THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Maturation of the infant respiratory microbiota, environmental drivers and health consequences: a prospective cohort study

OPEN ACCESS

1 **Title:** Maturation of the infant respiratory microbiota, environmental drivers and health

2 consequences: a prospective cohort study

3

4 **Authors:** Astrid A.T.M. Bosch[1,2†], Wouter A.A. de Steenhuijsen Piters[1,3,4†], Marlies A. van

5 Houten[2], Mei Ling J.N. Chu[1,3], Giske Biesbroek[1,2], Jolanda Kool[5], Paula Pernet[6], Pieter-Kees

6 C.M. de Groot[6], Marinus J.C. Eijkemans[7], Bart J.F. Keijser[5,8], Elisabeth A.M. Sanders[1] and

7 Debby Bogaert[1,3,4*]

8

9 [†] These authors contributed equally to this work.

10 * Corresponding author. Email: D.Bogaert@ed.ac.uk. Telephone: 0131 242 6582.

11

12 **Affiliations**

13 [1] Department of Paediatric Immunology and Infectious Diseases, Wilhelmina Children's

14 Hospital/University Medical Center Utrecht, Lundlaan 6, 3584 EA, Utrecht, The Netherlands.

15 [2] Spaarne Gasthuis Academy, Spaarnepoort 1, 2134 TM Hoofddorp, The Netherlands.

16 [3] Department of Medical Microbiology, University Medical Center Utrecht, Heidelberglaan

17 100, 3584 CX Utrecht, The Netherlands.

18 [4] Medical Research Council/University of Edinburgh Centre for Inflammation Research,

19 Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, EH16

20 4TJ Edinburgh, United Kingdom.

21 [5] Microbiology and Systems Biology Group, Netherlands Organisation for Applied Scientific

22 Research (TNO), Utrechtseweg 48, 3704 HE Zeist, The Netherlands.

23 [6] Department of Obstetrics and Gynaecology, Spaarne Gasthuis, Spaarnepoort 1, 2134 TM

24 Hoofddorp, The Netherlands.

25 [7] Biostatistics and Research Support, Julius Center for Health Sciences and Primary Care,

26 University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands.

27 [8] Department of Preventive Dentistry, Academic Center for Dentistry Amsterdam (ACTA),

28 University of Amsterdam and Vrije University Amsterdam, Gustav Mahlerlaan 3004, 1081 LA

29 Amsterdam, The Netherlands.

30

31 **Author contributions**

32 MAvH, EAMS, and DB designed the experiments, AATMB, MAvH, GB, EAMS, and DB

33 wrote the study protocols. AATMB, PP and PCMdG were responsible for patient recruitment.

34 AATMB and MAvH were responsible for clinical data collection. MLC was responsible for

35 sample preparation and MLC, JK, and BK for 16S-rRNA gene amplicon sequencing.

36 WAAdSP, MJCE and DB were responsible for bioinformatic processing and statistical

37 analyses. WAAdSP, AATMB, and DB wrote the paper. All authors significantly contributed to

38 interpreting the results, critically revised the manuscript for important intellectual content, and

39 approved the final manuscript.

46

47 **Running title:** Early-life respiratory microbiota development related to infections

48

55   **At a glance commentary**

56   *What is the current scientific knowledge on this subject?*

57   Factors affecting the risk of respiratory tract infections have been well characterized, however

58   it is unknown how these factors might impact respiratory microbiota development and thereby

59   susceptibility to respiratory tract infections (RTIs). Studies in mice suggest that timely

60   microbial cues contribute to healthy immune development, in turn enforcing the defense against

61   invading respiratory pathogens.

62

63   *What does this study add to the field?*

64   Using a longitudinal study design and high sampling resolution, we characterized the

65   nasopharyngeal microbiota maturation over the first year of life in 112 infants both during

66   health (11 sampling moments) and at the moment of RTIs. We observed differences in the

67   microbial community maturation in children who ultimately became more susceptible to

68   infections compared to children who were more resistant to infections. These changed dynamics

69   were related to shifts in the abundance of specific members of the microbiota and environmental

70   factors that are known to impact susceptibility to RTIs, such as mode of delivery, mode of

71   feeding, early antibiotic use and crowding. Altered microbiota maturation was evident from the

72   first month of life on and preceded factual RTIs, strongly suggesting that early-life microbiota

73   development impacts long-term respiratory health.

74

75   This article has an online data supplement, which is accessible from this issue's table of content

76   online at www.atsjournals.org

77

78

79

80

81 **Abstract**

82 **Rationale:** Perinatal and postnatal influences are presumed important drivers of the early-life

83 respiratory microbiome composition. We hypothesized that the respiratory microbiome

84 composition and development in infancy is affecting microbiome stability and thereby

85 resilience against respiratory tract infections (RTIs) over time.

86 **Objectives:** To investigate common environmental drivers, including birth mode, feeding type,

87 antibiotic exposure and crowding conditions, in relation to respiratory tract microbiota

88 maturation and stability, and consecutive risk of RTIs over the first year of life.

89 **Methods:** In a prospectively followed cohort of 112 infants, we characterized the

90 nasopharyngeal microbiota longitudinally from birth on (11 consecutive sample moments and

91 maximum three RTI samples per subject; in total n=1,121 samples) by 16S-rRNA gene

92 amplicon sequencing.

93 **Measurements and Main Results:** Using a microbiota-based machine-learning algorithm we

94 found that children experiencing a higher number of RTIs in the first year of life demonstrate

95 an aberrant microbial developmental trajectory already from the first month of life on as

96 compared to the reference group (0-2 RTIs/year). The altered microbiota maturation process

97 coincided with decreased microbial community stability, prolonged reduction of

98 *Corynebacterium* and *Dolosigranulum*., enrichment of *Moraxella* already very early in life,

99 followed by later enrichment of *Neisseria* and *Prevotella* spp. Independent drivers of these

100 aberrant developmental trajectories of respiratory microbiota members were mode of delivery,

101 infant feeding, crowding and recent antibiotic use.

102 **Conclusions:** Our results suggest that environmental drivers impact microbiota development

103 and consequently resilience against development of RTIs. This supports the idea that microbiota

104 form the mediator between early life environmental risk factors for and susceptibility to RTIs

105 over the first year of life.

106   **Key words:** respiratory microbiota, nasopharynx, respiratory tract infections, development,

107   risk factors.

108

Introduction

Acute respiratory tract infections (RTI) are a leading cause of childhood mortality, being responsible for ~0.9 million yearly deaths (15.5% of all deaths) worldwide in children <5 years (1). In addition, these infections are associated with significant morbidity (2) and are a major reason for antibiotic prescription (3), especially in young children. Although it is still unclear why one individual is more vulnerable to respiratory infections compared to another, it was previously hypothesized that - besides environmental and host-related influences - the respiratory microbiota may modulate susceptibility to disease.

Directly after birth, the mucosal surfaces of the respiratory tract of neonates are rapidly colonized with a variety of microbiota, that are swiftly moulded into niche-specific bacterial communities (4, 5). Over the first months to years of life, these communities are highly dynamic and heavily influenced by environmental factors, including mode of delivery (4, 6), season (7), feeding type (8), and antibiotic treatment (9). In previous studies we found that the microbial composition at the age of six weeks was indicative of microbiota stability and RTI susceptibility over the first two years of life (10, 11). This finding underscores the importance of direct postnatal environmental influences and subsequent early microbiota maturation on future respiratory health.

The healthy human respiratory microbiome is assumed to stimulate immune maturation (12, 13), promote epithelial integrity (14), and provide colonization resistance (15), thereby preventing overgrowth and invasion of potential pathogenic bacteria (16). In contrast, deviations from a healthy bacterial respiratory community composition have been associated with susceptibility to and/or severity of childhood respiratory diseases, including acute otitis

133   media (17, 18), respiratory syncytial virus (RSV) disease (19) and asthma development (20) in

134   various retrospective and cross-sectional studies (21).

135

136   We here postulate that alterations in the respiratory microbiome development early in life are a

137   consequence of changes in the abundance of specific bacterial biomarkers species. We

138   hypothesize that these alterations are controlled by known host-related and environmental

139   influences, and can ultimately lead to altered microbiota stability, in turn affecting RTI

140   susceptibility. Therefore, we prospectively investigated the nasopharyngeal microbiota

141   maturation of 112 unselected, healthy children with frequent, short interval sampling during the

142   first year of life as well as during RTI episodes. Hereby, we aimed to study respiratory

143   microbiota development early in life, and investigate its role as potential mediator between

144   early-life drivers and susceptibility to respiratory infectious disease.

145

146

147   **Methods**

148   Details on the study design, sample and data collection and bioinformatics/statistical methods

149   can be found in the online supplement Methods. Data have been deposited in the National

150   Center for Biotechnology Information GenBank database (accession number: SRP093519).

151   *Study population*

152   We enrolled in total 128 healthy children in an ongoing prospective birth cohort study aiming

153   to investigate the development of the infant microbiome during health and disease. Of 128

154   infants, 12 children were lost to follow-up (Figure E1). Details on the trial methods have been

155   described elsewhere (4). Written informed consent was obtained from both parents. The study

156 was approved by the Ethics Committee of Noord Holland, The Netherlands (M012-015,

157 NH012.394, NTR3986). Sequence data of part of the samples (≤6 months; n=743 samples of

158 101 children) were used for a study on the role of mode of delivery on respiratory microbiota

159 acquisition (4).

Data collection

161 For the current analyses, we included samples and data of 112/116 children who completed the

162 one-year follow-up and for whom we had ≥8 samples available for further analyses after

163 laboratory work-up (Figure E1). Home visits were conducted within two hours after birth, at

164 24 hours, at seven and 14 days, and at one, two, three, four, six, nine, and 12 months of age.

165 During each home visit, a trained doctor or research nurse obtained a nasopharyngeal swab

166 according to World Health Organization protocol (22) and completed an extensive survey on

167 the health status of the child, as well as on the presence or absence of potential risk and

168 environmental factors related to respiratory disease (4). Next to these regular visits, parents

169 were asked to contact the study team in case of an active RTI, defined as fever ≥38°C for >6

170 hours combined with malaise and presence of RTI symptoms. Following, a RTI visit was

171 planned within 48 hours after start of the fever to collect additional samples and to obtain more

172 detailed medical information.

16S-rRNA gene amplicon sequencing

174 Bacterial DNA of the nasopharyngeal samples was isolated, amplicon libraries of the 16S-

175 rRNA gene (V4 region) were generated, and sequencing was executed as previously described

176 (4, 23). Amplicon pools were paired-end sequenced in eight runs using an Illuminia MiSeq

177 instrument (Illumina Inc., San Diego, CA, USA). Bioinformatic processing included trimming,

178 error correction, assembly and 97%-identity clustering of reads into OTUs. Following removal

179   of chimeric reads, OTUs were taxonomically annotated using SILVA and BLASTN (Table E1).

180   We refer to OTUs using maximum genus level annotations, combined with a rank number based

181   on the abundance of each given OTU. Details on processing and quality control, including the

182   use of negative controls, are described in the online supplement Methods. After abundance-

183   filtering, a rarefied dataset was generated, and used for downstream analyses (24). α-diversity

184   measures were averaged over 100 rarefactions. β-diversity was assessed using the Bray-Curtis

185   dissimilarity metric.

Statistical analysis

187   All analyses were performed in the R version 3.3.0 within R studio version 0.99.902.

Random forest analysis

189   We hypothesized that the nasopharyngeal microbial succession patterns would be altered in

190   children who experienced more RTIs during their first year of life. Therefore, we stratified our

191   population into three groups based on the normal distribution of RTIs over the first year of life

192   (Figure E2); 39 children with 0-2 RTIs (reference group; *n*=372 samples), 52 children with 3-

193   4 RTIs (*n*=496 samples) and 21 children with 5-7 RTIs (*n*=197 samples). To identify OTUs

194   characteristic of a healthy microbiota maturation, we regressed the relative abundance of all

195   576 OTUs against chronological age in the reference group using the *randomForest* package,

196   and selected age-discriminatory OTUs using a step-wise backward 10-fold cross validation

197   procedure, see online supplement Methods and Figure E3A and E3B (24). This selection of

198   OTUs was subsequently used as input to a random forest model where we regressed the relative

199   abundance of these OTUs versus chronological age in the reference group. The resulting final

200   model was then used to predict chronological age, referred to as 'microbiota age', in samples

201   from individuals who experienced 3-4 and 5-7 RTIs and on the group of samples collected

202 during RTIs (*n*=56 samples). To generate accurate microbiota age estimates for the reference

203 group, we used a 10-fold cross-validation procedure. Relative microbiota age (RMA) was

204 calculated as follows: relative microbiota age = microbiota age of a given child − microbiota

205 age of the reference group at similar age as determined by a spline fit (24). As a post-hoc

206 analysis, we studied the effect of the *Moraxella*-genus on the performance of the microbiota

207 age model by excluding the OTUs belonging to the *Moraxella*-genus from the model while

208 monitoring the amount of variance explained.

209 *Associations between environmental factors and microbiota parameters*

210 'Environmental factors' used in the descriptions of the various models comprises birth mode,

211 breast feeding until three months of age, day care attendance, presence of siblings >five years

212 of age, antibiotic treatment in the previous four weeks and season of birth, if not specified

213 otherwise. If applicable, correction for multiple testing was performed using the Benjamini-

214 Hochberg procedure.

215 Microbial succession patterns were visualised using non-metric multidimensional

216 scaling (nMDS; *vegan* package) based on the Bray-Curtis dissimilarity matrix. We performed

217 two separate analyses based on permutational multivariate analysis of variance

218 (PERMANOVA)-tests and the Bray-Curtis dissimilarity matrix, to study the effect of 1)

219 environmental factors, age and subject, and 2) the number of RTIs experienced in the first year

220 of life, on the overall bacterial community structure. Permutations were constrained within

221 subjects to account for repeated measures. This analysis was repeated over 100 rarefactions to

222 assess the robustness of our results based on one rarefied set.

223 To complement the group-based analyses, we also assessed the microbial development

224 at the individual level using an unsupervised clustering approach. The proportion of samples

225 within each cluster at each time point was visualised using an alluvial diagram, stratified by the

226 number of RTIs that children experienced over the first year of life.

227

228   We used separate linear mixed models to assess the associations between (relative) microbiota

229   age and stability (α-/β-diversity measures) as dependent variables and 1) environmental factors

230   and 2) the number of RTIs (fixed effects), while adjusting for age and with the subject-variable

231   included as a random intercept (*lme4* package). In addition, the relationships between 1)

232   bacterial density and 2) relative abundance (dependent) and sampling moment (fixed) were

233   assessed using linear mixed models.

234

235   We used smoothing spline analysis of variance (SS-ANOVA; *metagenomeseq* package) for the

236   analyses of 1) the differences in abundance of age-discriminatory OTUs between RTI-groups,

237   and 2) the effects of birth mode and breastfeeding on the nasopharyngeal microbiota, as it

238   simultaneously tests for the existence and timing of differences in OTU-abundance. To confirm

239   associations between environmental factors and relative abundance of microbiota in a

240   multivariable manner, we used the Multivariate Association with Linear Models (MaAsLin)

241   (R-)package, adjusting for age and with subject as a random effect.

242

243

244   **Results**

245   *Baseline characteristics of the study population*

246   Baseline characteristics of the study population stratified by number of RTIs experienced in the

247   first year of life can be found in Table E2.

