ONLINE-ONLY SUPPLEMENTAL MATERIAL

MATERIALS AND METHODS

Infant morbidity

At the weekly visits, we assessed infant morbidity over the previous week using a forced-choice questionnaire on days affected by: fever; cough and/or difficult or rapid breathing (categorized as RTIs); diarrhoea (defined as ≥ 3 loose stools in a day); and/or other illness. At any time during the study, if an infant became ill, caregivers were instructed to bring the infant to the study clinic and the diagnosis and treatment were recorded. Diagnoses were categorized as RTIs (including upper respiratory symptoms, cough, influenza, difficulty breathing and/or pneumonia), diarrhoea, malaria, skin infections, or other.

Laboratory methods

We analysed triplicate samples of each batch (n=7) of the maize flour and the MNPs for iron concentration by using graphite-furnace atomic absorption spectrophotometry (AA240Z, Varian, Mulgrave, Australia) after mineralisation by microwave digestion. GOS concentration in the MNPs was measured by using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) at Friesland Campina, Wageningen, The Netherlands (data not shown).

We defined anaemia as an Hb <110 g/L[1], iron deficiency (ID) in this setting with a high infection burden as PF <30 mg/L[1] and/or elevated sTfR >8.3 μg/mL; and IDA as Hb <110

g/L and PF <30 mg/L and/or elevated sTfR >8.3 µg/mL. We defined C reactive protein and alpha-glycoprotein values >5 mg/L and >1 g/L as elevated, indicating inflammation. We calculated Z scores for height-for-age (HAZ), weight-for-age (WAZ) and weight-for-height (WHZ) using WHO Anthro software (Version 3.2.2.1). A Z-score indicates how many standard deviations an anthropometric value is from the mean for that age. Stunting was defined as HAZ, < -2 SDs. Underweight was defined as WAZ, < -2SDs. Wasting was defined as WHZ, < -2SDs.

Faecal samples were thawed at 4°C before further preparation for measurement of pH and calprotectin. For measurement of pH, 100 mg (±10%) of faeces were added to 1 mL nanopure water, vortexed for 30 sec and centrifuged for 3 min at 5000 rpm at 4°C; pH in the liquid phase was measured using a digital pH meter (Metrohm, Zofingen, Switzerland).

Faecal samples were thawed at 4°C before DNA extraction and quantification. 700 µL S.T.A.R. buffer (Roche, Indianapolis, IN, USA) and 250 (±10%) mg of faeces were added to 0.5 g of 0.1 mm sterilized zirconia beads in a 2.0 mL screw-cap tube. Lysis was performed in a FastPrep instrument (MP Biomedicals, Santa Ana, CA, USA) at 5.5 ms for 3 times 1 min at room temperature. Samples were then incubated, while shaking at 100 rpm, at 95°C for 15 min, after which samples were centrifuged at 16000 g for 5 min at 4°C. The supernatant was collected and kept on ice. The stool pellet was subjected to another lysis round as described above, except only 350 µL S.T.A.R. buffer was added. The resulting supernatant was pooled with the supernatant of the first lysis round. Purification of DNA was performed on the automated Maxwell instrument (Promega, Madison, WI, USA) by applying the Maxwell 16 Tissue LEV Total RNA Purification Kit (Promega) according to the manufacturer's instructions. 250 µL of the supernatant was added to the first well of the Maxwell cartridge

and 50 µL of RNAse/DNAse free water was used for elution of the DNA. Quantification was carried out with a Nanodrop ND-1000 Spectrophotometer (Witec AG, Luzern, Switzerland) at 260 nm.

Using qPCR, we targeted virulence and toxin genes of selected enteropathogenic bacteria: *B. cereus, C. difficile*, *C. perfringens,* EHEC, EPEC, and ETEC, *Salmonella* spp.*,* and *S. aureus*. The primers used for qPCR are shown in **S. Table 1**.

