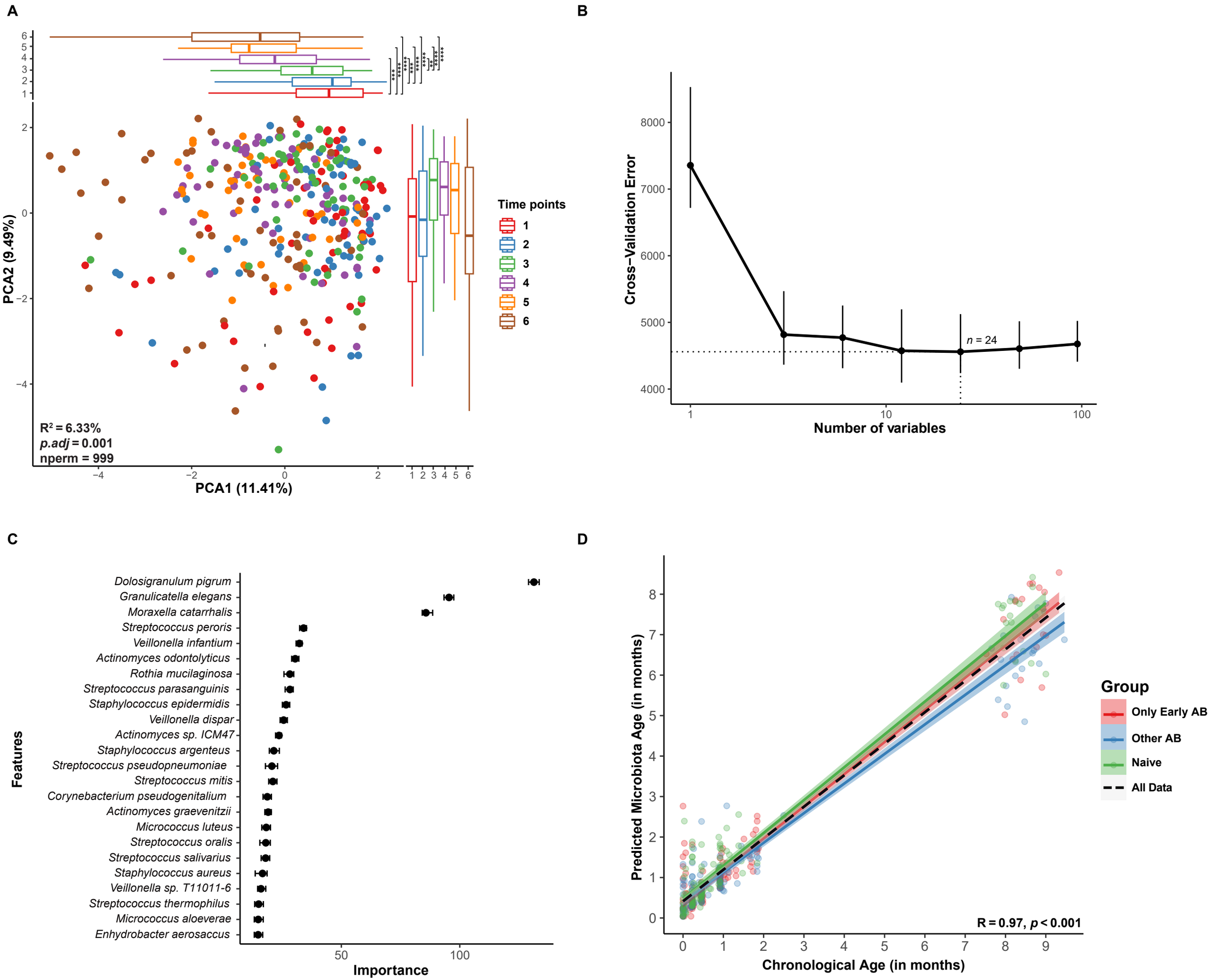
Apgar at 10 min (mean, SD)	9, 1	9, 1	8, 1	9, 1
Nutrition, n (%)				
Fully breastfed	7 (39%)	13 (30%)	5 (22%)	8 (40%)
> 50% MOM	3 (17%)	11 (26%)	7 (30%)	4 (20%)
> 50% formula	6 (33%)	8 (18%)	5(22%)	3(15%)
Only formula	2 (11%)	11 (26%)	6 (26%)	5 (25%)
Lost to follow up	3 (14%)	2 (4%)	1 (4%)	1(5%)

