

Title: The gut microbiome and early-life growth in a population with high prevalence of stunting

Authors: Ruairi C. Robertson^{1†}, Thaddeus J. Edens^{2†}, Lynnea Carr^{3†}, Kuda Mutasa⁴, Ceri Evans^{1,4}, Ethan K. Gough⁵, Hyun Min Geum⁶, Iman Baharmand⁶, Sandeep K. Gill⁶, Robert Ntozini⁴, Laura E Smith^{4,7}, Bernard Chasekwa⁴, Florence D. Majo⁴, Naume V. Tavengwa⁴, Batsirai Mutasa⁴, Freddy Francis⁸, Joice Tome⁴, Rebecca J. Stoltzfus⁹, Jean H. Humphrey⁵, Andrew J. Prendergast^{1,4,5}, Ameer R. Manges^{6,10}, and the SHINE Trial Team¹¹

Affiliations:

¹Blizard Institute, Queen Mary University of London, London, UK

²Devil's Staircase Consulting, West Vancouver, British Columbia, Canada

³Department of Microbiology and Immunology, University of British Columbia, Canada

⁴Zvitambo Institute for Maternal and Child Health Research, Harare, Zimbabwe

⁵Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

⁶School of Population and Public Health, University of British Columbia, Vancouver, Canada

⁷Department of Population Medicine and Diagnostics, Cornell University, Ithaca, NY, USA

⁸Department of Experimental Medicine, University of British Columbia, Canada

⁹Goshen College, Goshen, Indiana, USA

¹⁰British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada

¹¹Members of the SHINE Trial team who are not named authors are listed in https://academic.oup.com/cid/article/61/suppl_7/S685/358186

† These authors contributed equally to this work

***Corresponding author:** Ameer R. Manges, MPH, PhD, School of Population and Public Health, University of British Columbia, 314-2206 East Mall, Vancouver, BC Canada, amee.manges@ubc.ca

One sentence summary: The gut microbiome of rural Zimbabwean infants undergoes programmed maturation that is unresponsive to sanitation and nutrition interventions but is comprehensively modified by maternal HIV infection and can moderately predict linear growth.

1 **ABSTRACT**

2

3 Stunting affects one-in-five children globally and is associated with greater infectious
4 morbidity, mortality and neurodevelopmental deficits. Recent evidence suggests that the
5 early-life gut microbiome affects child growth through immune, metabolic and endocrine
6 pathways, and microbiome perturbations may contribute to undernutrition. We examined
7 early-life fecal microbiome composition and function in 875 stool samples collected
8 longitudinally in 335 children from 1-18 months of age in rural Zimbabwe, from a cluster-
9 randomized trial of improved water, sanitation, and hygiene (WASH), and improved infant
10 and young child feeding (IYCF). Using whole metagenome shotgun sequencing, we
11 examined the effect of the interventions, in addition to environmental or host factors
12 including maternal HIV infection, on the succession of the early-life gut microbiome, and
13 employed extreme gradient boosting machines (XGBoost) to model microbiome
14 maturation and to predict child growth. WASH and IYCF interventions had little impact on
15 the fecal microbiome, however children who were HIV-exposed but uninfected exhibited
16 over-diversification and over-maturity of the early-life gut microbiome in addition to
17 reduced abundance of *Bifidobacteria* species. Taxonomic microbiome features were
18 poorly predictive of linear and ponderal growth, however functional metagenomic
19 features, particularly B-vitamin and nucleotide biosynthesis pathways, moderately
20 predicted both attained linear and ponderal growth and growth velocity. We find that the
21 succession of the gut microbiome in a population at risk of stunting is unresponsive to
22 WASH and IYCF interventions, but is strongly associated with maternal HIV infection,
23 which may contribute to deficits in growth. New approaches targeting the gut microbiome
24 in early childhood may complement efforts to combat child undernutrition.

1 **Main Text:**

2

3 **INTRODUCTION**

4 Stunting, or linear growth failure, is a form of chronic undernutrition that affects 22% of
5 children under 5 years of age worldwide (1, 2). Stunting is associated with infectious
6 morbidity, reduced childhood survival and impaired cognitive development (3). The
7 lifelong impacts of poor growth contribute to an intergenerational cycle of stunting and
8 impaired development, lower educational attainment, and reduced adult economic
9 productivity (4). Nutritional interventions, however, only reduce stunting by approximately
10 12% (5), suggesting that other pathophysiological mechanisms contribute to chronic
11 undernutrition, which may inform new therapeutic strategies.

12

13 The determinants of stunting and other forms of child undernutrition are complex and
14 include a myriad of biological, environmental and social factors including breastfeeding
15 and complementary feeding practices, household water, sanitation and hygiene (WASH)
16 practices, birthweight, maternal HIV status, maternal anthropometry and maternal
17 education. However, growing evidence suggests that a subclinical disorder of the small
18 intestine, termed environmental enteric dysfunction (EED), may play a role in impaired
19 child growth (6). EED is characterized by blunted intestinal villi, increased gut
20 permeability, and microbial translocation into the circulatory system resulting in both local
21 and chronic inflammation and nutrient malabsorption (7, 8). It is hypothesized that high
22 enteric pathogen carriage, as seen in poor-hygiene, low-resource settings, contributes to
23 the pathophysiology of EED (9-11); however, interventions to improve WASH and reduce

1 the pathogen burden in children have failed to demonstrate improvements in linear growth
2 (12). Additionally, both enteric pathogen load and common biomarkers of EED are not
3 consistently associated with linear growth in different geographical cohorts (13-16),
4 suggesting that the pathway linking microbial exposures, impaired gut function and early-
5 life growth remains to be fully elucidated.

6
7 In addition to research investigating the influence of diarrheal pathogens on child
8 undernutrition and EED, emerging evidence supports the role of the commensal gut
9 microorganisms in mediating child growth. Healthy-growing children exhibit a patterned
10 ecological assembly of the gut microbiome through the first 2 years of life, which is defined
11 by delivery mode, breastfeeding and complementary feeding practices (17). This
12 microbial succession impacts a number of metabolic, immune and endocrine pathways
13 in early life that contribute to early-life growth and development (18). Disturbances to this
14 normal microbiome maturation therefore may impair these critical growth and
15 developmental pathways. Immaturity of the early-life gut microbiome is associated with
16 severe acute malnutrition (19), whilst reduced microbiome diversity is associated with
17 higher risk of future diarrheal episodes (20). Indeed, a 'malnourished' early-life gut
18 microbiome can recapitulate phenotypes of faltering growth and EED when transplanted
19 into germ-free mice and pigs (21, 22). Furthermore, nutritional interventions designed to
20 specifically target the impaired gut microbiome in acute malnutrition in both animal studies
21 and small-scale human trials have recently demonstrated a positive effect on ponderal
22 growth (23, 24), but not on linear growth.

23

1 Microbiome differences that may contribute to stunting are likely influenced by a number
2 of environmental factors including household WASH, infant feeding practices and
3 maternal HIV infection. To date, little research has investigated the effect of improved
4 WASH or infant feeding interventions on the assembly of the infant gut microbiome in low
5 resources settings. However, recent data show that children who are HIV-exposed but
6 uninfected (CHEU), consume breast-milk with an altered oligosaccharide composition
7 from their mothers (25), and may be exposed to abnormal microbiome profiles from their
8 mothers, which have been reported in people living with HIV (26, 27). CHEU also receive
9 prophylactic antibiotics, to prevent infectious morbidity associated with HIV exposure.
10 Each of these exposures may influence the seeding and succession of the gut
11 microbiome in CHEU (28, 29), which may contribute to the high prevalence of stunting
12 observed in CHEU (30). Evidence of the effect of other early-life environmental exposures
13 on the assembly of the infant gut microbiome in low resources settings is scarce but may
14 provide insights into the influence of microbial and microbiota-modifying exposures on
15 child growth in the context of undernutrition.

16
17 Previous cross-sectional data from sub-Saharan Africa hypothesized that
18 decompartmentalization of the gastrointestinal tract occurs in stunted children, as
19 demonstrated by the overgrowth of oropharyngeal bacterial taxa in the intestine (31),
20 whilst a handful of other cross-sectional studies report variations in gut microbiota
21 composition in stunted children that are inconsistent across geographical settings (32-
22 34). We previously reported that the maternal gut microbiome can predict birthweight and
23 neonatal growth in rural Zimbabwe (27). However, there are few studies mapping the

1 compositional and functional maturation of the gut microbiome throughout early
2 childhood, accounting for feeding, WASH, maternal HIV infection and other environmental
3 exposures, in populations from low-resource settings and at high risk of stunting.

4
5 Here, we characterize the succession and maturation of the fecal microbiome from 1 to
6 18 months of age in 335 children from rural Zimbabwe who were enrolled in the
7 Sanitation, Hygiene, Infant Nutrition Efficacy (SHINE) Trial (35). We explore gut
8 microbiome maturation and examine the influence of randomized WASH and nutrition
9 interventions and maternal HIV infection. Using compositional and functional
10 metagenomic data as well as extensive epidemiological information, we use machine
11 learning to test the ability of the early-life gut microbiome to predict both attained linear
12 and ponderal growth and growth velocity through the first 18 months of life.

13

14

15 **RESULTS**

16 ***Sub-study population characteristics***

17 The fecal microbiota was characterized in 875 fecal samples from 335 children from 1-18
18 months of age (Fig. S1). A mean (SD) of 2.6 (1.3) samples were analysed per child. The
19 children in the microbiome sub-study largely resembled the population of all live-born
20 infants in the overall SHINE trial cohort (Table S1), however, the microbiome sub-study
21 included a larger number of children who were born to women living with HIV (29.6%)
22 compared to the whole SHINE cohort (15.6%), due to the deliberate over-sampling of
23 HIV-positive mothers and their infants. In addition, the microbiome sub-study included

1 infants with slightly older mothers and longer gestational ages. The majority of infants
2 were born by vaginal delivery (94.5%) in an institution (89.9%) and were exclusively
3 breastfed (91% at 3 months). Prevalence of stunting (LAZ <-2) varied from 18-34% across
4 study time-points.

6 **Metagenome sequencing performance**

7 Overall, 875 unique infant-visit whole metagenome sequencing datasets were used. On
8 average, 12 million \pm 4.2 million quality-filtered read pairs were generated per sample.
9 Sixteen negative controls produced a mean of 655 quality-filtered reads (range = 149 to
10 1,425; SD = 456). The median percent of human reads detected was 0.05% but ranged
11 widely by age group and decreased over time (Fig. S2a). The median percent un-
12 annotatable reads detected in each sample was 58.6% and increased over time (Fig.
13 S2b). Thirty-six samples were subject to repeated extraction and metagenome
14 sequencing to assess technical variation. These samples originated from 4 unique
15 children, each with 3 visit samples, where each visit sample was extracted and
16 sequenced in 3 replicates. Principal coordinates analysis (PCoA) of Bray-Curtis distances
17 and phylum-level relative abundances revealed little variation between replicates (Fig.
18 S2c-d).