248   *Nasopharyngeal microbiota composition in the first year of life*

249 A median of 20,670 reads were generated per sample (range 3,911-97,870 reads), which were

250 binned into a total 576 operational taxonomic units (OTUs; after filtering), representing a total

251 of 14 bacterial phyla. Firmicutes was the most abundant phylum with a maximum abundance

252 of 65.4% at day one (mainly *Staphylococcus (3)*, *Dolosigranulum (4)* and *Streptococcus (5)*).

253 Later, Proteobacteria emerged and became predominant with a maximum abundance of 71.7%

254 at 12 months of life (mostly *Moraxella (1)*, *Haemophilus (6)* and *Moraxella (7)*; Figure 1,

255 Figure 2 and Figure E4). We observed major shifts in nasopharyngeal microbiota composition

256 between day 0 and day one and between day one and week one (Figure E5). The difference in

257 microbiota composition between day one and week one coincided with a strong increase in

258 absolute bacterial abundance, which then increased up to the age of ~1 month, after which it

259 stabilized (linear mixed model; q<0.001; Figure 3).

260 *Trajectories of microbial development*

261 We aimed to study whether nasopharyngeal microbiota development is different in infants

262 experiencing more RTIs in the first year of life compared to the low-burden infants. First, we

263 demonstrated that the microbial community composition was significantly associated with the

264 number of RTIs experienced in the first year of life (i.e. 1-7 RTIs; categorical), after adjusting

265 for age, using a PERMANOVA-test (Table E3A; 1.7% of the variance explained, p=0.001).

266 Subsequently, we stratified the study participants over three groups based on the number of

267 RTIs they experienced within the first 12 months of life (i.e. 0-2, 3-4 and 5-7 RTIs; Figure E2

268 and Table E2). To explore the microbial succession patterns at the individual level, we clustered

269 samples using an unsupervised clustering approach. The proportion of individuals in each

270 cluster at each time point was then visualised using an alluvial diagram stratified by the number

271 of RTIs experienced over the first year of life (figure E6). We identified 8 clusters over all time

272 points, of which the largest four were enriched for *Moraxella (1)* (MOR1, 38.5% of samples)

273   *Corynebacterium (2)* and *Dolosigranulum (4)* (CDG, 19.7%), *Staphylococcus (3)* (STA,

274   19.4%) and *Streptococcus (5)* (STR, 8.4%). In concordance with our previous observations, we

275   found that the CDG-cluster has a much more prominent and prolonged role in the reference

276   group compared to children who suffered from 5-7 RTIs. Instead, these children appear to 'skip'

277   the CDG-cluster altogether, transitioning directly from the early-life STA-cluster to the MOR1-

278   cluster (figure E6C), the latter of which is typically observed more often at later time points in

279   the reference cohort (figure E6A). In the children who experience 3-4 RTIs the cluster

280   distributions at each time point do resemble those of the reference group, although an early rise

281   of the *Haemophilus (6)* (HAE)-cluster was noted (figure E6B).

282   *Nasopharyngeal microbiota maturation in relation to susceptibility to RTI and identification of*

283   *age-discriminatory taxa*

284   To further assess these differences in microbiota dynamics we used a random forest regression

285   model. First, we identified age-discriminatory OTUs in the reference group (i.e. 0-2 RTIs;

286   Figure E3A and 3B) and regressed their relative abundance against chronological age, enabling

287   us to model healthy microbiota development (65.9% of variance explained, based on 10-fold

288   cross-validation, 100 repetitions). Then, the model was used to calculate predicted

289   chronological age or 'microbiota age' in children with 3-4 and 5-7 RTIs and in samples taken

290   during RTIs (58.1% variance explained), subsequently comparing these estimates to

291   chronological age. We first observed that children with 5-7 RTIs showed an accelerated

292   microbiota maturation when compared to the reference group from very early in life on (linear

293   mixed model; p=0.007). A similar, although non-significant trend was observed in children

294   with 3-4 vs reference group (linear mixed model; p=0.13; Figure 4A). The accelerated

295   microbiota developmental patterns in children with >2 RTIs were related to an early enrichment

296   of *Moraxella (1)* from just after birth on (SS-ANOVA; q=0.007), enrichment of *Neisseria*,

297 *Prevotella* and *Alloprevotella* spp. from month two onwards (SS-ANOVA; q≤0.021) and

298 (prolonged) absence of *Corynebacterium (2)* and *Corynebacterium (80)*, *Dolosigranulum (4)*

299 and *Streptococcus (10)* (SS-ANOVA; q≤0.039; Figure 4B, Figure E7 and Table E4A).

300 Subgroup analyses comparing either the 3-4 or 5-7 RTI groups to the reference group yielded

301 highly similar results (Table E4B and E4C).

302 To assess whether the above differences were predominantly driven by the *Moraxella* genus

303 rather than by the total group of biomarkers species, we assessed the impact of *Moraxella* spp.

304 on the performance of the microbiota age model by repeating the analyses including all

305 biomarker OTUs, except those belonging to the *Moraxella*-genus. This model, containing 18

306 OTUs, showed a confined effect of *Moraxella* spp., with a small reduction of performance in

307 the reference group (60.9% variance explained) and a slightly improved performance in

308 children who experienced 3-4 or 5-7 RTIs over the first year of life and in samples taken during

309 RTI (60.1% variance explained), compared with the model based on 22 OTUs.

310 *Relative microbiota age in relation to (susceptibility to) RTI*

311 By calculating the relative microbiota age (RMA; defined as the difference in microbiota age

312 between susceptible groups versus the reference group) we verified that microbiota age was

313 increased in children with 5-7 RTIs compared to the reference group (linear mixed model,

314 adjusted for age; p=0.007; Figure E8), which was already apparent in the first month of life

315 (p=0.011; linear mixed model; post-hoc analysis in children ≤1 month of age). This latter

316 finding was substantiated by a PERMANOVA-test, demonstrating that the microbiota

317 composition over the first month of life was significantly associated with the number of RTIs

318 over the first year of life (Table E3D; 0.8% of the variance explained, p=0.001). The RMA was

319 not significantly different between the group with 3-4 RTIs and the reference group (p=0.12).

320 Moreover, although the RMA was maximal during RTIs (median RMA +67.8 days in RTI

samples), we already observed an increase in RMA during the period preceding the factual RTI

(median RMA +37.1 days at the first time point preceding RTI [T = -1]; p=0.004), suggesting

that the microbiota maturation alterations precede RTIs. After recovery from an RTI, RMA

decreased towards the reference group, though did not normalize (median RMA +29.7 days T

= +1; p=0.04; Figure 4C). Although these changes in RMA appeared to be related to individual

OTUs (figure E9), these changes were not statically significant.

Nasopharyngeal microbiota stability over time

We next investigated whether bacterial community stability over time was different for children

who experienced 0-2, 3-4 and 5-7 RTIs over the first year of life. Community stability,

measured by the Bray-Curtis dissimilarity between consecutive time points, was significantly

different between children with 0-2 RTIs and those with 3-4 and 5-7 RTIs (linear mixed model;

p=0.005 and p=0.02, respectively). This phenomenon was apparent from the age of three

months on (Figure 5).

Impact of environmental drivers on bacterial community composition

We then aimed to assess the effect of environmental factors on nasopharyngeal microbiota

composition and succession. Using PERMANOVA tests, we found that factors with the largest

impact comprised subject (unadjusted R^2=18.7%), chronological age (10.4%) and

environmental drivers, including presence of siblings <five years of age (1.6%), day care

attendance (0.9%), season of birth (0.7%), breastfeeding for at least three months (0.5%), birth

mode (0.4%) and antibiotic usage in the previous month (0.3%; all p-values ≤0.016; Table E3B

and E3C).

Environmental drivers and their effects on microbiota maturation, stability and individual

bacterial taxa

344 After showing microbiota maturation is accelerated in children more susceptible to RTIs, we

345 next set out to determine the influence of environmental drivers on this process. We modelled

346 the RMA using a linear mixed model including environmental factors. We observed that

347 particularly the presence of young siblings and day care attendance are associated with an

348 increased RMA early in life (both $p<0.0005$). Similar associations were found when directly

349 modelling microbiota age instead of RMA versus environmental drivers (data not shown). In

350 contrast, the observed differences in microbiota stability between groups could not be explained

351 by environmental factors (linear mixed model; $p>0.05$) and did not relate to differences in α-

352 diversity measures between groups (linear mixed model; $p>0.05$, Figure E10). We also did not

353 detect differences in microbiota stability directly prior to, during or following a RTI episode.

354 We further tested the contribution of individual bacterial taxa to the associations

355 between environmental factors and microbiota maturation using MaAslin. With respect to age-

356 discriminatory taxa, we found that *Moraxella* spp. were positively and *Staphylococcus* spp.

357 were negatively associated with day care (both $q<0.0005$). Furthermore, we found that

358 *Corynebacterium (2)* and *Dolosigranulum (4)* were strongly reduced following antibiotic usage

359 ($q<0.03$). Additionally, we observed many associations between environmental drivers and

360 bacterial taxa that were not previously assigned as age-discriminatory biomarkers. Notably, the

361 presence of siblings was associated with increased abundance of the family *Pasteurellaceae*

362 ($q=0.003$), which includes the *Haemophilus* genus (Table E5).

363 *Temporal effects of mode of delivery and feeding type on bacterial taxa*

364 Since MaAsLin is not suited to identify temporary effects and the timeframes within which they

365 occur, we additionally studied the impact of early life drivers, such as mode of delivery and

366 feeding type, on the microbial succession patterns using SS-ANOVA. Of the age-

367 discriminatory taxa, early and/or prolonged predominance of *Corynebacterium (2)*,

368   *Corynebacterium (8)* and *Dolosigranulum (4)* (q≤0.03) and late enrichment of *Moraxella* spp.

369   (q<0.05; from ~month 3 on) were associated with vaginal birth and/or breastfeeding.

370   Contrariwise, in formula fed and/or caesarian born children we observed a high abundance of

371   *Gemella (9)* and *Streptococcus (10)* (q≤0.012) from birth on, and prolonged (4-11 months)

372   predominance of *Neisseria* spp. and (facultative) anaerobes including *(Allo)prevotella*,

373   *Granulicatella* and *Actinomyces* spp. (q<0.05) after the first month of life. Abundance of the

374   age-discriminatory taxum *Staphylococcus (3)* was related to birth by caesarian section in the

375   first month of life only (q=0.016). Besides, although not directly linked to microbiota

376   maturation, we found that the additional early enrichment of *Streptococcus (5)* was associated

377   with caesarian section and/or formula feeding (from birth on; q≤0.026), which could be

378   confirmed using MaAsLin (Table E5). Additionally, we observed temporal enrichment of oral

379   type of bacteria including streptococci and facultative anaerobic bacteria like *Prevotella*,

380   *Porphyromonas* and *Veillonella* spp., in formula fed children (from ~month 1-2 onwards) and

381   early abundance of *Dolosigranulum (4)* in breastfed children (Table E6 and E7 and Figure E11

382   and E12).

383

384

385   **Discussion**

386   Microbial colonization of the upper respiratory tract occurs directly after birth and develops

387   rapidly towards niche-specific profiles during the first weeks of life (4, 5, 10, 25). Several cross-

388   sectional case-control studies have shown differences in respiratory microbial profiles between

389   children with and without acute otitis media (18, 26), and between infants with mild, moderate

390   and severe RSV (19). Longitudinal studies, linking respiratory microbiota development and

391 maturation and (risk of) RTIs, however, are sparse, lack detailed information, and are only

392 retrospectively executed (10, 20).

393

394 Our results suggest that microbiota maturation in healthy children who experience a limited

395 number of 0-2 RTIs in the first year of life (reference group), is associated with a specific timing

396 of colonization events accompanied by the consecutive appearance and disappearance of

397 specific community members. In general, we observed that during the first week of life, the

398 microbiota development is typified by a strong increase in absolute bacterial abundance. In the

399 reference group, this coincides with the initial expansion of *Streptococcus* spp. at day one,

400 supplanted by rapid niche-differentiation at one week of life, initially driven by staphylococcal

401 predominance, but quickly followed by the establishment of multiple *Corynebacterium* and

402 *Dolosigranulum* spp.: a process which is strongly related to vaginal delivery (4) as well as

403 breastfeeding. Although *Moraxella* spp. become predominant community members over time

404 in most children, in the reference group they only become the main community members from

405 2-3 months of life on. From that age on, *Moraxella* spp. may still co-occur with

406 *Corynebacterium* and *Dolosigranulum* spp. in a mixed community profile or they can truly

407 dominate all other community members in a *Moraxella* spp. dominated community profile (4).

408 This natural process of consecutive events coincides with normalization of ecological stability

409 from the age of three months on and fewer infections.

410 In contrast, children with high susceptibility to RTIs over the first year of life exhibit an

411 accelerated bacterial community maturation from as early as the first month of life on, i.e. prior

412 to development of their first RTIs. This pattern was characterised by diminished and less

413 prolonged establishment of *Corynebacterium* and *Dolosigranulum* spp. coinciding with

414 premature predominance of *Moraxella* spp. colonization, and more abundant and prolonged

415 presence of oral types of bacteria in the nasopharyngeal niche, including *Neisseria* and

416   *Prevotella* spp. The observed aberrant microbial succession in children with more RTIs also

417   coincided with decreased bacterial community stability over time, which is in line with previous

418   observations and support the ecological theory that more stable microbiota are more resistant

419   to RTIs (10). Interestingly, we could also show that acceleration of microbiota age preceded

420   the factual RTIs, supporting the hypothesis that microbiota changes forego a clinically

421   symptomatic RTI. Conjointly, these findings support our hypothesis that the initial early

422   colonization after birth and subsequent development of URT microbiota over the first months

423   of life impact respiratory health.

424

425   Our data, in line with others, show that prolonged abundance of *Corynebacterium* and

426   *Dolosigranulum* spp. are linked to healthy microbiota development and microbiota stability

427   (10, 17, 20, 26), and are related to breastfeeding and vaginal delivery (4, 8, 27). Their co-

428   occurrence may be explained by the ability of *Dolosigranulum* spp. to produce lactic acid,

429   which plausibly selects for *Corynebacterium* spp. outgrowth (21). Antagonism between

430   *Corynebacterium* spp., and *Streptococcus pneumoniae*, a known respiratory pathogen, may at

431   least in part explain their association with respiratory health (17, 26, 28). Since we and others

432   (20, 29) showed that antibiotic use in infancy is associated with depletion of *Corynebacterium*

433   and *Dolosigranulum* spp., routinely used antibiotics may therefore have more (prolonged)

434   consequences for microbiota-driven resilience against RTIs than currently is thought.

435   Conversely, accelerated microbial succession patterns in children with more RTIs were

436   characterized by enrichment of *Neisseria* spp. and (facultative) anaearobic, mainly oral species,

437   including *Prevotella* spp., which in turn were linked to formula feeding. Similar findings have

438   been reported previously (10, 30), and imply a loss of topography within the upper respiratory

439   tract, suggesting that the host or the local ecosystem is unable to restrain oral microbiota within

440   their niche early in life. As presence of these bacteria is linked to RTI susceptibility, further

441   studies on their role in respiratory health is warranted.

