

1 Association between maternal breastmilk microbiota composition and rotavirus vaccine 2 response in African, Asian, and European infants: a prospective cohort study

3
4 Jonathan Mandolo^{1,2,3*}, Edward P. K. Parker^{4*}, Christina Bronowski⁵, Kulandaipalayam Natarajan C. Sindhu⁶
5 Alistair C. Darby⁵, Nigel A. Cunliffe^{5,7}, Gagandeep Kang^{6†}, Miren Iturriza-Gómara^{8†}, Arox W. Kamng'ona^{1,3†},
6 Khuzwayo C. Jere^{1,5,7,9†} on behalf of the RoVI study group⁸

7
8 * These authors contributed equally
9 † These authors jointly supervised this work

10
11 ¹ Virology Research Group, Malawi–Liverpool–Wellcome Trust Clinical Research Programme, Blantyre,
12 312225, Malawi

13 ² Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, L3 5QA, UK

14 ³ Department of Biomedical Sciences, School of Life Sciences and Allied Health Professions, Kamuzu
15 University of Health Sciences, Blantyre, 312225, Malawi.

16 ⁴ The Vaccine Centre, Department of Clinical Research, London School of Hygiene and Tropical Medicine,
17 London, WC1E 7HT, UK

18 ⁵ Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool L69 7BE, UK

19 ⁶ Wellcome Trust Research Laboratory, Division of Gastrointestinal Sciences, Christian Medical College,
20 Vellore, Tamil Nadu, India

21 ⁷ National Institute for Health and Care Research, Health Protection Research Unit in Gastrointestinal
22 Infections, University of Liverpool, UK

23 ⁸ Centre for Vaccine Innovation and Access, Program for Appropriate Technology in Health (PATH), Geneva,
24 1218, Switzerland.

25 ⁹ Department of Medical Laboratory Sciences, School of Life Sciences and Allied Health Professions, Kamuzu
26 University of Health Sciences, Blantyre, 312225, Malawi.

27
28
29
30 *Corresponding authors:*

31 Khuzwayo C. Jere (khuzwayo.jere@liverpool.ac.uk)

32
33 [§] Other members of the RoVI study group: Sudhir Babji, Blossom Benny, Noelia Carmona-Vicente, Nedson
34 Chasweka, End Chinyama, Queen Dube, Sidhartha Giri, Nicholas C. Grassly, Annai Gunasekaran, Deborah
35 Howarth, Sushil Immanuel, Jenna Lowe, Ira Praharaj, Bakthavatsalam Sandya Rani, Sophia Silas, Vivek
36 Kumar Srinivasan, Mark Turner, Srinivasan Venugopal, Valsan Philip Verghese.

37
38 **ABSTRACT** (192 words)

39 **Background.** Maternal breastmilk is a source of pre- and pro-biotics that impact neonatal gut
40 microbiota colonisation. Since oral rotavirus vaccines (ORVs) are administered at a time when
41 infants are often breastfed, breastmilk microbiota composition may have a direct or indirect
42 influence on vaccine take and immunogenicity.

43 **Methods.** Using standardised methods across sites, we compared breastmilk microbiota
44 composition in relation to geographic location and ORV response in cohorts prospectively followed
45 up from birth to 18 weeks of age in India ($n = 307$), Malawi ($n = 119$), and the UK ($n = 60$).

46 **Results.** Breastmilk microbiota diversity was higher in India and Malawi than the UK across three
47 longitudinal samples spanning weeks of life 1 to 13. Dominant taxa such as *Streptococcus* and
48 *Staphylococcus* were consistent across cohorts; however, significant geographic differences were
49 observed in the prevalence and abundance of common and rare genera throughout follow-up. No
50 significant associations were identified between breastmilk microbiota composition and ORV
51 outcomes including seroconversion, post-dose 1 vaccine shedding, and/or post-vaccination
52 rotavirus-specific IgA level.

53 **Conclusions.** Our findings suggest that breastmilk microbiota composition may not be a key factor
54 in shaping trends in ORV response within or between countries.

55

56 **Keywords.** breastmilk; rotavirus; microbiota; immunogenicity

57

58 **Manuscript: 3495 words**

59 INTRODUCTION

60 Maternal breastmilk is a key source of nutrition for newborn infants. It is enriched with a variety of
61 macro- and micro-nutrients vital for infant growth, and contains immunoglobulins, growth
62 hormones, and oligosaccharides that perform critical functions in infant gut homeostasis and
63 immune development [1]. The *Bifidobacteriaceae*, *Pseudomonadaceae*, *Streptococcaceae*, *Enterococcaceae*
64 and *Staphylococcaceae* bacterial families have consistently been identified as core constituents of the
65 breastmilk microbiota [2–4]. These and other bacteria in breastmilk may act as a source of
66 commensal bacteria, seeding the infant gut microbiota at a critical stage of neonatal development
67 [5]. Geographic region, mode of delivery, maternal health, and genetic factors are among the factors
68 associated with maternal breastmilk microbiota composition [2,3,6–9].

69
70 Rotavirus remains a major cause of severe gastroenteritis among children worldwide. More than
71 100 countries have incorporated oral rotavirus vaccine (ORV) into their national immunization
72 programs [10]. Malawi and the UK introduced the live-attenuated monovalent G1P[8] Rotarix
73 vaccine into their national immunisation programs in 2012 and 2013, respectively [11,12]. India
74 introduced a live-attenuated, monovalent vaccine containing a G9P[11] human strain into its
75 immunisation program in 2016 [13]. These vaccines have reduced the burden of rotavirus, although
76 in India and Malawi the estimated mortality burden due to rotavirus remained significant as of 2016
77 (9.2 and 31.2 per 100,000, respectively, compared to 0.1 per 100,000 in England) [14]. As reported for
78 other live oral vaccines such as oral poliovirus vaccine, ORV immunogenicity and efficacy is
79 significantly reduced in low- and middle-income (LMIC) compared with high-income countries
80 [15]. Given that LMICs account for approximately 95% of all rotavirus deaths worldwide [16], the
81 public health burden associated with impaired ORV response is considerable.

82
83 Several mechanisms may contribute to the impaired performance of ORV in LMICs. In Malawi and
84 India, we reported infant gut microbiota diversity to be negatively correlated with ORV response
85 [17,18]. Maternal rotavirus-specific IgG and IgA antibodies in breastmilk and serum were also
86 negatively correlated with ORV response [17,19], although similar correlations were absent among

87 infants in the UK [17]. Other factors which may impact ORV response include histo-blood group
88 antigen status, environmental enteric dysfunction (EED), and pre-vaccination rotavirus exposure
89 [20].

90

91 Since ORV is administered at a time when infants are often breastfeeding, we hypothesised that
92 maternal breastmilk microbiota composition may be associated with ORV response, either by
93 directly interacting with the vaccine viruses or indirectly via the developing infant gut microbiota.
94 We tested this hypothesis using standardised methods across cohorts in Malawi, India, and the UK
95 [17,18].

96

97 **MATERIALS AND METHODS**

98 **Study cohort**

99 This is a follow-up to the Rotavirus Vaccine Immunogenicity (RoVI) study – a multi-site
100 observational cohort study exploring the impact of maternal antibodies, microbiota development,
101 and EED on ORV response (CTRI/2015/11/006354). The study design, sample handling, lab assays,
102 and primary outcomes of the study have been described previously [17,18]. Briefly, pregnant
103 women were recruited across sites in Blantyre (Malawi), Vellore (India), and Liverpool (UK).
104 Infants received routine immunisations including two doses of Rotarix according to the national
105 immunisation schedule at each study site (weeks of life 6 and 10 in India and Malawi; weeks of life
106 8 and 12 in the UK). Rotavirus-specific IgA (RV-IgA) was measured in infant blood samples
107 collected pre- and 4 weeks post-vaccination. Rotavirus shedding was measured in six longitudinal
108 stool samples per infant, including 1 week after each ORV dose. Breastmilk samples were collected
109 in week of life 1 and in the week after each ORV dose (**Figure 1A**).

110

111 **Sample processing**

112 Breastmilk samples were collected in sterile sample pots by participants and transferred to the site-
113 specific laboratory by courier within 24 hours (and refrigerated throughout) in the UK or within 4
114 hours in India and Malawi. Upon receipt, samples were kept at 4°C for a maximum of 8 hours until
115 processing and subsequently stored in 2 ml aliquots in SuperLock tubes (Starlab) at -70 °C for a
116 maximum of 2 weeks prior to DNA extraction. DNA was extracted from 1 ml of breastmilk but
117 otherwise followed the stool-specific protocol previously described [17]. A negative extraction
118 control was included in each DNA extraction batch. DNA extracts from Malawi and India were
119 shipped on dry ice to the University of Liverpool for library preparation and sequencing.

120

121 **Microbiota sequencing**

122 Breastmilk microbiota composition was determined by sequencing the V3-V4 region of the 16S
123 rRNA gene. Amplicon generation, library preparation, and sequencing steps were performed as
124 previously described for stool [17], but with 15 cycles (as opposed to 10) for the initial amplicon
125 PCR and 20 cycles (as opposed to 15) for subsequent indexing PCR to ensure robust amplification
126 from the low-biomass samples. We sequenced amplicons for 1,301 separate breastmilk samples (894
127 from India, 275 from Malawi, and 132 from the UK) across 6 Illumina HiSeq2500 lanes (v2
128 chemistry with 600 cycles in rapid run mode). Samples from each participant were processed on the
129 same plate. Sequencing was batched by geographic location according to sample availability. Each
130 PCR plate included: a no-template PCR control; a breastmilk control sample provided by a mother
131 in the UK who was not enrolled in the study; DNA from a mock community (Zymo Research
132 D6306); and a pool of extraction controls corresponding to the samples contained on each plate for
133 samples from India and the UK. Due to shipment challenges, extraction controls corresponding to
134 141/243 (58%) of samples from Malawi were included in the extraction pools. To better define the
135 amplicon profile of extraction controls, we sequenced a further 49 pools containing 1-5 controls
136 from extraction batches performed in India or the UK. Final libraries contained up to four 96-well
137 PCR plates (384 amplicons). Breastmilk DNA samples were amplified on separate plates to stool
138 samples, though we allowed mixing of stool and breastmilk PCR plates in a given library. To

139 validate the robustness of the sequencing protocol, 90 breastmilk DNA samples (30 per cohort, all
140 collected in week of life 1) were transferred to Imperial College London and sequenced according to
141 the methods above with minor modifications, as previously described [17].