20 ***Succession of gut microbiome composition in early childhood***

21 After prevalence and relative abundance threshold filtering, 161 annotated bacterial
22 species were identified. Seven Eukaryotic and 4 Archaeal species were detected in a
23 small proportion of samples, but these did not meet the prevalence thresholds (Table S2).

1 *Bifidobacterium longum* was the predominant species at all time-points up to 12 months
2 of age. Four other *Bifidobacteria* species (*B. breve*, *B. bifidum*, *B. pseudocatenulatum*,
3 and *B. kashiwanohense*), *Escherichia coli*, *Bacteroides fragilis* and *Veillonella* species
4 were consistently amongst the most abundant species at the earlier time points before
5 being outnumbered by *Faecalibacterium prausnitzii* and *Prevotella copri* at 12 and 18
6 months of age. Taxonomic α -diversity metrics and gene richness tended to decline or
7 remain stable over the first 4-6 months of life, during exclusive breast-feeding, but
8 increased as expected with infant age from 6-18 months of age (Fig. S3a-b), with the
9 introduction of complementary feeds. A large proportion of variation in both compositional
10 ($R^2 = 0.198$, $P < 0.001$) and functional β -diversity ($R^2 = 0.378$, $P < 0.001$) was explained by
11 age (Fig. 1a-b and S3 c-d).

12
13 We employed extreme gradient boosting machines (XGBoost), a machine learning
14 approach, to train and test a model of compositional and functional microbiome
15 maturation, with child age as an outcome. Children who were born to HIV-negative
16 mothers, who were non-stunted at 18 months ($LAZ > -2$) and had at least 2 stool samples
17 collected were used as a 'training set', which was then used to predict child age in the
18 remaining samples (test set). Using species composition, the microbiome was highly
19 predictive of child age (Model pseudo- $R^2 = 0.77$, MAE = 1.4 months). This 'microbiota
20 age' score was also strongly correlated with biological age in the subset of children from
21 the test set who were also non-stunted and born to HIV-negative mothers (pseudo- $R^2 =$
22 0.67). The species most strongly predictive of age included *Faecalibacterium prausnitzii*,
23 *Blautia wexlerae*, *Prevotella copri*, *Staphylococcus hominis*, *Dorea formicigenerans*,

1 *Bifidobacterium longum*, *Agathobaculum butyriciproducens*, *Bifidobacterium bifidum*,
2 *Bacteroides thetaiotaomicron*, *Streptococcus vestibularis* and *Veillonella parvula* (Fig.
3 1c). Metagenomic pathways also predicted age with high accuracy (Model pseudo- R^2 =
4 0.68; MAE = 1.5; Fig. 1d). The pathways most strongly predictive of age included
5 methanogenesis from acetate (METH-ACETATE-PWY), multiple nucleotide and amino
6 acid metabolic pathways, including L-tryptophan biosynthesis (TRPSYN-PWY), purine
7 ribonucleosides degradation (PWY0-1296), pyrimidine deoxyribonucleotides de novo
8 biosynthesis I (PWY-7184), L-histidine degradation I (HISDEG-PWY), dTDP-L-rhamnose
9 biosynthesis I (DTDPRHAMSYN-PWY), flavin biosynthesis I (RIBOSYN2-PWY) and
10 nitrate reduction I (DENITRIFICATION-PWY). This ‘metagenome age’ was also highly
11 correlated with age in the subset of children from the test set who were also non-stunted
12 and born to HIV-negative mothers (pseudo- R^2 = 0.67). Using these models, we created
13 a microbiota-for-age Z-score (MAZ) and metagenome-for-age Z-score (MetAZ), which
14 accounted for variance of microbiota ages with respect to chronological ages at each
15 study visit (see Methods). The top 20 features contributing most strongly to age
16 predictions are plotted in Fig.1e-f.

17

18 ***WASH and IYCF interventions have little influence on the infant gut microbiome***

19 We have previously reported in the SHINE trial that the WASH intervention had no impact
20 on infant growth, whilst the IYCF intervention increased LAZ scores by 0.16, leading to a
21 23% reduction in stunting by 18 months of age. We tested whether these randomized
22 interventions impacted any metrics of gut microbiome diversity or maturity in each age
23 group. By performing principal coordinate analysis (PCoA) on Bray-Curtis distances of

1 taxonomic data and functional data we found no significant differences in β -diversity
2 between IYCF and non-IYCF arms at any time-point. There was a significant difference
3 in Bray-Curtis distances for microbiome composition between WASH and non-WASH
4 arms at the 3-month time-point (PERMANOVA, $P=0.013$; $R^2= 0.011$), but at no other time-
5 points (Fig.2a-d). No significant differences were observed in α -diversity metrics or gene
6 richness between WASH versus non-WASH arms nor IYCF versus non-IYCF arms at any
7 time-point (Fig.2e-f). Using multivariable regression analyses (MaAsLin2), there were
8 also no differences in the relative abundance of species or pathways between intervention
9 arms apart from a small number of features at 3 months in the WASH arms (increased
10 *Klebsiella pneumoniae*, reduced *Collinsella aerofaciens* and more abundant
11 metagenomic pathways involved in biotin and folate synthesis) and at 18 months in the
12 IYCF arms (reduced *Eubacterium siraeum*, *E. rectale* and *Agathobaculum*
13 *butyriciproducens*; Table S3 and S4).

14

15 ***HIV exposure comprehensively alters infant gut microbiome composition and*** 16 ***function***

17 We previously reported in this cohort that maternal HIV exposure significantly impacts
18 infant growth, whereby CHEU displayed significantly poorer linear growth compared with
19 children who are HIV-unexposed (CHU) (30). We assessed diversity metrics and
20 microbiome maturity in CHEU versus CHU and found that CHEU displayed significantly
21 greater diversity (Shannon index; Wilcoxon rank-sum test $P = 0.002$; Fig. 3a) and species
22 richness ($P = 0.01$; Fig. 3b) compared with CHU at 12 months of age, although the size
23 of the two groups was imbalanced within this age category ($n=91$ CHEU, $n=27$ CHU).

1 Other metrics of α -diversity were significantly lower in CHEU at 2 months of age
2 (evenness $P = 0.04$, Simpson index $P = 0.04$). Metagenomic gene richness was also
3 elevated in CHEU at 1, 3, 6 and 12 months (Wilcoxon rank-sum test $P = 0.002$, $P = 0.059$,
4 $P = 0.007$, $P < 0.001$, respectively; Fig. 3c). Analysis of taxonomic β -diversity between
5 CHEU and CHU also revealed significant differences at 1, 3, 6, 12 and 18 months
6 suggesting that *in utero* HIV exposure was significantly associated with gut microbiome
7 succession and development throughout the first 18 months of life (Fig. 3d-e). HIV
8 exposure also explained a significant proportion of the variation in metagenome pathway
9 β -diversity (PERMANOVA, $P = 0.002$, $R^2 = 0.033$) at 1 month of age (Fig. 3h-i), but not
10 at later ages. We next tested the association between HIV exposure and gut microbiome
11 maturity. We found that CHEU displayed greater microbiome age and MAZ, and hence
12 microbiota over-maturity, compared with CHU at 12 months of age (median 14.7 vs 9.2
13 months; $\beta = 3.18$, $P < 0.001$), which tended to be higher in CHEU at 1 months and at 6
14 months also ($P = 0.074$ and $P = 0.059$, respectively; Fig.3 f-g). However, at 18 months
15 CHEU displayed lower microbiome age (15.3 vs 16.4 months; $\beta = -1.1$, $P = 0.045$
16 respectively). Similarly, CHEU displayed significantly greater metagenome ages and
17 MetAZ scores compared with CHU at 1 month ($\beta = 1.1$, $p = 0.039$) and 12 months of age
18 ($\beta = 2.2$, $P = 0.04$; Fig. 3j-k), suggesting that HIV exposure drives both compositional and
19 functional microbiome over-maturity. The proportion of CHEU receiving prophylactic
20 cotrimoxazole ranged from 56-83% across the 3, 6, 12 and 18-month study visits.

21
22 We next explored which species were differentially abundant between CHEU and CHU
23 by performing multivariable regression analyses, adjusting for age at stool sample

1 collection, exclusive breastfeeding status, delivery mode, and randomised trial arm.
2 Between 1-3 months of age, two Bifidobacteria species, *B. longum* and *B. bifidum* (Fig.
3 4a-b), in addition to *Veillonella seminalis* were significantly less abundant in CHEU versus
4 CHU ($q < 0.1$; Table S3). Conversely, *Flavonifractor plautii* was significantly more
5 abundant in CHEU at 3 months ($q = 0.02$). At 18 months, *B. bifidum* was again significantly
6 less abundant in CHEU ($q = 0.04$), whilst two other Bifidobacteria species were also
7 weakly associated with CHEU, whereby *B. breve* was lower and *B. pseudocatenuatum*
8 higher ($q < 0.25$). Regression analyses of metagenomic pathways with CHEU status
9 generated similar outcomes. Following adjustment for covariates, significant associations
10 for metagenomic pathways were only present at 1 and 3 months of age. At 1 month of
11 age, these included significant negative associations between CHEU and amino acid
12 synthetic pathways (superpathway of L-threonine biosynthesis, superpathway of L-
13 isoleucine biosynthesis I and L-lysine biosynthesis I; Fig. 4c-e) and positive associations
14 with pathways involved in the degradation of sugar derivatives, including fructuronate,
15 glucuronate and galacturonate (PWY 7242 D-fructuronate degradation, PWY 6507 4-
16 deoxy-L-threo-hex-4-enopyranuronate degradation, GALACTUROCAT PWY D-
17 galacturonate degradation I, GALACT GLUCUROCAT PWY superpathway of
18 hexuronide and hexuronate degradation and GLUCUROCAT PWY superpathway of
19 beta D-glucuronide and D-glucuronate degradation; Fig. 4f-h, Table S4). A handful of
20 pathways involved in fatty acid oxidation and fermentation (fatty acid beta oxidation
21 peroxisome and succinate fermentation to butanoate) were also significantly positively
22 associated with CHEU at 3 months of age.