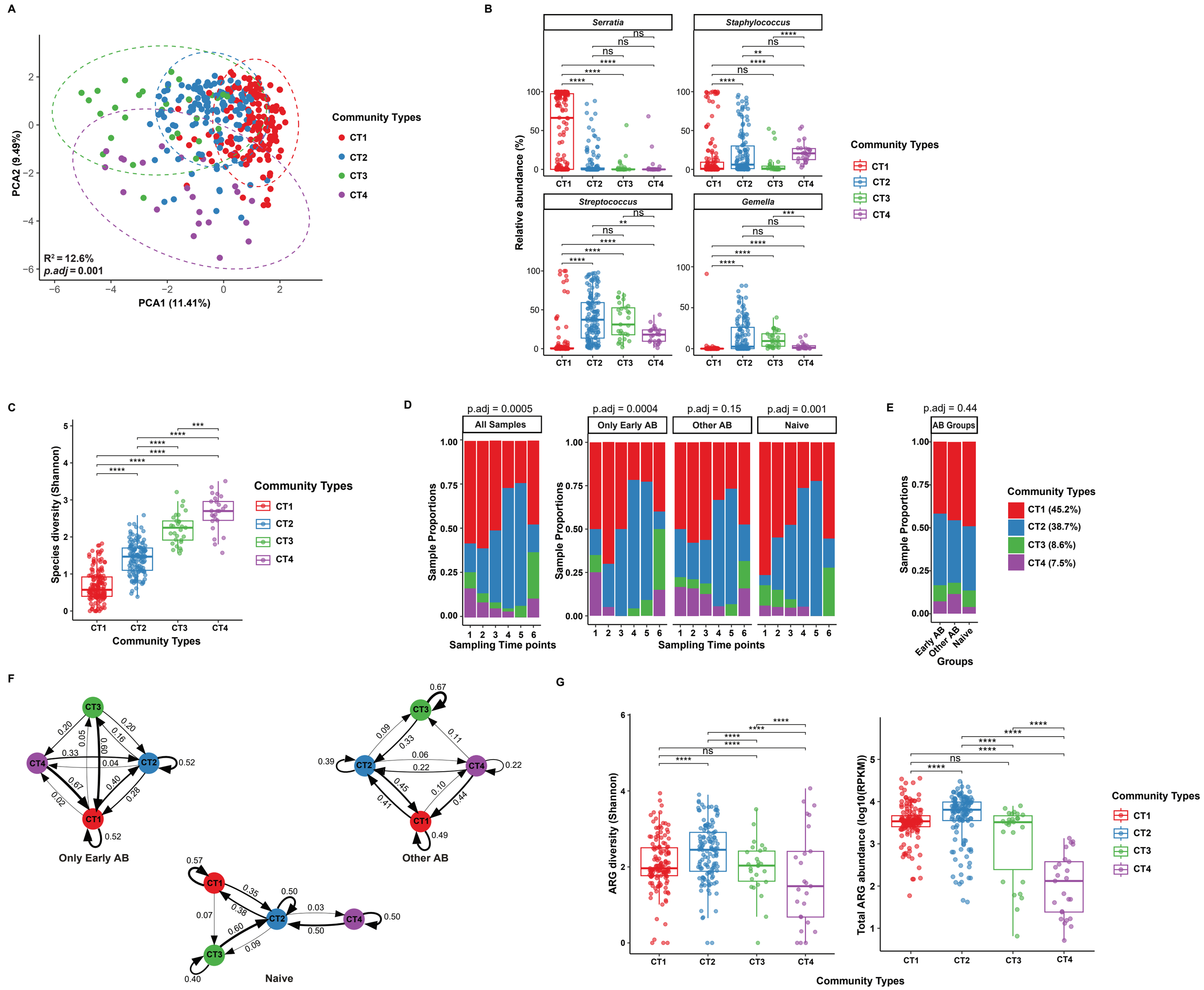
929 **Table 1. Cohort demographics.** AB: antibiotic; GA: gestational age; BW: birth weight; LBW:
930 low birth weight (from 1500g to < 2500g); VLBW: very low body weight (from 1000g to <
931 1500g); ELBW: extremely low body weight (< 1000g); ROM: rupture of membranes; MOM:
932 mothers' own milk. * $p<0.05$, ** $p<0.001$.