Supplemental Table 1. qPCR primers

Target	Target gene (description)	PCR primer and sequence (5'-3')	Product size	Ref.
Enteropathogenic *Escherichia coli* (EPEC)	*eaeA* (*E. coli* attaching and effacing)	EAE-a, ATGCTTAGTGCTGGTTTAGG EAE-b, GCCTTCATCATTTCGCTTTC	248	[2]
Enterohemorrhagic *Escherichia coli* (EHEC)	stx_1 (shiga toxin)	JMS1F, GTCACAGTAACAAACCGTAACA JMS1R, TCGTTGACTACTTCTTATCTGGA	95	[2]
Enterohemorrhagic *Escherichia coli* (EHEC)	stx_2 (shiga toxin)	JMS2F, CGACCCCTCTTGAACATA JMS2G, GATAGACATCAAGCCCTCGT	108	[2]
Enterotoxigenic *Escherichia coli* (ETEC)	*LT* (heat-labile enterotoxin)	LT-1, AGCAGGTTTCCCACCGGATCACCA LT-2, GTGCTCAGATTCTGGGTCTC	132	[2]
Enterotoxigenic *Escherichia coli* (ETEC)	*ST* (heat-stable enterotoxin)	STa-F, GCTAATGTTGGCAATTTTTATTTCTGTA STa-R, AGGATTACAACAAAGTTCACAGCAGTAA	190	[2]
Staphylococcus aureus	Nuclease	SA-1, GCGATTGATGGTGATACGGTT SA-2, CAAGCCTTGACGAACTAAAGC	276	[2]

Bacillus cereus	Hemolysin	BC-1, CTGTAGCGAATCGTACGTATC BC-2, TACTGCTCCAGCCACATTAC	185	[2]
Salmonella spp.	*invA* (invasion)	invA, 139, GTGAAATTATCGCCACGTTCGGGCAA invA, 141, TCATCGCACCGTCAAAGGAACC	284	[2]
Clostridium difficile	16S rRNA gene	cdF, TTGAGCGATTTACTTCGGTAAAGA cdR, CCATCCTGTACTGGCTCACCT	157	[3]
Clostridium perfringens	*pcl* (alpha toxin)	plcF, AAGTTACCTTTGCTGCATAATCCC plcR, ATAGATACTCCATATCATCCTGCT	283	[4]

59

60

61 For preparation of the DNA for qPCR we performed pre-amplification using primers shown
62 in S. Table 1, Fluidigm Preamp Master Mix (Fluidigm, PN 100-5580, California, United
63 States) and the 12 recommended cycles as described by the manufacturer (Fluidigm Quick
64 Reference for Pre-amplification of cDNA for Gene Expression with Delta Gene Assay, PN
65 100-5875 C1, Fluidigm). The final reaction products were diluted 5-fold. We prepared
66 sample pre-mix and assay mix and run the qPCR with BioMark 96.96 Gene Expression
67 Dynamic Arrays (PN BMK-M-96.69, Fluidigm, South San Francisco, USA) as described by
68 the manufacturer (Advanced Development Protocol 41: Using EvaGreen DNA Binding Dye
69 for Gene Expression with the 48.48 and 96.96 Dynamic Array IFCs, Fluidigm). To generate
70 standards, target genes were amplified from representative strains. Amplicons were purified
71 and were cloned into PGEMT Easy Vector and heterologously expressed in *E. coli* according
72 to supplier instructions (Promega). Standard curves were prepared from 10-fold dilutions of
73 linearised plasmids harbouring the gene of interest. Concentrations of the plasmids were
74 measured using a Qubit dsDNA BR assay kit (Q32850, Thermo Fisher Scientific) in triplicate
75 on a Spark M10 plate reader (Tecan Group Ltd., Maennedorf, Switzerland). We processed
76 data with the Fluidigm Real-Time PCR Analysis Software including melting curve analysis.
77 We used Excel (Microsoft Office 2010) for analysis of the qPCR data. The standard curves

were generated by linear regression analysis of the C_T (threshold cycle) values versus the amounts of the template DNA of the standard (log gene copies/µL). The goodness of fit (r^2) was calculated for each linear regression. Primer efficiency was calculated for each target. The limit of quantification (LOQ) was set at the last point of the standard line falling in the linear range. Samples below the LOQ were assigned the log gene copies/g faeces defined as LOQ/2 and defined as not detected. A summary variable was created for the toxin and virulence genes of the targeted pathogens. To correct for multiple 16S rDNA genes in *C.difficile*, gene copy numbers were divided by 11. Another summary variable was created for virulence and toxin genes of pathogenic *E. coli* (*eaeA*, *LT*, *ST*, *stx1*, *stx2*).