23

1 ***Microbiome functionality, but not composition, predicts linear and ponderal***
2 ***growth.***

3 We next examined the relationship between taxonomic and functional features of the gut
4 microbiome and attained growth (LAZ and WHZ) and growth velocity (WHZ and LAZ
5 velocity, increase in z-score increments per day between visits) from 1-18 months of age
6 using XGBoost models. Models were run separately for each of the six age groups,
7 stratified by maternal HIV status and run in two combinations of predictive features: (i)
8 microbiome features alone (species or pathways); and (ii) microbiome features and
9 epidemiological variables, which included maternal anthropometry, baseline WASH and
10 infant diet variables, amongst others (Table S5). In models combining microbiome
11 features with epidemiological features, birthweight, maternal height, maternal mid-upper
12 arm circumference, and household wealth were all important predictors of attained infant
13 LAZ and WHZ and growth velocity. Models including microbiome taxonomic features
14 (species) alone performed poorly for both attained and growth velocity at every age
15 category and regardless of HIV exposure status (Fig. 5a and 6a), with a majority of models
16 resulting in pseudo-R² values < 0. Model performance for linear growth improved when
17 epidemiological features were included, suggesting that gut microbiota composition alone
18 was poorly predictive of growth. Taxonomic features were weakly predictive of WHZ
19 velocity at 1-3 months (pseudo-R² 0.09-0.12) and attained WHZ at 6 months (pseudo-R²
20 = 0.22), but only in children born to HIV-negative mothers (Fig. 6a). Conversely, models
21 containing functional metagenomic pathways were moderately predictive of both attained
22 and future growth throughout 18 months of age (pseudo-R² = 0-0.55; Fig. 5a and 6a)
23 albeit with relatively large mean absolute errors (MAE) for both linear (0.48-0.95 LAZ) and

1 ponderal growth models (0.6-1.18 WHZ). The inclusion of epidemiological variables in the
2 metagenomic models added little to performance suggesting that pathway features were
3 independently predictive of both linear and ponderal growth. Models predicted WHZ
4 better than LAZ and models including children born to HIV-negative mothers also tended
5 to perform better. MAE decreased in all models as age increased (Fig. S4 a-d).

6

7 ***Microbiome features associated with linear growth***

8 In all but the 2-month age group, birthweight contributed most strongly to prediction of
9 attained LAZ. Metagenomic pathways consistently performed better as predictors than
10 other epidemiological features including maternal height. The most predictive pathways
11 were largely similar between infants born to HIV-positive and HIV-negative mothers. At 1
12 month and 2-months, metagenomic pathways encoding enzyme co-factor biosynthesis,
13 nucleotide degradation and biosynthesis and amino acid biosynthesis were consistently
14 predictive of both attained LAZ and LAZ velocity (Fig. 5b). Accumulated local effects
15 (ALE) (36, 37) plots show the average effect of some of the most important features on
16 model outcomes (Fig. 6c). At 3 months and 6 months, pathways encoding fermentation
17 and carbohydrate biosynthesis were consistently predictive of attained LAZ and LAZ
18 velocity, whilst amino acid degradation pathways, amongst others, were predictive of
19 growth at the oldest age groups. In particular, pathways involved in vitamin B biosynthesis
20 (flavin, folate, biotin, thiazole and cobalamin biosynthetic pathways) were consistently
21 predictive of attained LAZ, and included flavin biosynthesis I, 6-hydroxymethyl-
22 dihydropterin diphosphate biosynthesis, superpathway of tetrahydrofolate biosynthesis,
23 adenosylcobalamin salvage from cobinamide I, biotin biosynthesis II and thiazole

1 biosynthesis I. Increasing abundances of B vitamin biosynthesizing genes contributed to
2 increasing predicted growth in ALE plots, apart from Thiamin biosynthesis, which
3 predicted lower LAZ at 12 months (Fig. 5c). At 12 months of age, 4-coumarate
4 degradation (anaerobic), a pathway involved in plant polysaccharide degradation, was
5 the most predictive pathway of LAZ in children born to HIV-positive mothers, with greater
6 abundance associated with greater LAZ. We also assessed pathways predicting growth
7 velocity and found similar results to that of attained growth (Fig. S5a). At 2 months of age,
8 folate biosynthetic pathways (folate transformations II and N10-formyl-tetrahydrofolate
9 biosynthesis) were the top two predictive features of linear growth velocity in children born
10 to HIV-negative mothers, whilst at 3 months, purine and pyrimidine pathways were
11 amongst the most predictive features of linear growth velocity. Similarly, as per the
12 attained growth models, amino acid and fatty acid biosynthetic pathways were also
13 strongly predictive. Glycogen biosynthesis pathways were also consistently predictive of
14 linear growth velocity at all ages from 2 months onwards.

15

16 ***Microbiome features associated with ponderal growth***

17 In the few models incorporating taxonomic features that weakly predicted WHZ velocity
18 in infants born to HIV-negative mothers, *Escherichia coli* at 2 months and *Bacteroides*
19 *fragilis* and *Veillonella atypica* at 3 months were amongst the most predictive features
20 (Fig. S6). Many of the same categories of biosynthetic microbiota pathways were
21 predictive of both WHZ and LAZ, including amino acid and nucleotide (especially purine
22 biosynthesis) biosynthetic pathways in addition to a number of lipid synthesis pathways
23 at older age groups (Fig. 6b-c). The most predictive pathways were largely similar

1 between infants born to HIV-positive and HIV-negative mothers. O-antigen biosynthesis
2 pathways (PWY-7328 UDP-glucose-derived O-antigen building blocks biosynthesis &
3 PWY-7332 UDP-N-acetylglucosamine-derived O-antigen building blocks biosynthesis
4 and OANTIGEN-PWY pathway), which did not appear in LAZ models, were consistently
5 amongst the most predictive features of WHZ at 1, 6, 12 and 18 months, whereby greater
6 abundance was associated with reduced growth. Pyrimidine and purine synthetic
7 pathways were consistently the strongest predictors of WHZ velocity with varying
8 directions of association, including superpathway of pyrimidine deoxyribonucleotides de
9 novo biosynthesis (PWY0-166) which was the strongest predictive feature of WHZ
10 velocity at 12 months in children born to HIV-mothers (Fig. S5b). Similar to the attained
11 WHZ models, O-antigen biosynthesis pathways, amino acid synthetic pathways and
12 glycogen biosynthesis pathways were all strongly predictive of ponderal growth velocity.

13

14

15 **DISCUSSION**

16 We report the succession and maturation of the early-life gut microbiome in a cohort of
17 335 children from rural Zimbabwe through the first 18 months after birth. We find that
18 taxonomic composition of the gut microbiome is poorly predictive of child growth, however
19 functional composition moderately predicts both attained LAZ/WHZ and LAZ/WHZ
20 velocity, with pathways including B vitamin and nucleotide biosynthesis genes amongst
21 the most predictive of child growth. We also report that randomized WASH and IYCF
22 interventions have little impact on early-life gut microbiome composition, whilst maternal
23 HIV infection, which is associated with impaired infant growth, is associated with over-

1 maturity of the gut microbiome, featuring a depletion in commensal *Bifidobacteria*
2 species. Collectively, these data suggest that disturbances in the functional potential of
3 the infant gut microbiome may contribute to poor infant growth and that interventions
4 targeting the infant gut microbiome may serve as novel solutions to combat child stunting,
5 particularly in CHEU.

6
7 Our previous data from the SHINE trial found that the WASH intervention had no impact
8 on linear growth, whilst IYCF improved growth by 0.16 LAZ (35). Furthermore, WASH had
9 no impact on carriage of enteropathogens or diarrheal incidence (14). Here, we show that
10 the improved WASH and IYCF interventions also had little impact on the infant gut
11 microbiome throughout 18 months after birth. Our results support those from high-income
12 settings, showing a structured, programmed assembly of the gut microbiome in healthy
13 children who are born by vaginal delivery and exclusively breastfed (17, 38). These data
14 suggest that this programmed microbial maturation is robust to changes in WASH and
15 complementary feeding, as delivered in this trial, and that potential microbiome-mediated
16 pathways affecting early-life growth occur independently of these specific interventions.
17 Improvements in growth as a result of the IYCF intervention are not driven by the
18 microbiome, as supported by previous reports showing that the gut microbiome does not
19 mediate the effect of lipid-based IYCF nutrient supplements on child growth (39). More
20 intensive interventions that target WASH, microbial exposures, nutrient intake and
21 microbiota-directed foods during the first 2 years of life may be required to modify this
22 programmed trajectory of gut microbiome succession.

23

1 The data reported here complement previous research associating the gut microbiome
2 with infant growth. The composition and maturity of the gut microbiota has been shown
3 to be disturbed during severe acute malnutrition (SAM) and could be used to predict
4 growth recovery (19). More recently, an “ecogroup” of 15 bacterial taxa has been
5 identified that exhibits consistent covariation, thereby representing microbiota maturation,
6 throughout the first 2 years after birth across different geographical cohorts (40).
7 However, little research has examined microbiome maturation in the context of child
8 stunting. We report similar maturation of the early-life microbiome in this stunting cohort,
9 driven by many of the same age-predictive taxa as previously reported, notably
10 *Faecalibacterium prausnitzii* as the species most predictive of age. We extend this to
11 report functional maturation of the early-life gut microbiota and find that, in addition to
12 amino acid and B-vitamin biosynthetic pathways, methanogenesis from acetate (METH-
13 ACETATE-PWY) was the pathway most predictive of age, despite the apparent lack of
14 methanogens. The predictive strength of this pathway may reflect accumulation of
15 acetogenic species, including *Blautia wexlerae*, that feed into reactions upstream of the
16 METH-ACETATE-PWY. However, our results contrast with previous cross-sectional
17 studies reporting an association between the taxonomic composition of the gut
18 microbiome and stunting (31-33). A previous study from sub-Saharan Africa (Afribiota)
19 reported significant differences in the fecal microbiome of stunted and non-stunted
20 children between 2-5 years of age, hypothesizing that decompartmentalization of the
21 gastrointestinal tract and overgrowth of oropharyngeal taxa are associated with stunting
22 (31). We report here no association between the taxonomic composition of the gut
23 microbiome and linear growth, however our study examined children at younger ages to

1 that of the Afribiota cohort, suggesting that differences in the taxonomic composition of
2 the gut microbiome mediating linear growth may only manifest later in childhood.

3

4 We identified a range of metagenomic pathways that predicted linear and ponderal growth
5 through 18 months suggesting that the potential influence of an altered gut microbiome
6 on child growth is dependent upon a number of interacting metagenomic pathways. This
7 discrepancy in the ability of functional metagenomic features versus taxonomic features
8 to predict growth may suggest that metagenomic pathways contributing to differences in
9 early-life growth may be harboured across a number of functionally redundant species.
10 Intriguingly, many of the pathways that we found predictive of growth in infants, were also
11 found to be predictive of infant birthweight and neonatal growth in analysis of maternal
12 gut microbiomes from the same study cohort (27). Glycogen synthase was the pathway
13 most predictive of birthweight in maternal microbiomes, whereby higher abundance
14 predicted lower birthweight. Here we identified 2 related pathways (PWY-622 starch
15 biosynthesis, GLYCOGENSYNTH-PWY glycogen biosynthesis I (from ADP-D-Glucose))
16 that were consistently ranked as highly predictive of both linear and ponderal growth
17 velocity. Glycogen synthesis occurs as a starvation response in bacteria which facilitates
18 transition into a biofilm state (41, 42). These data suggest that microbiome starvation
19 responses are associated with growth as early as the first months after birth, which may
20 have downstream implications for host metabolism and associated growth pathways. It is
21 plausible that these pathways are a result of altered nutrient composition in breastmilk of
22 mothers of stunted infants, thereby providing insufficient substrates for infant microbiome
23 maturation, as has been identified previously (43). Alternatively, these signatures may be

1 a consequence of a host-induced effect in poorly growing infants on gut microbiome
2 function.