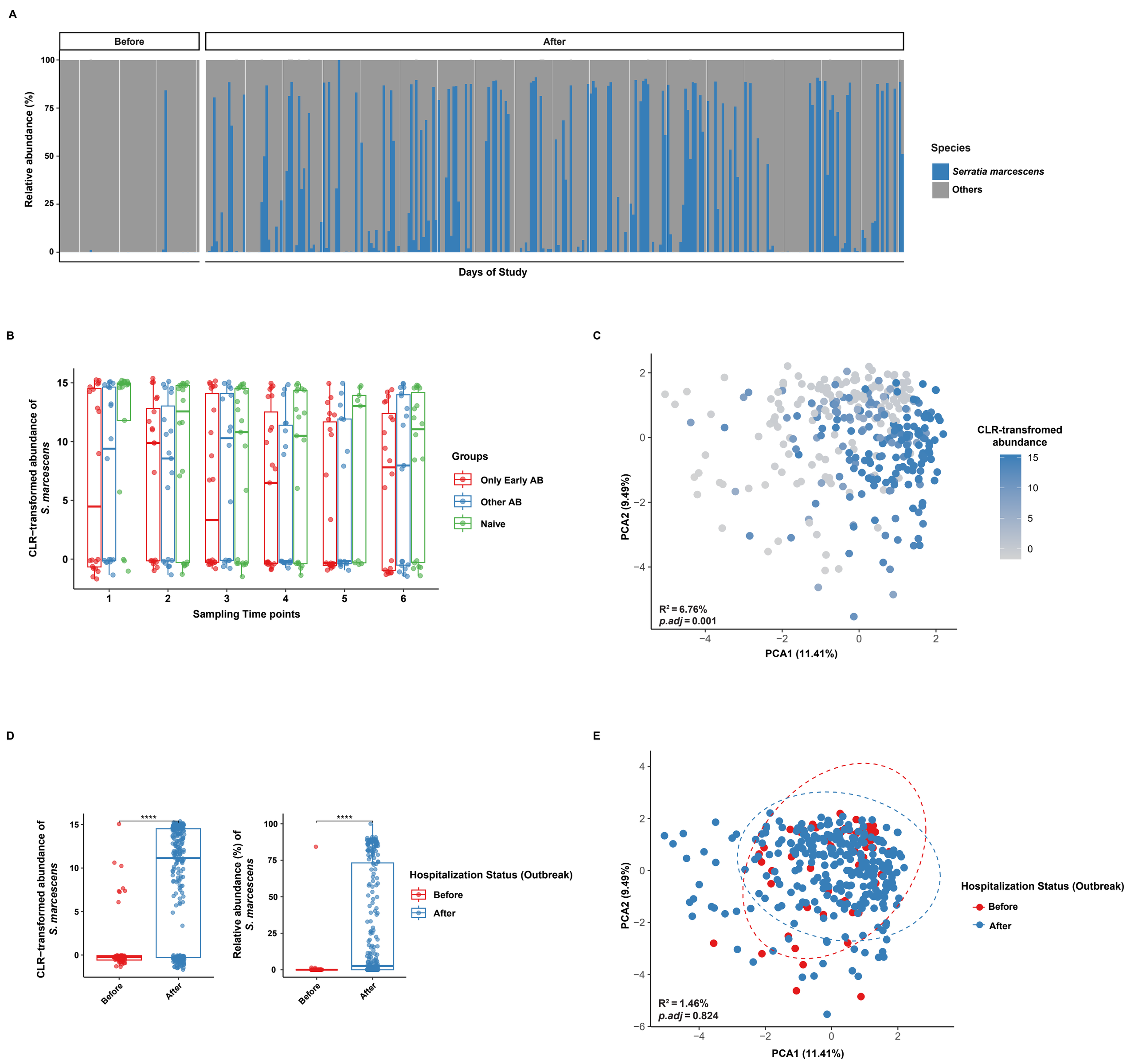

A Patient inclusion
Eligible infants approached (n=84)
Excluded (n=15)
Declined to participate (n=12)
Other reasons (n=3)
Enrolled (n=69)
Dropped out (n=3)
Available for follow-up (n=66)
Lost to follow-up (n=5)
Declined further participation (n=3)
Other reasons (n=2)
Completed follow up (n=61)
B Sample processing
Nasopharyngeal aspirate samples (n=369)
DNA extraction and qPCR (n=369)
Excluded prior to sequencing:
Negative 16S signal with qPCR (n=24)
WMS (n=345)
Excluded prior to downstream analysis:
Samples removed due to only contaminants taxa (n=3)
Samples with double sampling time points (n=6)
Analysis (n=336)