142

143 **Bioinformatic processing**

144 Adapters were trimmed from raw sequences using cutadapt version 1.18 [21]. We merged, filtered,
145 and denoised the amplicon sequences using the DADA2 pipeline in QIIME2 (version 2018.11) [22].
146 Forward and reverse reads were truncated to 270 bp and 200 bp, respectively. Taxonomic
147 assignment was performed via the *dada2* package (version 1.14.1) using the RDP naïve Bayesian
148 classifier trained on the Silva rRNA database (version 132). Ribosomal sequence variants (RSVs)
149 were retained if they were 390–440 bp in length, assigned as bacterial, detectable at •0.1%
150 abundance in at least one sample, and passed frequency-based contamination filtering using the
151 *decontam* package in R (version 3.6.1) [23]. Nanodrop readings (ng/•l) were used to define
152 concentration of the input template.

153

154 Given the additional amplification involved in library preparation for breastmilk samples, reads
155 were frequently detected in extraction controls ($n = 56$ individual or pooled controls with >10,000
156 reads after the filtering steps above). Several additional filtering steps were therefore included. First,
157 we retained RSVs if they were detectable at •0.1% abundance in •1% of breastmilk samples from at
158 least one country. Second, we applied prevalence-based filtering using the *decontam* package with a
159 p value threshold of 0.05 to exclude RSVs that were more common in extraction controls. Finally, we
160 removed samples if their mean Bray-Curtis distance (based on either weighted or unweighted
161 metrics) from breastmilk extraction controls was smaller than their mean distance from other
162 breastmilk samples collected from the same country (**Supplementary Figure 1**).

163

164 **Outcomes**

165 We compared breastmilk microbiota composition by country and ORV response. Our primary
166 indicator of ORV response was seroconversion status – defined as a 4-fold increase in RV-IgA
167 concentration or detection of antibodies at ≥ 20 IU/ml in previously seronegative infants. Secondary
168 outcomes included post-vaccination RV-IgA concentration (as a continuous variable) and rotavirus
169 shedding 1 week after the first dose of ORV (as an indicator of vaccine virus take). Shedding was
170 detected via real-time PCR targeting the Rotarix *NSP2* gene [24]. We also performed an exploratory
171 analysis of alpha and beta diversity to identify demographic and clinical factors associated with
172 breastmilk composition.

173

174 **Statistical analysis**

175 Analyses were performed in the programming language R following the statistical pipeline
176 previously described for stool samples with minor modifications [17]. Alpha and beta diversity
177 were calculated at a rarefaction depth of 15,000 sequences per sample. We performed cross-sectional
178 analyses of alpha diversity via analysis of variance (ANOVA), logistic regression (binary ORV
179 outcomes), Pearson's r with two-sided hypothesis testing (log-transformed RV-IgA), and linear
180 regression (exploratory covariates). We assesses beta diversity using permutational multivariate
181 ANOVA (PERMANOVA) with 999 permutations based on genus-level unweighted Bray–Curtis
182 distances. For binary outcomes, discriminant genera and RSVs were identified via two-sided
183 Fisher's exact test (differences in prevalence) and Aldex2 (two-sided Wilcoxon rank-sum test of
184 centred log-ratio transformed sequence counts), with taxa classified as discriminant if they had a p
185 value of <0.05 based on either method after Benjamini–Hochberg false discovery rate (FDR)
186 adjustment. Aldex2 was used to identify taxa correlated with log-transformed RV-IgA (FDR-
187 adjusted p value of <0.05 based on two-sided Spearman's rank test). Taxa were included if they
188 were detected with a prevalence of $>5\%$ in at least one of the groups being compared. We
189 supplemented cross-sectional analyses with longitudinal mixed-effects models of Shannon index
190 and taxon abundances (zero-inflated negative binomial models of genus-level sequence counts),

191 including week of life as a covariate and study ID as a random effect. Genera were included in
192 longitudinal models if they were present in 20% of samples in a given country.

193

194 We applied Random Forests in a series of cross-sectional analyses to predict country and ORV
195 outcome based on genus or RSV relative abundances. For each analysis, we performed 20 iterations
196 of 5-fold cross-validation. For binary outcomes, we standardised the baseline accuracy of
197 classification models at 50% by fitting each iteration of cross-validation on a random subset of 50
198 samples per group (or the number of samples in the minority group if this was <50). Models were
199 excluded if there were <10 samples in the minority group. For regression models, accuracy was
200 quantified by using linear regression to determine the out-of-bag R^2 values for predicted vs
201 observed RV-IgA values.

202

203 For positive controls and technical replicates, we used linear regression (alpha diversity and
204 common genera) and PERMANOVA (beta diversity) to quantify the proportion of variance
205 explained by sample ID.

206

207 The raw sequence data for this study have been deposited in the European Nucleotide Archive
208 under accession code PRJEB38948. Processed data and analysis code are available on Github
209 (<https://github.com/eparker12/RoVI>).

210

211 **Ethics approval**

212 The study was approved by the Institutional Review Board at the Christian Medical College (CMC)
213 in Vellore (IRB No. 9472/24.06.2015), the College of Medicine Research and Ethics Committee in
214 Blantyre (P.01/16/1853), and the North West—Liverpool Central Research Ethics Committee in
215 Liverpool (15/NW/0924).

216

217 **RESULTS**

218 **Study cohort**

219 Overall, 664 mother–infant pairs (395 in India, 187 in Malawi, and 82 in the UK) were enrolled in the
220 study and the primary endpoint (measurement of seroconversion or dose 1 shedding) was reached
221 for 484 (307 in India, 119 in Malawi, and 60 in the UK). Baseline characteristics, wild-type rotavirus
222 infection status, EED biomarker levels, and infant stool microbiota composition have previously
223 been compared by country and ORV outcome [17]. Exclusive breastfeeding was reported by
224 265/307 (86%) mothers in India, 108/119 (91%) in Malawi, and 26/60 (43%) in the UK, with partial
225 breastfeeding reported by a further 32/307 (10%) in India, 11/119 (9%) in Malawi, and 20/60 (33%)
226 in the UK. Exclusive breastfeeding was positively correlated with ORV seroconversion and post-
227 vaccination infant RV-IgA levels in India but not in other cohorts. Breastmilk RV-IgA levels were
228 negatively correlated with infant RV-IgA levels in India and Malawi [17].

229

230 **ORV shedding and immunogenicity**

231 As previously reported [17], seroconversion was observed in 27/51 (53%) infants in the UK, 85/305
232 (28%) in India, and 24/103 (23%) in Malawi. Rotavirus shedding 1 week after the first dose of ORV
233 was detected in 55/60 (92%) infants in the UK, 82/305 (27%) in India, and 56/101 (55%) in Malawi.
234 Geometric mean concentrations (GMCs) of RV-IgA (IU/ml) after vaccination were 27 (17–45) in the
235 UK, 20 (95% CI 16–25) in India, and 9 (6–12) in Malawi.

236

237 Indian infants were characterised by high rates of neonatal rotavirus infection, defined as detection
238 of wild-type rotavirus shedding in week 1 of life or baseline seropositivity (pre-vaccination RV-IgA
239 •20 IU/ml). This was observed in 166/304 (55%) infants in India, 10/90 (11%) in Malawi, and 2/54
240 (4%) in the UK. Given the potential impact of neonatal infection on ORV shedding and
241 immunogenicity [17], we report results for the Indian cohort overall and stratified by neonatal
242 infection status below.

243

244 **Geographic differences in breastmilk microbiota composition**

245 Of 1,301 breastmilk samples sequenced from this study population, 1,124 yielded high-quality
246 microbiota profiles (• 15,000 sequences after quality filtering; $95,075 \pm 113,894$ [mean \pm s.d.]
247 sequences per sample). Microbiota profiles of positive controls and technical replicates were
248 consistent across sequencing runs and facilities (**Supplementary Figure 2**).

249

250 There were marked differences in breastmilk microbiota composition between cohorts. Microbiota
251 diversity was significantly lower in UK than both other cohorts at all timepoints. Diversity was
252 similar in India and Malawi at week of life 1, but higher in India than Malawi at weeks of life 7 and
253 11 (**Figure 1B**). Samples clustered by individual (PERMANOVA $R^2 = 49\%$, $p = 0.001$), with country
254 accounting for 6–9% of variation depending on age (**Figure 1C**). Although 350 genera were detected
255 overall, a small proportion were dominant in each cohort (**Figure 1C** and **Supplementary Figure 3**).
256 Among dominant genera, *Streptococcus* was depleted in India compared with both other cohorts,
257 while *Acinetobacter* and *Corynebacterium* were enriched. *Staphylococcus* followed a parallel trajectory
258 in each cohort, peaking in week of life 1, and was less abundant in Malawi than both other cohorts.
259 *Bifidobacterium* was observed at lower abundance in the UK, reflecting the pattern previously
260 reported for stool samples [17], while *Gemella* was enriched in this cohort (**Figure 1D**).