Barcoded amplicons from the V3-V4 region of 16S rRNA genes were generated using a 2-step PCR, and universal primers appended with Illumina adaptors were used for initial amplification of the V3-V4 part of the 16S rRNA gene with the following sequences: forward primer, '5-CCTACGGGAGGCAGCAG-3' (broadly conserved bacterial primer 357F); reverse primer, '5-TACNVGGGTATCTAAKCC' (broadly conserved bacterial primer (with adaptations) 802R) appended with Illumina adaptor sequences. PCR amplification mixture contained: 1 µL fecal sample DNA, 1 µL bar-coded forward primer, 15 µL master mix (1 µL KOD Hot Start DNA Polymerase (1 U/µL; Novagen, Madison, WI, USA), 5 µL KOD-buffer (10×), 3 µL $MgSO_4$ (25 mM), 5 µL dNTP mix (2 mM each), 1 µL (10 µM) of reverse primer) and 33 µL sterile water (total volume 50 µL). PCR conditions were: 95°C for 2 min followed by 35 cycles of 95°C for 20 sec, 55°C for 10 sec, and 70°C for 15 sec. The approximately 500 bp PCR amplicon was subsequently purified using the MSB Spin PCRapace kit (Invitek, Berlin, Germany) and concentration and quality was subsequently checked with a Qubit fluorometer (Thermo Fisher Scientific). Purified PCR products were shipped to BaseClear BV (Leiden, The Netherlands) and used for the second PCR in combination with sample-

specific barcoded primers. PCR products were purified, checked on a Bioanalyzer (Agilent) and quantified, followed by multiplexing, clustering, and sequencing on an Illumina MiSeq with the paired-end (2x) 300 bp protocol and indexing. The sequencing run was analysed with the Illumina CASAVA pipeline (v1.8.3) with demultiplexing based on sample-specific barcodes. The raw sequencing data produced was processed removing the sequence reads of too low quality (only "passing filter" reads were selected) and discarding reads containing adaptor sequences or PhiX control with an in-house filtering protocol. A quality assessment on the remaining reads was performed using the FASTQC quality control tool version 0.10.0.

Statistical methods

We analysed biochemical and anthropometric data using the R statistical programming environment (R Version 3.2.3, 2015). To test for baseline differences: we used one-factor ANOVA with Tukey post hoc tests to assess between group differences for variables of iron status, inflammation, anthropometrics, faecal calprotectin, I-FABP and pH; we used generalized linear model (glm) analysis for binary data and we used Wilcoxon rank sum test for all targeted bacteria by qPCR. To test the intervention effect: we used ANCOVA with baseline values as covariate to test for between group differences at 4 months for variables of iron status, inflammation, anthropometrics, faecal calprotectin, I-FABP and pH; we used glm analysis for binary data. For targeted bacteria, we used Mann-Whitney tests to test between group differences at baseline, week 3 and 4 months, and Wilcoxon rank sum tests to test within group changes across these time points.