442   In literature, conflicting results have been reported regarding the role of *Moraxella* spp.

443   in the pathogenesis of RTIs. Some studies found that *Moraxella* spp. colonization was

444   associated with respiratory infections including pneumonia and bronchiolitis (11), while others

445   reported that the *Moraxella*-dominated profile was associated with bacterial community

446   stability (10, 20) and fewer RTI episodes (10). Although in our study, development from a

447   *Staphylococcus-* into a *Corynebacterium/Dolosigranulum-*, towards a *Moraxella*-dominated

448   profile eventually occurs in the great majority of children, we here show that especially lack of

449   *Corynebacterium*/*Dolosigranulum* spp. establishment coincides with a premature transition

450   from *Staphylococcus-* towards a *Moraxella*-dominated profile, which is associated with influx

451   of oral bacterial species and an increased risk of RTIs (20). In line, several studies in mice have

452   demonstrated that the neonatal immune system requires cues from the respiratory microbiota

453   for its development within a specific time frame (12, 13). Indeed, premature *Moraxella* spp.

454   colonization is shown to induce a mixed pro-inflammatory immune response (31), although

455   data on the effects of *Moraxella* spp. colonization at later age are lacking. In addition, it

456   deserves further study whether the required microbial triggers might be species and/or strain

457   specific.

458

459   In our prospective, birth cohort study we collected frequent nasopharyngeal samples of a large

460   number of healthy children at regular intervals over the first year of life as well as during RTIs,

461   allowing us to study the microbial development during health, preceding and during RTI

462   episodes. More importantly, it allowed us to explore microbiota dynamics and drivers of

463   susceptibility to RTIs. Strengths of our study include the frequency of sampling and the

464   consistency in data and sample collection by trained doctors and research nurses. We made a

rigorous effort to minimize the potential effect of environmental contamination on low-density nasopharyngeal samples collected from children at very early age. Last, we used non-parametric, machine-learning techniques combined with (multivariable) spline-based mixed models to explore specific age-dependent patterns in microbial succession.

Our study also has limitations. First, parents were asked to contact the research team in case of a RTI. Therefore, likely not all RTI episodes may have been captured for in depth analyses. Exhaustive efforts were however made to obtain detailed information on all experienced RTIs when questionnaires were filled out during regular home visits to minimize reporting bias in our multivariable analyses (Bosch *et al*, unpublished data). Second, despite frequent sampling, our samples capture snapshots of a highly dynamic and developing microbiome, therefore we can only make assumptions about the dynamics in between sampling moments. Third, although we observed that microbiota changes seem to forego RTIs and are associated with RTI susceptibility, our study design precludes any definite statements on causality.

We here provide evidence that accelerated microbiota maturation is associated with microbiota instability and number of RTIs over the first year of life. These changed dynamics could be observed as early as within the first month of age, i.e. prior to the first RTI experiences. We also were able to link the impact of known important drivers such birth mode, feeding type, the presence of siblings, early day-care attendance, and recent use of antimicrobial therapy, via altered microbiota development to susceptibility to RTIs. The potential implications of these findings for our understanding of pathogenesis of disease, as well as diagnostic and preventive strategies, deserves further investigation.

References

1. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, Lawn JE, Cousens S, Mathers C, Black RE. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet* 2016;388:3027–3035.

2. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:743–800.

3. Donnelly JP, Baddley JW, Wang HE. Antibiotic utilization for acute respiratory tract infections in U.S. emergency departments. *Antimicrob Agents Chemother* 2014;58:1451–1457.

4. Bosch AATM, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, de Steenhuijsen Piters WAA, de Groot P-KCM, Pernet P, Keijser BJF, Sanders EAM, Bogaert D. Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. *EBioMedicine* 2016;9:336–345.

5. Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat Med* 2017;23:314–326.

6. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, D Lieber A, Wu F, Perez-Perez GI, Chen Y, Schweizer W, Zheng X, Contreras M, Dominguez-Bello MG, Blaser MJ. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* 2016;8:343ra82–343ra82.

7. Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E, Bruin J, Montijn R, Bonten M, Sanders E. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS ONE* 2011;6:e17035.

8. Biesbroek G, Bosch AATM, Wang X, Keijser BJF, Veenhoven RH, Sanders EAM, Bogaert D. The Impact of Breastfeeding on Nasopharyngeal Microbial Communities in Infants. *Am J Respir Crit Care Med* 2014;190:298–308.

9. Blaser MJ. Antibiotic use and its consequences for the normal microbiome. *Science* 2016;352:544–545.

10. Biesbroek G, Tsivtsivadze E, Sanders EAM, Montijn R, Veenhoven RH, Keijser BJF, Bogaert D. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am J Respir Crit Care Med* 2014;190:1283–1292.

11. Vissing NH, Chawes BLK, Bisgaard H. Increased risk of pneumonia and bronchiolitis after bacterial colonization of the airways as neonates. *Am J Respir Crit Care Med* 2013;188:1246–1252.

12. Gollwitzer ES, Saglani S, Trompette A, Yadava K, Sherburn R, McCoy KD, Nicod LP, Lloyd CM, Marsland BJ. Lung microbiota promotes tolerance to allergens in neonates via PD-L1. *Nat Med* 2014;20:642–647.

13. Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Siebert R, Baron RM, Kasper DL, Blumberg RS. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* 2012;336:489–493.

14. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science* 2012;336:1268–1273.

15. Kamada N, Chen GY, Inohara N, Núñez G. Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol* 2013;14:685–690.

16. Man WH, de Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* 2017;8:51.

558 17.   Pettigrew MM, Laufer AS, Gent JF, Kong Y, Fennie KP, Metlay JP. Upper respiratory
559   tract microbial communities, acute otitis media pathogens, and antibiotic use in healthy
560   and sick children. *Appl Environ Microbiol* 2012;78:6262–6270.
561 18.   Hilty M, Qi W, Brugger SD, Frei L, Agyeman P, Frey PM, Aebi S, Muhlemann K.
562   Nasopharyngeal microbiota in infants with acute otitis media. *J Infect Dis*
563   2012;205:1048–1055.
564 19.   de Steenhuijsen Piters WAA, Heinonen S, Hasrat R, Bunsow E, Smith B, Suarez-
565   Arrabal M-C, Chaussabel D, Cohen DM, Sanders EAM, Ramilo O, Bogaert D, Mejias
566   A. Nasopharyngeal Microbiota, Host Transcriptome, and Disease Severity in Children
567   with Respiratory Syncytial Virus Infection. *Am J Respir Crit Care Med*
568   2016;194:1104–1115.
569 20.   Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, Holt BJ, Hales BJ, Walker
570   ML, Hollams E, Bochkov YA, Grindle K, Johnston SL, Gern JE, Sly PD, Holt PG,
571   Holt KE, Inouye M. The infant nasopharyngeal microbiome impacts severity of lower
572   respiratory infection and risk of asthma development. *Cell Host and Microbe*
573   2015;17:704–715.
574 21.   de Steenhuijsen Piters WAA, Sanders EAM, Bogaert D. The role of the local microbial
575   ecosystem in respiratory health and disease. *Philosophical Transactions of the Royal*
576   *Society B: Biological Sciences* 2015;370:20140294.
577 22.   O'Brien KL, Nohynek H, World Health Organization Pneumococcal Vaccine Trials
578   Carraige Working Group. Report from a WHO working group: standard method for
579   detecting upper respiratory carriage of Streptococcus pneumoniae. *Pediatr Infect Dis J*
580   2003;22:133–140.
581 23.   Wyllie AL, Chu MLJN, Schellens MHB, van Engelsdorp Gastelaars J, Jansen MD, van
582   der Ende A, Bogaert D, Sanders EAM, Trzciński K. Streptococcus pneumoniae in
583   Saliva of Dutch Primary School Children. *PLoS ONE* 2014;9:e102045.
584 24.   Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, Benezra A,
585   DeStefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q,
586   Province MA, Petri WA, Ahmed T, Gordon JI. Persistent gut microbiota immaturity in
587   malnourished Bangladeshi children. *Nature* 2014;510:417–421.
588 25.   Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N,
589   Knight R. Delivery mode shapes the acquisition and structure of the initial microbiota
590   across multiple body habitats in newborns. *Proc Natl Acad Sci USA* 2010;107:11971–
591   11975.
592 26.   Laufer AS, Metlay JP, Gent JF, Fennie KP, Kong Y, Pettigrew MM. Microbial
593   communities of the upper respiratory tract and otitis media in children. *mBio*
594   2011;2:e00245–10.
595 27.   Hunt KM, Foster JA, Forney LJ, Schütte UM, Beck DL, Abdo Z, Fox LK, Williams
596   JE, McGuire MK, McGuire MA. Characterization of the diversity and temporal
597   stability of bacterial communities in human milk. *PLoS ONE* 2011;6:e21313.
598 28.   Bomar L, Brugger SD, Yost BH, Davies SS, Lemon KP. Corynebacterium accolens
599   Releases Antipneumococcal Free Fatty Acids from Human Nostril and Skin Surface
600   Triacylglycerols. *mBio* 2016;7:.
601 29.   Prevaes SMPJ, de Winter-de Groot KM, Janssens HM, de Steenhuijsen Piters WAA,
602   Tramper-Stranders GA, Wyllie AL, Hasrat R, Tiddens HA, van Westreenen M, van der
603   Ent CK, Sanders EAM, Bogaert D. Development of the Nasopharyngeal Microbiota in
604   Infants with Cystic Fibrosis. *Am J Respir Crit Care Med* 2016;193:504–515.
605 30.   Holgerson PL, Vestman NR, Claesson R, Ohman C, Domellöf M, Tanner ACR,
606   Hernell O, Johansson I. Oral microbial profile discriminates breast-fed from formula-
607   fed infants. *J Pediatr Gastroenterol Nutr* 2013;56:127–136.

608   31.   Følsgaard NV, Schjørring S, Chawes BL, Rasmussen MA, Krogfelt KA, Brix S,
609   Bisgaard H. Pathogenic bacteria colonizing the airways in asymptomatic neonates
610   stimulates topical inflammatory mediator release. *Am J Respir Crit Care Med*
611   2013;187:589–595.
612   32.   Clarke KR. Non-parametric multivariate analyses of changes in community structure.
613   *Australian journal of ecology* 1993;doi:10.1111/j.1442-9993.1993.tb00438.x/full.
614
615

Figure 1 – Microbiota development over the first year of life.

(A) Relative abundance of the 15 highest ranking OTUs over the first year of life (age in days) and during of RTIs. OTUs are colour coded as indicated in the figure legend, which was based on their phylum level taxonomic annotation: red, Firmicutes; yellow, Actinobacteria; blue, Proteobacteria and green, Bacteroidetes. We observed a high abundance of Firmicutes (*Staphylococcus (3)* and *Dolosigranulum (4)*) and Actinobacteria (*Corynebacterium* spp.) early in life, which was gradually replaced by Proteobacteria (*Moraxella (1)*, *Moraxella (7)*, *Haemophilus (6)* and *Neisseria* spp.). OTUs that were not among the 15 highest ranking were collapsed and referred to as 'Residuals', stratified by phylum for the five most abundant phyla.

(B) Relative abundance of the 15 highest ranking OTUs over the first two months of life. Visualisation of microbiota profiles per time point allows for a more detailed assessment of microbial dynamics at early time points. Over the first week of life, a relatively high abundance of *Streptococcus (5)*, *Janthinobacterium (13)* and *Neisseria* spp. and *Rothia (12)* was observed, apart from other OTUs belonging mainly to the Firmicutes, Proteobacteria and Actinobacteria phyla (See Figure E5). d = day; w = week; m = month.

Figure 2 – Non-metric multidimensional scaling (nMDS) plot visualizing the microbiota succession patterns in the first year of life.

Each point represents the microbial community composition of one sample. Samples taken during health (n=1,065) are coloured based on the age at which they were taken (colours ranging from yellow [day 0] to dark green [year 1]). In addition, samples taken during RTI are depicted (n=56; dark red). The standard deviation of data points within time point/RTI strata is shown by ellipses. The 15 highest ranked OTUs were simultaneously visualized (triangles).

The size of the triangles is relative to the mean relative abundance of the OTU it represents. The stress value indicates how well the high-dimensional data are captured in the two-dimensional space; a value of ~0.2 indicates that the representation of some points is potentially misleading and that a representation in a higher dimensional space might be more appropriate (see Figure E4 for detailed assessment) (32). d = day; w = week; m = month; RTI = respiratory tract infection.

Figure 3 – Absolute bacterial density over the first year of life.

Boxplots showing the absolute bacterial density (in pg/µL 16S-rRNA-gene) in blanks (n=55; blue), in samples taken during health at various time points (n=1,065; colours ranging from yellow [day 0] to dark green [year 1]) and during RTI (n=56; red). Bacterial density is particularly low at days 0 and 1, then gradually increases until the age of ~1 month, after which it remained largely stable. Box plots represent the 25th and 75th percentiles (lower and upper boundaries boxes, respectively), the median (middle horizontal line), and measurements that fall within 1.5 times the interquartile range (IQR; distance between 25th and 75th percentiles; whiskers) or outside 1.5 times the IQR (points). Q-values were derived from a linear mixed model with $\log_{10}$-transformed bacterial density as outcome variable, time point as fixed effect and subject as a random effect. Only samples taken at regular intervals were considered and each consecutive time point was compared to the previous time point using the *multcomp* package. ***, q-value <0.001; **, 0.001 ≤q-value <0.01. d = day; w = week; m = month; RTI = respiratory tract infection.

Figure 4 – Microbiota maturation and age-discriminatory taxa stratified by RTI susceptibility.

(A) Microbiota age estimates plotted against chronological age stratified by number of RTIs experienced during the first year of life. The curves represent smooth spline fits for each cohort.

662   P-values are based on a linear mixed model, including age (spline) and number of RTIs (i.e. 0-

663   2, 3-4 or 5-7 RTIs) as fixed effects and subject as random effect.

664   **(B)** Heatmap of the mean relative abundance of the 22 age-discriminatory OTUs against

665   moment of sampling in each cohort. OTUs are ordered vertically based on average linkage

666   hierarchical clustering using the Euclidean distance matrix. Colours correspond with row wise

667   normalized relative abundances (i.e. red indicates the maximum relative abundance of that OTU

668   over all cohorts, black indicates the minimum relative abundance). OTU-names are bold and

669   coloured green if they were significantly enriched in the reference group (0-2 RTIs) compared

670   to children with >2 RTIs. Red was used to denote the OTUs that were observed in higher

671   abundance in children with >2 RTIs (based on SS-ANOVA q-values; see Table 4A). d = day;

672   w = week; m = month; RTI = respiratory tract infection.

673   **(C)** Relative microbiota age (RMA) before (light green shades), during (red) and after RTI

674   (dark green). The relative microbiota age two time points before RTI ('-2'; n=51; on average -

675   104 days to RTI), one time point before RTI ('-1'; n=47; -50 days to RTI), at RTI ('RTI'; n=56;

676   mean age at sampling of 216 days) and after RTI ('+1'; n=41; +57 days after RTI) is depicted

677   in boxplots (see legend Figure 2). RMA already increased at time points preceding a factual

678   RTI (median RMA +7.3 days at T = -2, +37.1 days at T = -1, and +67.8 days at RTI). P-values

679   are based on a linear mixed model including timing of sampling (i.e. '-2', '-1', 'RTI' or '+1')

680   and age (continuous) as fixed effects and subject as random effect. The contrasts '-2' vs '-1', '-

681   1 vs 'RTI' and 'RTI' vs '+1' were tested (*multcomp* package). **, $0.001 \leq$ q-value <0.01; *,

682    $0.01 \leq$ q-value <0.05.