261

262 Additional discriminant taxa were identified when considering both common and rare genera via
263 longitudinal and cross-sectional models (**Supplementary Figure 4**). Based on longitudinal models,
264 17 genera were enriched in India compared with both other cohorts, including nine Proteobacteria
265 (e.g. *Aeromonas* and *Alishewenalla*), three Firmicutes (e.g. *Enterococcus* and *Aerococcus*) and five
266 Actinobacteria (e.g. *Dermacoccus*). Nine genera were enriched in Malawi compared with both other
267 cohorts, including the Bacteroidetes genus *Prevotalla* 9 alongside eight Firmicutes (e.g.
268 *Faecalibacterium* and *Lachnospiraceae*). Three genera – *Gemella*, *Haemophilus*, and *Enterobacter* – were
269 enriched in the UK compared with the other cohorts.

270

271 Random Forests discriminated samples by country with high accuracy based on genus relative
272 abundance (median cross-validation accuracies of 85–95%; baseline accuracy 50%; **Supplementary**

273 **Figure 5)**. Genera underlying the predictive accuracy (based on mean importance scores) were
274 consistent with the discriminant taxa described above (**Supplementary Table 1**).

275

276 We also assessed alpha and beta diversity of breastmilk samples in relation to individual-level
277 variables measured in each cohort (**Figure 2**). With the exception of infant serum • 1 acid
278 glycoprotein level (a marker of systemic inflammation), which was modestly associated with beta
279 diversity in Malawian samples (R^2 3.7%), no covariates were significantly associated with
280 breastmilk microbiota composition.

281

282 **Breastmilk microbiota composition versus ORV response**

283 Based on longitudinal models of Shannon index, we observed no significant differences in
284 microbiota diversity according to seroconversion status in any cohort (**Figure 3A**). This was also the
285 case for cross-sectional analyses, with the exception of comparisons in Malawi at week 7 of life (the
286 week after the first dose of ORV), wherein Shannon index was negatively correlated with
287 seroconversion. Beta diversity analyses based on genus-level unweighted Bray-Curtis distances did
288 not reveal any significant association between breastmilk microbiota composition and
289 seroconversion status (**Figure 3B**). Likewise, Random Forest models based on genus or
290 RSV abundances failed to accurately predict seroconversion (**Figure 3C**), and no discriminant taxa
291 were identified based on cross-sectional analyses of prevalence or abundance after FDR correction.
292 Longitudinal models of common genera (• 20%) revealed frequent age-associated changes in taxon
293 abundance but only one significant association with seroconversion (a negative correlation between
294 *Alloprevotella* abundance and seroconversion in Malawi; **Supplementary Table 2**).

295

296 Cross-sectional analyses of secondary ORV endpoints, including post-vaccination RV-IgA
297 concentration (**Supplementary Figure 6**) and dose 1 ORV shedding (**Supplementary Figure 7**) were
298 consistent with those for seroconversion, revealing no significant associations. Very few

299 discriminant genera were identified with respect to secondary outcomes based on longitudinal
300 models of genus abundance (**Supplementary Table 2**).

301

302 **DISCUSSION**

303 Breastmilk is a key source of pre- and pro-biotics that shape infant gut microbiota configuration.
304 This, in turn, plays a pivotal role in shaping immune development. We documented significant
305 differences in breastmilk microbiota composition between Malawi, India, and the UK across the
306 first 13 weeks of life. However, no consistent differences in breastmilk microbiota composition were
307 observed with respect to ORV response.

308

309 Despite the geographic differences in breastmilk microbiota composition, several genera of bacteria
310 were dominant across the three cohorts. Together, *Streptococcus*, *Staphylococcus*, *Acinetobacter*,
311 *Bifidobacterium*, *Veillonella*, *Gemella*, *Corynebacterium* and *Pseudomonas* formed approximately 75% for
312 the breastmilk microbiota as determined by 16S amplicon sequencing – consistent with the
313 dominant taxa reported in previous studies [2–4,25]. The relative abundances of these dominant
314 genera changed over time, with *Staphylococcus* declining in abundance from week 1 of life onwards,
315 while *Streptococcus* and *Veillonella* increased in abundance. This is similar to the trajectory in
316 breastmilk microbiota composition reported in Kenya [26]. The infant salivary microbiota is known
317 to be colonised by *Streptococcus* [27,28], such that the continued dominance of *Streptococcus* in
318 maternal breastmilk may partly reflect breastmilk–saliva interplay throughout early life. Skin-
319 associated genera including *Staphylococcus* and *Corynebacterium* were also among the dominant
320 genera in maternal breastmilk, consistent with previous findings [29].

321

322 Breastmilk microbiota diversity was higher in Malawi and India compared with the UK. This
323 contrasts with discrepancies we reported in stool microbiota diversity, which was significantly
324 higher in Malawi than both India and the UK at week 1 of life but converged over the ensuing 6–8
325 weeks [17]. In a previous cross-sectional study spanning 11 study sites, breastmilk microbiota

326 diversity was highest in rural Ethiopia and lowest in Ghana, with intermediate levels across other
327 sites in Africa, Europe, North America, and South America [8]. At genus level, *Streptococcus* was
328 more abundant in Malawi and the UK than in India, while *Bifidobacterium* was depleted in the UK
329 compared with both other cohorts. Prior studies have also highlighted geographically distinct
330 abundance profiles including depletion of *Bifidobacterium* in European compared with African
331 samples [8]. Together, these studies highlight the significant regional variation that occurs in
332 breastmilk microbiota diversity and composition. To delineate overarching global trends (e.g. urban
333 vs rural, high-income vs LMIC), future studies integrating well-powered representative data from
334 multiple countries, such as the present, are warranted.

335

336 We did not observe consistent associations between breastmilk microbiota composition and ORV
337 response. At the time of the first dose of ORV, breastmilk microbiota diversity in Malawi was
338 negatively correlated with ORV seroconversion – a correlation that was also apparent among infant
339 stool samples in this cohort [17]. However, while we reported consistent correlations between
340 diversity and seroconversion among Indian and Malawian infants' stool samples, there was no
341 consistent discrepancy across cohorts in breastmilk. A previous study in India documented higher
342 *Enterobacter/Klebsiella* abundance in breastmilk and infant stool samples of infants with symptomatic
343 rotavirus disease compared to those with asymptomatic or no infection [30]. However, no
344 significant discrepancies were observed between neonates with asymptomatic infection and those
345 lacking infection, which is consistent with the lack of association reported here in relation to
346 attenuated viral exposure via ORV.

347

348 To our knowledge, this is the first study to explore the link between breastmilk microbiota
349 composition and ORV response. Our study is strengthened by the use of standardised methods
350 across cohorts, and the exploration of multiple indicators of ORV response, including dose 1
351 shedding. Nonetheless, several limitations of the present study should be considered. Owing to
352 recruitment challenges in Malawi [17], we fell short of the target sample size in this cohort ($n = 119$
353 rather than 150), potentially undermining our ability to detect important associations between

354 breastmilk microbiota composition and ORV response in this cohort. Because of their low biomass,
355 breastmilk samples were subjected to extra rounds of PCR amplification to attain adequate material
356 for sequencing (35 cycles vs 25 used for stool), leading to amplification from extraction controls. We
357 accounted for this via stringent abundance- and prevalence-based filtering of potential
358 contaminants and excluded samples which clustered among extraction controls rather than other
359 breastmilk samples. Nonetheless, the potential contribution of contamination and site-specific batch
360 effects to the observed trends cannot be discounted.

361

362 Our findings suggest that breastmilk microbiota composition may not be a key factor shaping
363 trends in ORV response within or between countries. Other components of human milk were not
364 considered here and would be a valuable focus of future investigation. Human milk
365 oligosaccharides such as lacto-*N*-tetraose have previously been linked with symptomatic rotavirus
366 infection in Indian neonates, possibly via an effect on neonatal G10P[11] rotavirus infectivity [30] .
367 Future studies of the breastmilk metabolome may help discern whether similar factors influence the
368 immunogenicity and efficacy of ORV.

369

370 **Supplementary Materials**

371 Supplementary Figures S1 – S3 and Supplementary Table S1 – S7 are attached.

372

373 **Notes**

374 **Disclaimer**

375 The funders had no role in the study design, data collection and interpretation, or the decision to
376 submit the work for publication. The authors received no financial support or other form of
377 compensation related to the development of the manuscript. N.A.C. and K.C.J are affiliated with the
378 National Institute for Health and Care Research (NIHR) Health Protection Research Unit in
379 Gastrointestinal Infections at the University of Liverpool, a partnership with the UK Health Security
380 Agency (UKHSA), in collaboration with the University of Warwick. The views expressed are those

381 of the author(s) and not necessarily those of the NIHR, the Department of Health and Social Care or
382 the UKHSA.

383 **Financial support**

384 This work was supported by different funders per site. The UK and Malawi sites were funded by
385 the UK Medical Research Council and the UK Department for International Development (Newton
386 Fund MR/N006259/1). K.C.J. is funded by a Wellcome International Training Fellowship (number
387 201945/Z/16/Z). The site in India was funded by the Government of India's Department of
388 Biotechnology.

389 **Acknowledgements**

390 We thank all members of the clinical study teams in Vellore, Blantyre, and Liverpool, including
391 Falak Diab, Siobhan Holt, and the research midwives at the Liverpool Women's Hospital; Dawn
392 Redman and the team of research nurses at Alder Hey Children's Hospital; Uma Raman, Charlet,
393 Margaret, Jacklin, and the field research assistants at Christian Medical College, Vellore; and James
394 Tamani, Anna Ainani, Amisa Chisale, Bertha Masamba, Carlo Gondwe, and Evelyn Gondwe in
395 Blantyre, Malawi. Richard Eccles, Anita Lucaci, Richard Gregory, John Kenny, and other staff at the
396 Centre for Genomic Research (University of Liverpool) provided valuable support for the 16S
397 microbiota sequencing work. Above all, we are grateful to the families involved in the study.