3

4 Pathways encoding biosynthesis of B vitamins were consistently amongst the top
5 predictive features in models predicting both attained and LAZ and WHZ growth velocity.
6 Previous evidence supports the importance of B vitamins in early-life growth whereby
7 maternal folic acid supplementation increases infant birthweight (44). In infants, vitamin
8 B12 status is predictive of both linear and ponderal growth (45), however the largest
9 randomized trial of B12 supplementation on infant growth to date showed no effect (46,
10 47). The gut microbiome biosynthesizes and metabolises B vitamins, including cobalamin
11 (B12) and folate (B9), at levels similar to dietary intake, and abundance of B vitamin-
12 synthesizing genes in the infant gut microbiome differs by delivery mode, antibiotic
13 exposure (48), exclusive breastfeeding practices and geographic location, where vitamin
14 biosynthesis genes are greater in Western settings (38). Greater relative abundance of B
15 vitamin biosynthetic pathways such as thiazole, tetrahydrofolate and flavin biosynthesis
16 in the maternal gut microbiome predicted greater birthweight and neonatal growth in this
17 same cohort, whilst biotin biosynthesis predicted reduced birthweight (27). The gut
18 microbiome transferred from mother to infants may influence the metabolic capacity of
19 the infant microbiome to biosynthesize essential nutrients and influence downstream
20 growth pathways. Purine and pyrimidine biosynthetic pathways consistently contributed
21 to growth predictions across all age groups. In mothers from this same cohort, purine and
22 pyrimidine salvage pathways were associated with increasing birthweight (27). Meta-
23 analyses have also found that dietary nucleotide supplementation in infants significantly

1 increases head circumference and rate of weight gain (49), suggesting that microbiome-
2 derived nucleotide metabolism may play an important role in nutritional status in early
3 infancy.

4
5 An important observation, however, is that many of these pathways predicting growth,
6 including B vitamin and purine/pyrimidine biosynthesis, were also predictive of age and
7 hence microbiota maturation. There was a strong association between age and growth in
8 the SHINE cohort, whereby LAZ declined steadily between 1-18 months of age. This
9 highlights the difficulty in delineating the independent effect of the gut microbiome on
10 growth during infancy, when the microbiome is concurrently undergoing age-related
11 maturation, which is by far the strongest contributor to gut microbiome variability.
12 Although we attempted to account for age-related effects by examining samples within
13 specified age categories, the microbiome-growth relationship observed here may be
14 confounded by age. Previous studies, employing MAZ as a maturation index to account
15 for age have demonstrated microbiome maturation is disturbed in acutely malnourished
16 states (19) but is not associated with linear growth (50). Our observations that functional
17 microbiome characteristics moderately predict changes in linear growth add novel
18 findings to this literature but need to be replicated in other large cohorts examining the
19 functional maturation of the gut microbiome throughout early childhood in similar settings.

20
21 We report that maternal HIV infection had a significant impact on the infant microbiome
22 throughout the first 18 months after birth. We previously reported that CHEU have a 16%
23 higher prevalence of stunting, 40% higher risk of infant mortality and poorer cognitive

1 development compared with CHU (30). The results presented here raise the intriguing
2 possibility that altered succession and assembly of the infant gut microbiome may drive
3 some of these poorer clinical outcomes in CHEU. These findings are in line with previous
4 reports of disturbed gut microbiome composition in CHEU (25, 29). A number of factors
5 may explain these differences. Firstly, CHEU receive prophylactic cotrimoxazole from 6
6 weeks of age, which may impact gut microbiome succession throughout childhood (51,
7 52). However, we found the largest differences in gut microbiome composition and
8 function in samples from infants < 6 weeks of age, suggesting that these findings were
9 independent of antibiotic prophylaxis and that CHEU may acquire an altered microbiome
10 from their mothers. We saw relatively minor differences, however, in the gut microbiome
11 of mothers living with HIV or without HIV in this same cohort (27). Although there were
12 significant differences in compositional beta diversity, the only species that differed in
13 abundance was *Treponema berlinense*, which was significantly less abundant in mothers
14 living with HIV. Exclusivity of breast-feeding is one of the most impactful factors
15 determining infant gut microbiome composition, however there was no significant
16 difference in EBF rates between CHEU and CHU in this cohort, suggesting that exclusivity
17 of breast-feeding was not responsible for these differences. Previous research has shown
18 that the HMO content of breast milk differs between mothers living with and without HIV
19 (25). HMOs are among the primary substrates for digestion by the infant gut
20 microbiome thereby fundamentally determining gut microbiome composition. Indeed, we
21 found that *Bifidobacteria* species, which are primary degraders of HMOs were
22 significantly less abundant in CHEU, as were genes involved in amino acid biosynthesis.
23 Previous, in-depth profiling of infant immune development has found that a lack of

1 *Bifidobacteria* in infancy is associated with systemic inflammation and immune
2 dysregulation (21) which are also observed in CHEU (53-55), suggesting that the lack of
3 commensal *Bifidobacteria* may mediate some of the poor immune, growth and clinical
4 outcomes observed in CHEU.

5
6 This study is strengthened by the large birth cohort of healthy infants from two rural
7 districts in sub-Saharan Africa. This population is underrepresented in microbiome
8 research to date. The use of whole metagenome shotgun sequencing strengthens the
9 study, providing a unique dataset in which to examine both compositional and functional
10 microbiome maturation in early childhood. Furthermore, the machine learning approach
11 (XGBoost) and comprehensive clinical and epidemiologic data allows us to account for
12 environmental exposures influencing gut microbiome succession and infant growth in this
13 rural, low-resource settings. However, there are several limitations to this analysis: (i) the
14 SHINE microbiome sub-study included more HIV-positive mothers than the main SHINE
15 trial (30% versus 15%), and the proportion of HIV-exposed infants varied by age group.
16 This resulted in some small sub-groups in some of our age categories by HIV exposure
17 analyses, which likely resulted in unstable predictions in certain XGBoost models; (ii) a
18 significant proportion of the sequencing reads included in our datasets were not
19 annotatable (median 58.6%) using the specified bioinformatic pipelines (MetaPhlan3 and
20 HUMAnN3). This large abundance of unknown sequences is common in samples derived
21 from non-Western populations (56) and leads to inferences solely being made from the
22 assignable fraction, potentially missing important microbiota features that are predictive
23 of infant growth but are currently not represented in databases; (iii) *Escherichia coli* was

1 one of the most prevalent bacterial species across all age groups; however, MetaPhlAn3
2 cannot differentiate between *E. coli* pathotypes. Different *E. coli* pathotypes, such as
3 enteropathogenic and enteroaggregative *E. coli* have been associated with intestinal
4 pathology and EED, however we previously reported in the same cohort that some of
5 these pathotypes were not associated with growth (14); (iv) we attempted to account for
6 the age-related confounding of the microbiome-growth relationship by predicting attained
7 growth and growth velocity in discrete age groups, but residual confounding may still be
8 present, influencing our ability to identify microbiome features independently associated
9 with growth; (v) data on infant antimicrobial use in the SHINE trial was incomplete, limiting
10 our ability to confidently assess this and other potential confounders; (vi) differences in
11 microbiome composition and function may also be driven by differences in intestinal
12 microbial load, motility and biogeography, which were not assessed; (vii) finally, we chose
13 mother-infant pairs with the most complete sample collection during follow-up, in order to
14 strengthen our inferences about development of the gut microbiome over time. Baseline
15 characteristics of the microbiome sub-cohort were largely similar to those of the larger
16 trial, suggesting that the microbiome sub-study cohort studied was largely representative
17 of the larger SHINE trial.

18
19 Collectively, these data suggest that HIV exposure shapes maturation of the infant gut
20 microbiota, and that the functional composition of the infant gut microbiome is moderately
21 predictive of infant growth in a population at high risk of stunting. Novel therapeutic
22 approaches targeting the gut microbiome may mitigate the poor clinical outcomes that
23 are observed in CHEU, a growing population of children in sub-Saharan Africa. By

1 contrast, current WASH and IYCF interventions fail to impact the infant gut microbiome
2 and therefore transformative WASH and microbiome-targeted dietary interventions may
3 prove to be more successful approaches to target the microbial pathways mediating early-
4 life growth.

5
6

7 **MATERIALS AND METHODS**

8 ***SHINE trial design***

9 The study design and methods for The Sanitation Hygiene Infant Nutrition Efficacy
10 (SHINE) trial and for the corresponding microbiome analyses, have been reported
11 previously (57, 58). Briefly, SHINE was a 2x2 cluster-randomized trial, conducted
12 between 2012 and 2017, to determine the independent and combined effects of improved
13 infant and young child feeding (IYCF) and WASH on child stunting and anaemia in two
14 rural Zimbabwean districts (NCT01824940). 5280 pregnant women were cluster-
15 randomized to one of four interventions: WASH, IYCF, WASH+IYCF, and Standard of
16 Care (SOC). The SOC interventions, included in all trial arms, comprised exclusive
17 breastfeeding promotion for all infants up to 6 months and strengthened prevention of
18 mother to child transmission (PMTCT) of HIV services. The household WASH intervention
19 was initiated during pregnancy and was designed to reduce exposure to human and
20 animal feces, including, at the household level: construction of a ventilated improved pit
21 latrine, installation of two hand-washing stations plus monthly delivery of liquid soap and
22 water chlorination solution, provision of a play space for the infant, and hygiene
23 counseling. The IYCF intervention was designed to improve infant diets using a small-

1 quantity lipid-based nutrient supplement (SQ-LNS), provided to the infant from 6-18
2 months, and educational interventions promoting the use of age-appropriate, locally
3 available foods and dietary diversity. Lastly, a combined trial arm, WASH+IYCF,
4 evaluated the effects of both improved WASH and infant nutrition.

