

1 **Fecal biomarkers of environmental enteric dysfunction and the gut microbiota of**  
2 **rural Malawian children: an observational study**

3  
4 **Author names and affiliations**

5 David Chaima<sup>1,2</sup>, Harry Pickering<sup>1</sup>, John D. Hart<sup>1</sup>, Sarah E. Burr<sup>1,2</sup>, Kenneth M. Maleta<sup>3</sup>,  
6 Khumbo Kalua<sup>4,5</sup>, Robin L. Bailey<sup>1</sup>, \*Martin J. Holland<sup>1</sup>

7  
8 <sup>1</sup>Clinical Research Department, Faculty of Infectious and Tropical Diseases, London School  
9 of Hygiene and Tropical Medicine, London, United Kingdom

10 <sup>2</sup>Department of Pathology, Microbiology Unit, College of Medicine, University of Malawi,  
11 Blantyre, Malawi

12 <sup>3</sup>School of Public Health and Family Medicine, College of Medicine, University of Malawi,  
13 Blantyre, Malawi

14 <sup>4</sup>Blantyre Institute of Community Outreach, Blantyre, Malawi

15 <sup>5</sup>Department of Ophthalmology, College of Medicine, University of Malawi, Blantyre, Malawi

16  
17 **Corresponding author**

18 Martin Holland; Clinical Research Department, Faculty of Infectious and Tropical Diseases,  
19 London School of Hygiene and Tropical Medicine, London, United Kingdom,  
20 +442079272573, [Martin.Holland@lshtm.ac.uk](mailto:Martin.Holland@lshtm.ac.uk)

21  
22 **Highlights**

- 23 • In Malawian children, fecal concentrations of myeloperoxidase, alpha-1 antitrypsin  
24 and neopterin decreased with age
- 25 • A marginal negative association was found between alpha diversity of the gut  
26 microbiota and fecal neopterin concentration
- 27 • A higher abundance of *Succinivibrio* was found in children with lower concentrations  
28 of biomarker of environmental enteric dysfunction
- 29 • Fatty acid biosynthesis, tetrapyrrole biosynthesis and pyrimidine nucleotide  
30 degradation pathways were associated with environmental enteric dysfunction  
31 biomarker score

32

33

## 34 **Summary**

35 Environmental enteric dysfunction (EED) is a subclinical condition of the gut characterized  
36 by changes in morphology and function with underlying chronic inflammatory responses.  
37 This study characterized composition and diversity of the gut microbiota in rural Malawian  
38 children with and without signs of EED. Fecal samples were collected from children aged 1-  
39 59 months. Neopterin, myeloperoxidase and alpha-1 antitrypsin concentrations were  
40 quantified by ELISA and combined to form a composite EED score using principal  
41 component analysis. DNA was extracted from fecal samples and V4-16S rRNA sequencing  
42 was used to characterize the gut microbiota. The concentrations of all three biomarkers  
43 decreased with increasing age, which is consistent with other studies of children living in  
44 similar low-income settings. *Firmicutes*, *Bacteroidetes*, *Proteobacteria* and *Actinobacteria*  
45 were the dominant phyla while *Faecalibacterium* and *Bifidobacterium* were the most  
46 prevalent genera. Increased alpha diversity was associated with a reduction in neopterin  
47 concentration. Microbiota composition was associated with the composite EED score.  
48 Increased abundance of *Succinivibrio* was associated with reduced composite EED scores.

## 49 **Introduction**

50 The gut microbiota is essential in maintaining normal physiological and metabolic  
51 processes. Changes in the gut microbiota composition and function have been associated  
52 with a number of conditions including environmental enteric dysfunction (EED)<sup>1,2</sup>. EED is  
53 characterized by partial villous atrophy, impaired enterocyte tight junctions and underlying  
54 chronic inflammatory responses. The specific cause of EED has not yet been clearly  
55 established, however, one model suggests that ingestion of contaminated food or water  
56 changes the composition and function of the gut microbiota, resulting in intestinal  
57 inflammation and morphological changes to the intestinal epithelium thereby affecting  
58 normal gut function<sup>2,3</sup>. EED affects more than two-thirds of children aged between 0-5 years  
59 living in low-income settings<sup>4,5</sup> and has been linked with growth faltering and reduced  
60 efficacy of oral vaccines<sup>6</sup>.

61  
62 Confirmation of a diagnosis of EED requires visualization of the intestinal epithelium using  
63 endoscopy or biopsy, however, these are invasive tests that would be unethical to perform  
64 in asymptomatic children. The urinary lactulose: mannitol (L:M) ratio test has been  
65 extensively used in the past as a non-invasive test to diagnose EED however, the test is  
66 inconsistent; procedural details, including fasting prior to ingestion, sugar dosage, time of  
67 urine collection and method of detection, vary between studies making comparison of  
68 results difficult. Leakage of urine sample from pediatric urine bags and urine contamination  
69 by stool are other additional challenges in infants. Potential biomarkers that can easily be  
70 quantified in fecal samples and blood using ELISA techniques have been tested as  
71 biomarkers of EED<sup>7-11</sup>. These biomarkers are thought to measure the various processes  
72 involved in EED, such as intestinal epithelial damage, intestinal inflammation, intestinal  
73 permeability, microbial translocation and systemic inflammation<sup>2</sup>. Myeloperoxidase (MPO),

74 alpha-1 antitrypsin (AAT) and neopterin (NEO) are three potential biomarkers that have  
75 been evaluated for use in EED diagnosis, although available data suggests that no single  
76 biomarker is sufficient to indicate EED<sup>6,12</sup>. Accordingly, a multi-site, longitudinal study by  
77 Kosek *et al.*<sup>13</sup> combined the absolute measured concentrations of AAT, MPO and NEO into  
78 a single metric, using principal component analysis, to form a composite EED score. They  
79 then explored the relationship between the composite EED score and linear growth in  
80 infants from resource limited settings. That study reported an association between higher  
81 composite EED scores and growth faltering. In the present study, we used fecal samples  
82 collected from rural Malawian children to characterize the fecal microbiota defined by V4  
83 16S rRNA gene sequencing and explored associations with EED indicated by fecal levels of  
84 MPO, AAT and NEO.

85

## 86 Results

### 87 Distribution of individual biomarker concentration and the composite EED score

88 One hundred and two baseline samples were analyzed by 16S rRNA sequencing as  
89 detailed elsewhere<sup>14</sup>. Of these, 75 had sufficient sample remaining to quantify MPO, 69 for  
90 AAT and 64 for NEO. A total of 60 fecal samples had ELISA results for all 3 biomarkers.  
91 The age and sex of all 102 children who provided fecal samples for sequencing were not  
92 markedly different from those of the children who had sufficient sample to quantify all the 3  
93 biomarkers or any of the 3 biomarkers (**Table 1**).