398 **Author contributions**

399 Conceptualisation, M.I.G., A.C.D., E.P.K.P., K.C.J., A.W.K., and G.K.; Methodology, J.M., C.B.,
400 E.P.K.P., A.C.D., M.I.G., and G.K.; Software, E.P.K.P.; Validation, E.P.K.P.; Formal Analysis, J.M.
401 and E.P.K.P.; Investigation, J.M., C.B., E.P.K.P., A.W.K., and K.C.J.; Data Curation, C.B., E.P.K.P.,
402 and J.M.; Writing – Original Draft, J.M. and E.P.K.P.; Writing – Review & Editing, A.C.D., N.A.C.,
403 M.I.G., A.W.K. and K.C.J.; Visualisation, E.P.K.P and J.M.; Supervision, N.A.K. and K.J.; Project
404 Administration, C.B., M.I.G., K.N.S., K.C.J., and G.K.; Funding Acquisition, M.I.G., and G.K. All
405 authors read and approved the manuscript.

406

407 **Conference presentation**

408 Part of this work was presented at the 14th International dsRNA Virus Symposium 2022, Banff,
409 Alberta, Canada, 10th October - 14th October 2022 and the 1st Kamuzu University of Health (KUHeS)
410 Research Dissemination Conference, Blantyre, Malawi, 24th - 25th November 2022

411

412 **Declaration of Interests**

413 M.I.G. has received research grants from GSK and Merck, and has provided expert advice to GSK.

414 K.C.J. has received investigator-initiated research grant support from GSK.

415

416 **References**

- 417 1. Bode L. Human milk oligosaccharides: Every baby needs a sugar mama. *Glycobiology*. **2012**; 22(9):1147–1162.
- 418 2. Kumar H, Toit E du, Kulkarni A, et al. Distinct patterns in human milk microbiota and fatty acid profiles across
419 specific geographic locations. *Front Microbiol*. **2016**; 7(OCT).
- 420 3. Li SW, Watanabe K, Hsu CC, et al. Bacterial composition and diversity in breast milk samples from mothers living
421 in Taiwan and Mainland China. *Front Microbiol*. **2017**; 8(MAY):1–15.
- 422 4. Ding M, Qi C, Yang Z, et al. Geographical location specific composition of cultured microbiota and: *Lactobacillus*
423 occurrence in human breast milk in China. *Food Funct* [Internet]. **2019**; 10(2):554–564. Available from:
424 <https://doi.org/10.1039/C8FO02182A>
- 425 5. Martín V, Maldonado-Barragán A, Moles L, et al. Sharing of bacterial strains between breast milk and infant feces.
426 *J Hum Lact*. **2012**; 28(1):36–44.
- 427 6. Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. The human milk microbiome changes
428 over lactation and is shaped by maternal weight and mode of delivery. *Am J Clin Nutr*. **2012**; 96(3):544–551.
- 429 7. Ding M, Qi C, Yang Z, et al. Composition of gut microbiota and its influence on the immunogenicity of oral
430 rotavirus vaccines. *Food Funct*. Elsevier Ltd; **2019**; 36(24):554–564.
- 431 8. Lackey KA, Williams JE, Meehan CL, et al. What's normal? Microbiomes in human milk and infant feces are
432 related to each other but vary geographically: The inspire study. *Front Nutr*. **2019**; 6(April).
- 433 9. Khodayar-Pardo P, Mira-Pascual L, Collado MC, Martínez-Costa C. Impact of lactation stage, gestational age and
434 mode of delivery on breast milk microbiota. *J Perinatol*. **2014**; 34:599–605.
- 435 10. Hallowell BD, Tate J, Parashar U. An overview of rotavirus vaccination programs in developing countries. *Expert*
436 *Rev Vaccines* [Internet]. Taylor & Francis; **2020**; 19(6):529–537. Available from:
437 <https://doi.org/10.1080/14760584.2020.1775079>
- 438 11. Bar-Zeev N, King C, Phiri T, et al. Impact of monovalent rotavirus vaccine on diarrhoea-associated post-neonatal
439 infant mortality in rural communities in Malawi: a population-based birth cohort study. *Lancet Glob Heal*. **2018**;

- 440 6(9):e1036–e1044.
- 441 12. Thomas SL, Walker JL, Fenty J, et al. Impact of the national rotavirus vaccination programme on acute
442 gastroenteritis in England and associated costs averted. *Vaccine* [Internet]. The Authors; **2017**; 35(4):680–686.
443 Available from: <http://dx.doi.org/10.1016/j.vaccine.2016.11.057>
- 444 13. Nair NP, Reddy SN, Giri S, et al. Rotavirus vaccine impact assessment surveillance in India: Protocol and
445 methods. *BMJ Open*. **2019**; 9(4):1–8.
- 446 14. Troeger C, Khalil IA, Rao PC, et al. Rotavirus Vaccination and the Global Burden of Rotavirus Diarrhea among
447 Children Younger Than 5 Years. *JAMA Pediatr*. **2018**; 172(10):958–965.
- 448 15. Walker CLF, Black RE. Rotavirus vaccine and diarrhea mortality: quantifying regional variation in effect size.
449 (Special Issue: Technical inputs, enhancements and applications of the Lives Saved Tool (LiST)). *BMC Public*
450 *Health*. **2011**; 11(Suppl. 3):5–11.
- 451 16. Tate JE, Burton AH, Boschi-Pinto C, Steele AD, Duque J, Parashar UD. 2008 estimate of worldwide rotavirus-
452 associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination
453 programmes: a systematic review and meta-analysis. *Lancet Infect Dis* [Internet]. Elsevier; **2012**; 12(2):136–141.
454 Available from: [https://doi.org/10.1016/S1473-3099\(11\)70253-5](https://doi.org/10.1016/S1473-3099(11)70253-5)
- 455 17. Parker EPK, Bronowski C, Sindhu KNC, et al. Impact of maternal antibodies and microbiota development on the
456 immunogenicity of oral rotavirus vaccine in African, Indian, and European infants. *Nat Commun*. Springer US;
457 **2021**; 12(1).
- 458 18. Sindhu KNC, Cunliffe N, Peak M, et al. Impact of maternal antibodies and infant gut microbiota on the
459 immunogenicity of rotavirus vaccines in African, Indian and European infants: Protocol for a prospective cohort
460 study. *BMJ Open*. **2017**; 7(3):1–10.
- 461 19. Kazimbaya KM, Chisenga CC, Simuyandi M, et al. In-vitro inhibitory effect of maternal breastmilk components
462 on rotavirus vaccine replication and association with infant seroconversion to live oral rotavirus vaccine. *PLoS*
463 *One* [Internet]. **2020**; 15(11 November):1–13. Available from: <http://dx.doi.org/10.1371/journal.pone.0240714>
- 464 20. Parker EPK, Ramani S, Lopman BA, et al. Causes of impaired oral vaccine efficacy in developing countries. *Future*
465 *Microbiol*. **2018**; 13(1):97–118.
- 466 21. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. **2011**;
467 17(1):10–12.
- 468 22. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High resolution sample
469 inference from Illumina amplicon data. *Nat Methods*. **2016**; 13(7):581–583.
- 470 23. Davis NM, Proctor DiM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of
471 contaminant sequences in marker-gene and metagenomics data. *Microbiome*. **2018**; 6(1):1–14.
- 472 24. Gautam R, Mijatovic-Rustempasic S, Esona MD, Tam KI, Quaye O, Bowen MD. One-step multiplex real-time RT-
473 PCR assay for detecting and genotyping wild-type group A rotavirus strains and vaccine strains (Rotarix® and

- 474 RotaTeq®) in stool samples. *PeerJ*. **2016**; 2016(1).
- 475 25. Hunt KM, Foster JA, Forney LJ, et al. Characterization of the diversity and temporal stability of bacterial
476 communities in human milk. *PLoS One*. **2011**; 6(6):1–8.
- 477 26. Maqsood R, Skidmore PT, Holland LA, et al. Dynamic Changes in Breast Milk Microbiome in the Early
478 Postpartum Period of Kenyan Women Living with HIV Are Influenced by Antibiotics but Not Antiretrovirals.
479 *Microbiol Spectr. American Society for Microbiology*; **2022**; 10(2):1–10.
- 480 27. Aas JA, Bruce J. P, Lauren N. S, Dewh FE, Aas JA. Defining the Normal Bacterial Flora of the Oral Cavity. *J Clin*
481 *Microbiol*. **2005**; 43(11):5721–5732.
- 482 28. Davé V, Street K, Francis S, et al. Bacterial microbiome of breast milk and child saliva from low-income Mexican-
483 American women and children. *Pediatr Res*. **2016**; 79(6):846–854.
- 484 29. Hunt KM, Foster JA, Forney LJ, et al. Characterization of the diversity and temporal stability of bacterial
485 communities in human milk. *PLoS One*. **2011**; 6(6):1–8.
- 486 30. Ramani S, Stewart CJ, Laucirica DR, et al. Human milk oligosaccharides, milk microbiome and infant gut
487 microbiome modulate neonatal rotavirus infection. *Nat Commun [Internet]. Springer US*; **2018**; 9(1):1–12.
488 Available from: <http://dx.doi.org/10.1038/s41467-018-07476-4>
- 489