5
6 Infants were followed up at study visits at 1, 3, 6, 12 and 18 months of age. Length and
7 weight were measured at each infant visit, as described previously (35). Length-for-age
8 z scores (LAZ) and weight-for-height z scores (WHZ) were calculated from length and
9 weight measurements at each visit according to WHO Child Growth Standards.
10 Epidemiologic data for the infants was collected from the baseline and follow-up visits
11 using trial questionnaires that included maternal anthropometry, birth outcomes, baseline
12 household WASH facilities, household wealth, maternal education, religion, parity,
13 household size, dietary diversity, changes in breastfeeding and complementary feeding
14 practices, food security, 7-day and 3-month infant health status, and antimicrobial use.

16 ***HIV testing***

17 HIV testing was conducted on mothers at the baseline visit using a rapid test algorithm
18 (Alere Determine HIV1/2 test, followed by INSTI HIV-1/2 test if positive). Those testing
19 positive for HIV had CD4 counts measured (Alere Pima Analyser) and referral to local
20 clinics; women were encouraged to begin co-trimoxazole prophylaxis and ART, to
21 exclusively breastfeed, and to attend clinic at 6 weeks postpartum for early infant
22 diagnosis and infant co-trimoxazole prophylaxis. Women testing negative for HIV were
23 offered retesting at 32 gestational weeks and 18 months postpartum. Children of mothers

1 living with HIV were offered testing for HIV at each of the study visits. Those who tested
2 positive were referred to local clinics for ART. HIV was diagnosed using DNA PCR on
3 dried blood-spot samples or RNA PCR on plasma in samples collected prior to 18 months.
4 In samples collected after 18 months, HIV was diagnosed by PCR or rapid test algorithm,
5 depending on samples provided. Children born to HIV-positive mothers and whom tested
6 negative at 18 months were classified as HIV-exposed uninfected (CHEU). Inconclusive
7 or discordant results were re-tested; if no further samples were available or repeat testing
8 was inconclusive, children were classified as HIV-unknown.

9

10 ***Microbiome sub-study***

11 All CHEU and a subgroup of CHU from the SHINE study were enrolled into an
12 Environmental Enteric Dysfunction (EED) sub-study (n=1,656 mother-child pairs); these
13 infants underwent intensive biological specimen collection at 1, 3, 6, 12 and 18 months
14 of age (58). The EED sub-study was therefore enriched for mothers living with HIV, by
15 design. Sample selection for inclusion into the current microbiome study was conducted
16 to enhance longitudinal profiling of the mother and infants gut microbiota. Of the mother-
17 infant pairs within the EED sub-study, those with least one maternal fecal specimen (of 2
18 possible) and at least 2 infant fecal specimens (of 5 possible) were included in the gut
19 microbiome analyses. An additional 94 samples collected at the 1 and 3-month visits, that
20 did not meet these criteria, but had microbiome sequencing data available from a
21 separate study examining rotavirus vaccine immunogenicity in the SHINE trial (59), were
22 also included in these analyses. Infant ages varied at each study visit due to the allowable
23 window around the visit date for the larger SHINE trial. Therefore, for this microbiome

1 study, each stool samples were re-categorized into 6 age groups corresponding to
2 important stages in infant microbiome development: “1 month” (0-6 weeks), “2 months”
3 (7 weeks – <3 months), “3 months” (3-6 months), “6 months” (6-9 months), “12 months”
4 (9-15 months), and “18 months” (15-20 months).

5 6 ***Sample collection***

7 Study visits were conducted by trained study nurses in participants' homes. Sterile stool
8 collection tubes were provided to mothers, who collected stool samples from their infants
9 on the morning of each study visit. Samples were placed in cool boxes immediately upon
10 collection by study nurses and transported by motorbike to field laboratories where they
11 were aliquoted and stored at -80°C within 6 hours of collection before subsequent
12 transport to the central laboratory in Harare for long-term storage at -80°C. An aliquot of
13 each stool sample was shipped on dry ice by courier to the British Columbia Centre for
14 Disease Control in Vancouver, Canada. A strict cold chain was maintained throughout
15 transport, ensuring no freeze-thaw cycles occurred between sample collection and
16 processing.

17 18 ***Whole metagenome library preparation and sequencing***

19 DNA was extracted from 100-200mg of stool samples using the Qiagen DNeasy
20 PowerSoil Kit as per the manufacturer's instructions. DNA quantity was assessed by
21 fluorometry (QuBit) and quality confirmed by spectrophotometry (SimpliNano). 1µg DNA
22 was subsequently used as input for metagenomic sequencing library preparation using
23 the Illumina TruSeq PCR-free library preparation protocol, using custom end-repair,

1 adenylation and ligation enzyme premixes (New England Biolabs). The concentration and
2 size of constructed libraries were assessed by qPCR and by TapeStation (Agilent). DNA-
3 free negative controls and positive controls (ZymoBIOMICS) were included in all DNA
4 extraction and library preparation steps. Libraries were pooled in random batches of 48
5 samples including one negative control. A set of specimens were subject to replicate DNA
6 extraction, library preparation and, sequencing to estimate the magnitude of technical
7 variability among samples. Whole metagenome sequencing was performed with 125-
8 nucleotide paired-end reads using either the Illumina HiSeq 2500 or HiSeqX platforms at
9 Canada's Michael Smith Genome Sciences Centre, Vancouver, Canada.

10

11 **Bioinformatics**

12 Sequenced reads were trimmed of adapters and filtered to remove low-quality, short
13 (<70% raw read length), and duplicate reads, as well as those of human, other animal or
14 plant origin, using *KneadData* with default settings. Species composition was determined
15 by identifying clade-specific markers from reads using MetaPhlan3 with default settings
16 (60). Relative abundance estimates were obtained from known assigned reads, and
17 unknown read proportions were estimated from total, assigned and unassigned, reads.
18 Percent human DNA was estimated from *KneadData* output, using the proportion of
19 quality-filtered reads that align to the human genome. Given the smaller viral genome
20 sizes, sequencing depth, and limitations of MetaPhlan3 for virus identification, we did not
21 include viruses in our current analyses. We applied a minimum threshold of >0.1% relative
22 abundance and $\geq 5\%$ prevalence for all detected species. Metabolic pathway composition
23 was determined using HUMAnN3 with default settings against the UniRef90 database

1 (60). Pathway abundance estimates were normalized using reads per kilobase per million
2 mapped reads (RPKM) and then re-normalized to relative abundance. We applied a
3 minimum relative abundance threshold of $3 \times 10^{-7}\%$ and $\geq 5\%$ prevalence for all
4 metagenomic pathways.

5 6 **Statistical analysis**

7 All data were analysed using R (v.4.0.5). Microbiome data were handled using the
8 phyloseq package. Alpha diversity metrics were calculated using the vegan package.
9 Beta-diversity was estimated using the Bray-Curtis dissimilarity index and analysed by
10 permutation analysis of variance (PERMANOVA). Differential abundance analysis of
11 species or functional pathways was assessed using multiple regression analyses using
12 the *MaAsLin2* package (61). Four covariates were chosen for adjustment in *MaAsLin2*
13 regression models and included age at stool sample collection, exclusive breastfeeding
14 status (recorded at 3 months old), delivery mode, and randomised trial arm. These
15 covariates were chosen based on biological plausibility and previous evidence of their
16 influence on gut microbiome composition in large birth cohorts (17). Adjustment for
17 multiple comparisons was performed using the Benjamini-Hochberg false discovery rate
18 (FDR).

19
20 The SHINE trial did not observe an interaction between the randomized WASH and IYCF
21 interventions and growth; therefore, randomised trial arms were combined into WASH
22 versus non-WASH arms and IYCF versus non-IYCF arms for specified analyses. We
23 restricted the IYCF analysis to the 6, 12 and 18 month visits, corresponding to the period

1 during which supplemental infant feeding was introduced (from 6 months of age). All
2 children, regardless of HIV status or exposure status were included in the growth
3 analyses (875 total stool samples). 16 samples had missing ages and were excluded from
4 age prediction models. Stool samples collected from children classified as HIV-unknown
5 (24 samples) or HIV-positive (4 samples) at 18 months were excluded from direct
6 comparisons of CHEU vs CHU infants.

7

8 ***XGBoost models***

9 Relationships between the infant microbiome and age or growth (attained LAZ/WHZ and
10 WHZ/LAZ velocity) were evaluated using extreme gradient boosting machines
11 (XGBoost). XGBoost builds an optimized predictive model by creating an ensemble from
12 a series of weakly predictive models. XGBoost is also non-parametric, can capture non-
13 linear relationships, and can accommodate high-dimensional data (62). The XGBoost
14 models were developed using microbiome relative abundances (species or pathways)
15 and all epidemiologic variables.

16

17 XGBoost model selection was performed in 3 stages as previously described (27). In
18 brief, *BayesianOptimization* function of the *rBayesianOptimization* package was used
19 with 10-fold cross-validation to select model hyperparameters by minimizing the mean
20 squared error (MSE). Models with the lowest MSE (in the 5th percentile) were retained,
21 and from these models the variables that contributed to the top 95% of variable
22 importance by proportion were retained. In stage two, all epidemiologic variables were
23 included with the microbiome variables obtained in stage one, and *BayesianOptimization*

1 was re-run as for stage one but using leave-one-out cross-validation. Microbiome
2 variables that contributed to the top 95% of variable importance by proportion were
3 retained. In stage three, all epidemiologic variables, microbiome features, and
4 hyperparameters selected in stage two were used to fit our final models, using leave-one-
5 out cross-validation to minimize the MSE. This 3-stage hyperparameter tuning and model
6 building was performed for two feature sets, one comprising microbiome features and a
7 second comprising microbiome plus epidemiologic features; this was done to assess
8 model performance and to examine the contribution of epidemiologic versus microbiome
9 features. Separate models were built for attained LAZ and WHZ and growth velocity
10 outcomes (13). We assessed microbiota composition and functional pathways
11 separately. XGBoost models were fit using the H2O.ai engine and *h2o* R package
12 interface with the *XGBoost* package. XGBoost model performance was evaluated using
13 pseudo- R^2 and mean absolute error (MAE). Pseudo- R^2 values < 0 indicated that the
14 prediction of the model was worse than the mean response. Scaled relative importance
15 for each model feature was used to identify the twenty most informative variables for
16 further interpretation, where the most important variable is ranked first, and the
17 importance of subsequent variables are relative to the first variable. The marginal
18 relationships between the twenty most important features and each growth outcome were
19 visualized for interpretation (36) using accumulated local effects plots (ALE). ALE plots
20 can be interpreted as showing a marginal effect, adjusted for all covariates retained in the
21 final model, showing the expected change in the outcome variable per increment in a
22 model feature. The resulting effect sizes are plotted cumulatively and centered about the
23 average effect size (37). ALEs were generated using the *ALEplot* package, modified to

1 compute confidence intervals, and were plotted using *ggplot2*. Standard deviations were
2 calculated per increment and were used to calculate and plot increment-wise 95%
3 confidence intervals.