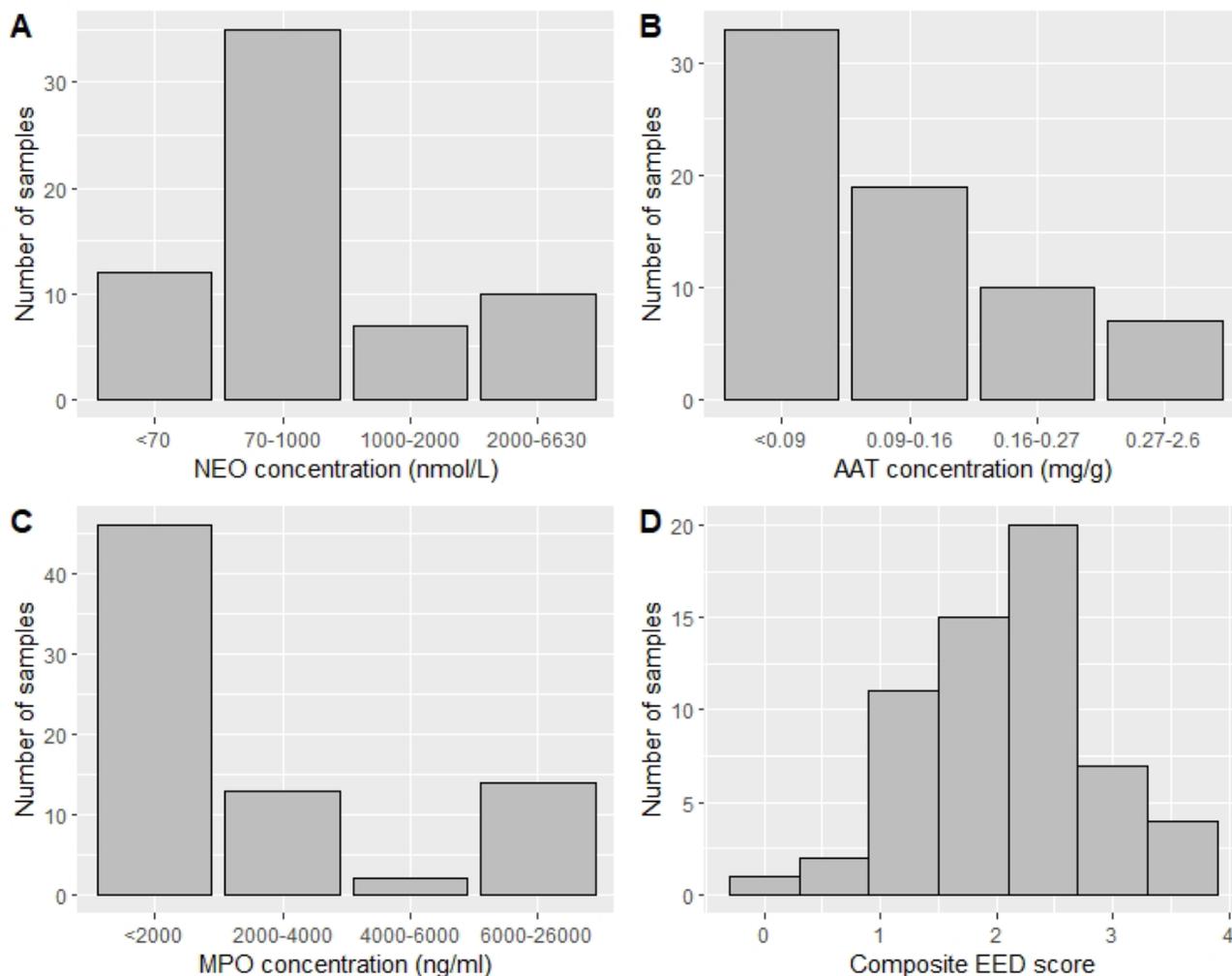
94

95 **Table 1:** Age and sex distribution of all children involved in 16S sequencing and those who  
96 had sufficient sample available for MPO, AAT and NEO assays

Variable	All children (n=102)	All three biomarkers (n=60)	MPO (n=75)	AAT (n=69)	NEO (n=64)
Mean (SD) age in months	23.0 (13.5)	24.3 (13.8)	23.3 (13.8)	23.6 (14.1)	24.0 (13.7)
Male sex, n (%)	52 (51)	34 (57)	41 (55)	38 (55)	35 (55)

97

98 The distribution of fecal biomarker concentrations was calculated in all available samples  
99 and compared by age. The concentration of all three biomarkers was skewed (**Figure 2**). A  
100 large proportion of children (81%) had elevated NEO levels relative to reference values  
101 (**Figure 2A**). Ten percent of the children had elevated levels of fecal AAT (**Figure 2B**) and  
102 38.9% (29/75) of children had elevated MPO levels (**Figure 2C**). The 3 biomarkers were  
103 combined to form a composite EED score as described and the range was plotted (**Figure**  
104 **2D**). EED scores ranged from 0 (lowest score) to 3.46 (highest score) with a median and  
105 inter quartile range (IQR) of 2.26 (0.61).



106  
 107 **Figure 2** Distribution of individual biomarker concentrations and the composite EED  
 108 score. (A) NEO, (B) AAT, (C) MPO and (D) composite EED score. Normal values for the  
 109 individual biomarkers were based on values reported in the literature (NEO  $\leq 70$  nmol/L,  
 110 AAT  $\leq 0.27$  mg/g and MPO  $\leq 2000$  ng/ml)<sup>6,12,13</sup>.

111 Effect of age on biomarker concentration

112 Age in months was categorized into 4 groups as follows: 0-12, 13-24, 25-36, and 37-60 and  
 113 the Kruskal-Wallis test was used to test for the differences in biomarker concentration  
 114 among the groups. Median fecal concentration for all the 3 biomarkers decreased with age  
 115 (Table 2).

116

117 **Table 2:** Median biomarker concentration by age

*Biomarker concentration	Age (months)				<sup>a</sup> p value
	0 -12	13 - 24	25 - 36	37 - 48	

NEO (nmol/L)	2100 (2290), n=18	236 (402), n=18	155 (340), n=12	187 (175), n=16	0.0003
AAT (mg/g)	0.16 (0.21), n=21	0.08 (0.09), n=18	0.06 (0.07), n=14	0.09 (0.14), n=16	0.006
MPO (ng/ml)	3591 (15998), n=22	1313 (2852), n=21	1114(1248), n=15	581 (979), n=17	0.0009

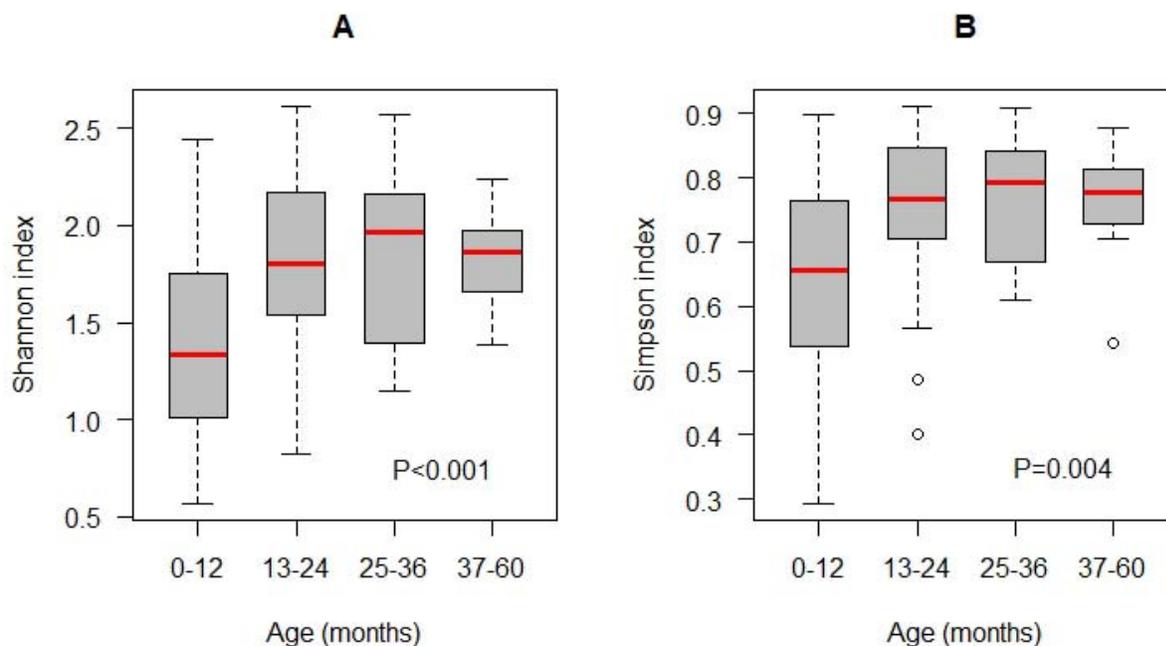
118 \*Biomarker concentrations are presented as median (IQR). <sup>a</sup>p value obtained from a  
119 Kruskal-Wallis test.