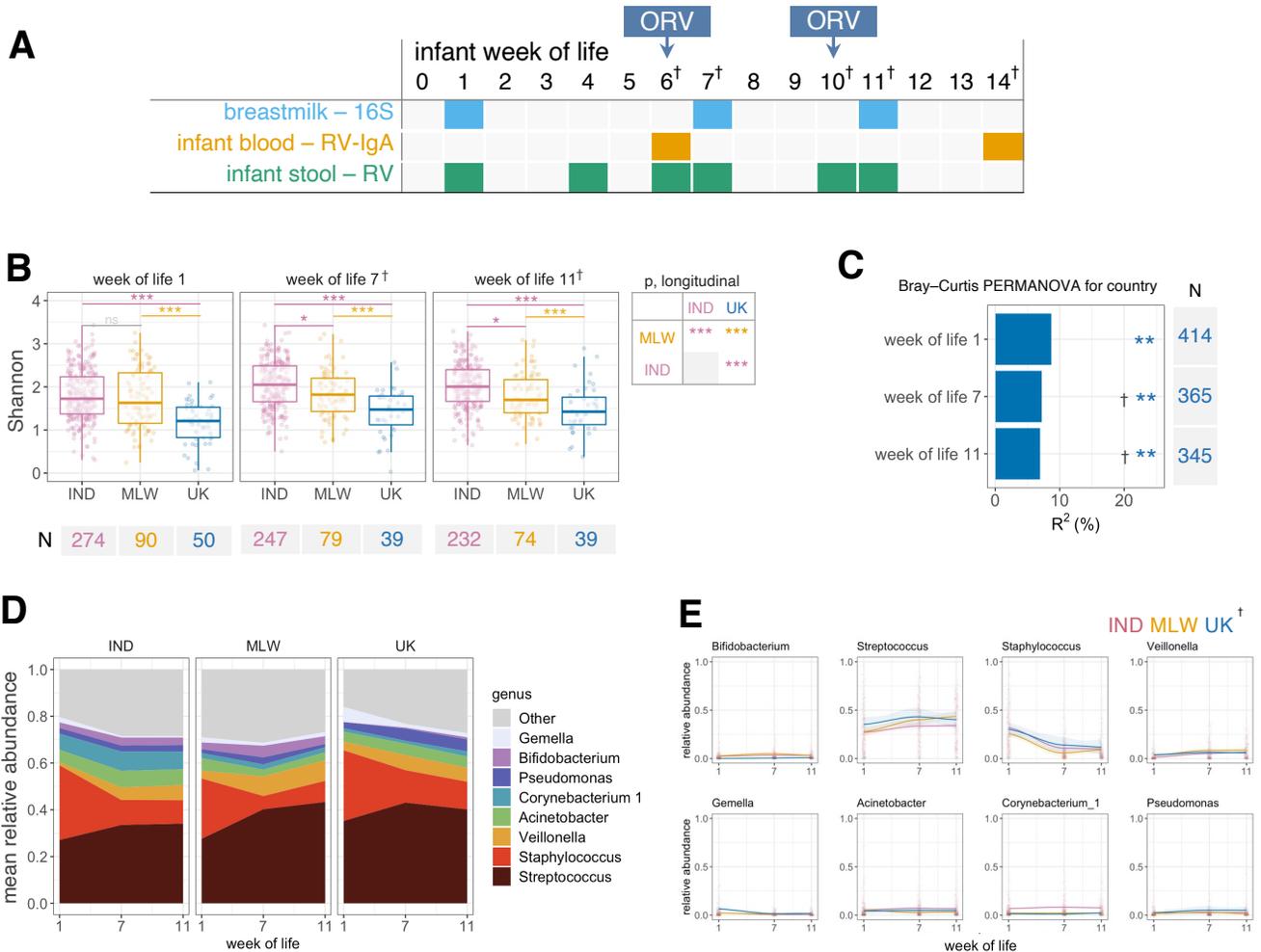


Figure 1. Geographic differences in breastmilk microbiota composition. (A) Sample collection strategy. **(B)** Analysis of alpha diversity, based on genus-level Shannon index. Cross-sectional comparisons were performed using ANOVA with post-hoc Tukey tests. Longitudinal comparisons were performed using mixed-effects regressions with false discovery rate correction of pairwise comparisons. **(C)** Proportion of variation in microbiota composition associated with country, calculated via PERMANOVA using genus-level unweighted Bray–Curtis distances. **(D)** Longitudinal plot of mean genus abundances. Genera are included if present with a mean relative abundance of $\geq 5\%$ in at least one country at one or more timepoints. **(E)** Longitudinal relative abundance plots for major genera by country. Lines show local weighted regression (loess) fits with 95% confidence intervals. IND, India; MLW, Malawi; ns, not significant; †, +2 weeks samples collected at weeks of life 7 and 11 in the UK due to later vaccination schedule; * $p < 0.05$; ** $p = 0.001$; *** $p < 0.0005$.

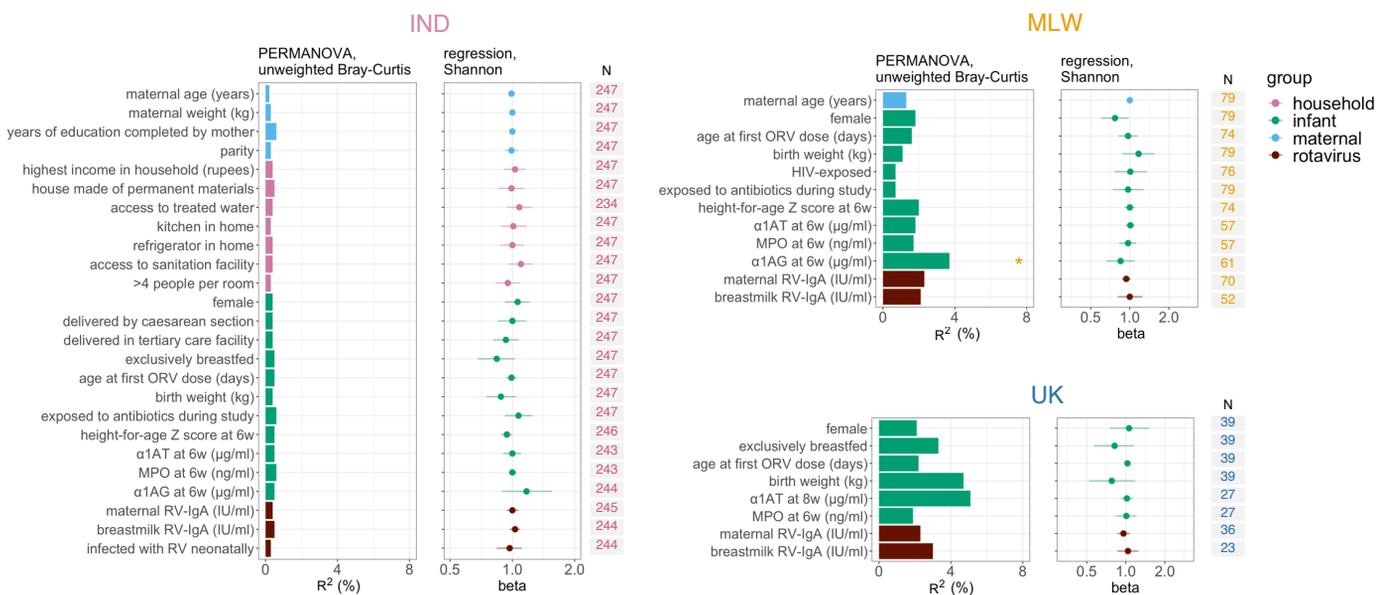


Figure 2. Cofactors associated with breastmilk microbiota composition. Samples collected 1 week after the first dose of oral rotavirus vaccine were included (week of life 7 in India and Malawi; week of life 9 in the UK). The left panel, presenting data for Indian samples ($n = 247$), contains the full list of exploratory variables (with the exception of HIV exposure status, which was also assessed for Malawi). For analyses of samples from Malawi and the UK (right panels; $n = 79$ and 39 , respectively), variables were excluded if they were not measured or exhibited limited variability ($n < 10$ in either comparison group). PERMANOVA was performed using genus-level unweighted Bray–Curtis distances. Shannon index was calculated at genus level and assessed as an outcome variable via linear regression. α1AT, α1-antitrypsin; α1AG, α1 acid glycoprotein; IND, India; MLW, Malawi; MPO, myeloperoxidase; ORV, oral rotavirus vaccine; RV, rotavirus; * FDR $p < 0.05$.