4

5 ***Microbiome age***

6 We investigated microbiome maturation by building an age prediction model using
7 XGBoost and microbiome features only (species or pathways). We partitioned the infants
8 and their corresponding datasets into three groups to train and test a model of microbiome
9 age: (1) CHU with LAZ > -2 at 18 month of age, who contributed >1 dataset (healthy
10 training set), (2) remaining CHU infants with LAZ > -2 at 18 month of age, who contributed
11 a single metagenomic dataset (healthy test set); and (3) CHEU or children with LAZ ≤ -2
12 at 18 months of age (unhealthy test set). Age was log transformed as a response in the
13 XGBoost model. We performed the same 3-stage tuning and model building procedure,
14 as described above. We generated model performance metrics, including pseudo-R²,
15 mean absolute error (MAE) and mean squared error (MSE) for the three sets. For the
16 training set, we used the cross-validation, hold-out predictions to generate the metrics,
17 and for the two test sets, we used the predicted values from the final to calculate the
18 model performance metrics. We exponentiated the predicted log transformed ages and
19 plotted these values against the observed age. The predicted age using these models is
20 referred to as ‘microbiota age’ for the models trained using species and ‘metagenome
21 age’ for models trained using pathways. To account for variance of microbiota ages with
22 respect to chronological age within the age range of each study visit, a microbiota for age
23 Z-score (MAZ) and metagenome-for-age Z-score (MetAZ) was also created using the

1 microbiome age and metagenome ages as previously described (19). A Z-score was
2 calculated to account for variation in ages within each study visit using the following
3 formula:

4
5
$$\frac{\text{(Microbiota age of child - median microbiota age of 'healthy' child at same study visit)}}{\text{standard deviation of microbiota age of 'healthy' child at same study visit.}}$$

7

8

9 **SUPPLEMENTARY MATERIALS:**

10

11 **Fig. S1. Participants and samples included in the study.**

12

13 **Fig. S2. Whole metagenome sequencing performance.**

14

15 **Fig. S3. Diversity metrics in entire dataset.**

16

17 **Fig. S4. XGBoost model performance metrics as assessed by mean absolute error.**

18

19 **Fig. S5. Top ranked pathways in XGBoost models predicting growth velocity.**

20

21 **Fig. S6. Figure S5. Top ranked species in XGBoost models predicting growth.**

22

1 **Table S1. Baseline characteristics of infants in SHINE trial and microbiome sub-**
2 **study**

3
4 **Table S2. Detected Eukaryota and Archaea species prior to prevalence filtering**

5
6 **Table S3. Multivariate regression analysis examining the effect of maternal HIV**
7 **infection, age (days), exclusive breastfeeding status, delivery mode and trial arm**
8 **on taxonomic microbiome composition**

9
10 **Table S4. Multivariate regression analysis examining the effect of maternal HIV**
11 **infection, age (days), exclusive breastfeeding status, delivery mode and trial arm**
12 **on gene pathway microbiome composition**

13
14 **Table S5. Epidemiological variables included in XGBoost models**

15
16

17 **REFERENCES**

- 18 1. Victora CG, Christian P, Vdaletti LP, Gatica-Domínguez G, Menon P, Black RE. Revisiting
19 maternal and child undernutrition in low-income and middle-income countries: variable progress
20 towards an unfinished agenda. *Lancet*. 2021;397(10282):1388-99.
- 21 2. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, Ezzati M, Grantham-
22 McGregor S, Katz J, Martorell R, Uauy R, Group MaCNS. Maternal and child undernutrition and
23 overweight in low-income and middle-income countries. *Lancet*. 2013;382(9890):427-51.

- 1 3. Prendergast AJ, Humphrey JH. The stunting syndrome in developing countries. *Paediatr*
2 *Int Child Health*. 2014;34(4):250-65.
- 3 4. Dewey KG, Begum K. Long-term consequences of stunting in early life. *Matern Child Nutr*.
4 2011;7 Suppl 3:5-18.
- 5 5. Dewey KG, Stewart CP, Wessells KR, Prado EL, Arnold CD. Small-quantity lipid-based
6 nutrient supplements for the prevention of child malnutrition and promotion of healthy
7 development: overview of individual participant data meta-analysis and programmatic
8 implications. *Am J Clin Nutr*. 2021;114(Suppl 1):3S-14S.
- 9 6. Prendergast A, Kelly P. Enteropathies in the developing world: neglected effects on global
10 health. *Am J Trop Med Hyg*. 2012;86(5):756-63.
- 11 7. Harper KM, Mutasa M, Prendergast AJ, Humphrey J, Manges AR. Environmental enteric
12 dysfunction pathways and child stunting: A systematic review. *PLoS Negl Trop Dis*.
13 2018;12(1):e0006205.
- 14 8. Crane RJ, Jones KD, Berkley JA. Environmental enteric dysfunction: an overview. *Food*
15 *Nutr Bull*. 2015;36(1 Suppl):S76-87.
- 16 9. Kosek MN, Investigators M-EN. Causal Pathways from Enteropathogens to
17 Environmental Enteropathy: Findings from the MAL-ED Birth Cohort Study. *EBioMedicine*.
18 2017;18:109-17.
- 19 10. Prendergast AJ, Kelly P. Interactions between intestinal pathogens, enteropathy and
20 malnutrition in developing countries. *Curr Opin Infect Dis*. 2016;29(3):229-36.
- 21 11. Amadi B, Zyambo K, Chandwe K, Besa E, Mulenga C, Mwakamui S, Siyumbwa S, Croft
22 S, Banda R, Chipunza M, Chifunda K, Kazhila L, VanBuskirk K, Kelly P. Adaptation of the small
23 intestine to microbial enteropathogens in Zambian children with stunting. *Nat Microbiol*.
24 2021;6(4):445-54.
- 25 12. Pickering AJ, Null C, Winch PJ, Mangwadu G, Arnold BF, Prendergast AJ, Njenga SM,
26 Rahman M, Ntozini R, Benjamin-Chung J, Stewart CP, Huda TMN, Moulton LH, Colford JM, Luby
27 SP, Humphrey JH. The WASH Benefits and SHINE trials: interpretation of WASH intervention
28 effects on linear growth and diarrhoea. *Lancet Glob Health*. 2019;7(8):e1139-e46.

- 1 13. Mutasa K, Ntozini R, Mbuya MNN, Rukobo S, Govha M, Majo FD, Tavengwa N, Smith
2 LE, Caulfield L, Swann JR, Stoltzfus RJ, Moulton LH, Humphrey JH, Gough EK, Prendergast AJ.
3 Biomarkers of environmental enteric dysfunction are not consistently associated with linear
4 growth velocity in rural Zimbabwean infants. *Am J Clin Nutr.* 2021;113(5):1185-98.
- 5 14. Rogawski McQuade ET, Platts-Mills JA, Gratz J, Zhang J, Moulton LH, Mutasa K, Majo
6 FD, Tavengwa N, Ntozini R, Prendergast AJ, Humphrey JH, Liu J, Houpt ER. Impact of Water
7 Quality, Sanitation, Handwashing, and Nutritional Interventions on Enteric Infections in Rural
8 Zimbabwe: The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) Trial. *J Infect Dis.*
9 2020;221(8):1379-86.
- 10 15. Rogawski ET, Liu J, Platts-Mills JA, Kabir F, Lertsethtakarn P, Siguas M, Khan SS,
11 Praharaj I, Murei A, Nshama R, Mujaga B, Havt A, Maciel IA, Operario DJ, Taniuchi M, Gratz J,
12 Stroup SE, Roberts JH, Kalam A, Aziz F, Qureshi S, Islam MO, Sakpaisal P, Silapong S, Yori PP,
13 Rajendiran R, Benny B, McGrath M, Seidman JC, Lang D, Gottlieb M, Guerrant RL, Lima AAM,
14 Leite JP, Samie A, Bessong PO, Page N, Bodhidatta L, Mason C, Shrestha S, Kiwelu I, Mduma
15 ER, Iqbal NT, Bhutta ZA, Ahmed T, Haque R, Kang G, Kosek MN, Houpt ER, Investigators M-
16 EN. Use of quantitative molecular diagnostic methods to investigate the effect of enteropathogen
17 infections on linear growth in children in low-resource settings: longitudinal analysis of results
18 from the MAL-ED cohort study. *Lancet Glob Health.* 2018;6(12):e1319-e28.
- 19 16. Richard SA, McCormick BJJ, Murray-Kolb LE, Lee GO, Seidman JC, Mahfuz M, Ahmed
20 T, Guerrant RL, Petri WA, Rogawski ET, Houpt E, Kang G, Mduma E, Kosek MN, Lima AAM,
21 Shrestha SK, Chandyo RK, Bhutta Z, Bessong P, Caulfield LE, Investigators M-EN. Enteric
22 dysfunction and other factors associated with attained size at 5 years: MAL-ED birth cohort study
23 findings. *Am J Clin Nutr.* 2019;110(1):131-8.
- 24 17. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, Ross MC, Lloyd
25 RE, Doddapaneni H, Metcalf GA, Muzny D, Gibbs RA, Vatanen T, Huttenhower C, Xavier RJ,
26 Rewers M, Hagopian W, Toppari J, Ziegler AG, She JX, Akolkar B, Lernmark A, Hyoty H, Vehik
27 K, Krischer JP, Petrosino JF. Temporal development of the gut microbiome in early childhood
28 from the TEDDY study. *Nature.* 2018;562(7728):583-8.
- 29 18. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. The Human Microbiome and
30 Child Growth - First 1000 Days and Beyond. *Trends Microbiol.* 2019;27(2):131-47.