### 120 Gut microbiota and the effect of age

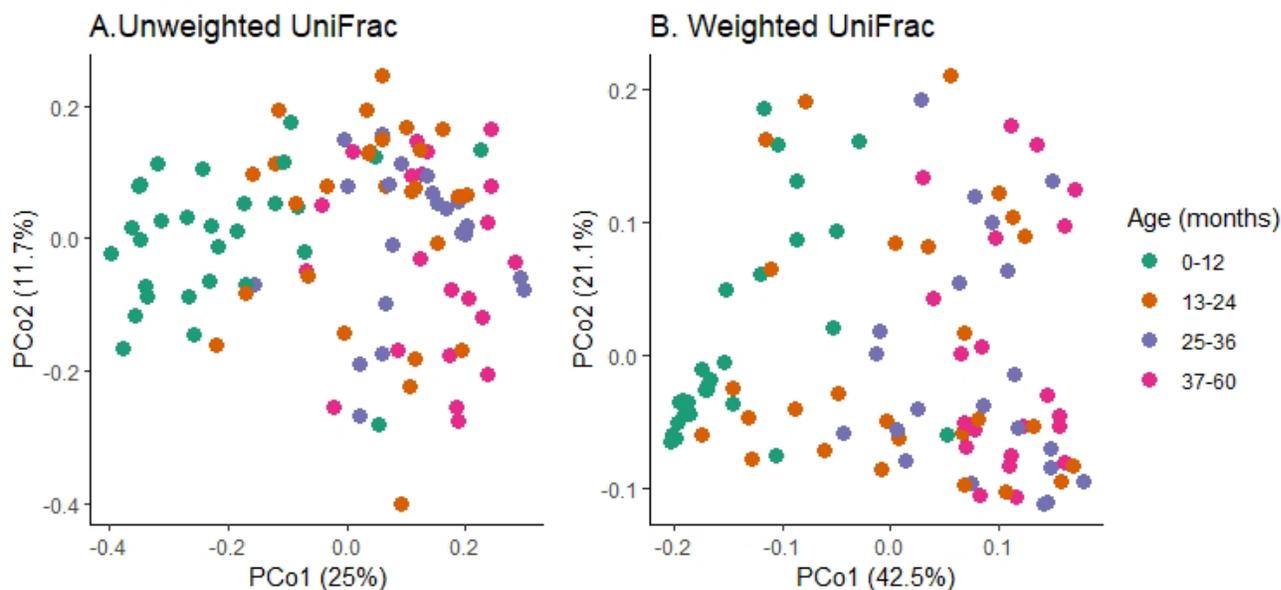
121 Characterization of the microbiota of this sample set has recently been described<sup>14</sup>. Briefly,  
122 the 102 samples generated a total of 1,685,792 reads with an average read depth per  
123 sample of 16,527±11,071. *Firmicutes*, *Bacteroidetes*, *Proteobacteria* and *Actinobacteria*  
124 were the most dominant phyla while *Faecalibacterium*, *Streptococcus*, *Bacteroides*,  
125 *Haemophilus*, *Subdoligranulum*, *Bifidobacterium* and *Escherichia-Shigella* were the most  
126 dominant genera. The mean (SD) Shannon index was 1.72 (0.45) and median (IQR)  
127 Simpson index was 0.75 (0.16).

128 In the present study, the effect of age on Shannon or Simpson diversity indices and the  
129 composition of the microbiota was explored among the 4 age groups. Shannon and  
130 Simpson indices were significantly lower in fecal samples from younger children compared  
131 to older children, indicating increasing diversity with age (**Figure 3A**, ANOVA  $p<0.001$  and  
132 **Figure 3B**, Kruskal-Wallis  $p=0.004$  respectively).

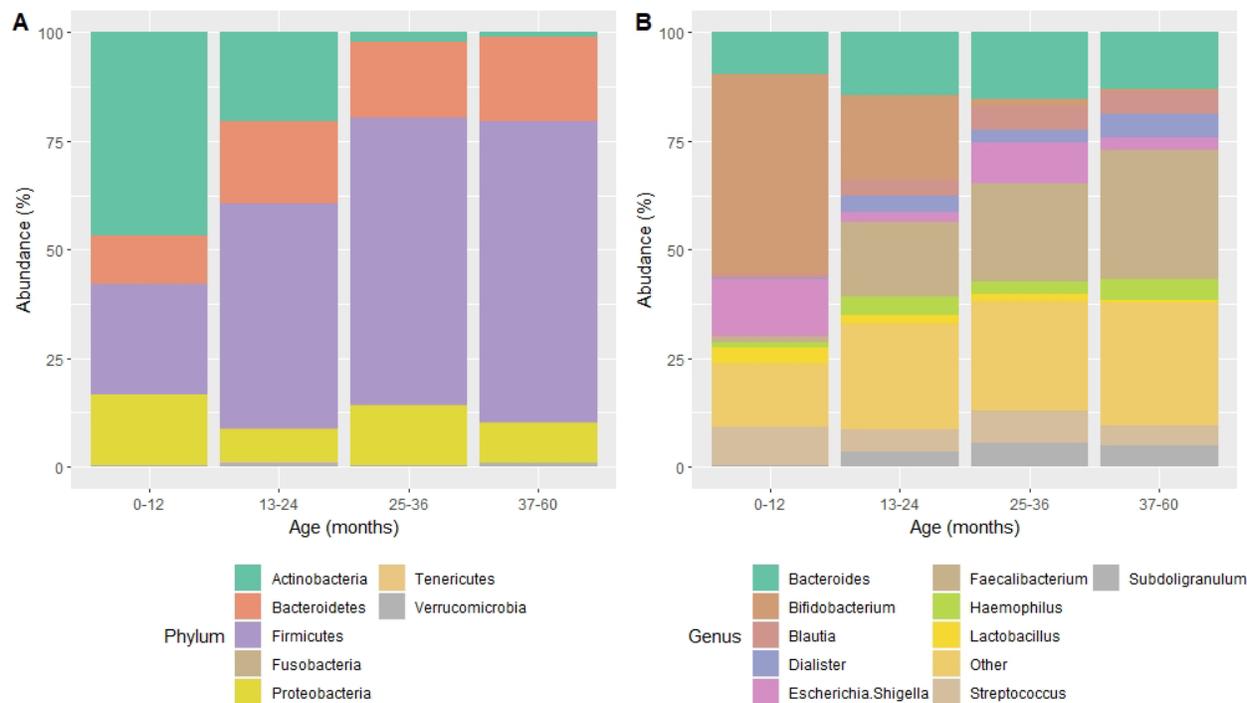
133  
134 Microbiota composition by age group was also examined using PERMANOVA.  
135 Comparisons of weighted and unweighted UniFrac distance showed significant differences  
136 in clustering of microbial communities by age ( $R^2 = 0.17$ ,  $F=6.4$ ,  $p=0.001$  and  $R^2=0.29$ ,  
137  $F=12.8$ ,  $p=0.001$  respectively) (**Figure 4**). Additionally, a comparison of relative abundance  
138 of microbial taxa by age showed that fecal samples from younger children (0-12 months)  
139 had a higher abundance of *Actinobacteria* and a lower abundance of *Bacteroidetes* than  
140 samples from older children (Kruskal-Wallis test  $p=0.01$ , **Figure 5A**). At the genus level, a  
141 higher abundance of *Bifidobacterium* and a lower abundance of genus *Faecalibacterium*  
142 were observed in younger children (0-12 months) compared to the older age groups  
143 (Kruskal-Wallis test  $p=0.01$ , **Figure 5B**).



144  
145 **Figure 3 Comparison of the distribution of (A) Shannon and (B) Simpson indices by**  
146 **age group.  $p$  value for Shannon index ( $p < 0.001$ ) was obtained from ANOVA while  $p$  value**  
147 **for Simpson index ( $p = 0.004$ ) was obtained from the Kruskal-Wallis test.**



148  
149 **Figure 4 Overall microbiota composition differences by age. (A) PCoA plot of**  
150 **unweighted UniFrac distances ( $R^2 = 0.17$ ,  $F = 6.4$ ,  $p = 0.001$ ) (B) PCoA plot of weighted**  
151 **UniFrac distances ( $R^2 = 0.29$ ,  $F = 12.8$ ,  $p = 0.001$ ).**



152

153 **Figure 5 Relative abundance of taxa by age.** (A) Relative abundance of phyla (B)  
 154 Relative abundance of genera. The stacked bar plot shows the 10 most abundant genera;  
 155 the remaining genera were grouped as “Other”.

156 Associations between alpha diversity and fecal biomarkers of EED

157 Linear regression analysis adjusted for age and sex was used to examine the relationship  
 158 between Shannon or Simpson diversity and fecal concentrations of NEO, AAT and MPO.  
 159 There was no relationship between Shannon diversity and AAT or MPO concentrations, or  
 160 between Simpson diversity and AAT or MPO concentrations. (**Table 3**). However, there  
 161 were marginal negative associations between both Shannon and Simpson diversity, and  
 162 NEO concentrations; increased Shannon or Simpson diversity was associated with a  
 163 decrease in NEO concentration (**Table 3**). There was no association between Shannon or  
 164 Simpson diversities and the composite EED scores ( $p=0.14$  and  $P=0.33$  respectively).