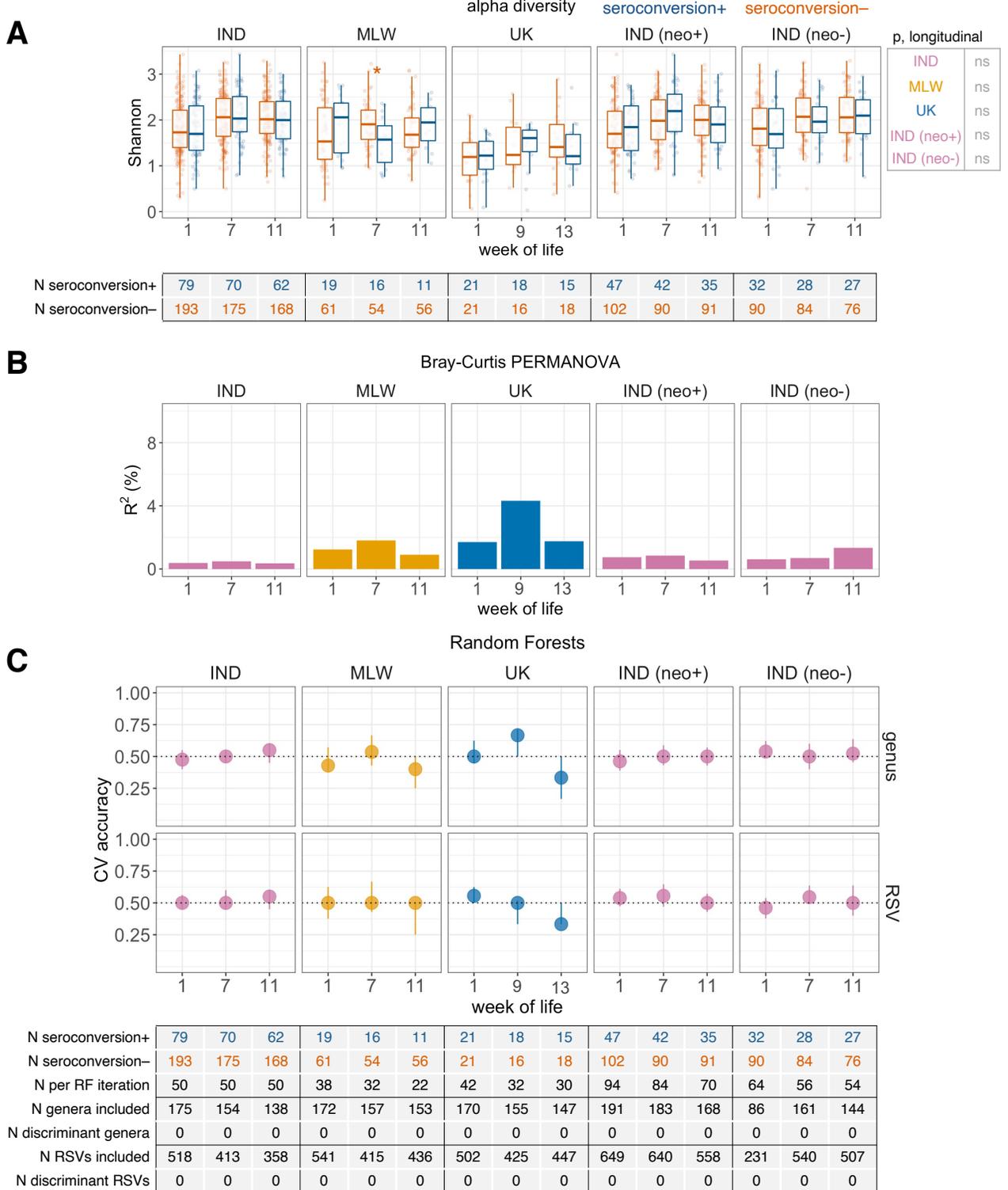
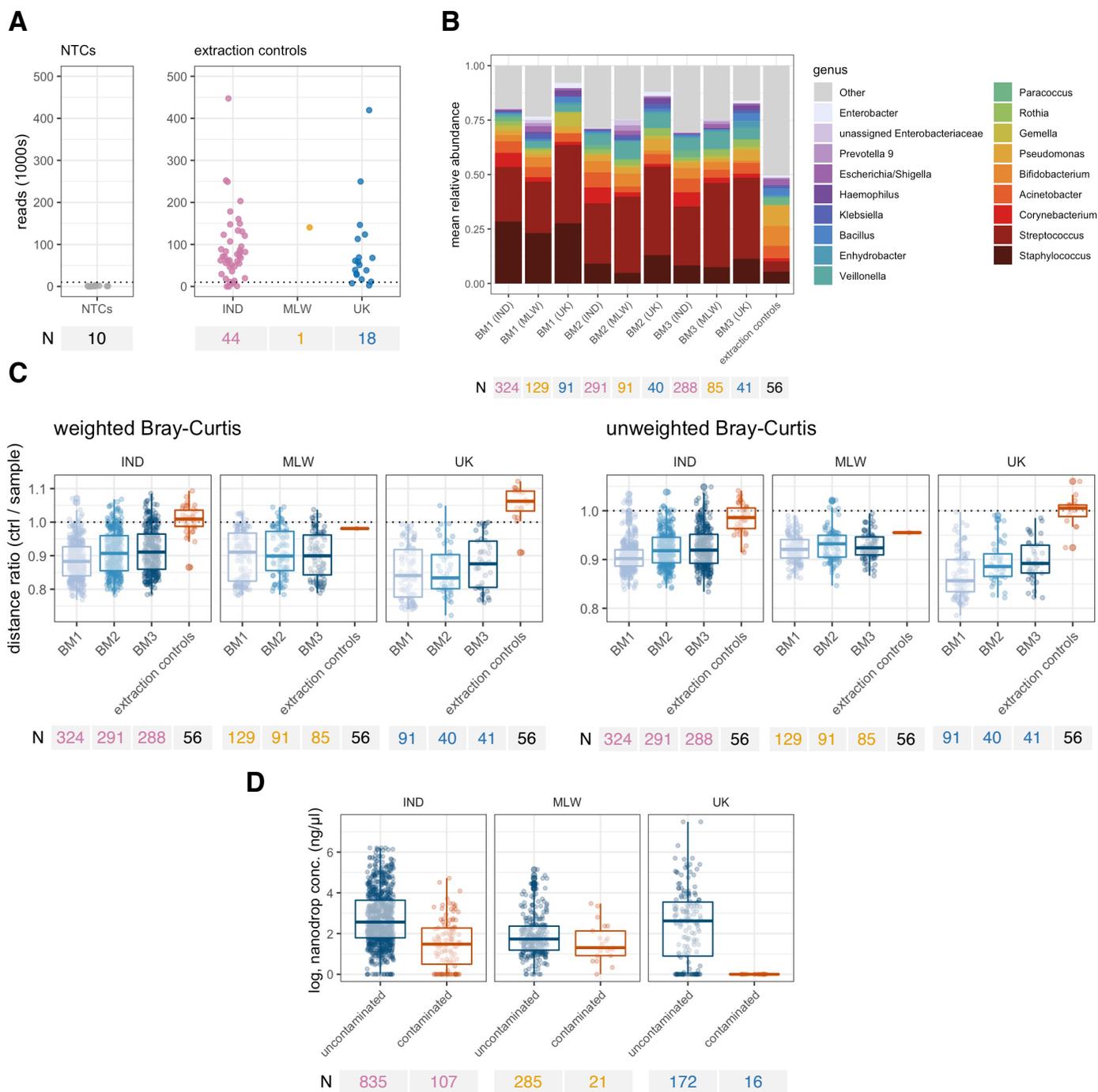
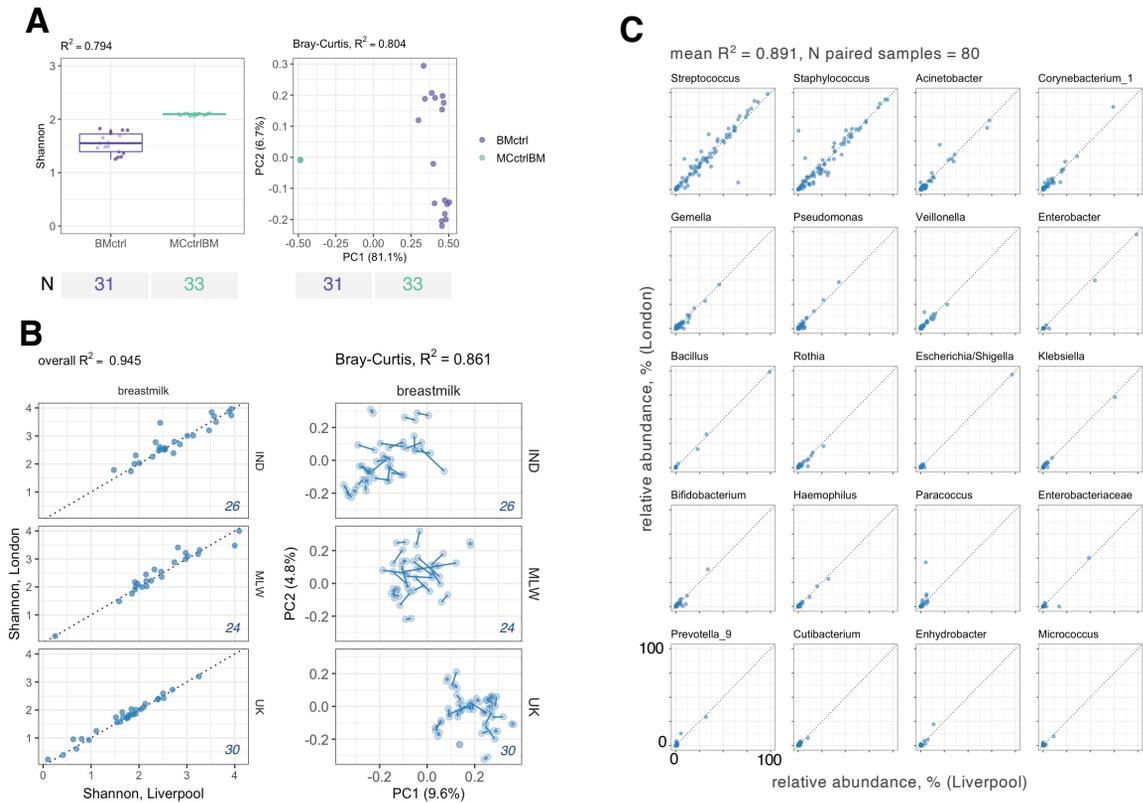