- 1 19. Subramanian S, Huq S, Yatsunenکو T, Haque R, Mahfuz M, Alam MA, Benezra A,
2 DeStefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q, Province MA, Petri
3 WA, Ahmed T, Gordon JI. Persistent gut microbiota immaturity in malnourished Bangladeshi
4 children. *Nature*. 2014;510(7505):417-21.
- 5 20. Rouhani S, Griffin NW, Yori PP, Gehrig JL, Olortegui MP, Salas MS, Trigo DR, Moulton
6 LH, Hout E, Barratt MJ, Kosek MN, Gordon JI. Diarrhea as a Potential Cause and
7 Consequence of Reduced Gut Microbial Diversity Among Undernourished Children in Peru. *Clin
8 Infect Dis*. 2020;71(4):989-99.
- 9 21. Blanton LV, Charbonneau MR, Salih T, Barratt MJ, Venkatesh S, Ilkaveya O,
10 Subramanian S, Manary MJ, Trehan I, Jorgensen JM, Fan YM, Henrissat B, Leyn SA, Rodionov
11 DA, Osterman AL, Maleta KM, Newgard CB, Ashorn P, Dewey KG, Gordon JI. Gut bacteria that
12 prevent growth impairments transmitted by microbiota from malnourished children. *Science*.
13 2016;351(6275).
- 14 22. Smith MI, Yatsunenکو T, Manary MJ, Trehan I, Mkakosya R, Cheng J, Kau AL, Rich SS,
15 Concannon P, Mychaleckyj JC, Liu J, Hout E, Li JV, Holmes E, Nicholson J, Knights D, Ursell
16 LK, Knight R, Gordon JI. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor.
17 *Science*. 2013;339(6119):548-54.
- 18 23. Chen RY, Mostafa I, Hibberd MC, Das S, Mahfuz M, Naila NN, Islam MM, Huq S, Alam
19 MA, Zaman MU, Raman AS, Webber D, Zhou C, Sundaresan V, Ahsan K, Meier MF, Barratt MJ,
20 Ahmed T, Gordon JI. A Microbiota-Directed Food Intervention for Undernourished Children. *N
21 Engl J Med*. 2021;384(16):1517-28.
- 22 24. Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Kung VL, Cheng J, Chen RY,
23 Subramanian S, Cowardin CA, Meier MF, O'Donnell D, Talcott M, Spears LD, Semenkovich CF,
24 Henrissat B, Giannone RJ, Hettich RL, Ilkayeva O, Muehlbauer M, Newgard CB, Sawyer C, Head
25 RD, Rodionov DA, Arzamasov AA, Leyn SA, Osterman AL, Hossain MI, Islam M, Choudhury N,
26 Sarker SA, Huq S, Mahmud I, Mostafa I, Mahfuz M, Barratt MJ, Ahmed T, Gordon JI. Effects of
27 microbiota-directed foods in gnotobiotic animals and undernourished children. *Science*.
28 2019;365(6449).

- 1 25. Bender JM, Li F, Martelly S, Byrt E, Rouzier V, Leo M, Tobin N, Pannaraj PS, Adisetiyo
2 H, Rollie A, Santiskulvong C, Wang S, Autran C, Bode L, Fitzgerald D, Kuhn L, Aldrovandi GM.
3 Maternal HIV infection influences the microbiome of HIV-uninfected infants. *Sci Transl Med*.
4 2016;8(349):349ra100.
- 5 26. Lozupone CA, Li M, Campbell TB, Flores SC, Linderman D, Gebert MJ, Knight R, Fontenot
6 AP, Palmer BE. Alterations in the gut microbiota associated with HIV-1 infection. *Cell Host*
7 *Microbe*. 2013;14(3):329-39.
- 8 27. Gough EK, Edens TJ, Geum HM, Baharmand I, Gill SK, Robertson RC, Mutasa K, Ntozini
9 R, Smith LE, Chasekwa B, Majo FD, Tavengwa NV, Mutasa B, Francis F, Carr L, Tome J,
10 Stoltzfus RJ, Moulton LH, Prendergast AJ, Humphrey JH, Manges AR, Team ST. Maternal fecal
11 microbiome predicts gestational age, birth weight and neonatal growth in rural Zimbabwe.
12 *EBioMedicine*. 2021;68:103421.
- 13 28. Amenogbe N, Dimitriu P, Cho P, Ruck C, Fortuno ES, Cai B, Alimenti A, Côté HCF, Maan
14 EJ, Slogrove AL, Esser M, Marchant A, Goetghebuer T, Shannon CP, Tebbutt SJ, Kollmann TR,
15 Mohn WW, Smolen KK. Innate Immune Responses and Gut Microbiomes Distinguish HIV-
16 Exposed from HIV-Unexposed Children in a Population-Specific Manner. *J Immunol*.
17 2020;205(10):2618-28.
- 18 29. Machiavelli A, Duarte RTD, Pires MMS, Zárata-Bladés CR, Pinto AR. The impact of *in*
19 *utero* HIV exposure on gut microbiota, inflammation, and microbial translocation. *Gut Microbes*.
20 2019;10(5):599-614.
- 21 30. Evans C, Chasekwa B, Ntozini R, Majo FD, Mutasa K, Tavengwa N, Mutasa B, Mbuya
22 MNN, Smith LE, Stoltzfus RJ, Moulton LH, Humphrey JH, Prendergast AJ, Team SHINEST.
23 Mortality, Human Immunodeficiency Virus (HIV) Transmission, and Growth in Children Exposed
24 to HIV in Rural Zimbabwe. *Clin Infect Dis*. 2021;72(4):586-94.
- 25 31. Vonaesch P, Morien E, Andrianonimiadana L, Sanke H, Mbecko JR, Huus KE,
26 Naharimanananirina T, Gondje BP, Nigatoloum SN, Vondo SS, Kaleb Kandou JE,
27 Randremanana R, Rakotondrainipiana M, Mazel F, Djorie SG, Gody JC, Finlay BB, Rubbo PA,
28 Wegener Parfrey L, Collard JM, Sansonetti PJ, Investigators A. Stunted childhood growth is

- 1 associated with decompartmentalization of the gastrointestinal tract and overgrowth of
2 oropharyngeal taxa. *Proc Natl Acad Sci U S A*. 2018;115(36):E8489-E98.
- 3 32. Dinh DM, Ramadass B, Kattula D, Sarkar R, Braunstein P, Tai A, Wanke CA, Hassoun S,
4 Kane AV, Naumova EN, Kang G, Ward HD. Longitudinal Analysis of the Intestinal Microbiota in
5 Persistently Stunted Young Children in South India. *PLoS One*. 2016;11(5):e0155405.
- 6 33. Gough EK, Stephens DA, Moodie EE, Prendergast AJ, Stoltzfus RJ, Humphrey JH,
7 Manges AR. Linear growth faltering in infants is associated with *Acidaminococcus* sp. and
8 community-level changes in the gut microbiota. *Microbiome*. 2015;3:24.
- 9 34. Ordiz MI, Stephenson K, Agapova S, Wylie KM, Maleta K, Martin J, Trehan I, Tarr PI,
10 Manary MJ. Environmental Enteric Dysfunction and the Fecal Microbiota in Malawian Children.
11 *Am J Trop Med Hyg*. 2017;96(2):473-6.
- 12 35. Humphrey JH, Mbuya MNN, Ntozini R, Moulton LH, Stoltzfus RJ, Tavengwa NV, Mutasa
13 K, Majo F, Mutasa B, Mangwadu G, Chasokela CM, Chigumira A, Chasekwa B, Smith LE, Tielsch
14 JM, Jones AD, Manges AR, Maluccio JA, Prendergast AJ, Team SHINEST. Independent and
15 combined effects of improved water, sanitation, and hygiene, and improved complementary
16 feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. *Lancet Glob
17 Health*. 2019;7(1):e132-e47.
- 18 36. Zhao Q, Hastie T. Causal interpretations of black-box models. *J Bus Econ Stat*.
19 2019;2019.
- 20 37. Apley D, Zhu J. Visualizing the effects of predictor variables in black box supervised
21 learning models. *J R Stat Soc Ser B (Statistical Methodol)*. 2020;82:1059-86.
- 22 38. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris
23 M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J,
24 Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. Human
25 gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-7.
- 26 39. Hughes RL, Arnold CD, Young RR, Ashorn P, Maleta K, Fan YM, Ashorn U, Chaima D,
27 Malamba-Banda C, Kable ME, Dewey KG. Infant gut microbiota characteristics generally do not

- 1 modify effects of lipid-based nutrient supplementation on growth or inflammation: secondary
2 analysis of a randomized controlled trial in Malawi. *Sci Rep.* 2020;10(1):14861.
- 3 40. Raman AS, Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Subramanian S, Kang G,
4 Bessong PO, Lima AAM, Kosek MN, Petri WA, Rodionov DA, Arzamasov AA, Leyn SA, Osterman
5 AL, Huq S, Mostafa I, Islam M, Mahfuz M, Haque R, Ahmed T, Barratt MJ, Gordon JI. A sparse
6 covarying unit that describes healthy and impaired human gut microbiota development. *Science.*
7 2019;365(6449).
- 8 41. Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, Viale AM,
9 Pozueta-Romero J. Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiol*
10 *Rev.* 2010;34(6):952-85.
- 11 42. Sekar K, Linker SM, Nguyen J, Grünhagen A, Stocker R, Sauer U. Bacterial Glycogen
12 Provides Short-Term Benefits in Changing Environments. *Appl Environ Microbiol.* 2020;86(9).
- 13 43. Charbonneau MR, O'Donnell D, Blanton LV, Totten SM, Davis JC, Barratt MJ, Cheng J,
14 Guruge J, Talcott M, Bain JR, Muehlbauer MJ, Ilkayeva O, Wu C, Struckmeyer T, Barile D,
15 Mangani C, Jorgensen J, Fan YM, Maleta K, Dewey KG, Ashorn P, Newgard CB, Lebrilla C, Mills
16 DA, Gordon JI. Sialylated Milk Oligosaccharides Promote Microbiota-Dependent Growth in
17 Models of Infant Undernutrition. *Cell.* 2016;164(5):859-71.
- 18 44. Jonker H, Capelle N, Lanes A, Wen SW, Walker M, Corsi DJ. Maternal folic acid
19 supplementation and infant birthweight in low- and middle-income countries: A systematic review.
20 *Matern Child Nutr.* 2020;16(1):e12895.
- 21 45. Strand TA, Ulak M, Kvestad I, Henjum S, Ulvik A, Shrestha M, Thorne-Lyman AL, Ueland
22 PM, Shrestha PS, Chandyo RK. Maternal and infant vitamin B12 status during infancy predict
23 linear growth at 5 years. *Pediatr Res.* 2018;84(5):611-8.
- 24 46. Strand TA, Taneja S, Kumar T, Manger MS, Refsum H, Yajnik CS, Bhandari N. Vitamin
25 B-12, folic acid, and growth in 6- to 30-month-old children: a randomized controlled trial.
26 *Pediatrics.* 2015;135(4):e918-26.
- 27 47. Strand TA, Ulak M, Hysing M, Ranjitkar S, Kvestad I, Shrestha M, Ueland PM, McCann
28 A, Shrestha PS, Shrestha LS, Chandyo RK. Effects of vitamin B12 supplementation on