683   **Figure 5** – Microbiota stability over time stratified by RTI susceptibility.

684   Bray-Curtis dissimilarities were calculated within each subject between each pair of

685   consecutive time points. The bacterial community stability was significantly lower in children

686   with 3-4 (p=0.005) or 5-7 RTIs (p=0.02) compared to the reference group of children

687   experiencing 0-2 RTIs within the first year of life. P-values are based on a linear mixed model,

688   including age (spline) and number of RTIs as fixed effects and subject as random effect. The

689   shaded area around each smoothing spline represents the 95% confidence interval.

690

Legends Online Supplement

Methods – Online supplement methods.

Figure E1 – Flow chart study.

Flow chart showing the number of initially enrolled women and the reasons for exclusion of
participants.

Figure E2 – Distribution of respiratory tract infections within the cohort.

Histogram of the number of RTIs versus their frequency. 'N' denotes the number of individuals,
'n' gives the number of samples. The cohort was divided in RTI groups based on the distribution
of RTIs; each sub cohort corresponds with a tertile.

Figure E3 – OTU selection procedure.

(A) Plot showing the 10-fold cross-validation error (mean ±standard deviation) as a function of
the number of OTUs used to regress against chronological age in the reference cohort (children
with 0-2 RTIs). An optimal trade-off between the mean squared error (MSE; i.e. cross-
validation error) and number of OTUs in the model was observed at 22 OTUs.

(B) Age-discriminatory OTUs ranked in descending order based on their importance to the
accuracy of the model. OTU importance was estimated by calculating the increase in mean-
squared error (MSE) of the microbiota age prediction after randomly permuting the relative
abundance values of each given OTU (mean ±standard deviation, 100 replicates).

Figure E4 – nMDS diagnostic plots and three-dimensional nMDS.

712 **(A)** Scree plot to depict the relationship between the number of (nMDS)-dimensions and stress.

713 Naturally, the stress will reduce by increasing the number of dimensions, however only a

714 maximum number of three dimensions can reasonably be interpreted. Using three dimensions

715 the stress-value drops well below 0.2 (32), suggesting that a decent ordination of the data is

716 possible in this number of dimensions.

717 **(B)** Three-dimensional nMDS plot. The main data structure visualized using the two-

718 dimensional plot appears to be preserved when plotting the same data in three dimensions.

719 **Figure E5** – Relative abundance of early colonizing bacteria.

720 Bar plots visualizing the relative abundance (mean ±standard error of the mean) of the 10

721 highest ranking OTUs at each (early) time point (only considering day 0 and 1 and week 1 and

722 2). For each OTU, we calculated the significance of change in relative abundance for each pair

723 of consecutive time points (i.e. day 0 vs day 1, day 1 vs week 1 and week 1 vs week 2) using

724 mixed linear models including subject as random effect. Significant differences between

725 contrasts were determined using the *multcomp*-package. A Benjamini-Hochberg procedure was

726 used to correct for multiple comparisons (simultaneously considering all OTUs/contrasts). ***,

727 q-value <0.001; **, 0.001 ≤q-value <0.01; *, 0.01 ≤q-value <0.05.

728 **Figure E6** – Individual microbial developmental trajectories in time.

729 Using average linkage hierarchical clustering based on the Bray-Curtis dissimilarity matrix

730 samples were binned into 8 clusters consisting of ≥10 samples. These clusters were enriched

731 for *Moraxella (1)* (MOR1) *Corynebacterium (2)* and *Dolosigranulum (4)* (CDG),

732 *Staphylococcus (3)* (STA) and *Streptococcus (5)* (STR), *Moraxella (7)* (MOR7), *Haemophilus*

733 *(6)* (HAE), *Corynebacterium (8)* (COR8) and *Neisseria* spp. (NEI). The number of individuals

734 in each cluster at each time point was visualised in alluvial diagrams, which were stratified by

735 the number of RTIs experienced over the first year of life (i.e. **(A)** 0-2 RTIs, **(B)** 3-4 RTIs and

736 **(C)** 5-7 RTIs). The height of the figures corresponds with the total number of samples within

737 that group. In addition, the height of the nodes and the thickness of the lines connecting the

738 nodes is proportional to the number of samples. We observed that the CDG-cluster is

739 underrepresented in children who experienced 5-7 RTIs over time. Instead, the early-life STA-

740 cluster rapidly transitions into the MOR1-cluster, which is associated with older ages.

741 **Figure E7** – Relative abundance of age-discriminatory taxa at each time point. The line plots

742 indicate the microbiota development for each age-discriminatory taxum. Dots represent mean

743 relative abundance at a given time point within the stratum and whiskers depict the standard

744 error of the mean. See Table E4 for statistical assessment. d = day; w = week; m = month.

745 **Figure E8** – Relative microbiota age stratified by time point and RTI cohort.

746 Boxplots (see legend Figure 3) depicting relative microbiota age (RMA) for each cohort. The

747 RMA was significantly higher in children who experienced 5-7 RTIs compared to the reference

748 group, after adjusting for either age or sampling moment (both p=0.007).

749 **Figure E9** – Relative abundance of age-discriminatory taxa before (light green shades), during

750 (red) and after RTI (dark green; see also legend figure 4C). Relative abundances were depicted

751 using boxplots (see legend Figure 3). We tested the statistical significance of differences in

752 microbial abundance between sampling moments using a linear mixed model including timing

753 of sampling (i.e. '-2', '-1', 'RTI' or '+1') and age (continuous) as fixed effects and subject as

754 random effect. The contrasts '-2' vs '-1', '-1 vs 'RTI' and 'RTI' vs '+1' were tested (*multcomp*

755 package). Although we did observe changes in abundance of individual OTUs that appeared to

756 be related to changes in RMA, these changes were not statistically significant (after adjusting

757 for multiple testing).

758   **Figure E10** – α-diversity measures stratified by RTI susceptibility.

759   We tested the number of observed species, Simpson and Shannon diversity indices. No

760   significant differences between RTI groups were observed. Points represent mean values and

761   whiskers depict the standard error of the mean. P-values were derived from mixed linear models

762   with subject as random effect and adjusted for age (spline); $p > 0.05$). d = day; w = week; m =

763   month.

764   **Figure E11** – Relative abundance of the 15 highest ranking OTUs during the first year of life

765   stratified by birth mode and feeding type - flow diagram.

766   See legend Figure 1A. We observed an increased relative abundance of *Corynebacterium (2)*

767   and *Dolosigranulum (4)* until the age of five months and late *Moraxella* spp. enrichment in

768   children vaginally delivered and/or breastfed. Birth by caesarean section was associated with

769   early *Staphylococcus (3)* predominance. Feeding type was studied as a categorical variable

770   indicating whether children were exclusively breastfed (BF) up to the age of three months (3m).

771   See Table E6 and E7 for statistical assessment.

772   **Figure E12** – Relative abundance of the 15 highest ranking OTUs during the first year of life

773   stratified by birth mode and feeding type - line plots.

774   **(A)** Line plots indicating the microbiota succession patterns of abundant taxa, stratified by birth

775   mode (caesarean section vs vaginal). Points represent means and whiskers represent standard

776   errors of the mean. See Table E6 for statistical assessment. d = day; w = week; m = month; RTI

777   = respiratory tract infection.

778   **(B)** Line plots indicating the microbiota succession patterns of abundant taxa, stratified by

779   feeding type (exclusive breastfeeding up to the age of three months yes/no). Points represent

780   means and whiskers represent standard errors of the mean. See Table E10 for statistical

781   assessment. d = day; w = week; m = month; RTI = respiratory tract infection.

Figure 1

A

B
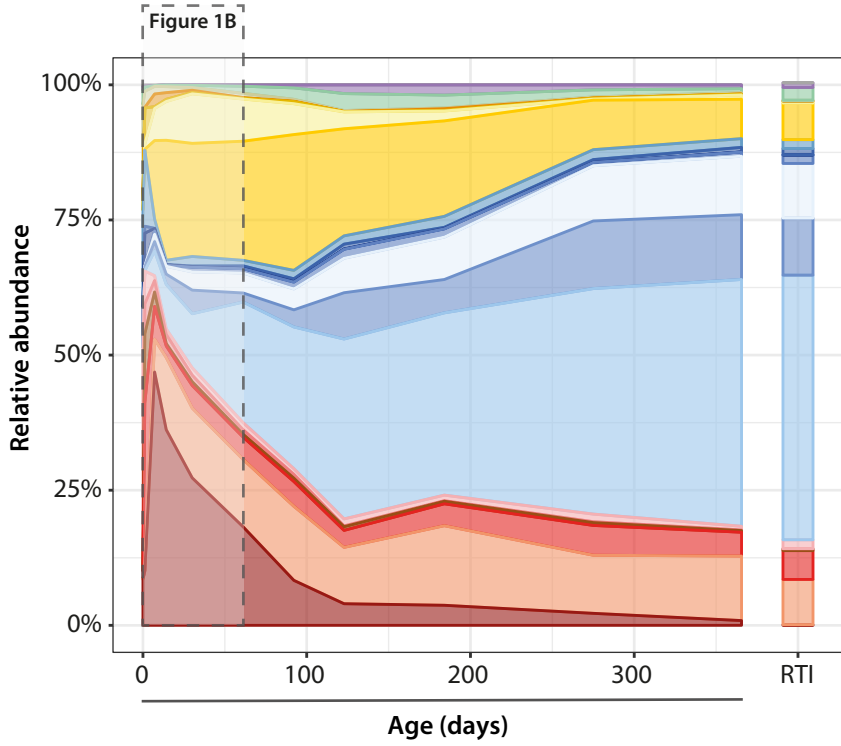
Firmicutes
- *Staphylococcus (3)*
- *Dolosigranulum (4)*
- *Streptococcus (5)*
- *Gemella (9)*
- *Streptococcus (10)*
- Residuals Firmicutes

Proteobacteria
- *Moraxella (1)*
- *Haemophilus (6)*
- *Moraxella (7)*
- *Neisseria (11)*
- *Janthinobacterium (13)*
- *Neisseria (14)*
- Residuals Proteobacteria
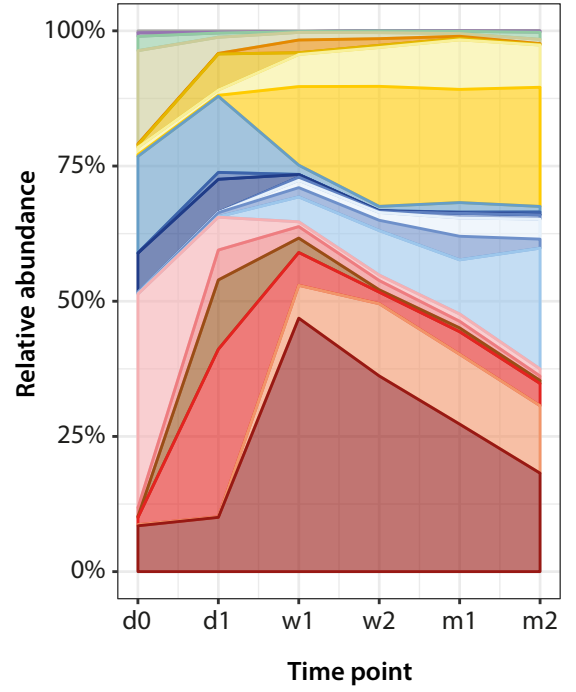
Actinobacteria
- *Corynebacterium (2)*
- *Corynebacterium (8)*
- *Rothia (12)*
- *Corynebacterium (15)*
- Residuals Actinobacteria
- Residuals Bacteroidetes
- Residuals Fusobacteria
- Residuals

Figure 2

Figure 3

Figure 4

A

3–4 vs 0–2: $p = 0.127$
5–7 vs 0–2: $p = 0.007$

Microbiota age (days) vs Chronological age (days)

Cohort: 0–2, 3–4, 5–7, RTI

C

RMA (days) vs Time point (−2, −1, RTI, +1)

**

*

normalised OTU abundance

Min. Mean Max.

B

OTU vs Time point

0–2 RTIs | 3–4 RTIs | 5–7 RTIs

Gemella (9)
Streptococcus (10)
Actinomyces (78)
Acidovorax (52)
Janthinobacterium (13)
Staphylococcus (3)
Corynebacterium (8)
Dolosigranulum (4)
Corynebacterium (80)
Corynebacterium (2)
Leptotrichia (81)
Alloprevotella (19)
Prevotella (39)
Leptotrichia (104)
Neisseria (14)
Neisseria (11)
Neisseriaceae (23)
Granulicatella (40)
Moraxella (49)
Moraxella (74)
Moraxella (7)
Moraxella (1)

d0 d1 w1 w2 m1 m2 m3 m4 m6 m9 m12 RTI
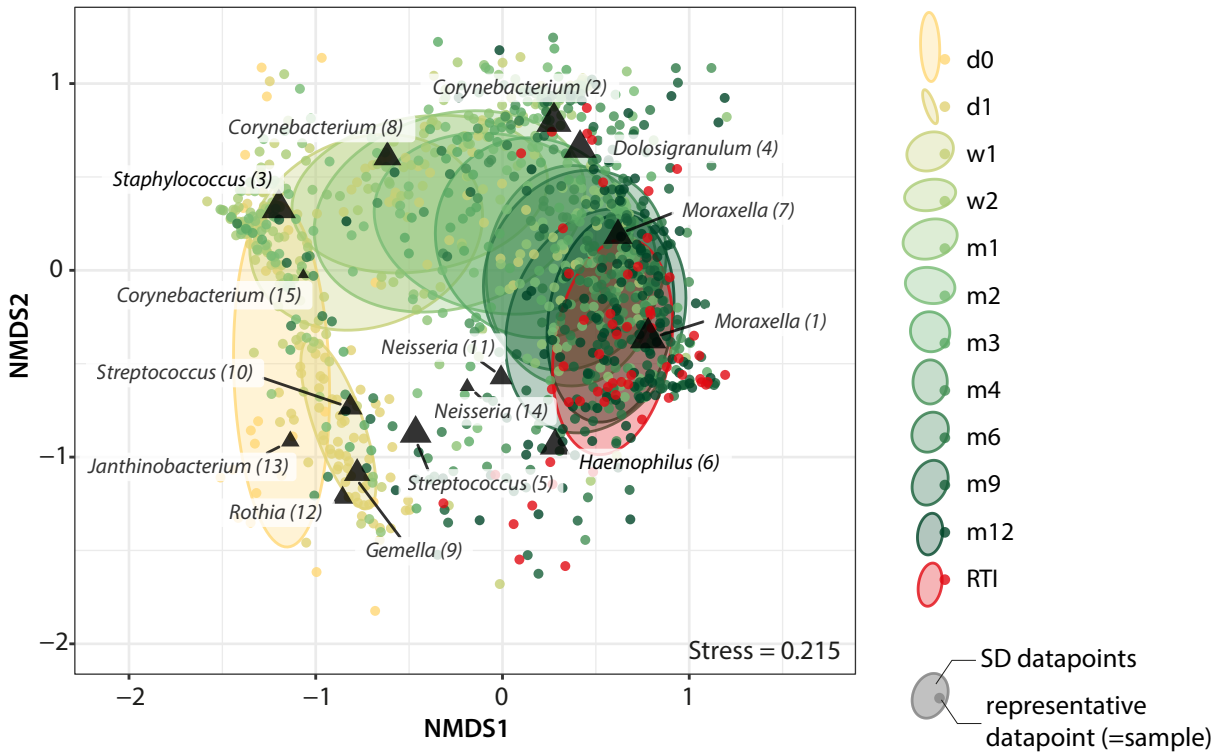
Figure 5

1 **ONLINE SUPPLEMENT**

2

3 **Title:** Maturation of the infant respiratory microbiota, environmental drivers and health

4 consequences: a prospective cohort study

5

6 **Authors:** Astrid A.T.M. Bosch[†], Wouter A.A. de Steenhuijsen Piters[†], Marlies A. van Houten,

7 Mei Ling J.N. Chu, Giske Biesbroek, Jolanda Kool, Paula Pernet, Pieter-Kees C.M. de Groot,

8 Marinus J.C. Eijkemans, Bart J.F. Keijser, Elisabeth A.M. Sanders and Debby Bogaert*

9

10 [†] These authors contributed equally to this work

11 * Corresponding author. Email: D.Bogaert@ed.ac.uk.