To assess morbidity, we performed glm analysis (CRAN package glmer) using a weekly binary response variable from the weekly monitoring and the interaction time*group as fixed effects, and subject as a random effect. We performed subgroup analysis to test for a time

effect within each group. We used glm analysis to test for differences in infants treated for RTI, diarrhoea, malaria, skin infection and other diagnosis using as response variable infant treated for the respective diagnosis at least once (binary data) and the group as fixed effect. We included in a complete model gender, age, change in Hb from baseline to 4 months and change in the sum of the VTGs of all pathogens as covariates. Gender, age or change in the sum of the VTGs of the pathogens did not improve the model and were removed. The change in Hb was a significant cofactor: a greater increase in Hb was associated with a reduced RTI incidence. We performed glm analysis using a monthly binary response variable of infants treated at least once for a RTI during that month, with time as a fixed effect and subject as a random effect. Statistical correlation was tested using Spearman's rank correlation coefficient (r_s); for correlations with gut bacterial groups, we included only samples with detectable gene copy numbers. Baseline and 3 week data from the 10 infants that dropped out after 3 weeks were included in the analysis.

16S rRNA gene sequences were analysed using a workflow based on Qiime 1.8[5]. Operational taxonomic unit (OTU) clustering (open reference), taxonomic assignment and reference alignment were done with the pick_open_reference_otus.py workflow script of Qiime, using uclust as clustering method (97% identity) and GreenGenes v13.8 as reference database for taxonomic assignment. Reference-based chimera removal was done with Uchime[6]. The RDP classifier version 2.2 was performed for taxonomic classification[7]. The fold difference was calculated as the 2log of the ratio of the relative abundance between groups (0=no difference between group, 1=twice as abundant, etc.). Statistical significance between contrasts for taxonomy abundances was tested by the non-parametric Mann-Whitney U test. Comparisons of targets of our primary interest (*Lactobacillaceae*, *Bifidobacteriaceae, Enterobacteria*ceae, *Bacteroidetes* and *Clostridiales* members *Roseburia* and *Clostridium)*

were not corrected for multiple testing. Comparisons of relative abundances of other taxa were explorative and therefore also not corrected for multiple testing. Statistical tests were performed as implemented in SciPy (https://www.scipy.org/), downstream of QIIME analyses as described above. Visualization of differences in relative abundance of taxa between the three study groups was done in Cytoscape[8]. Multivariate Redundancy Analysis (RDA) was performed in Canoco version 5.0.4 using default settings of the analysis type "Constrained"[9]. For the RDA, relative abundance values of OTUs were used as response data and the MNP group as explanatory variable. For visualization purposes families or genera (and not OTUs) were plotted as supplementary variables. RDA calculates p-values by permutating (Monte Carlo) the sample status. Longitudinal effects of intervention were tested by calculating 2log ratios in which the relative abundance of an OTU at 4 months or 3 weeks was divided by the relative abundance of an OTU at baseline. These ratios were used as response variables in RDAs, and were weighted based on the average relative abundance of each OTU in all infants. Taxa names in brackets are annotations supplied by the Greengenes database and are not officially accepted by the Society for General Microbiology. The phylogenetic distances (weighted UniFrac) from baseline to 3 weeks and from baseline to 4 months for each individual were calculated, and intervention groups were compared by Kruskal-Wallis test with Dunn's posthoc test, as implemented in Graphpad Prism 5.01. *P* values <0.1 were considered statistical trends, and *P* values <0.05 as statistically significant.

RESULTS

Supplemental Table 2. Anthropometric variables, by group. Differences in these variables among Kenyan infants (n=155) at baseline and after 4 months of consuming daily a

MNP containing either: no iron (control); 5 mg of iron (Fe); or 5 mg of iron and 7.5 g GOS (FeGOS).

	Control	Fe	FeGOS
Height (cm)			
Baseline	68.5 (66.0-70.5) [1]	68.0 (66.0-69.6)	68.8 (67.0-70.0)
4 months	72.0 (70.0-74.0)	72.0 (70.0-74.0)	73.0 (70.9-75.0)
Weight (kg)			
Baseline	8.0 (7.2-8.5)	7.7 (7.0-8.4)	7.5 (7.0-8.4)
4 months	8.9 (8.0-9.5)	8.5 (7.6-9.5)	8.5 (8.0-9.2)
HAZ			
Baseline	-0.30±1.14[2]	-0.49±1.27	-0.10±1.01
4 months	-0.81±1.00	-0.79±1.32	-0.44±1.27
Stunting, n (%)			
Baseline	3(6)	6(12)	0(0)
4 months	5(10)	6(12)	4(8)
WAZ			
Baseline	-0.29±1.02	-0.57±1.27	-0.51±0.94
4 months	-0.41±1.15	-0.50±1.33	-0.54±1.02
Underweight, n (%)			
Baseline	1(2)	6(12)	3(6)
4 months	3(6)	6(12)	5(10)
WHZ			
Baseline	-0.07±1.18	-0.35±1.22	-0.51±1.03
4 months	0.05±1.31	-0.12±1.29	-0.42±1.15
Wasting, n (%)			
Baseline	2(4)	4(8)	6(12)
4 months	4(8)	3(6)	5(10)