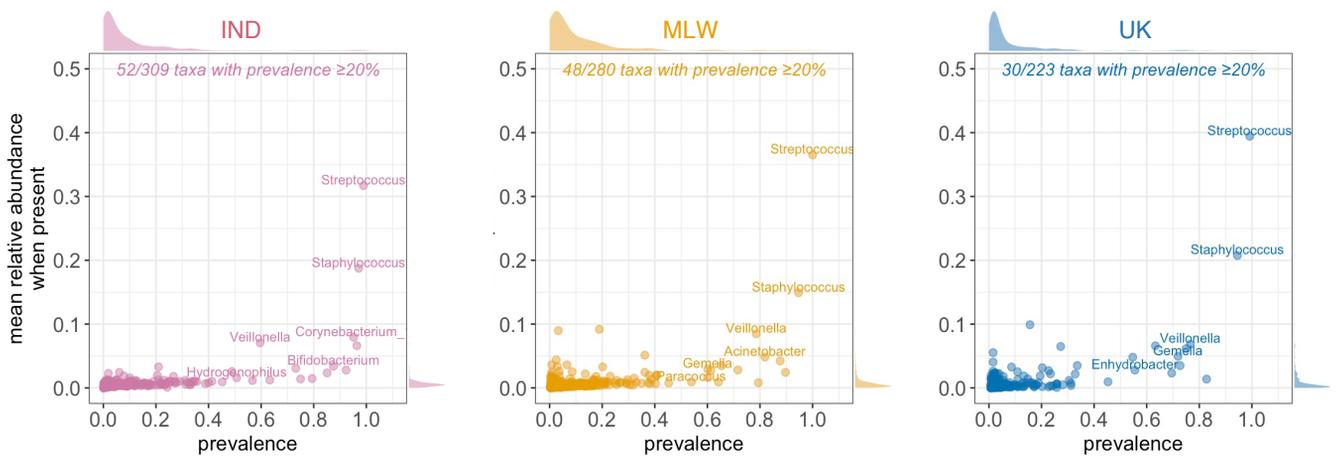
Figure 3. Association between breastmilk microbiota composition and oral rotavirus vaccine seroconversion. (A) Analysis of alpha diversity, based on genus-level Shannon index. Cross-sectional comparisons were performed using logistic regression. Longitudinal comparisons were performed using mixed-effects models. **(B)** Proportion of variation in microbiota composition associated with seroconversion, calculated via PERMANOVA using genus-level unweighted Bray-Curtis distances. **(C)** Cross-validation accuracy of Random Forests for prediction of seroconversion. Median out-of-bag accuracy (proportion correctly assigned) and interquartile range across 20 iterations of 5-fold cross-validation are displayed. Each iteration included an equal number of responders and non-responders (50 per group where possible, or else the number in the minority group if this was <50). Taxa were classified as discriminant if they had an FDR-adjusted p value of <0.05 based on either two-sided Fisher's exact test (differences in prevalence) or Aldex2 with two-sided Wilcoxon rank-sum test (differences in abundance). CV, cross-validation; IND, India; MLW, Malawi; neo+, infected with rotavirus neonatally (defined by detection of rotavirus shedding in week of life 1 or baseline seropositivity); neo-, uninfected with rotavirus neonatally; ns, not significant; RF, Random Forests; RSV, ribosomal sequence variant; * $p < 0.05$.



Supplementary Figure 1. Contaminant filtering steps for breastmilk samples. (A) Read counts for negative controls. Counts from were consistently above 10,000 (dotted line) for pooled or individual extraction controls but not NTCs. (B) Mean genus abundance profile by sample type and country. Extraction controls displayed a distinct genus abundance profile with notable enrichment of rare taxa (labelled 'other'). Samples or controls with at least 10,000 sequences were included. (C) Identification of samples with a contaminant profile. For each sample, the mean distance was calculated from other breastmilk samples from the same country and all breastmilk extraction controls. If the sample clusters more closely with other samples on average, the ratio of these distances will be <1 . (D) Nanodrop concentrations of samples identified as potentially contaminated based on either weighted or unweighted Bray-Curtis (ratio >1 in panel [C]). A pseudocount of 1 was added before log transformation.

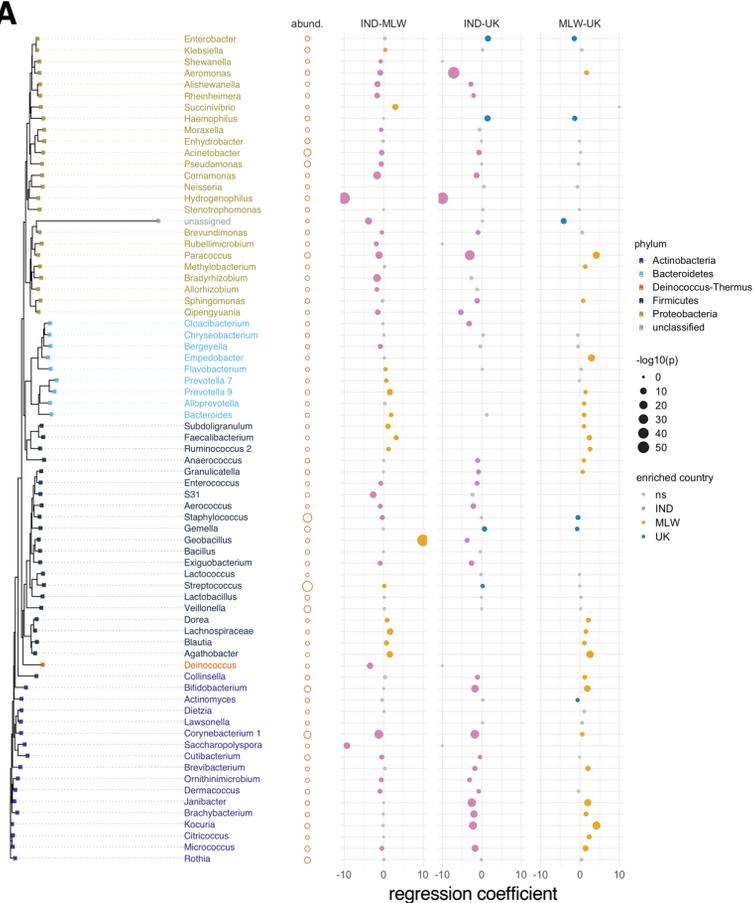


Supplementary Figure 2. Technical replicate profile. Alpha diversity and beta diversity for **(A)** positive controls and **(B)** validation replicates at ribosomal sequence variant level. Positive controls were included on each PCR plate and included a breastmilk sample (BMctrl) and a mock bacterial community (MCctrl). Validation samples were processed at a separate sequencing facility. These were evenly distributed across the study sites (30 per site per sample group) and randomised across a single sequencing plate. Sample pairs were retained in the analysis of both technical replicates had $\geq 15,000$ sequences after quality filtering (80/90 [89%]). The proportion of variation attributable to sample type was determined by linear regression (for Shannon index) and PERMANOVA of unweighted Bray–Curtis distances (for beta diversity). In **(B)**, technical replicates are linked by a line in the right-hand panel. Sample counts are indicated in italics. **(C)** Relative abundances of major genera in validation samples. R^2 was determined based on linear regression.



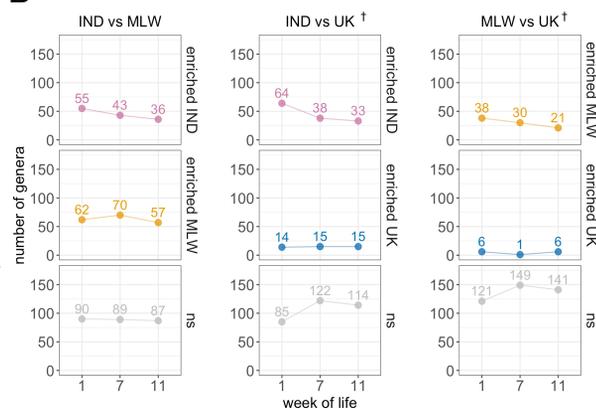
Supplementary Figure 3. Genus abundance and prevalence profile of breastmilk samples. All breastmilk samples from each cohort were included in the prevalence and abundance calculations (n = 753, 243, and 128 for India, Malawi, and the UK, respectively). Margins display density plots. IND, India; MLW, Malawi.

A

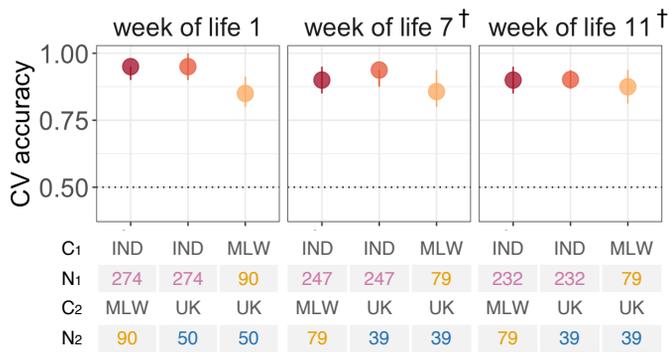
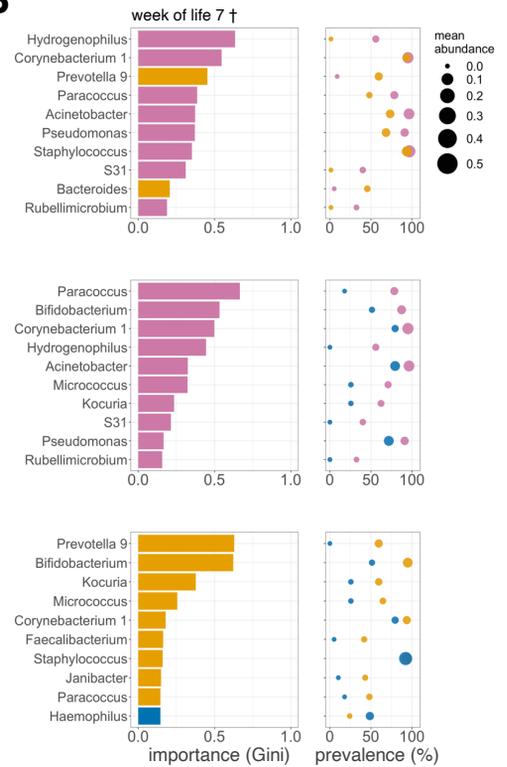