A
Total Bacterial load
log 16S qPCR (ng)
Sampling Time points
(DOL range)
(0-3)
(4-9)
(12-19)
(20-31)
(32-76)
(229-288)
Time points
B
Total Bacterial load
log 16S qPCR (ng)
Sampling Time points
Groups
Only Early AB
Other AB
Naive
C
Relative abundance
Sampling Time points
D
Only Early AB
Other AB
Naive
Relative abundance
Sampling Time points
Genus
Streptococcus
Serratia
Staphylococcus
Gemella
Klebsiella
Others
Dolosigranulum
Corynebacterium
Escherichia
Moraxella
Neisseria
Enterobacter
Veillonella
Rothia
Enterococcus
Bifidobacterium
Fusobacterium
Actinomyces
Haemophilus
Prevotella
Granulicatella
E
Relative abundance
Sampling Time points
F
Only Early AB
Other AB
Naive
Relative abundance
Sampling Time points
Species
Serratia marcescens
Others
Streptococcus mitis
Staphylococcus epidermidis
Gemella haemolysans
Staphylococcus aureus
Escherichia coli
Dolosigranulum pigrum
Streptococcus oralis
Klebsiella pneumoniae
Serratia nematodiphila
Corynebacterium accolens
Moraxella catarrhalis
Klebsiella variicola
Neisseria flavescens
Streptococcus pseudopneumoniae
Streptococcus vestibularis
Klebsiella quasipneumoniae
Streptococcus sp M334
Enterobacter cloacae complex
Streptococcus pneumoniae


A
Species diversity (Shannon)
Sampling Time points
(DOL range)
1 (0-3)
2 (4-9)
3 (12-19)
4 (20-31)
5 (32-76)
6 (229-288)
Time points
B
Only Early AB
Other AB
Naive
Sampling Time points
C
Time point 2
PCA2 (11.36%)
PCA1 (12.86%)
R² = 2.74%
p.adj = 0.016
Early AB exposure
No
Yes
D
Time point 3
PCA2 (9.64%)
PCA1 (13%)
R² = 2.91%
p.adj = 0.014
Late AB exposure
No
Yes
E
Aitchison Dissimilarity
ns
ns
**
Groups
Only Early AB
Other AB
Naive
F
Aitchison Dissimilarity
Time point Pairs
1–2
2–3
3–4
4–5
5–6
Groups
Only Early AB
Other AB
Naive


A
PCA2 (9.49%)
PCA1 (11.41%)
R² = 6.33%
p.adj = 0.001
nperm = 999
Time points
1
2
3
4
5
6
B
Cross−Validation Error
Number of variables
n = 24
C
Features
Dolosigranulum pigrum
Granulicatella elegans
Moraxella catarrhalis
Streptococcus peroris
Veillonella infantium
Actinomyces odontolyticus
Rothia mucilaginosa
Streptococcus parasanguinis
Staphylococcus epidermidis
Veillonella dispar
Actinomyces sp. ICM47
Staphylococcus argenteus
Streptococcus pseudopneumoniae
Streptococcus mitis
Corynebacterium pseudogenitalium
Actinomyces graevenitzii
Micrococcus luteus
Streptococcus oralis
Streptococcus salivarius
Staphylococcus aureus
Veillonella sp. T11011-6
Streptococcus thermophilus
Micrococcus aloeverae
Enhydrobacter aerosaccus
Importance
D
Predicted Microbiota Age (in months)
Chronological Age (in months)
R = 0.97, p < 0.001
Group
Only Early AB
Other AB
Naive
All Data


A
PCA2 (9.49%)
PCA1 (11.41%)
R² = 12.6%
p.adj = 0.001
Community Types
CT1
CT2
CT3
CT4
B
Serratia
Staphylococcus
Streptococcus
Gemella
Relative abundance (%)
C
Species diversity (Shannon)
Community Types
D
p.adj = 0.0005
All Samples
p.adj = 0.0004
Only Early AB
p.adj = 0.15
Other AB
p.adj = 0.001
Naive
Sample Proportions
Sampling Time points
E
p.adj = 0.44
AB Groups
Early AB
Other AB
Naive
Groups
CT1 (45.2%)
CT2 (38.7%)
CT3 (8.6%)
CT4 (7.5%)
F
Only Early AB
Other AB
Naive
G
ARG diversity (Shannon)
Total ARG abundance (log10(RPKM))
Community Types


A
Before
After
Relative abundance (%)
100
75
50
25
0
Days of Study
Species
Serratia marcescens
Others
B
CLR−transformed abundance of S. marcescens
15
10
5
0
1
2
3
4
5
6
Sampling Time points
Groups
Only Early AB
Other AB
Naive
C
PCA2 (9.49%)
PCA1 (11.41%)
R² = 6.76%
p.adj = 0.001
CLR-transfromed abundance
15
10
5
0
D
CLR−transformed abundance of S. marcescens
****
Before
After
Relative abundance (%) of S. marcescens
****
Hospitalization Status (Outbreak)
Before
After
E
PCA2 (9.49%)
PCA1 (11.41%)
R² = 1.46%
p.adj = 0.824
Hospitalization Status (Outbreak)
Before
After