165 **Table 3:** Relationship between alpha diversity indices and individual biomarkers of EED

Variable	AAT		MPO		NEO	
	Coefficient (95% CI)	p value	Coefficient (95% CI)	p value	Coefficient (95% CI)	p value
Shannon index	-0.15 (-0.34,0.05)	0.14	-1187 (-4377,2001)	0.46	-758.87 (-1496,-21)	0.04

Simpson index	-0.61		-607		-2710
	(-1.31,0.09)	0.09	(-11935,10721)	0.92	(-5386,-34)

166 *CI= confidence interval*

167 *Coefficients (95%CI) and p values were obtained from linear regression models that were*  
 168 *adjusted for age and sex*

169

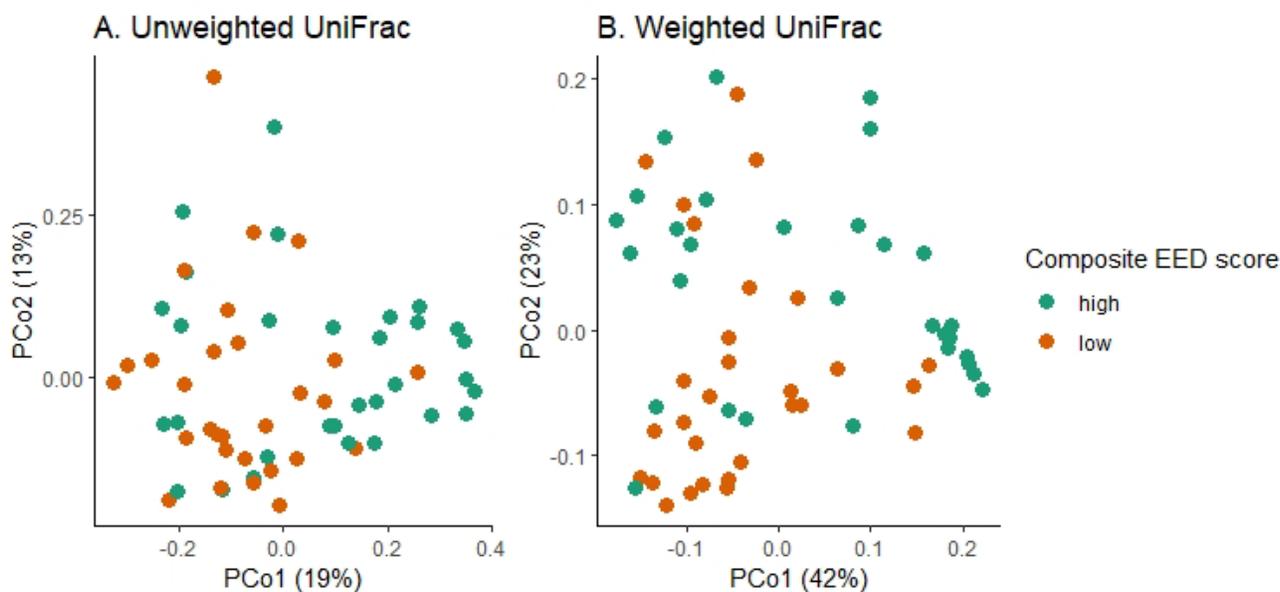
170 *Associations between fecal bacterial community structure and biomarkers of EED*

171 PERMANOVA did not detect any differences in bacterial community composition between  
 172 normal and elevated biomarker concentration for each of the 3 biomarkers (**Table 4**).  
 173 However, PERMANOVA detected significant differences in bacterial community  
 174 composition between fecal samples with low and high composite EED scores using  
 175 unweighted ( $R^2 = 0.04$ ,  $F=2.53$ ,  $p=0.011$ ) and weighted UniFrac ( $R^2 = 0.05$ ,  $F=3.35$ ,  $p=0.017$ )  
 176 (**Figures 6A and B**).

177 **Table 4:** Association between bacterial community composition and individual biomarkers  
 178 of EED.

Variable	Unweighted UniFrac		Weighted UniFrac	
	R <sup>2</sup>	p value	R <sup>2</sup>	p value
<b>MPO</b>	0.02	0.07	0.01	0.6
<b>AAT</b>	0.02	0.18	0.03	0.07
<b>NEO</b>	0.02	0.21	0.02	0.39

179



180

181 **Figure 6 Overall microbiota composition differences by composite EED score.** (A)  
 182 *PCoA plot of unweighted UniFrac* ( $R^2 = 0.04$ ,  $F=2.53$ ,  $p=0.011$ ) (B) *PCoA plot of weighted*  
 183 *UniFrac* ( $R^2 = 0.05$ ,  $F=3.35$ ,  $p=0.017$ ).

184 Differential abundance analysis at the genus level, both unadjusted and adjusted for  
 185 multiple comparison using Benjamin-Hochberg p value correction, was performed to identify  
 186 genera differentially abundant between low and high EED scores. Three genera,  
 187 *Succinivibrio*, *Butyrivibrio* and *Enterococcus*, were differentially abundant in fecal samples  
 188 with high composite EED score compared to fecal samples with low composite EED score  
 189 (**Table 5**). An age and sex-adjusted zero-inflated negative binomial regression model was  
 190 used to further explore this relationship. In the adjusted analysis, only *Succinivibrio* was  
 191 associated with composite EED score. An increase in the abundance of *Succinivibrio* was  
 192 associated with 0.14 reduced odds of a having a high composite EED score (**Table 5**).

193 **Table 5: Differentially abundant genera in fecal samples with high versus**  
 194 **low composite EED scores.**

Genus	log <sub>2</sub> Fold Change	lfcSE	p value <sup>a</sup>	padj <sup>b</sup>	OR <sup>c</sup> (95%CI)	p value <sup>c</sup>
<i>Succinivibrio</i>	-26.3	2.6	<0.001	<0.001	0.14 (0.08,0.23)	<0.001
<i>Butyrivibrio</i>	-25.6	2.9	<0.001	<0.001	3.52 (0.68,18.16)	0.791
<i>Enterococcus</i>	23.5	2.9	<0.001	<0.001	1.82 (0.25,1.2)	0.619

195 <sup>a</sup>Denotes unadjusted p value while <sup>b</sup>denotes p value obtained after adjusting for multiplicity  
 196 with Benjamin-Hochberg p value correction.

197 OR=odds ratio.

198 CI=confidence interval.

199 <sup>c</sup>Denotes OR (95%CI) and p values obtained with zero-inflated negative binomial  
 200 regression (conditional model) after adjusting for age and sex.

201 lfcSE = log fold change Standard Error.

202

203 Associations between predicted functional metagenome of the microbiota and the  
 204 composite EED score

205 We then used PICRUSt2<sup>15</sup> analysis to predict metagenomes from the 16S sequence data of  
 206 all 102 baseline fecal samples. This method uses evolutionary modelling to predict which

207 gene families are present and then combines gene families to estimate the composite  
208 metagenome and their metabolic pathways. PICRUSt2 analysis predicted that the  
209 metagenome of these 102 fecal samples accounted for 431 metabolic pathways, which  
210 were collapsed to 67 superpathways for downstream analysis (**Supplementary table S1**).  
211 To explore associations between metabolic pathways identified by PICRUSt2 and the  
212 composite EED score, we used the subset of 60 samples that had ELISA results for all the  
213 three biomarkers. This sub-set of 60 samples was predicated to contain 52 of the 67  
214 superpathways identified in the 102 sequenced samples, the remaining 15 superpathways  
215 were excluded from further analysis. We then used a lasso regression model to determine  
216 which of these metabolic pathways were most important for predicting the composite EED  
217 score, in which the model coefficient for uninformative variables is reduced to zero. Of the  
218 52 pathways included in the final analysis, the lasso regression model retained five  
219 pathways with non-zero coefficients (**Table 6**). We then explored the relationship between  
220 these five pathways, and the composite EED score using age-adjusted logistic regression  
221 analysis. Three of the five pathways were significantly associated with reduced risk of high  
222 EED scores. A unit increase in the abundance of both the fatty acid biosynthesis and  
223 pyrimidine nucleotide degradation pathways was associated with 0.47 reduced odds of high  
224 composite EED score while a unit increase in the abundance of tetrapyrrole biosynthesis  
225 pathway was associated 0.52 reduced odds of high composite EED score. There were no  
226 significant associations between coenzyme B or catechol degradation pathways and the  
227 composite EED score in this adjusted analysis (**Table 6**).