12

13 **This file includes:**

14 (Online supplement) Methods

15

16 **Other supplementary materials for this manuscript include the following:**

17 Figures E1-E12, as a separate pdf-file.

18 Tables E1-E7, including captions, as a separate excel spreadsheet.

19

20

Methods

Study population

Nasopharyngeal swabs were collected from healthy children who participated in an ongoing prospective birth cohort study. The primary aim of this population-based study is to investigate the development and dynamics of the microbiota in infants during health and disease, with special interest in the impact of mode of delivery on microbial succession. Since approximately 15% of the Dutch children are born by caesarian section (E1), the cohort is enriched by caesarian section deliveries with the aim to obtain a ~50/50 distribution between caesarian section born children and vaginally delivered children. The study is conducted in the Netherlands, a small country (approximately 17 million inhabitants) in North-Europe with high socio-economic standards and a moderate sea climate characterized by cool summers and mild winters.

The trials' methods have been described elsewhere (E2). In short, healthy, term born newborns (gestational age >37 weeks) were enrolled in the study directly after birth. Exclusion criteria at baseline were major congenital anomalies, severe maternal or neonatal complications during birth, language barrier, intention to move outside the research area, or parents under the age of 18 years. Written informed consent was obtained from both parents before birth of the child. Participants did not receive any financial compensation. An acknowledged national Ethics Committee in the Netherlands (METC Noord-Holland, committee on research involving human subjects) approved the study (M012-015, NH012.394, NTR3986). The study was conducted in accordance with the European Statements for Good Clinical Practice. We estimated 10-20% loss to follow-up, therefore we had ethical approval to replace participants in case they dropped out of the study before six months of follow-up. Eventually we had complete datasets up to one

45   year of age of 116 participants. These infants were born between December 19th, 2012 and

46   November 2st, 2014.

47   Of these 116 children, we had at least eight samples of good quality available for 112 children

48   after laboratory work-up (Figure E1).

49

50   Our study was powered to detect differences in microbial communities between vaginally born

51   children and children born by caesarian-section, which was the primary aim of the cohort study.

52   We performed power calculations aiming to be able to detect at least two-fold differences in at

53   least the top 25 most common bacteria after correction for multiple testing. Given the variability

54   and spread in abundance of OTUs we calculated that 40 children per group would give us

55   sufficient power (>80%) to address our primary research question. Because the inclusion rate

56   of caesarian-born children was lower than expected, we were allowed to extend the enrollment

57   period, resulting in a much larger sample size (N=128) than initially expected, enabling us to

58   thoroughly investigate secondary outcomes, such as the association between microbiota

59   differences and the number of RTIs, again providing us with sufficient power to analyze group

60   sizes of approximately 40 children per group.

61

62   *Data collection*

63   Home visits were conducted directly after birth, 24 hours after birth, at seven days, 14 days,

64   and one, two, three, four, six, nine, and 12 of months of age. Postpartum visits were all

65   performed within two hours from birth and all day one samples were obtained within 24-36

66   hours after delivery. For logistic reasons (sampling preferably during office hours and

67   considering parental vacations), we allowed some flexibility for the remaining sample

68   moments: all week one samples were obtained within 5-9 days (mean 7); all week two samples

69   between 12-17 days (mean 14); month one samples between 23-27 days (mean 30), months two

70   samples between 49-73 days (mean 61), months three samples between 83-11 days (mean 92),

71   months four samples between 112-133 days (mean 123), months six between 177-197 days

72   (mean 184), months nine samples between 260-288 (mean 275), months 12 samples between

73   358-382 (mean 366) days postpartum, resulting in no overlap between sample moments (see

74   Figure 4A).

75

76   Each home visit, nasopharyngeal samples were obtained by trained doctors and research nurses

77   in a semi-sterile setting as previously described (E2). In short, deep nasopharyngeal swabs were

78   collected trans nasally using a flexible, sterile swab (Copan eSwab, 484CE). Directly after

79   sampling, the swabs were snap-frozen and stored in a sterile, filtered solution (10% Glycerol

80   (VWR international BV 1.04093.1000) in 0.1% DEPC water (SERVA Electrophoresis,

81   39798.03). The swabs were transported on dry ice and stored at -80°C until further analyses. In

82   addition, the research team completed an extensive survey on the health status of the child and

83   environmental factors, including breastfeeding, crowding conditions, and medication use.

84

85   Next to these regular and frequent visits, parents were asked to contact the study team in case

86   of an active respiratory tract infections, defined as fever ≥38°C (per rectal measurement) for >6

87   hours combined with general unwell feeling and presence of RTI symptoms, including earache,

88   cough, hoarseness, wheeze, dyspnoea and/or runny nose. During an extra home visit (RTI visit

89   within 48 hours after start of the fever), we collected additional nasopharyngeal samples using

90   the same procedure as described above and obtained information about the duration of the fever,

91   RTI symptoms, and antibiotic use. In addition, the research team called parents two to four

92   weeks after the RTI visit to complete the questionnaire. Since 15 of the children had respiratory

93   symptoms with fever during one of the regular visits, these were also considered as a RTI

94   episode in the analyses.

95

Bacterial DNA isolation and quantification

97 Bacterial DNA from 200 μ l sample was isolated by bead-beating in phenol (E3) and quantified

98 using a qPCR with primers directed at the 16S-rRNA gene (E4, 5). DNA was then eluted in

99 two aliquots of 25 µl elution buffer and stored at -20°C until further analyses.

100

16S-rRNA gene amplicon sequencing

102 PCR amplicon libraries were generated by amplification of the 16S ribosomal RNA gene using

103 barcoded primers directed at the V4 hypervariable region, as previously described (E2). Primer

104 pair 533F/806R was used for amplification. Amplicon pools from samples and controls were

105 sequenced in eight runs using an Illuminia MiSeq instrument, resulting in paired-end 200 or

106 250 nucleotide reads. We first trimmed all reads to a length of 200 nucleotides (Fastx toolkit,

107 version 0.0.13) and then applied an adaptive, window-based trimming algorithm (Sickle,

108 version 1.33) (E6) using a quality threshold of Q30 and a length threshold of 150 nucleotides

109 to filter out low quality reads/nucleotides. We aimed to further reduce the number of sequence

110 errors in the reads by applying an error correction algorithm (BayesHammer, SPAdes genome

111 assembler toolkit, version 3.5.0) (E7). After quality filtering and error correction, reads were

112 assembled into contigs (PANDAseq, version 2.9) (E8, 9) and demultiplexed (Qiime version

113 1.9.1; split_libraries.py) (E10). We removed singleton sequences (1.4%) and identified

114 chimeras using both *de novo* and reference chimera identification (UCHIME; 3.2%). After

115 removal of chimeric sequences, VSEARCH abundance-based greedy clustering was used to

116 pick OTUs at a 97% identity threshold (E11). OTUs were then annotated by the Naïve Bayesian

117 RDP classifier (version 2.2) (E12) with a classification confidence of 50% (default) (E13) and

118 annotations were based on the 97% identity SILVA 119 release reference database (E14). The

119 SILVA-annotations for the most abundant/age-discriminatory taxa were verified using

120   BLASTN (E15) (Table E1). In the main text we further refer to OTUs using maximum genus

121   level annotations, combined with a rank number based on the abundance of each given OTU.

122

123   *Data normalisation and filtering*

124   We generated an abundance-filtered dataset by including only those OTUs that were present at

125   or above a confident level of detection (0.1% relative abundance) in at least two samples,

126   retaining 576 OTUs (0.3% of reads excluded) (E16). We generated a rarefied OTU-table at a

127   sequence depth of 3,500 reads, calculated the relative abundance of OTUs and used this table

128   as input for downstream analyses, including visualisations, random forest modelling and

129   stability analyses. α-diversity measures were calculated for 100 rarefactions at a sequencing

130   depth of 3500 reads and averaged. Raw read counts were normalised intrinsically using

131   cumulative sum scaling (CSS) if modelling was performed using the *metagenomeSeq* package

132   and the *fitTimeSeries* function (E17). Using this function, the temporal associations between

133   each of the 22 age-discriminatory taxa and risk of RTIs were assessed; only significant results

134   were reported. For the analyses on the temporal effects of birth mode and feeding type, OTUs

135   with >10 sequences in ≥50 samples were included. Similarly, for analyses based on Multivariate

136   Association with Linear Models (MaAsLin), we selected OTUs from the rarefied OTU-table

137   with a relative abundance of >0.1% in ≥50 samples. Next, de OTU-table was expanded by

138   calculating the cumulative relative abundance of the selected OTUs at all taxonomic levels (i.e.

139   ranging from species/OTU-level to kingdom level). β-diversity was assed using the Bray-Curtis

140   dissimilarity metric.

141

142   *Quality control of 16S-rRNA gene amplicon sequencing*

143   URT samples, especially in very young children, are typically low in bacterial density (E18),

144   and therefore measures to control for potential contamination with environmental of DNA are

145   of vital importance. Since we were particularly interested in the initial colonization patterns of

146   the children in our cohort, we set out to discern samples with a high likelihood of environmental

147   contamination, from those samples that did not resemble negative DNA blanks through an

148   unsupervised clustering approach. Both low DNA samples (0.2 pg/µl-0.5 pg/µl) and blanks

149   (n=50; 30 excluded because of too low sequence depth) were rarefied to a depth of 2,000 reads

150   and subjected to average linkage hierarchical clustering based on the Bray-Curtis dissimilarity

151   (100 repeats). For each repeat, we used the maximum Silhouette index to determine the optimal

152   number of clusters (up to 20 clusters tested). Samples that co-clustered with DNA blanks in

153   >5% of the repeats were excluded from subsequent analyses, together with samples that were

154   sequenced twice, samples with a density of <0.2 pg/µl or read counts <3,500 sequences, and

155   samples of individuals that were lost to follow-up and/or had <8 samples available (in total 211

156   samples excluded), resulting in 1,121 samples from 112 individuals. Sequence data of part of

157   the samples (≤6 months) of part of the children (743 samples, 101 individuals) were used for a

158   previous study on the role of mode of delivery on early respiratory microbiota development

159   (E2).

160

161   In addition, we included 14 mock communities, consisting of 12 bacterial species commonly

162   observed in the upper respiratory tract (i.e. *Bacteroides fragilis*, *Haemophilus influenzae*,

163   *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Klebsiella oxytoca*, *Klebsiella*

164   *pneumoniae,* haemolytic *Streptococcus* group A*, Pseudomonias aeruginosa*, *Staphylococcus*

165   *epidermidis*, *Staphylococcus aureus* and *Moraxella catarrhalis*). Equivalent amounts of DNA

166   isolated from these species were combined and included as internal controls in the Illumina

167   MiSeq runs.

168

169   *Statistical analysis*

170   All analyses were performed in the R version 3.3.0 within R studio version 0.99.902. All figures

171   were created using the *ggplot2* R-package and edited using Illustrator CC. We corrected for

172   multiple testing if applicable using the Benjamini-Hochberg procedure (resulting in corrected

173   P-values or q-values; *p.adjust* function). 'Environmental factors' used in the descriptions of the

174   various models below comprises birth mode, breast feeding until three months of age, day care

175   attendance, presence of siblings under five years of age, antibiotic treatment in the previous

176   four weeks and season of birth, if not specified otherwise.

177

178   *Baseline tables*

179   Baseline tables were created using the *tableone* package (E19). Continuous variables were

180   tested for normality using a Shapiro-Wilk test. Variables with a non-normal distribution were

181   characterised using a median and interquartile range and the statistical significance of

182   differences between groups was calculated using a Mann-Whitney U or Kruskal-Wallis test.

183   Normally distributed variables were summarised by a mean and standard deviation and

184   differences were tested for significance using a Student's t-test/analysis of variance (ANOVA).

185   For categorical variables, we used a Chi-square to test for statistically significant differences

186   between groups. A Fisher's exact test was used for categorical variables if the expected cell

187   count was less than five.

188

189   *Non-metric multidimensional scaling and multivariate modelling*

190   Microbial succession patterns were visualised using non-metric multidimensional scaling

191   (nMDS; *metaMDS* function in the *vegan* package; trymax=1,000) (E20) based on the Bray-

192   Curtis dissimilarity matrix. Ellipses were calculated using the *veganCovEllipse* function and

193   represent the standard deviation of data points. Stress-values, which indicate how well the

194   ordination captured the high-dimensional data (i.e. a measure of goodness-of-fit), were

195 reported. We tested whether a nMDS-visualisation in a higher dimensional space would

196 decrease the stress of the ordination using a scree plot (1-6 dimensions tested). Based on our

197 findings (balancing number of dimensions, reduction in stress-value and interpretability of the

198 plot) we decided to provide a three-dimensional nMDS plot as a supplementary figure.

199

200 To quantify the effect of environmental variables and number of RTIs on the overall microbiota

201 composition we performed permutational multivariate analysis of variance (PERMANOVA)-

202 tests (*adonis* function of the *vegan* package; Bray-Curtis dissimilarity, 999 permutations). To

203 assess the robustness of our findings based on one rarefied OTU-table, we reran the same

204 PERMANOVA-tests on 100 rarefied OTU-tables and compared the effect size of the variables

205 under consideration across rarefactions (Table E3A and E3C).

206

207 *Clustering and alluvial diagram*

208 To complement our findings based on our group-level analyses, we additionally assessed

209 microbial development at the individual level. We first clustered individuals using unsupervised

210 average linkage hierarchical clustering based on the Bray-Curtis dissimilarity matrix. The

211 number of clusters was determined based on the Silhouette and Calinski-Harabasz indices (*fpc*

212 package) (E21). Clusters consisting ≥10 samples were considered for subsequent analyses. The

213 proportion of samples within each cluster at each time point was visualised using an alluvial

214 diagram (*ggvisSankey*-function within the *googleVis* package) (E22).

215 The alluvial diagram was stratified into three groups based on the normal distribution of RTIs

216 in the population; 39 children with 0-2 RTIs (reference group), 52 children with 3-4 RTIs and

217 21 children with 5-7 RTIs over the first year of life.