[1] Median (IQR), all such values. [2] Mean±SD, all such values. For continuous variables, between group differences at baseline were tested using one-factor ANOVA. Between group differences at study endpoint were tested using ANCOVA with baseline values as covariates. For categorical variables, between group differences at baseline and endpoint were tested using generalized linear models. There were no significant baseline or endpoint differences between any of the variables. HAZ, height-for-age Z score; WAZ, weight-for-age Z score; WHZ, weight-for-height Z score.

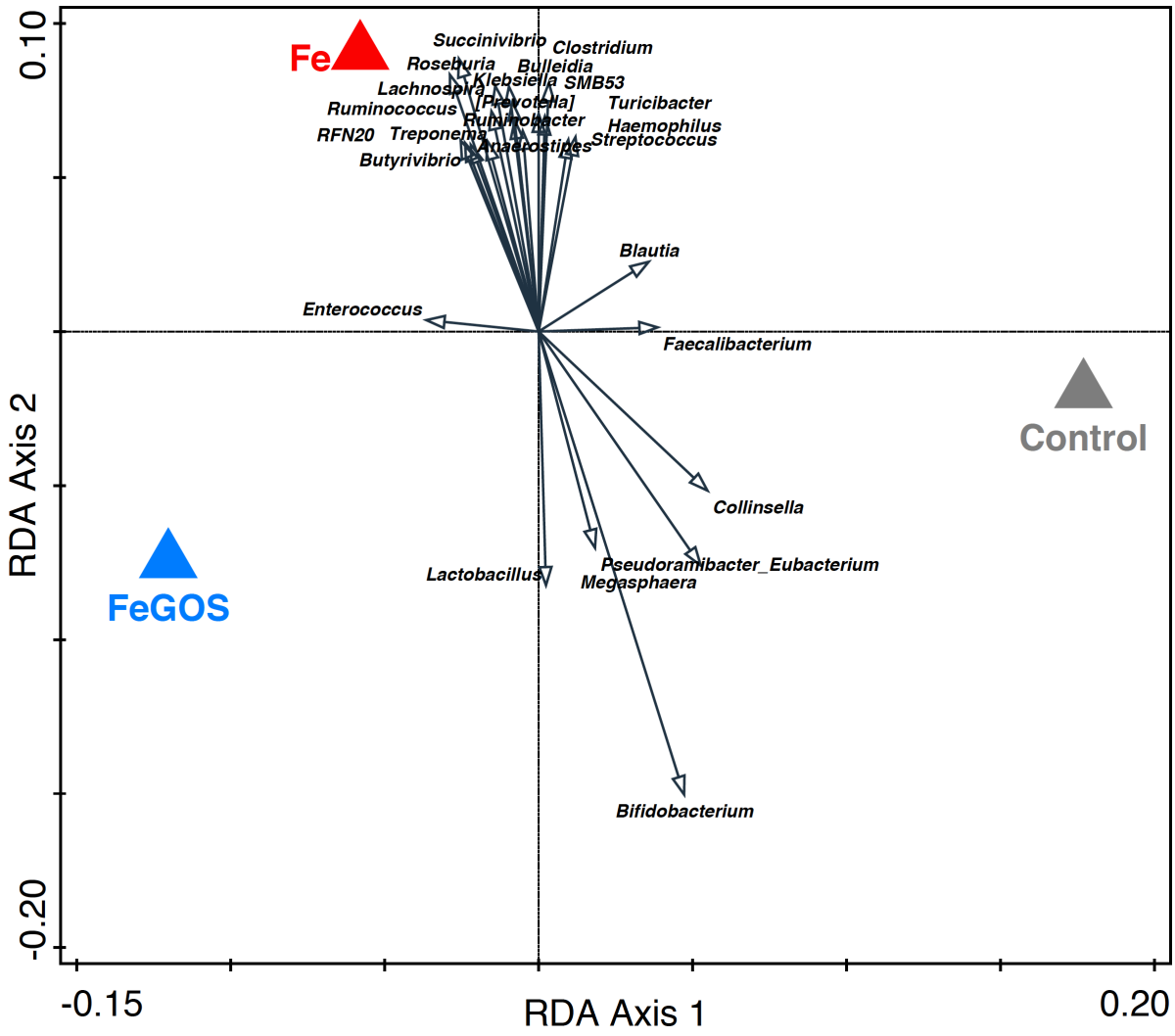
174

175 **Correlations between inflammation and enteropathogens**

176 At baseline, there were correlations between faecal calprotectin and C reactive protein
177 (r=0.240; *P*=0.003), alpha-glycoprotein (r=0.169; *P*=0.036), the VTGs of all pathogens
178 (r=0.206; *P*=0.014) and the VTGs of all pathogenic *E. coli* (r=0.184, *P*=0.045). At 3 weeks,
179 there were correlations between faecal calprotectin and the VTGs of all pathogens (r=0.214,
180 *P*=0.011) and of the VTGs of all pathogenic *E. coli* (r=0.246, *P*=0.009). At 4 months, there
181 were correlations between faecal calprotectin and the VTGs of all pathogens (r=0.249;
182 *P*=0.004) and the VTGs of all pathogenic *E. coli* (r=0.310; *P*=0.002). In contrast, at baseline,
183 3 weeks or 4 months, C reactive protein, alpha-glycoprotein and I-FABP were not
184 significantly correlated with the VTGs of all pathogens or the VTGs of all pathogenic *E. coli*.

185

186



187

188 **Supplemental Figure 1. Redundancy analysis (RDA) at 4 months, by group, with**

189 **bacterial genera plotted as supplementary variables.** RDA in Kenyan infants (n=155)