228 **Table 6:** Associations between metabolic pathways and composite EED score

Metabolic pathway	<sup>a</sup> Variable importance score	OR (95% CI)	<sup>b</sup> p value
Pyrimidine nucleotide degradation	7.80E-05	0.47 (0.23,0.81)	0.006
Fatty Acid biosynthesis	4.79E-07	0.47 (0.23,0.86)	0.015
Tetrapyrrole biosynthesis	2.36E-05	0.52 (0.27,0.87)	0.015
Coenzyme B	7.52E-05	0.99 (0.87,1.05)	0.556
Catechol degradation	2.43E-04	0.99 (0.81,1.25)	0.982

229 *CI= confidence interval, OR=odds ratio. <sup>a</sup>Pathways with non-zero coefficients retained by*  
230 *lasso regression. <sup>b</sup>Permuted (10,000) p value from an age-adjusted logistic regression.*

231

## 232 Discussion

233 We estimated the magnitude of intestinal inflammation and permeability in rural Malawian  
234 children using fecal biomarkers and explored associations with the fecal microbiota,  
235 characterized by 16S rRNA sequencing. We found elevated fecal levels of NEO in almost  
236 two-thirds of the children; MPO was raised in one-third of children whilst AAT was elevated  
237 in 10% of children. All biomarker levels decreased with age, consistent with trends shown

238 by previous studies from similar low-income settings. In studies by Naylor *et al.*<sup>6</sup>, Campbell  
239 *et al.*<sup>16</sup> and McCormick *et al.*<sup>12</sup>, more than 80% of infant fecal samples from low income  
240 settings had elevated concentrations of MPO, AAT and NEO relative to values from high  
241 income settings<sup>6,16</sup> and all biomarker concentrations decreased with age.

242 We also found an effect of age on the microbiota. Microbiota diversity was higher in older  
243 age groups compared to younger children. Younger children were shown to have a higher  
244 abundance of *Actinobacteria* and *Bifidobacterium*, and a lower abundance of the  
245 *Bacteroidetes* and *Faecalibacterium* compared to the older age groups. These findings are  
246 consistent with trends reported by recent longitudinal studies<sup>17,18</sup>. The increase in diversity  
247 and changes in intestinal microbiota seen with the different age groups are also consistent  
248 with changes that have been reported during the transition from breast-feeding to  
249 weaning<sup>19,20</sup>.

250 We did not find a relationship between alpha diversity and fecal levels of AAT or MPO,  
251 however a marginal negative association was seen between alpha diversity and NEO, a  
252 biomarker of intestinal inflammation. To date, there is no other published data available  
253 supporting a relationship between fecal microbiota diversity and intestinal inflammation in  
254 EED, however, a reduction in fecal microbiota diversity has been reported previously in  
255 inflammatory bowel disease<sup>21-23</sup>, which is consistent with our findings. These results  
256 suggest a potential role of the gut microbiota in modulating inflammatory responses.

257 There is limited published data indicating associations between intestinal bacterial  
258 community structure and EED. Recently, Ordiz *et al.* found differences in intestinal bacterial  
259 composition between children with increased versus normal intestinal permeability as  
260 measured by the L:M ratio test<sup>24</sup>; the genera *Succinivibrio*, *Klebsiella*, and *Clostridium\_XI*  
261 were less prevalent in children with increased intestinal permeability compared to children  
262 with normal levels. Similarly, we found an association between higher abundance of  
263 *Succinivibrio* and low composite EED score. *Succinivibrio* is a gram-negative anaerobic  
264 commensal of the human colon that has been associated with fiber-degradation<sup>25,26</sup>. The  
265 association between *Succinivibrio* abundance and the composite EED score, which is  
266 consistent with the results of others<sup>24</sup>, indicates a potential role for *Succinivibrio* in gut  
267 function.

268  
269 Using the 16SrRNA sequencing data generated, PICRUSt2 analysis predicted 67 metabolic  
270 pathways of which 5 were identified by the lasso regression model as important for  
271 predicting EED. Among the 5 pathways, the fatty acid biosynthesis, pyrimidine nucleotide  
272 degradation and tetrapyrrole biosynthesis pathways were significantly associated with  
273 composite EED score in an age-adjusted logistic regression model, which to an extent  
274 supports the prediction accuracy of the lasso regression model. Fatty acid synthase, an  
275 enzyme that catalyzes fatty acid biosynthesis, has been shown to help maintain the  
276 intestinal mucus barrier by regulating mucin 2 in murine models<sup>27</sup>. In the present study, we

277 found that an increase in the relative abundance of the fatty acid biosynthesis pathway was  
278 associated with reduced chance of high composite EED scores. As such, we can postulate  
279 that the fatty acid biosynthesis pathway plays a role in the maintenance of gut integrity in  
280 humans. For their part, pyrimidines serve important roles in human metabolism, such as  
281 ribonucleotide bases in RNA (uracil and cytosine) and deoxyribonucleotide bases in DNA  
282 (cytosine and thymine). Some bacterial species are known to catabolize pyrimidines via a  
283 reductive pathway leading to the formation of  $\beta$ -amino acids<sup>28,29</sup>. Dietary supplementation of  
284  $\alpha$ -amino acids has been shown to improve intestinal mucosa development in animals<sup>30</sup>,  
285 however, there is limited data on the role of  $\beta$ -amino acids formed from the pyrimidine  
286 nucleotide degradation pathway in gut function. Thus, the role of the pyrimidine nucleotide  
287 degradation pathway in gut function should be explored further. The tetrapyrrole  
288 biosynthesis pathway includes pathways in porphyrin (heme) and corrinoid (cobalamin)  
289 metabolism. Some porphyrins have been shown to have anti-inflammatory properties  
290 through the inhibition of the activity of Fyn, a non-receptor Src-family tyrosine kinase,  
291 triggering anti-inflammatory events associated with down-regulation of T-cell receptor signal  
292 transduction, leading to inhibition of tumor necrosis factor alpha (TNF- $\alpha$ ) production<sup>31</sup>. In the  
293 present study, enrichment of the tetrapyrrole biosynthesis pathway was associated with a  
294 reduced chance of high composite EED score, a result consistent with the anti-inflammatory  
295 properties of components of the tetrapyrrole pathway.

296  
297 The present study also found a higher abundance of *Succinivibrio* in children with low  
298 composite EED scores. This bacterium has been associated with metabolism of  
299 nondigestible dietary carbohydrates in humans and animals<sup>25,26</sup>. Among the 67 metabolic  
300 pathways predicted by PICRUST2 in the present study, there were pathways that are  
301 involved in carbohydrate metabolism such as the sugar degradation pathway, however  
302 these pathways were not identified as being important in predicting EED using the lasso  
303 regression model. Further studies are needed to explore the correlations between  
304 *Succinivibrio* abundance, carbohydrate metabolism and gut function.

305  
306 In conclusion, our data suggest that there is an association between intestinal microbiota  
307 diversity and intestinal inflammation as measured by fecal NEO. The genus *Succinivibrio*,  
308 which was reported by Ordiz *et al.*<sup>24</sup> to be reduced in EED, was also associated with low  
309 composite EED scores in our study. This bacterium has been associated with fiber-  
310 degradation<sup>25</sup> and has recently been shown to be more abundant in children consuming a  
311 diet rich in fiber in low-income settings<sup>26</sup>. Future studies should further investigate the role  
312 of this bacterium and the predicted metabolic pathways important in EED.

313  
314 **Limitations of the study**  
315 One of the limitations of this study is the use of EED biomarker values from high income  
316 countries to define cut-off limits for normal or elevated biomarker concentration. This

317 limitation is not unique to this study; the lack of reference values for the biomarkers of EED  
318 in children living in low income settings has been cited by others<sup>6,12</sup>. There is therefore a  
319 need for large-scale studies to define reference values of the biomarkers of EED in children  
320 living in low income settings. A further limitation was that this was an opportunistic study  
321 that used previously collected samples and therefore may not have had adequate statistical  
322 power to identify significant associations. A population-based survey would likely have been  
323 more robust.

324

## 325 **Resource availability**

### 326 Lead contact

327 Further information and requests for resources should be directed to and will be fulfilled by  
328 the lead contact, Martin Holland ([Martin.Holland@lshtm.ac.uk](mailto:Martin.Holland@lshtm.ac.uk)).

329

### 330 Materials availability

331 The study did not generate new unique reagents or materials.

332

### 333 Data and code availability

334 The datasets generated from the present study are available from the corresponding author  
335 on reasonable request. The V4 16S rRNA sequence data for this study have been  
336 deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number  
337 PRJEB41238 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB41238>).