218

219   *Random forest modelling*

220   We hypothesized that the microbial succession patterns in the upper respiratory tract would be

221   altered in children who are more susceptible to RTIs. To investigate this hypothesis, we used a

222   machine learning technique referred to as random forest, which consists of an ensemble of

223   decision trees, each of which is built based on random partition of the data, using a random

224   selection of predictors (E23). We chose a random forest-approach over a more traditional,

225   reductionist approach where we would model individual OTUs, as we did not want to make

226   any assumptions on the highly variable relationships between specific OTUs and age (figure

227   E7). Also, OTU-abundance data is usually very sparse and overdispersed, which hinders the

228   application of traditional statistical techniques. Last, the random forest approach enabled us to

229   simultaneously model these challenging data, as well as reduce the dimensionality of the data,

230   the latter of which is essential to microbiota analysis.

231   To identify OTUs characteristic of a healthy microbiota maturation, we regressed the relative

232   abundance of the 576 OTUs observed against chronological age in the reference group (i.e. 0-

233   2 RTIs) using the *randomForest* package, (ntree=10,000, default mtry, defined as the number

234   of variables in the model divided by 3) (E24), as previously described (E16). The optimal

235   number of age-discriminatory taxa required for the prediction of microbiota age was determined

236   by calculating the cross-validated prediction performance of models with a sequentially reduced

237   numbers of variables (ranked by importance measured by the mean increase of squared error if

238   that variable would be removed from the model; *caret* package (E25); 100 iterations; Figure

239   E3A and E3B). This selection of OTUs was subsequently used as input to a random forest

240   model used to regress the relative abundance versus chronological age in the reference group

241   (resulting in the final model). We determined the importance of the reduced set of variables

242   based on the percentage increase in mean squared error after permuting the values for each

243   OTU (100 iterations). The final model was then used to predict chronological age, referred to

244 as 'microbiota age', in individuals who experienced 3-4 and 5-7 RTIs and on the group of

245 samples collected during RTIs. We used the *train* function in the 'caret' package (E25) to

246 determine cross-validated predictions of microbiota age for the healthy cohort (10 folds, 100

247 iterations, default mtry) to avoid reporting overfitted estimates (Figure 4A). The importance of

248 the age-discriminatory OTUs was visualised per cohort at each time point using a heatmap.

249 OTUs were vertically ordered based on an average linkage hierarchical clustering to visualise

250 the interrelations between OTUs. The colours of the heatmap were row-wise normalized (i.e.

251 red indicates the highest relative abundance of that OTU, black indicates the lowest value.). As

252 a post-hoc analysis, we studied the effect of the *Moraxella*-genus on the performance of the

253 microbiota age model by excluding the OTUs belonging to the *Moraxella*-genus from the model

254 while monitoring the amount of variance explained.

255

256 Since the relationship between chronological age and microbiota age was not linear, we

257 calculated the relative microbiota age as described before (E16). Relative microbiota age

258 (RMA) was calculated as follows: relative microbiota age = microbiota age of a given child –

259 microbiota age of children of similar age in the reference group (determined by a spline fit)

260 (E16).

261

262 *Linear mixed models*

263 Linear mixed models were used to assess the effect of fixed variables on a continuous dependent

264 variable, while including subject as a random intercept to adequately control for repeated

265 measures (*lmer* function of the *lme4* package) (E26). Separate models were used study the effect

266 of 1) environmental variables and 2) RTI susceptibility (defined as having experienced 0-2, 3-

267 4 or 5-7 RTIs during the first year of life) on relative microbiota age, Bray-Curtis-dissimilarity

268 and α-diversity measures. If a non-linear relationship between age and the dependent variable

269 was suspected, age was included in the model as a natural spline fit with five degrees of freedom

270 (*ns* function of the *splines* R-package). In addition, we assessed influence of sampling moment

271 on bacterial density ($\log_{10}$-transformed) and relative abundance (only first four time points)

272 using linear mixed models. Furthermore, we investigated the changes in RMA and the relative

273 abundance of age-discriminatory taxa at two time points before RTI, during RTI and at one

274 time point after RTI using a mixed linear model with RMA/OTU-abundance as outcome

275 variables and including timing of sampling (i.e. '-2', '-1', 'RTI' or '+1'), age as fixed effects

276 and subject as random effect. We did not consider interactions between variables in our models.

277 Post-hoc tests on contrasts of interest were performed using the *multcomp* package (E27).

278 Contrasts as specified in the main text were included and we adjusted for multiple testing using

279 the 'single-step' procedure (*multcomp* default), except when stated otherwise.

280

281 *Time series modelling*

282 To assess differences in abundance of OTUs between groups, we used smoothing spline

283 ANOVA as implemented in the *fitTimeSeries* function (E28) of the *metagenomeseq* R-package

284 (E17), which aims to model the differences in OTU-abundances between groups over time and

285 is able to not only test if differences exist, but also to evaluate the timing of these differences.

286 In addition, this function allows for the inclusion of a 'class'-effect, to adequately control for

287 repeated measures. Smoothing spline ANOVA models were used to study the (timing of)

288 differential abundance of age-discriminatory taxa determined by random forest between

289 children with 0-2 versus 3-4 RTIs and 0-2 versus 5-7 RTIs over the first year of life. In addition,

290 these models were used to assess the effect of birth mode and exclusive breastfeeding until

291 three months on the abundance of OTUs that passed the abundance filter, as these variables

292 likely have a temporary effect on microbial abundance. P-values were determined based on

293 1,000 permutations.

294

295 *Multivariable modelling*

296 To identify significant associations between environmental variables (as defined before) and

297 the relative abundance of OTUs in a multivariable manner, we used Multivariate Association

298 with Linear Models (MaAsLin). Age was included as a natural spline with five degrees of

299 freedom. Taxonomic entities simultaneously included in the models were OTUs that passed the

300 abundance filtering criterion and OTUs binned together at higher taxonomic levels (i.e. genus,

301 family, class, order, phylum and kingdom). We included subject as a random effect and ran the

302 models using default settings.

303

304 **References**

305 E1. Delnord M, Blondel B, Drewniak N, Klungsøyr K, Bolumar F, Mohangoo A, Gissler
306   M, Szamotulska K, Lack N, Nijhuis J, Velebil P, Sakkeus L, Chalmers J, Zeitlin J,
307   Euro-Peristat Preterm Group. Varying gestational age patterns in cesarean delivery:
308   an international comparison. *BMC Pregnancy Childbirth* 2014;14:321.
309 E2. Bosch AATM, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, de
310   Steenhuijsen Piters WAA, de Groot P-KCM, Pernet P, Keijser BJF, Sanders EAM,
311   Bogaert D. Development of Upper Respiratory Tract Microbiota in Infancy is
312   Affected by Mode of Delivery. *EBioMedicine* 2016;9:336–345.
313 E3. Wyllie AL, Chu MLJN, Schellens MHB, van Engelsdorp Gastelaars J, Jansen MD,
314   van der Ende A, Bogaert D, Sanders EAM, Trzciński K. Streptococcus pneumoniae
315   in Saliva of Dutch Primary School Children. *PLoS ONE* 2014;9:e102045.
316 E4. Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E, Bruin J, Montijn
317   R, Bonten M, Sanders E. Variability and diversity of nasopharyngeal microbiota in
318   children: a metagenomic analysis. *PLoS ONE* 2011;6:e17035.
319 E5. Biesbroek G, Tsivtsivadze E, Sanders EAM, Montijn R, Veenhoven RH, Keijser BJF,
320   Bogaert D. Early respiratory microbiota composition determines bacterial succession
321   patterns and respiratory health in children. *Am J Respir Crit Care Med*
322   2014;190:1283–1292.
323 E6. Joshi NA, Fass JN. Sickle: A sliding-window, adaptive, quality-based trimming tool
324   for FastQ files. 1st ed. at <https://github.com/najoshi/sickle.>.
325 E7. Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering
326   for error correction in single-cell sequencing. *BMC Genomics* 2013;14 Suppl 1:S7.
327 E8. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq:
328   paired-end assembler for illumina sequences. *BMC Bioinformatics* 2012;13:31.
329 E9. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. Insight into biases
330   and sequencing errors for amplicon sequencing with the Illumina MiSeq platform.
331   *Nucleic Acids Res* 2015;43:e37–e37.

332 E10. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK,
333   Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D,
334   Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J,
335   Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J,
336   Knight R. QIIME allows analysis of high-throughput community sequencing data.
337   *Nat Methods* 2010;7:335–336.
338 E11. Westcott SL, Schloss PD. De novo clustering methods outperform reference-based
339   methods for assigning 16S rRNA gene sequences to operational taxonomic units.
340   *PeerJ* 2015;3:e1487.
341 E12. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid
342   assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ*
343   *Microbiol* 2007;73:5261–5267.
344 E13. Claesson MJ, O'Sullivan O, Wang Q, Nikkilä J, Marchesi JR, Smidt H, de Vos WM,
345   Ross RP, O'Toole PW. Comparative analysis of pyrosequencing and a phylogenetic
346   microarray for exploring microbial community structures in the human distal
347   intestine. *PLoS ONE* 2009;4:e6669.
348 E14. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner
349   FO. The SILVA ribosomal RNA gene database project: improved data processing
350   and web-based tools. *Nucleic Acids Res* 2013;41:D590–6.
351 E15. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA
352   sequences. *J Comput Biol* 2000;7:203–214.
353 E16. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, Benezra A,
354   DeStefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q,
355   Province MA, Petri WA, Ahmed T, Gordon JI. Persistent gut microbiota immaturity
356   in malnourished Bangladeshi children. *Nature* 2014;510:417–421.
357 E17. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for
358   microbial marker-gene surveys. *Nat Methods* 2013;10:1200–1202.
359 E18. Biesbroek G, Sanders EAM, Roeselers G, Wang X, Caspers MPM, Trzciński K,
360   Bogaert D, Keijser BJF. Deep sequencing analyses of low density microbial
361   communities: working at the boundary of accurate microbiota detection. *PLoS ONE*
362   2012;7:e32942.
363 E19. Yoshida K, Bohn J. tableone: Create "Table 1" to Describe Baseline Characteristics.
364   0 ed. 2015;at <https://CRAN.R-project.org/package=tableone>.
365 E20. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson
366   GL, Solymos P, Stevens M, Wagner H. vegan: Community Ecology Package. 2015;at
367   <http://CRAN.R-project.org/package=vegan>.
368 E21. Hennig C. fpc: Flexible Procedures for Clustering. 2015;at <http://CRAN.R-
369   project.org/package=fpc>.
370 E22. Gesmann M, de Castillo D. Using the Google Visualisation API with R. *R Journal*
371   2011;3:40–44.
372 E23. Breiman L. Random forests. *Machine learning* 2001;
373 E24. Liaw A, Wiener M. Classification and Regression by randomForest. *R news*
374   2002;2:18–22.
375 E25. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw*
376   2008;28:.
377 E26. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models
378   Using lme4. *J Stat Softw* 2015;67:.
379 E27. Hothorn T, Bretz F, Westfall P, Heiberger RM. Package "multcomp." 2016;
380 E28. Talukder H, Paulson JN, Bravo HC. fitTimeSeries: Longitudinal differential
381   abundance analysis for marker-gene surveys. *bioconductorjp*
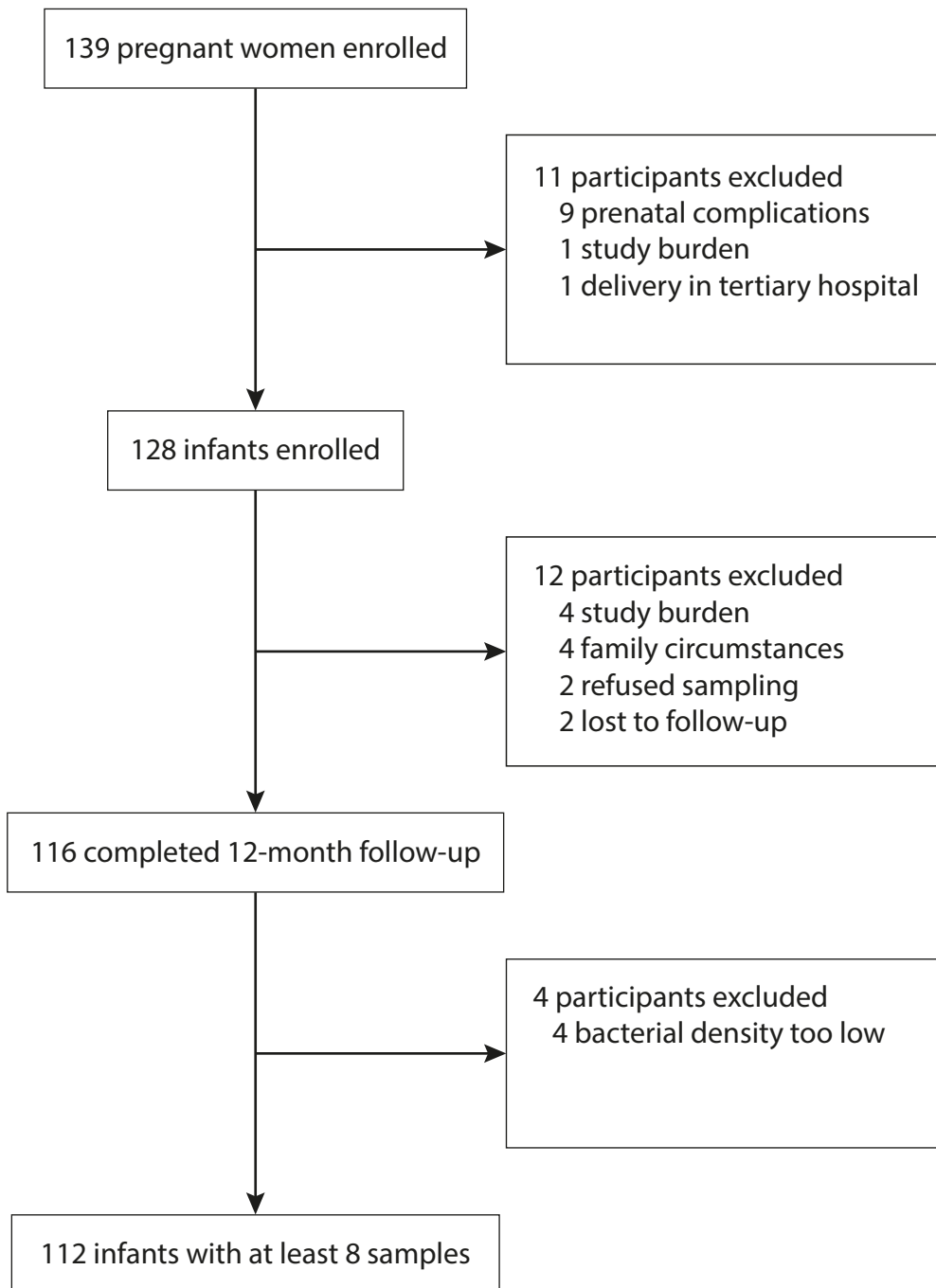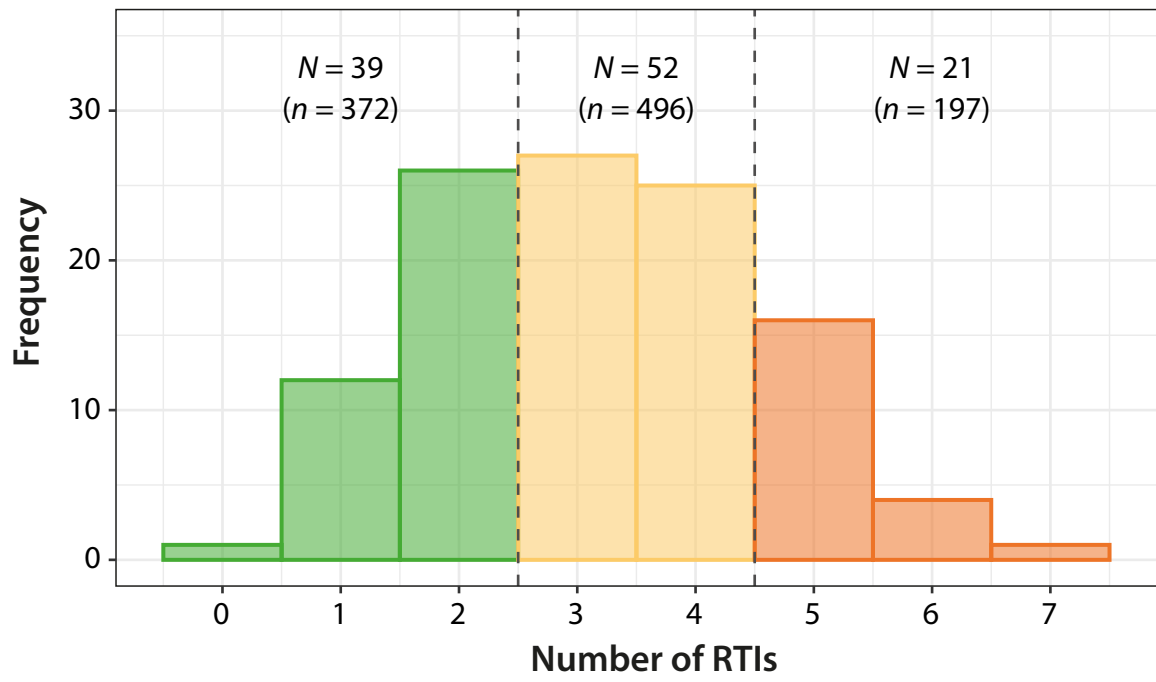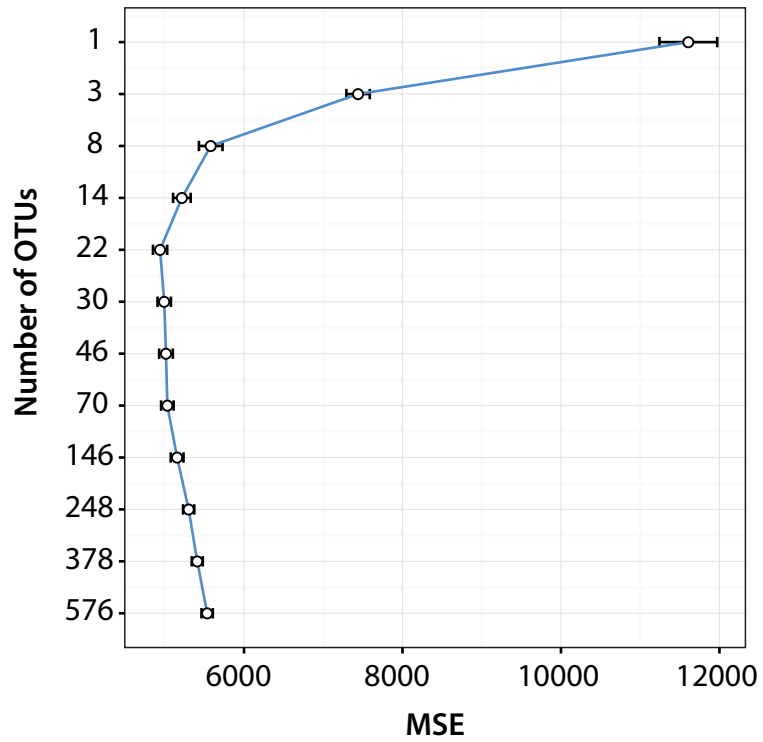
382
383
384