190 after 4 months of receiving daily a MNP containing either: no iron (control); 5 mg of iron

191 (Fe); or 5 mg of iron and 7.5 g of GOS (FeGOS). The variation explained by group was 0.8%

192 (p=0.034).

193

194 **References**

195

196 1 World Health Organization. Iron deficiency anaemia: assessment, prevention and

197 control. World Health Organization, Geneva, 2001.

198 2 Fukushima H, Tsunomori Y, Seki R. Duplex real-time SYBR Green PCR assays for
199 detection of 17 species of food- or waterborne pathogens in stools. *J Clin Microbiol*
200 2003;41:5134–5146.
201 3 Rinttila T, Kassinen A, Malinen E, Krogius L, Palva A. Development of an extensive
202 set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in
203 faecal samples by real-time PCR. *J Appl Microbiol* 2004;97:1166–1177.
204 4 Rinttilä T, Lyra A, Krogius-Kurikka L, Palva A. Real-time PCR analysis of enteric
205 pathogens from fecal samples of irritable bowel syndrome subjects. *Gut Pathogens* 2011;3:6.
206 5 Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-
207 throughput community sequencing data. *Nature Methods* 2010;7:335-336.
208 6 Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves
209 sensitivity and speed of chimera detection. *Bioinformatics* 2011;27:2194-2200.
210 7 Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved
211 alignments and new tools for rRNA analysis. *Nucleic Acids Res* 2009;37:D141-145.
212 8 Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for
213 integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-2504.
214 9 Braak CJ, Smilauer P. Canoco Reference manual and user's guide to Canoco for
215 Windows: Software for Canonical Community Ordination, Microcomputer Power, Ithaca,
216 NY, USA, 2012.
217