338

## 339 **Methods**

340 All methods can be found in the accompanying Transparent Methods supplemental file.

341

## 342 **Acknowledgments**

343 We would like to acknowledge the support of the field staff at the Blantyre Institute of  
344 Community Outreach, Blantyre, Malawi. Funding for this work was provided by the Bill and  
345 Melinda Gates Foundation, under grant numbers OPP1066930 and OP1032340.

346

## 347 **Author contributions**

348 Conceptualization, SEB, KK, RLB and MJH; Methodology, DC, HP, and JDH; Formal  
349 Analysis, DC and HP; Resources, SEB, KK, RLB and MJH; Data Curation, DC, JDH and  
350 KK; Writing – Original Draft Preparation, DC; Writing – Review & Editing, DC, HP, SEB,  
351 KMM, KK, RLB and MJH; Supervision, SEB, KK, RLB and MJH; Project Administration; KK  
352 and RLB; Funding Acquisition, SEB, KK, RLB and MJH. All authors read and approved the  
353 final manuscript.

354

## 355 **Declaration of interests**

356 The authors declare no competing interests.

357

## 358 **References**

- 359 1 Schirmer, M., Franzosa, E. A., Lloyd-Price, J., McIver, L. J., Schwager, R., Poon, T.  
360 W., Ananthakrishnan, A. N., Andrews, E., Barron, G., Lake, K., Prasad, M., Sauk, J.,  
361 Stevens, B., Wilson, R. G., Braun, J., Denson, L. A., Kugathasan, S., McGovern, D.  
362 P. B., Vlamakis, H. *et al.* Dynamics of metatranscription in the inflammatory bowel  
363 disease gut microbiome. *Nature microbiology* **3**, 337-346, doi:10.1038/s41564-017-  
364 0089-z (2018).
- 365 2 Prendergast AJ, H. J., Mutasa K Assessment of Environmental Enteric Dysfunction  
366 in the SHINE Trial: Methods and Challenges. *Clin Infect Dis* **61**, S726–S732,  
367 doi:10.1093/cid/civ848 (2015).
- 368 3 Lunn, P. G. The impact of infection and nutrition on gut function and growth in  
369 childhood. *The Proceedings of the Nutrition Society* **59**, 147-154 (2000).
- 370 4 Galpin, L., Manary, M. J., Fleming, K., Ou, C. N., Ashorn, P. & Shulman, R. J. Effect  
371 of Lactobacillus GG on intestinal integrity in Malawian children at risk of tropical  
372 enteropathy. *The American journal of clinical nutrition* **82**, 1040-1045,  
373 doi:10.1093/ajcn/82.5.1040 (2005).
- 374 5 Campbell, D. I., McPhail, G., Lunn, P. G., Elia, M. & Jeffries, D. J. Intestinal  
375 inflammation measured by fecal neopterin in Gambian children with enteropathy:  
376 association with growth failure, *Giardia lamblia*, and intestinal permeability. *Journal of*  
377 *pediatric gastroenterology and nutrition* **39**, 153-157 (2004).
- 378 6 Naylor, C., Lu, M., Haque, R., Mondal, D., Buonomo, E., Nayak, U., Mychaleckyj, J.  
379 C., Kirkpatrick, B., Colgate, R., Carmolli, M., Dickson, D., van der Klis, F., Weldon,  
380 W., Steven Oberste, M., teams, P. s., Ma, J. Z. & Petri, W. A., Jr. Environmental  
381 Enteropathy, Oral Vaccine Failure and Growth Faltering in Infants in Bangladesh.  
382 *EBioMedicine* **2**, 1759-1766, doi:10.1016/j.ebiom.2015.09.036 (2015).
- 383 7 Lin, A., Arnold, B. F., Afreen, S., Goto, R., Huda, T. M. N., Haque, R., Raqib, R.,  
384 Unicomb, L., Ahmed, T., Colford, J. M. & Luby, S. P. Household environmental  
385 conditions are associated with enteropathy and impaired growth in rural Bangladesh.  
386 *The American journal of tropical medicine and hygiene* **89**, 130-137,  
387 doi:10.4269/ajtmh.12-0629 (2013).
- 388 8 Jones, K. D., Hunten-Kirsch, B., Laving, A. M., Munyi, C. W., Ngari, M., Mikusa, J.,  
389 Mulongo, M. M., Odera, D., Nassir, H. S., Timbwa, M., Owino, M., Fegan, G., Murch,  
390 S. H., Sullivan, P. B., Warner, J. O. & Berkley, J. A. Mesalazine in the initial  
391 management of severely acutely malnourished children with environmental enteric  
392 dysfunction: a pilot randomized controlled trial. *BMC medicine* **12**, 133,  
393 doi:10.1186/s12916-014-0133-2 (2014).
- 394 9 George, C. M., Oldja, L., Biswas, S., Perin, J., Lee, G. O., Kosek, M., Sack, R. B.,  
395 Ahmed, S., Haque, R., Parvin, T., Azmi, I. J., Bhuyian, S. I., Talukder, K. A.,  
396 Mohammad, S. & Faruque, A. G. Geophagy is associated with environmental

- 397 enteropathy and stunting in children in rural Bangladesh. *The American journal of*  
398 *tropical medicine and hygiene* **92**, 1117-1124, doi:10.4269/ajtmh.14-0672 (2015).
- 399 10 Donowitz, J. R., Haque, R., Kirkpatrick, B. D., Alam, M., Lu, M., Kabir, M., Kakon, S.  
400 H., Islam, B. Z., Afreen, S., Musa, A., Khan, S. S., Colgate, E. R., Carmolli, M. P.,  
401 Ma, J. Z. & Petri, W. A., Jr. Small Intestine Bacterial Overgrowth and Environmental  
402 Enteropathy in Bangladeshi Children. *mBio* **7**, e02102-02115,  
403 doi:10.1128/mBio.02102-15 (2016).
- 404 11 Benzoni, N., Korpe, P., Thakwalakwa, C., Maleta, K., Stephenson, K., Manary, M. &  
405 Manary, M. Plasma endotoxin core antibody concentration and linear growth are  
406 unrelated in rural Malawian children aged 2-5 years. *BMC research notes* **8**, 258,  
407 doi:10.1186/s13104-015-1238-1 (2015).
- 408 12 McCormick BJJ, L. G., Seidman JC, et al. Dynamics and Trends in Fecal Biomarkers  
409 of Gut Function in Children from 1-24 Months in the MAL-ED Study. *Am J Trop Med*  
410 *Hyg.* **96**, 465-472, doi:10.4269/ajtmh.16-0496 (2017).
- 411 13 Kosek, M., Haque, R., Lima, A., Babji, S., Shrestha, S., Qureshi, S., Amidou, S.,  
412 Mduma, E., Lee, G., Yori, P. P., Guerrant, R. L., Bhutta, Z., Mason, C., Kang, G.,  
413 Kabir, M., Amour, C., Bessong, P., Turab, A., Seidman, J. *et al.* Fecal markers of  
414 intestinal inflammation and permeability associated with the subsequent acquisition  
415 of linear growth deficits in infants. *The American journal of tropical medicine and*  
416 *hygiene* **88**, 390-396, doi:10.4269/ajtmh.2012.12-0549 (2013).
- 417 14 Chaima, D. P., H.; Hart, J.D.; Burr, S.E.; Houghton, J.; Maleta, K.; Kalua, K.; Bailey,  
418 R.L.; Holland, M.J. . Up to Four Biannual Administrations of Mass Azithromycin  
419 Treatment Are Associated with Modest Changes in the Gut Microbiota of Rural  
420 Malawian Children. *Preprints* **2020**, doi:doi: 10.20944/preprints202010.0138.v1  
421 (2020).
- 422 15 Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C.  
423 M., Huttenhower, C. & Langille, M. G. I. PICRUSt2 for prediction of metagenome  
424 functions. *Nature Biotechnology* **38**, 685-688, doi:10.1038/s41587-020-0548-6  
425 (2020).
- 426 16 Campbell, R. K., Schulze, K. J., Shaikh, S., Mehra, S., Ali, H., Wu, L., Raqib, R.,  
427 Baker, S., Labrique, A., West, K. P., Jr. & Christian, P. Biomarkers of Environmental  
428 Enteric Dysfunction Among Children in Rural Bangladesh. *Journal of pediatric*  
429 *gastroenterology and nutrition* **65**, 40-46, doi:10.1097/mpg.0000000000001557  
430 (2017).
- 431 17 Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J.-z., Abe,  
432 F. & Osawa, R. Age-related changes in gut microbiota composition from newborn to  
433 centenarian: a cross-sectional study. *BMC microbiology* **16**, 90, doi:10.1186/s12866-  
434 016-0708-5 (2016).
- 435 18 Yatsunencko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G.,  
436 Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A.