	C1	IND	IND	MLW
N ¹ samples (N ¹ infants)		753 (302)	753 (302)	243 (108)
C2		MLW	UK	UK
N ² samples (N ² infants)		243 (108)	128 (51)	128 (51)
N taxa tested		70	58	54
enriched IND (FDR p<0.05)		29	28	–
enriched MLW (FDR p<0.05)		15	–	26
enriched UK (FDR p<0.05)		–	4	6

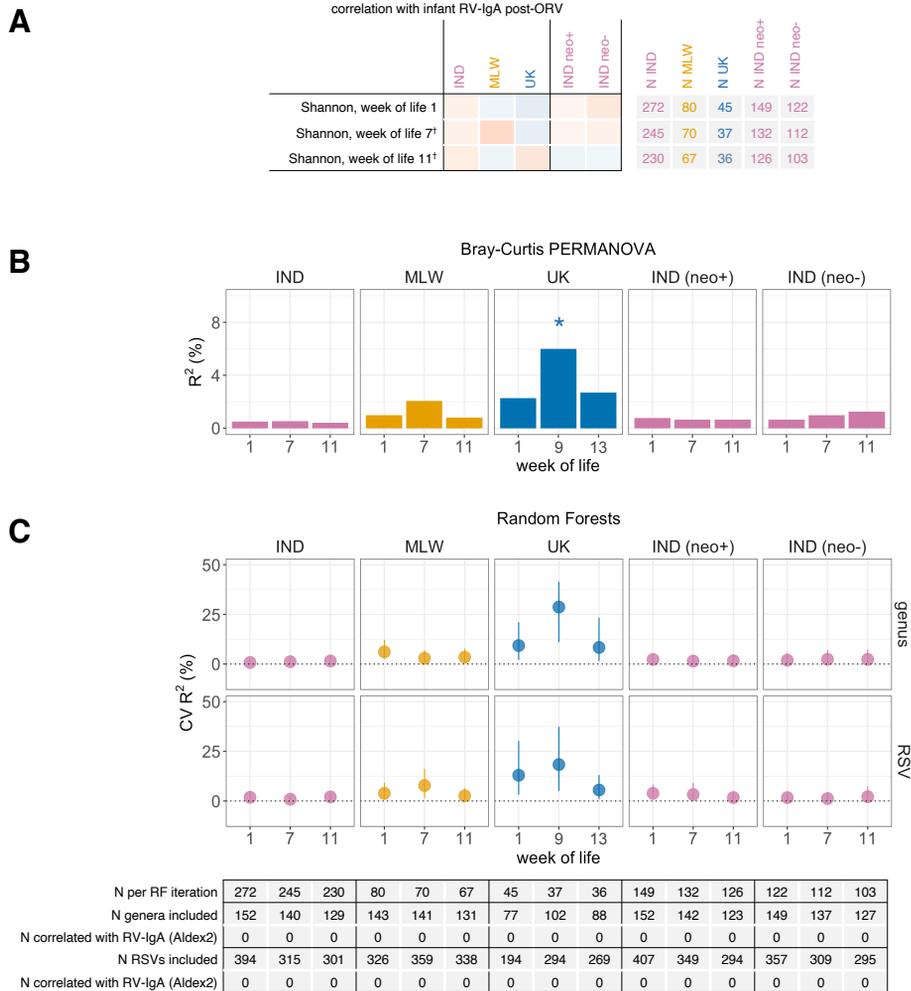
B



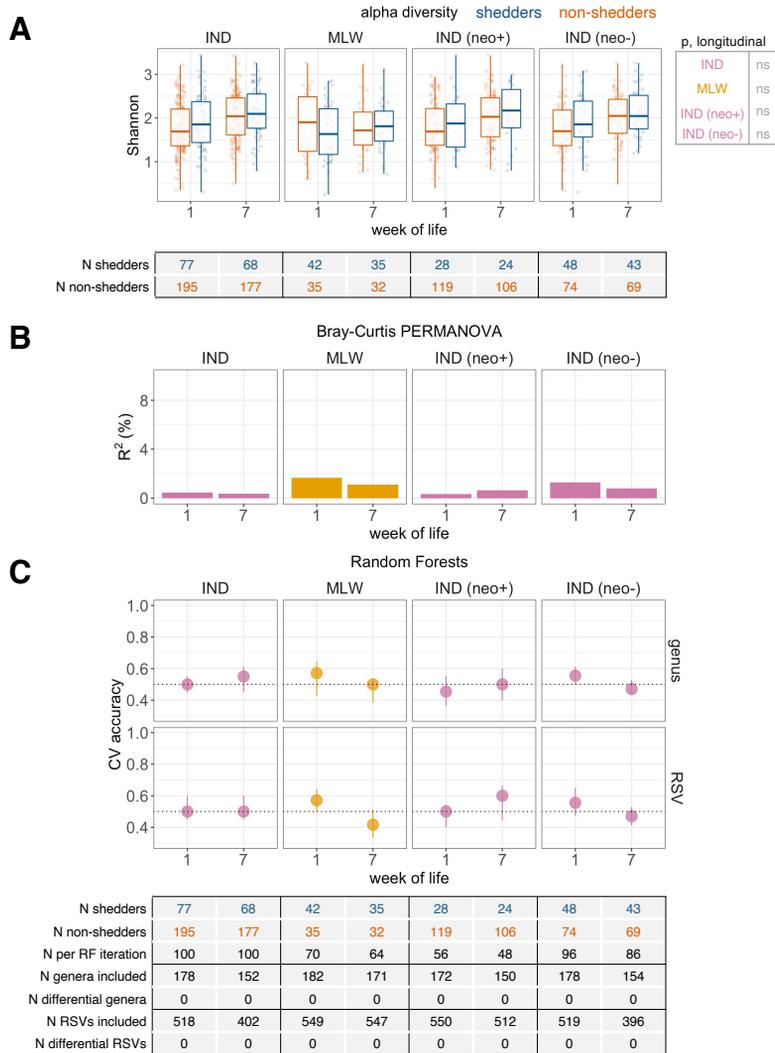
Supplementary Figure 4. (A) Longitudinal models of genus abundance by country. Mixed-effects zero-inflated negative binomial models regressions were used to identify discriminant genera. Genera were included if present in at least 20% of samples from at least one country being compared. Regression coefficients are displayed with point size scaled by p value. Genus order is based on a neighbour-joining tree derived from JC69 distances, with the most abundant ribosomal sequence variant serving as the reference sequence for each genus. Circles to the right of the tree are scaled by mean relative abundance across infant samples (following arcsine square root transformation). **(B)** Cross-sectional comparisons of genus abundance by country. Discriminant genera were identified based on two-sided Fisher’s exact test (differences in prevalence) and Aldex2 with two-sided Wilcoxon rank sum test (differences in abundance). The number of genera with an FDR-adjusted p value of <0.05 based on either method is highlighted for each pairwise cross-sectional comparison. C, country; FDR, false discovery rate; IND, India; MLW, Malawi; ns, not significant; †, +2 weeks samples collected at weeks of life 7 and 11 in the UK due to later vaccination schedule. See **Supplementary Table 1** for full details of discriminant taxa.

A**B**

Supplementary Figure 5. Prediction of country by Random Forests. (A) Cross-validation accuracy of Random Forests. Median out-of-bag accuracy (proportion correctly assigned) and interquartile range across 20 iterations of 5-fold cross-validation are displayed. A random subset of 50 samples per country was used for each iteration. **(B)** The 10 most important genera selected by Random Forests for discriminating infants by country in the week after the first dose of oral rotavirus vaccine. Mean cross-validation importance scores based on Gini index are depicted alongside the prevalence and mean abundance of the corresponding genera. C, country; CV, cross-validation; IND, India; MLW, Malawi; †, +2 weeks in UK due to later vaccination schedule.



Supplementary Figure 6. Association between breastmilk microbiota composition and post-vaccination rotavirus-specific IgA concentration. (A) Analysis of alpha diversity, based on genus-level Shannon index. Shannon index was compared with log-transformed RV-IgA values using Pearson's correlation coefficient (r) with two-sided hypothesis testing. (B) Proportion of variation in microbiota composition associated with RV-IgA, calculated via PERMANOVA using genus-level unweighted Bray-Curtis distances. (C) Cross-validation accuracy of Random Forests for prediction of post-vaccination RV-IgA. Median out-of-bag R^2 and interquartile range are displayed for predicted vs observed RV-IgA across 20 iterations of 5-fold cross-validation. Correlations between log-ratio transformed taxon abundance counts and RV-IgA were determined via Aldex2 with two-sided Spearman's rank test. Taxa were classified as discriminant if they had an FDR-adjusted p value of <0.05 . IND, India; MLW, Malawi; neo+, infected with rotavirus neonatally (defined by detection of rotavirus shedding in week of life 1 or baseline seropositivity); neo-, uninfected with rotavirus neonatally; ns, not significant; RF, Random Forests; RSV, ribosomal sequence variant; * $p < 0.05$.



Supplementary Figure 7. Association between breastmilk microbiota composition and dose 1 oral rotavirus vaccine shedding. See Figure 3 for details; the same analyses of (A) alpha diversity, (B) beta diversity, and (C) Random Forests cross-validation accuracy are presented here with shedding 1 week after the first dose of oral rotavirus vaccine as outcome. Comparisons were not performed for the UK due to the small number of non-shedders (5 out of 60 infants) in this cohort. CV, cross-validation; IND, India; MLW, Malawi; neo+, infected with rotavirus neonatally (defined by detection of rotavirus shedding in week of life 1 or baseline seropositivity); neo-, uninfected with rotavirus neonatally; ns, not significant; RF, Random Forests; RSV, ribosomal sequence variant; * $p < 0.05$.