Figure E1

```
                    ┌─────────────────────────────┐
                    │  139 pregnant women enrolled │
                    └──────────────┬──────────────┘
                                   │              ┌──────────────────────────────────┐
                                   │              │ 11 participants excluded          │
                                   ├─────────────▶│    9 prenatal complications       │
                                   │              │    1 study burden                 │
                                   │              │    1 delivery in tertiary hospital│
                                   │              └──────────────────────────────────┘
                                   ▼
                    ┌─────────────────────────────┐
                    │    128 infants enrolled      │
                    └──────────────┬──────────────┘
                                   │              ┌──────────────────────────────────┐
                                   │              │ 12 participants excluded          │
                                   ├─────────────▶│    4 study burden                 │
                                   │              │    4 family circumstances         │
                                   │              │    2 refused sampling             │
                                   │              │    2 lost to follow-up            │
                                   ▼              └──────────────────────────────────┘
                    ┌─────────────────────────────┐
                    │ 116 completed 12-month follow-up │
                    └──────────────┬──────────────┘
                                   │              ┌──────────────────────────────────┐
                                   │              │ 4 participants excluded           │
                                   ├─────────────▶│    4 bacterial density too low    │
                                   │              └──────────────────────────────────┘
                                   ▼
                    ┌─────────────────────────────┐
                    │ 112 infants with at least 8 samples │
                    └─────────────────────────────┘
```

Figure E2

Figure E3

A

B

Figure E4

A

B

Legend:
- d0
- d1
- w1
- w2
- m1
- m2
- m3
- m4
- m6
- m9
- m12
- RTI
- SD datapoints
- representative datapoint (=sample)

Figure E5

Figure E6

A

Time point

B
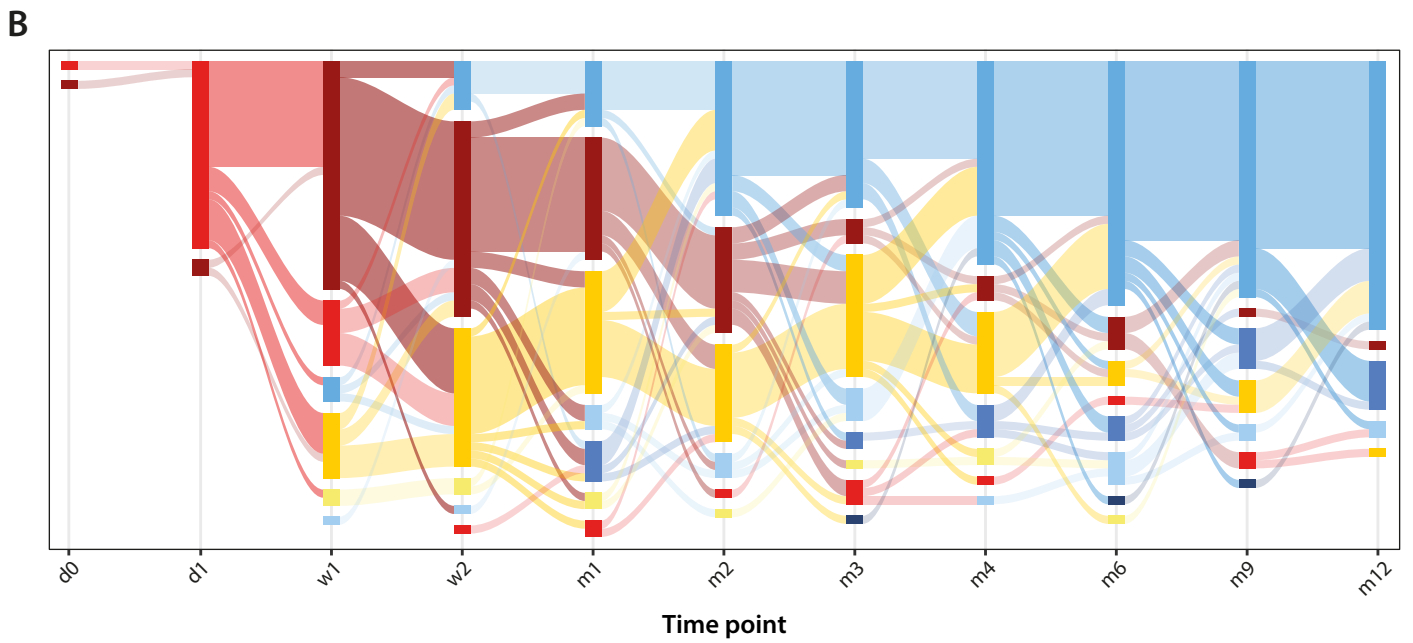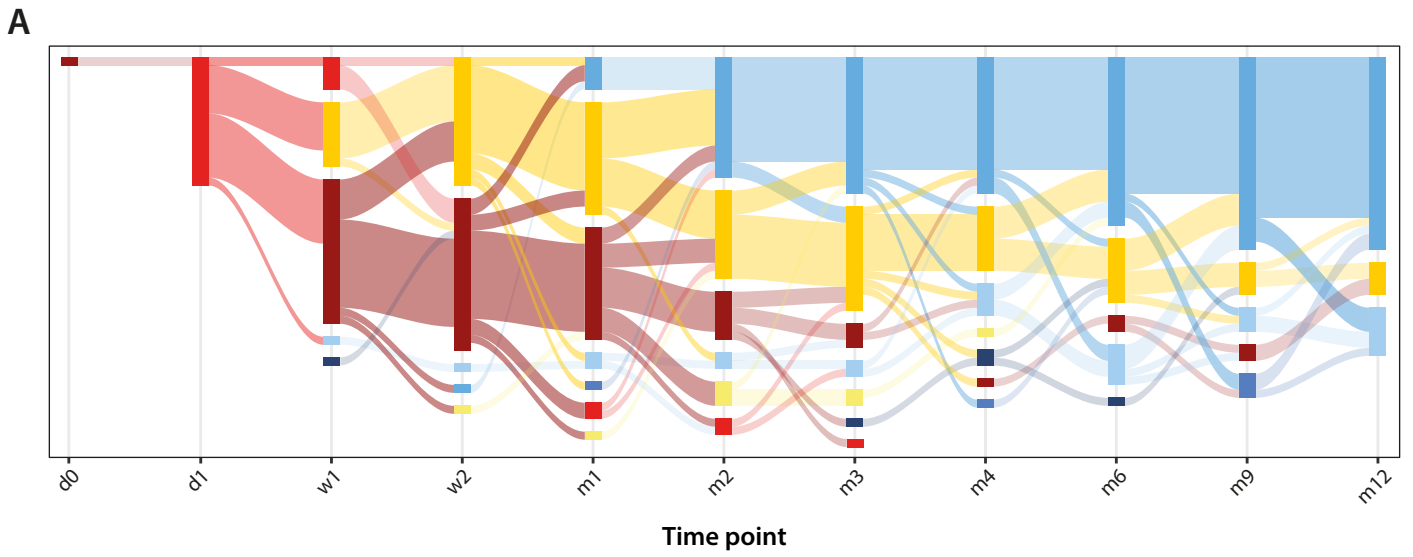
Time point