- 437 C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber,  
438 C., Clemente, J. C., Knights, D. *et al.* Human gut microbiome viewed across age and  
439 geography. *Nature* **486**, 222-227, doi:10.1038/nature11053 (2012).
- 440 19 De Cruz, P., Prideaux, L., Wagner, J., Ng, S. C., McSweeney, C., Kirkwood, C.,  
441 Morrison, M. & Kamm, M. A. Characterization of the gastrointestinal microbiota in  
442 health and inflammatory bowel disease. *Inflammatory bowel diseases* **18**, 372-390,  
443 doi:10.1002/ibd.21751 (2012).
- 444 20 Kleessen, B., Bunke, H., Tovar, K., Noack, J. & Sawatzki, G. Influence of two infant  
445 formulas and human milk on the development of the faecal flora in newborn infants.  
446 *Acta paediatrica (Oslo, Norway : 1992)* **84**, 1347-1356 (1995).
- 447 21 Michail, S., Durbin, M., Turner, D., Griffiths, A. M., Mack, D. R., Hyams, J., Leleiko,  
448 N., Kenche, H., Stolfi, A. & Wine, E. Alterations in the gut microbiome of children with  
449 severe ulcerative colitis. *Inflammatory bowel diseases* **18**, 1799-1808,  
450 doi:10.1002/ibd.22860 (2012).
- 451 22 Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L.,  
452 Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J. & Dore, J. Reduced diversity  
453 of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut*  
454 **55**, 205-211, doi:10.1136/gut.2005.073817 (2006).
- 455 23 Dicksved, J., Halfvarson, J., Rosenquist, M., Järnerot, G., Tysk, C., Apajalahti, J.,  
456 Engstrand, L. & Jansson, J. K. Molecular analysis of the gut microbiota of identical  
457 twins with Crohn's disease. *The ISME journal* **2**, 716-727, doi:10.1038/ismej.2008.37  
458 (2008).
- 459 24 Ordiz, M. I. e. a. Environmental Enteric Dysfunction and the Fecal microbiota in  
460 Malawian Children. *The American journal of tropical medicine and hygiene* **96**, 473-  
461 476 (2017).
- 462 25 BRYANT, M. P. Normal Flora—Rumen Bacteria. *The American journal of clinical*  
463 *nutrition* **23**, 1440-1450, doi:10.1093/ajcn/23.11.1440 (1970).
- 464 26 De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S.,  
465 Collini, S., Pieraccini, G. & Lionetti, P. Impact of diet in shaping gut microbiota  
466 revealed by a comparative study in children from Europe and rural Africa.  
467 *Proceedings of the National Academy of Sciences of the United States of America*  
468 **107**, 14691-14696, doi:10.1073/pnas.1005963107 (2010).
- 469 27 Wei, X., Yang, Z., Rey, F. E., Ridaura, V. K., Davidson, N. O., Gordon, J. I. &  
470 Semenkovich, C. F. Fatty acid synthase modulates intestinal barrier function through  
471 palmitoylation of mucin 2. *Cell host & microbe* **11**, 140-152,  
472 doi:10.1016/j.chom.2011.12.006 (2012).
- 473 28 Kim, S. & West, T. P. Pyrimidine catabolism in *Pseudomonas aeruginosa*. *FEMS*  
474 *microbiology letters* **61**, 175-179, doi:10.1016/0378-1097(91)90547-n (1991).
- 475 29 Patel, B. N. & West, T. P. Degradation of the pyrimidine bases uracil and thymine by  
476 *Escherichia coli* B. *Microbios* **49**, 107-113 (1987).

- 477 30 Yuping Shan, A. S., Jianping Li, Changlu Zhou. Dietary supplementation of arginine  
478 and glutamine enhances the growth and intestinal mucosa development of weaned  
479 piglets. *Livestock Science* **Volume 150**, Pages 369-373,  
480 doi:<https://doi.org/10.1016/j.livsci.2012.10.006>. (2012).
- 481 31 Jelić, D., Tatić, I., Trzun, M., Hrvačić, B., Brajša, K., Verbanac, D., Tomašković, M.,  
482 Čulić, O., Antolović, R., Glojnarić, I., Weygand-Đurašević, I., Vladimir-Knežević, S. &  
483 Mildner, B. Porphyrins as new endogenous anti-inflammatory agents. *European*  
484 *journal of pharmacology* **691**, 251-260, doi:10.1016/j.ejphar.2012.05.049 (2012).
- 485 32 Hart, J. D., Samikwa, L., Sikina, F., Kalua, K., Keenan, J. D., Lietman, T. M., Burr, S.  
486 E. & Bailey, R. L. Effects of Biannual Azithromycin Mass Drug Administration on  
487 Malaria in Malawian Children: A Cluster-Randomized Trial. *The American journal of*  
488 *tropical medicine and hygiene* **103**, 1329-1334, doi:10.4269/ajtmh.19-0619 (2020).
- 489 33 Keenan, J. D., Bailey, R. L., West, S. K., Arzika, A. M., Hart, J., Weaver, J., Kalua,  
490 K., Mrango, Z., Ray, K. J., Cook, C., Lebas, E., O'Brien, K. S., Emerson, P. M.,  
491 Porco, T. C. & Lietman, T. M. Azithromycin to Reduce Childhood Mortality in Sub-  
492 Saharan Africa. *New England Journal of Medicine* **378**, 1583-1592,  
493 doi:10.1056/NEJMoa1715474 (2018).
- 494 34 Illumina. *Miseq System User Guide*. (2013).
- 495 35 Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A.,  
496 Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger,  
497 K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez,  
498 A. M., Chase, J. *et al.* QIIME 2: Reproducible, interactive, scalable, and extensible  
499 microbiome data science. *PeerJ Preprints* **6**, e27295v27292,  
500 doi:10.7287/peerj.preprints.27295v2 (2018).
- 501 36 Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H.,  
502 Whitman, W. B., Euzéby, J., Amann, R. & Rosselló-Móra, R. Uniting the  
503 classification of cultured and uncultured bacteria and archaea using 16S rRNA gene  
504 sequences. *Nature Reviews Microbiology* **12**, 635, doi:10.1038/nrmicro3330  
505 <https://www.nature.com/articles/nrmicro3330#supplementary-information> (2014).
- 506 37 Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R.,  
507 Mills, D. A. & Caporaso, J. G. Quality-filtering vastly improves diversity estimates  
508 from Illumina amplicon sequencing. *Nat Methods* **10**, 57-59, doi:10.1038/nmeth.2276  
509 (2013).
- 510 38 McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive  
511 Analysis and Graphics of Microbiome Census Data. *PloS one* **8**, e61217,  
512 doi:10.1371/journal.pone.0061217 (2013).
- 513 39 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and  
514 dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550,  
515 doi:10.1186/s13059-014-0550-8 (2014).

## 516 **Supplemental Information**

517

### 518 **Transparent methods**

#### 519 Study design

520 This study utilized baseline samples from a nested cohort study, conducted within the  
521 framework of a randomized controlled trial (NCT02047981). The study design and  
522 methodology for sample selection of both the nested cohort and the parent trial have been  
523 previously described<sup>14,32,33</sup>. Within the nested cohort study, 102 baseline samples from  
524 children residing in 30 clusters within the Mangochi district were analysed by 16S rRNA  
525 gene sequencing. All sampled children were aged between 1 and 59 months and weighed  
526 at least 3.8kg. Sample collection took place between May and July 2015.

#### 527 Fecal sample collection

528 Fecal samples for participating children were collected by their mothers or guardians, who  
529 were provided with sterile fecal collection bottles (Wheaton, UK) and given verbal  
530 instruction, by a nurse, on how to collect the sample. Fecal samples were returned to the  
531 field team as soon as possible after collection. These samples were put in a cool box with  
532 ice packs until the end of the day (not more than 8 hours after collection) and were brought  
533 to the laboratory where they were placed at -80°C for storage.

534

#### 535 Fecal biomarkers of EED

536 Commercially available ELISA kits for MPO (Immundiagnostik, Germany), AAT  
537 (Immundiagnostik) and NEO (Genway, CA, USA) were used for the quantification of each  
538 biomarker. Each assay was performed according to the kit instructions using stool diluted  
539 1/500 for MPO, 1/25,000 for AAT and 1/100 in phosphate buffered saline (PBS) for NEO<sup>12</sup>.  
540 Fecal samples with out-of-range concentration values were re-tested at appropriate  
541 dilutions until in-range values were obtained.