C

Time point

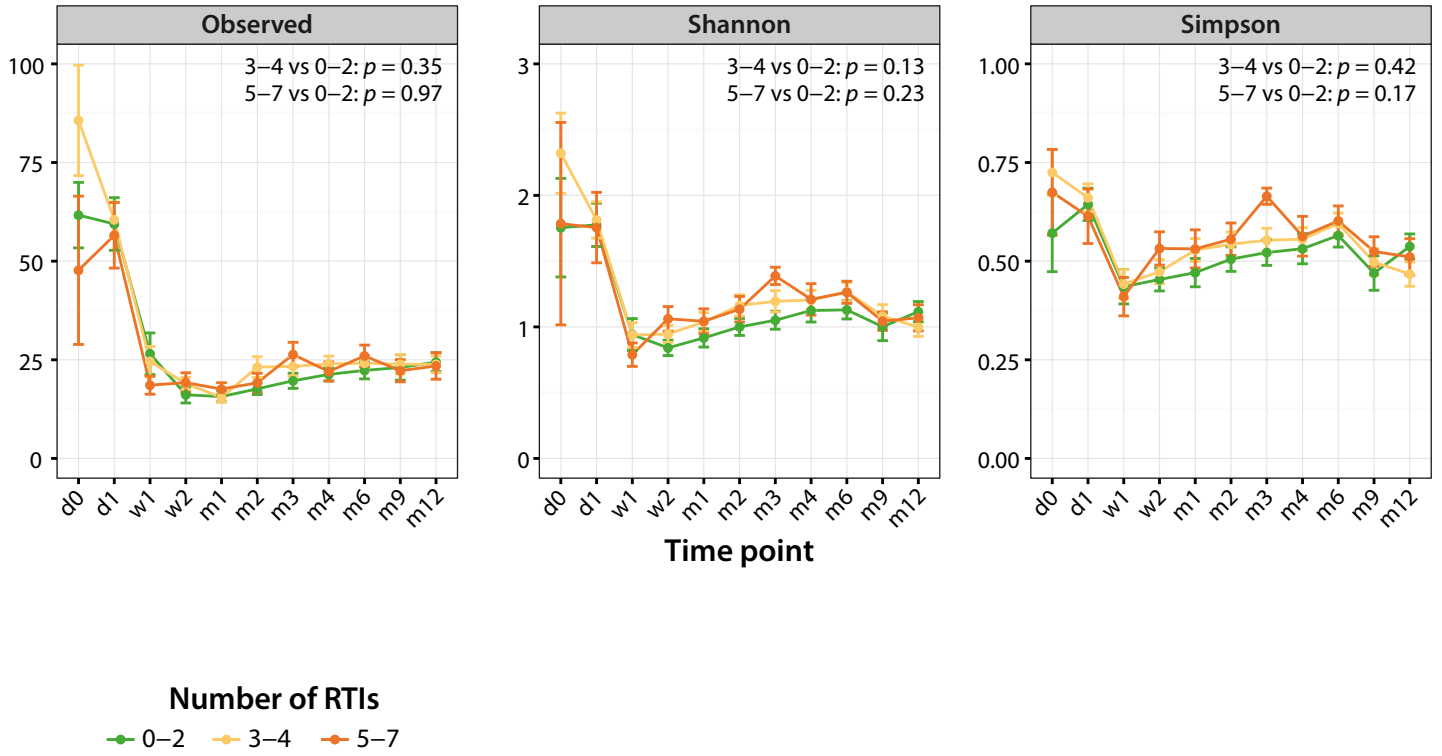| STR | CDG | MOR1 | HAE |
| STA | COR8 | MOR7 | NEI |
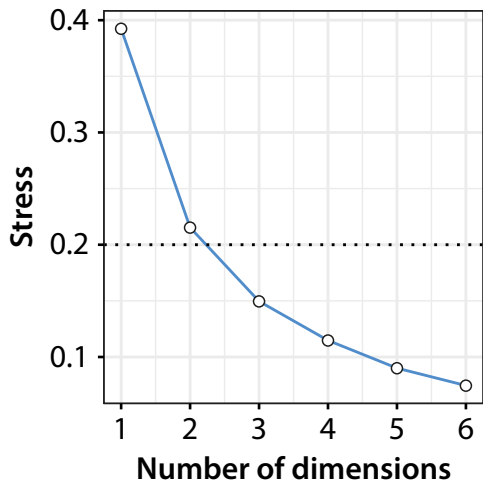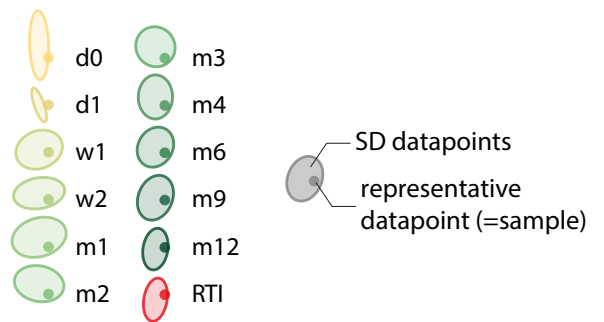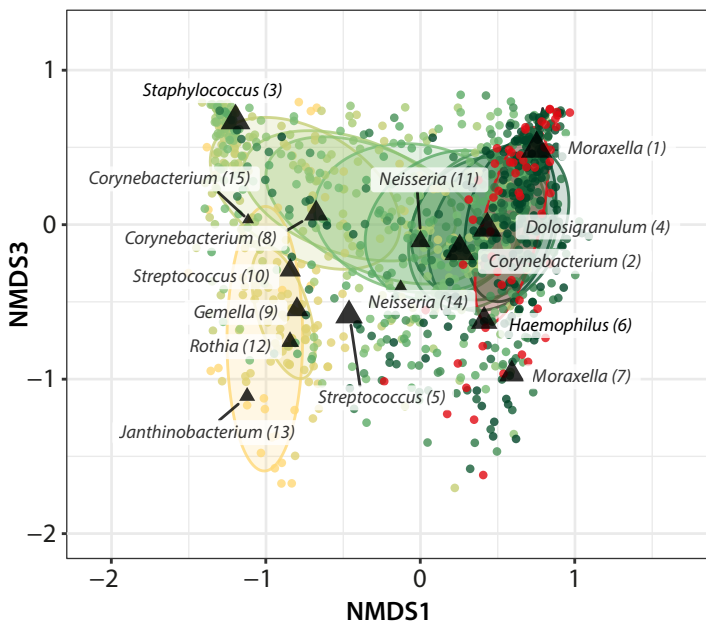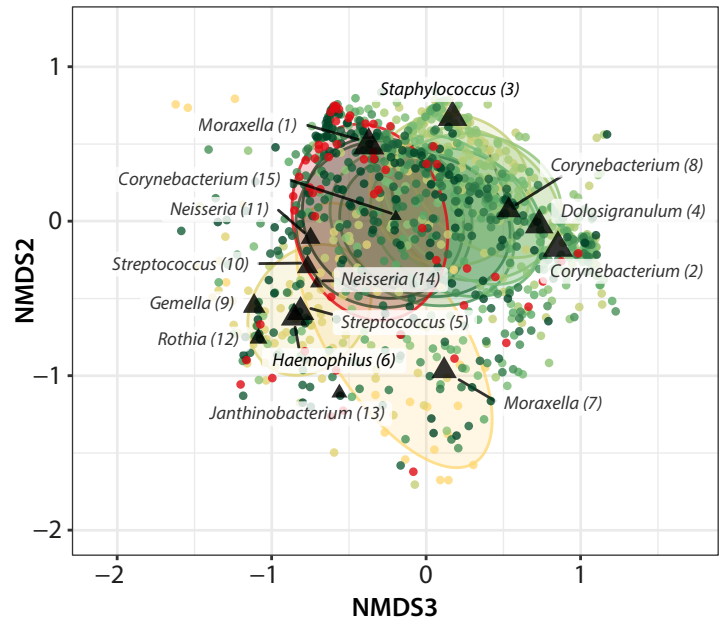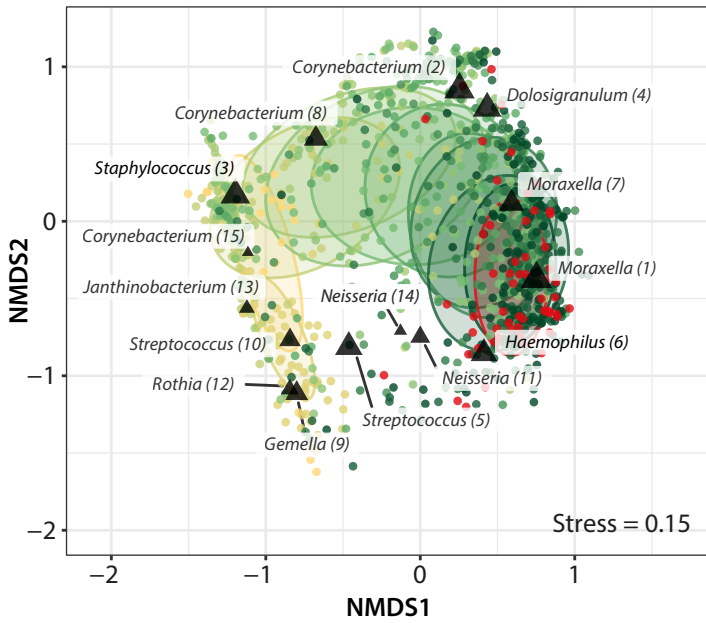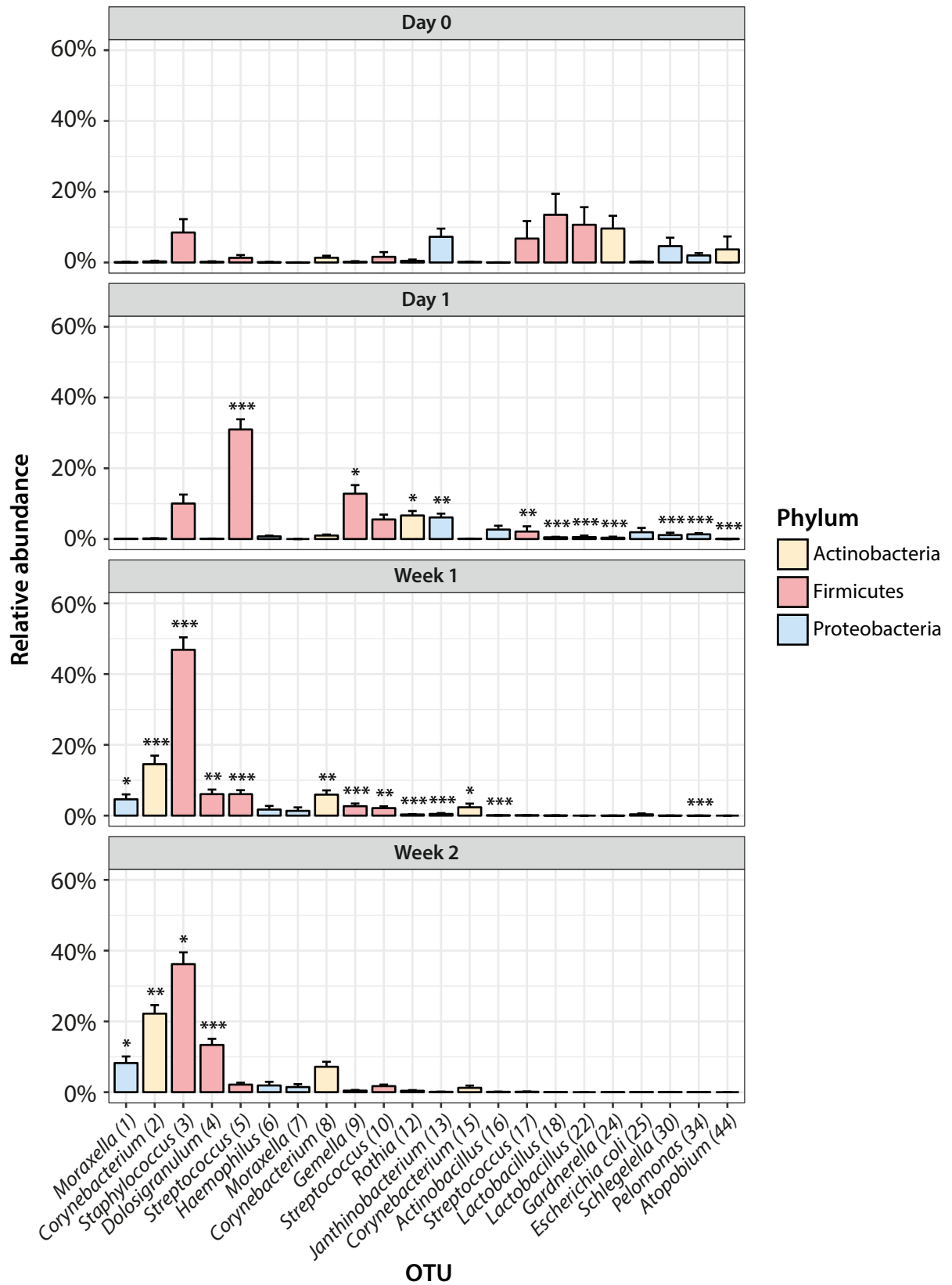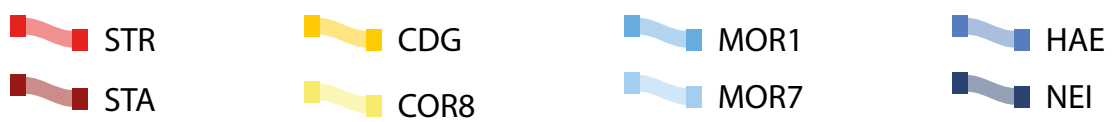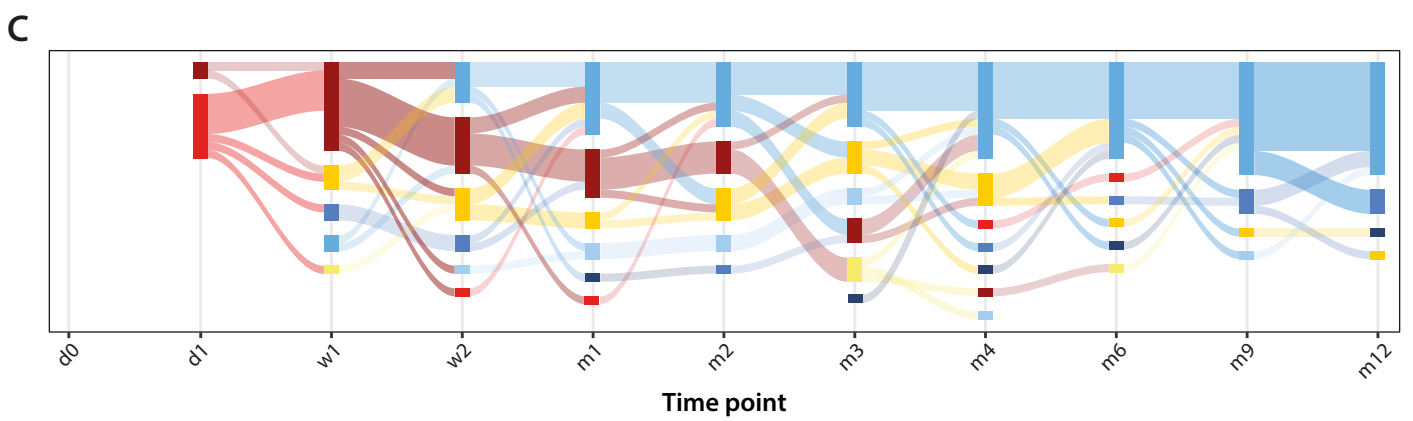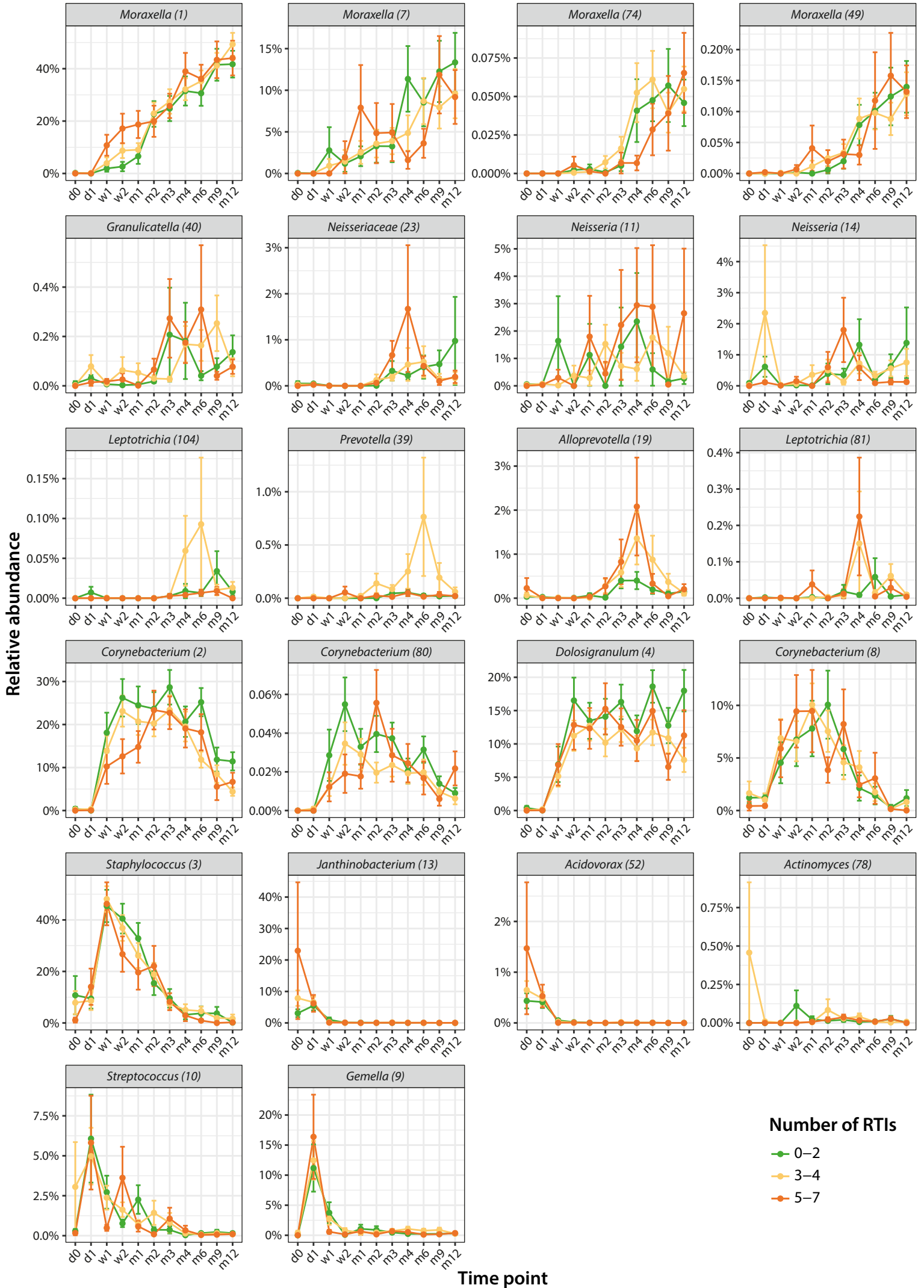
Figure E7

Figure E8

Figure E9

Figure E10

Observed

3–4 vs 0–2: $p = 0.35$
5–7 vs 0–2: $p = 0.97$

Shannon

3–4 vs 0–2: $p = 0.13$
5–7 vs 0–2: $p = 0.23$

Simpson

3–4 vs 0–2: $p = 0.42$
5–7 vs 0–2: $p = 0.17$

Time point

Number of RTIs

0–2   3–4   5–7

Figure E11

Figure E12

A

Birth mode — vaginal — caesarian section

B

Relative abundance

Time point

Breastfeeding <3m ● yes ● no