542

#### 543 16S rRNA gene sequencing and sequence processing

544 16S rRNA gene sequencing and sequence processing were performed as detailed  
545 elsewhere<sup>14</sup>. Briefly, DNA libraries were prepared by amplifying an approximately 390-bp  
546 V4 region of the 16S rRNA gene in the genomic DNA extracted from 250mg of fecal  
547 sample. The amplicons were purified and then quantified on a Qubit 2.0 Fluorometer  
548 (Invitrogen) with high-sensitivity reagents (Life Technologies) and thereafter pooled in  
549 equimolar amounts. A PhiX-spiked DNA library was then assayed by 2x300bp paired-end  
550 sequencing on the Illumina MiSeq platform for a total of 600 cycles using a standard  
551 protocol<sup>34</sup>.

552

553 Raw sequence data was processed in QIIME 2<sup>35</sup> as per standard practice<sup>34</sup> and *de novo*-  
554 clustered into operational taxonomic units (OTUs) at  $\geq 97\%$  identity. Taxonomy was  
555 assigned to the OTUs using a naïve Bayes classifier pre-trained on the SILVA 16S

556 database<sup>36</sup> subsequently generating an OTU table. To exclude spurious OTUs, only  
557 bacterial OTUs identified to the genus level, with sequences more than 0.005% of the total  
558 number of sequences<sup>37</sup> and a frequency of more than 0.01% in any sample were retained in  
559 the analyses.

560  
561 Statistical analysis  
562 Shannon (H) and Simpson (D) diversity indices were calculated from the OTUs to indicate  
563 alpha diversity using the phyloseq R package<sup>38</sup>. The differences in Shannon and Simpson  
564 distribution between groups were statistically tested using ANOVA and the non-parametric  
565 Kruskal-Wallis test respectively. Weighted and unweighted UniFrac distance matrices were  
566 also calculated to determine bacterial phylogenetic distance between samples.

567  
568 Median (25<sup>th</sup>, 75<sup>th</sup> quartile) concentration of individual fecal biomarkers of EED were used to  
569 determine their distribution. Published cut-off values for each of the 3 biomarkers<sup>12,13</sup> were  
570 used to determine proportions of fecal samples with elevated biomarkers. Concentrations of  
571 the 3 biomarkers were combined to form a composite EED score following the method of  
572 Kosek *et al.*<sup>13</sup> with the following modification: absolute weightings instead of rounded  
573 weightings from the principal component analysis<sup>13</sup> were used to generate continuous EED  
574 scores.

575  
576 Associations between alpha diversity indices and biomarker concentrations were  
577 investigated using regression analyses adjusted for age and sex. PERMANOVA with 1000  
578 permutation tests was run on weighted and unweighted UniFrac distances to compare  
579 bacteria community compositional differences between groups; differences in UniFrac  
580 clustering by groups were visualized using principal coordinate analysis (PCoA) plots.  
581 Differential abundance analysis at the genus level was performed to identify OTUs  
582 differentially abundant in EED using the DESeq2 package in R<sup>39</sup>. Statistical significance of  
583 log<sub>2</sub> fold changes was assessed using the default Wald test with Benjamin-Hochberg p  
584 value correction. Cut-off for all significant tests was set at p<0.01. Zero-inflated negative  
585 binomial regression analysis was conducted, adjusting for age and sex, to establish the  
586 relationship between the differentially abundant OTUs and the composite EED score.

587  
588 PICRUSt<sup>2</sup><sup>15</sup> analysis was used to predict metagenomes from 16S data and a reference  
589 genome database. We included OTU abundance data from 102 baseline samples  
590 that passed previously described quality control steps. Briefly, unique sequences  
591 representing the previously defined OTUs were added to reference multiple-sequence  
592 alignments and phylogeny containing 20000 full 16S rRNA genes of archaea and bacteria  
593 from the IMG database. Hidden-state prediction based on position of OTU sequences in the  
594 reference phylogeny was used to predict gene family abundances. To estimate functional  
595 profiles of each sample, gene family abundances per OTU were adjusted for OTU

596 abundance values per sample. Sample functional profiles were finally utilized to infer  
597 MetaCyc pathway abundances. PICRUST2 identified 431 metabolic pathways; these were  
598 collapsed to 67 superpathways for downstream analysis. To explore associations between  
599 metabolic pathways identified by PICRUST2 and EED, only the 60 samples that had ELISA  
600 results for all the three biomarkers from which we were able to calculate the composite EED  
601 score were included. In this subset of 60 samples, 15 super pathways had zero count score  
602 contribution and were excluded from further analysis. Lasso regression was then used to  
603 identify the inferred metabolic or functional properties of 16S rRNA community that are  
604 important for EED, the tuning parameter ( $\lambda$ ) was chosen by ten-fold cross-validation.  
605 Age-adjusted logistic regression analysis with permuted p values was used to explore  
606 associations between pathways predicted by PICRUST2, and identified as important by the  
607 lasso regression model, and the composite EED score.

### 608 Ethical considerations

609 This study was conducted in accordance with the Declaration of Helsinki. It was approved  
610 by the London School of Hygiene and Tropical Medicine Ethics Committee (UK) and the  
611 College of Medicine Research Ethics Committee (Malawi). Information and consent forms  
612 were translated into local languages (Yao and Chichewa) prior to their approval by the local  
613 ethics committee. Consent was first obtained at the community level through discussions  
614 with the village chief and community elders who then indicated the willingness, or  
615 unwillingness, of the community to participate through verbal consent. Written, informed  
616 consent (by thumbprint or signature) was then obtained from the parent or legal guardian of  
617 each child for inclusion before they participated in the study. During the consenting process,  
618 all parents and guardians were informed of their freedom to withdraw their child from the  
619 study at any time without giving reason for doing so.

### 620 **Supplementary table S1: Metabolic pathways predicted by PICRUST2**

<b>Metabolic pathway</b>	<b>Count score sum for each predicted pathway</b>
Proteinogenic Amino Acid Biosynthesis	15784472
Purine Nucleotide Biosynthesis	9236611
Vitamin Biosynthesis	7194008
Sugar Biosynthesis	5100685
Cell Wall Biosynthesis	4291038
Fatty Acid Biosynthesis	4143453
Phospholipid Biosynthesis	4110331
Pyrimidine Nucleotide Biosynthesis	3588915
2 Deoxyribonucleotide Biosynthesis	3315321
Sugar Degradation	2854327
Polysaccharide Degradation	1969759
Terpenoid Biosynthesis	1875327

Purine Nucleotide Degradation	1836367
Fermentation to Acetate	1602071
Chorismate Biosynthesis	1595960
Quinol and Quinone Biosynthesis	1354676
Metabolic Clusters	1349440
NAD Metabolism	1224545
Polyprenyl Biosynthesis	1204621
Coenzyme A Biosynthesis	1136914
Sugar Derivative Degradation	932614
Glycan Biosynthesis	850382
Sugar Acid Degradation	804311
CO <sub>2</sub> Fixation	597805
Other Amino Acid Biosynthesis	583747
Lipopolysaccharide Biosynthesis	569489
Tetrapyrrole Biosynthesis	496428
Pyrimidine Nucleotide Degradation	447537
Sulfur Compound Metabolism	384679
Porphyrin Compound Biosynthesis	327754
Proteinogenic Amino Acid Degradation	260546
Nitrogen Compound Metabolism	243591
8 Amino 7 oxononanoate Biosynthesis	213254
Formaldehyde Assimilation	88494
Phenolic Compound Degradation	64477
Generation of Precursor Metabolite and Energy	61787
Siderophore Biosynthesis	52992
Formaldehyde Oxidation	48703
4 Aminobutanoate Degradation	44988
Phosphorus Compound Metabolism	44779
Fatty Acid Degradation	42828
Toluene Degradation	40837
Allantoin Degradation	38424
Propanoate Degradation	29836
Glycerol Degradation	10073
Catechol Degradation	6892
Coenzyme B	2076
Coenzyme M Biosynthesis	1670
Coenzyme F420	1591
Reductant Biosynthesis	1197
Protocatechuate Degradation	507

## Gallate Degradation

194

---

621  
622 *The pathways have been ranked from those with the highest count score sum to the lowest.*