- 1 neurodevelopment and growth in Nepalese Infants: A randomized controlled trial. PLoS Med.
2 2020;17(12):e1003430.
- 3 48. Vänni P, Tejesvi MV, Ainonen S, Renko M, Korpela K, Salo J, Paalanen N, Tapiainen T.
4 Delivery mode and perinatal antibiotics influence the predicted metabolic pathways of the gut
5 microbiome. Sci Rep. 2021;11(1):17483.
- 6 49. Wang L, Mu S, Xu X, Shi Z, Shen L. Effects of dietary nucleotide supplementation on
7 growth in infants: a meta-analysis of randomized controlled trials. Eur J Nutr. 2019;58(3):1213-
8 21.
- 9 50. Kamng'ona AW, Young R, Arnold CD, Kortekangas E, Patson N, Jorgensen JM, Prado
10 EL, Chaima D, Malamba C, Ashorn U, Fan YM, Cheung YB, Ashorn P, Maleta K, Dewey KG. The
11 association of gut microbiota characteristics in Malawian infants with growth and inflammation.
12 Sci Rep. 2019;9(1):12893.
- 13 51. Bourke CD, Gough EK, Pimundu G, Shonhai A, Berejena C, Terry L, Baumard L,
14 Choudhry N, Karmali Y, Bwakura-Dangarembizi M, Musiime V, Lutaakome J, Kekitiinwa A,
15 Mutasa K, Szubert AJ, Spyer MJ, Deayton JR, Glass M, Geum HM, Pardieu C, Gibb DM, Klein
16 N, Edens TJ, Walker AS, Manges AR, Prendergast AJ. Cotrimoxazole reduces systemic
17 inflammation in HIV infection by altering the gut microbiome and immune activation. Sci Transl
18 Med. 2019;11(486).
- 19 52. D'Souza AW, Moodley-Govender E, Berla B, Kelkar T, Wang B, Sun X, Daniels B,
20 Coutoudis A, Trehan I, Dantas G. Cotrimoxazole Prophylaxis Increases Resistance Gene
21 Prevalence and α -Diversity but Decreases β -Diversity in the Gut Microbiome of Human
22 Immunodeficiency Virus-Exposed, Uninfected Infants. Clin Infect Dis. 2020;71(11):2858-68.
- 23 53. Prendergast AJ, Chasekwa B, Rukobo S, Govha M, Mutasa K, Ntozini R, Humphrey JH.
24 Intestinal Damage and Inflammatory Biomarkers in Human Immunodeficiency Virus (HIV)-
25 Exposed and HIV-Infected Zimbabwean Infants. J Infect Dis. 2017;216(6):651-61.
- 26 54. Jalbert E, Williamson KM, Kroehl ME, Johnson MJ, Cutland C, Madhi SA, Nunes MC,
27 Weinberg A. HIV-Exposed Uninfected Infants Have Increased Regulatory T Cells That Correlate
28 With Decreased T Cell Function. Front Immunol. 2019;10:595.

- 1 55. Dirajlal-Fargo S, Mussi-Pinhata MM, Weinberg A, Yu Q, Cohen R, Harris DR, Bowman E,
2 Gabriel J, Kulkarni M, Funderburg N, Chakhtoura N, McComsey GA. HIV-exposed-uninfected
3 infants have increased inflammation and monocyte activation. *AIDS*. 2019;33(5):845-53.
- 4 56. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett
5 A, Ghensi P, Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD, Quince C, Huttenhower
6 C, Segata N. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000
7 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*. 2019;176(3):649-
8 62.e20.
- 9 57. Gough EK, Prendergast AJ, Mutasa KE, Stoltzfus RJ, Manges AR, Team SHINEST.
10 Assessing the Intestinal Microbiota in the SHINE Trial. *Clin Infect Dis*. 2015;61 Suppl 7:S738-44.
- 11 58. Humphrey JH, Jones AD, Manges A, Mangwadu G, Maluccio JA, Mbuya MN, Moulton LH,
12 Ntozini R, Prendergast AJ, Stoltzfus RJ, Tielsch JM, Team SHINEST. The Sanitation Hygiene
13 Infant Nutrition Efficacy (SHINE) Trial: Rationale, Design, and Methods. *Clin Infect Dis*. 2015;61
14 Suppl 7:S685-702.
- 15 59. Robertson RC, Church JA, Edens TJ, Mutasa K, Min Geum H, Baharmand I, Gill SK,
16 Ntozini R, Chasekwa B, Carr L, Majo FD, Kirkpatrick BD, Lee B, Moulton LH, Humphrey JH,
17 Prendergast AJ, Manges AR, Team ST. The fecal microbiome and rotavirus vaccine
18 immunogenicity in rural Zimbabwean infants. *Vaccine*. 2021;39(38):5391-400.
- 19 60. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A,
20 Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower
21 C, Franzosa EA, Segata N. Integrating taxonomic, functional, and strain-level profiling of diverse
22 microbial communities with bioBakery 3. *Elife*. 2021;10.
- 23 61. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, Tickle TL, Weingart G,
24 Ren B, Schwager EH, Chatterjee S, Thompson KN, Wilkinson JE, Subramanian A, Lu Y, Waldron
25 L, Paulson JN, Franzosa EA, Bravo HC, Huttenhower C. Multivariable association discovery in
26 population-scale meta-omics studies. *PLoS Comput Biol*. 2021;17(11):e1009442.
- 27 62. Chen T, Guestrin C, editors. XGBoost: A Scalable Tree Boosting System. *Proceedings of*
28 *the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*;
29 2016; New York, NY, USA: ACM.

1 **Acknowledgements:** We thank all the mothers, babies, and their families who
2 participated in the SHINE trial, the leadership and staff of the Ministry of Health and Child
3 Care in Chirumanzu and Shurugwi districts and Midlands Province (especially
4 environmental health, nursing, and nutrition) for their roles in operationalization of the
5 study procedures, the Ministry of Local Government officials in each district who
6 supported and facilitated field operations, Phillipa Rambanepasi and her team for
7 proficient management of all the finances, Virginia Sauramba for management of
8 compliance issues, and the programme officers at the Gates Foundation and the
9 Department for International Development, who enthusiastically worked with us over a
10 long period to make SHINE happen.

11

12 **Funding:**

- 13 - Bill & Melinda Gates Foundation (OPP1021542 and OPP1143707; JHH and AJP),
14 with a subcontract to the University of British Columbia (20R25498; ARM)
- 15 - United Kingdom Department for International Development (DFID/UKAID; JHH
16 and AJP)
- 17 - Wellcome Trust (093768/Z/10/Z, 108065/Z/15/Z; AJP)
- 18 - Swiss Agency for Development and Cooperation (JHH and AJP)
- 19 - US National Institutes of Health (2R01HD060338-06; JHH)
- 20 - UNICEF (PCA-2017-0002; JHH and AJP).

21 The funders had no role in the design of the study and collection, analysis, and
22 interpretation of data and in writing the manuscript.

23

1 **Author contributions:**

2 Conceptualized and study design: ARM, LES, RJS, JHH and AJP.

3 Data and biospecimen collection: KM, RN, BC, FDM, NVT, JT, and BM

4 Biospecimen processing: HMG, IB, SKG, RCR, FF and LC

5 Bioinformatics and machine learning analyses: TJE

6 Data analysis and interpretation: RCR, ARM, TJE, LC, CE and EKG

7 Writing - original draft: RCR, ARM, LC and TJE

8 Writing - reviewing and editing: All authors

9 Supervision and verification of data: ARM, AJP and JHH

10

11 **Competing interests:** TJE was paid a scientific consulting fee in relation to the analysis

12 of the data presented here by Zvitambo Institute for Maternal and Child Health Research.

13 RCR declares remittance from Abbott Nutrition Health Institute (March 2022) and Nutricia

14 (May 2021) for public conference talks outside the submitted work. All other authors

15 declare that they have no competing interests.

16

17 **Data and materials availability:** Raw sequencing reads are deposited in the European

18 Bioinformatics Database under accession number PRJEB51728. Associated metadata

19 from the larger SHINE trial will be uploaded to <http://ClinEpiDB.org> following publication

20 of all primary manuscripts from the SHINE trial. Prior to that time, the data are housed on

21 the ClinEpiDB platform at the Zvitambo Institute for Maternal and Child Health Research

22 and available upon request from Ms. Virginia Sauramba (vsauramba@zvitambo.co.zw).

23

1 **Ethics approvals:** All SHINE mothers provided written informed consent. The Medical
2 Research Council of Zimbabwe (MRCZ/A/1675), Johns Hopkins Bloomberg School of
3 Public Health (JHU IRB # 4205.), and the University of British Columbia (H15-03074)
4 approved the study protocol, including the microbiome analyses. The SHINE trial is
5 registered at ClinicalTrials.gov (NCT01824940).

6

7

8

9

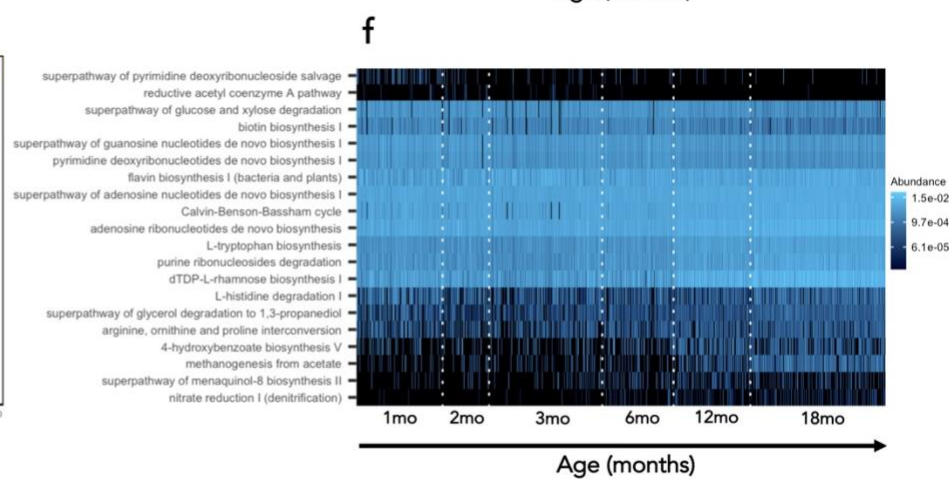
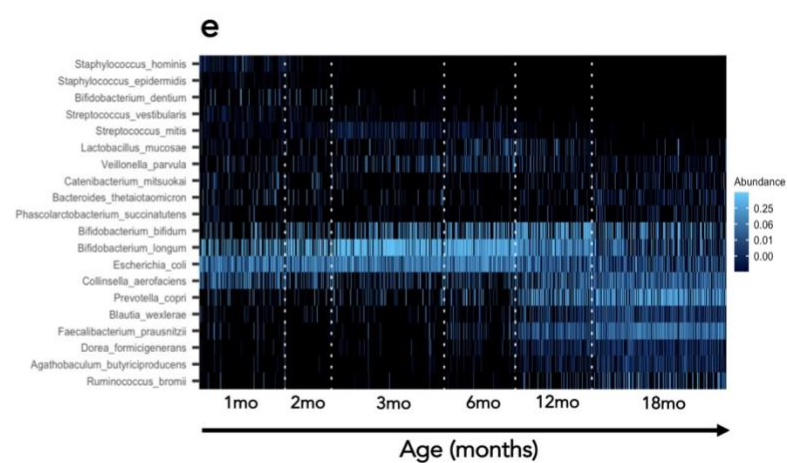
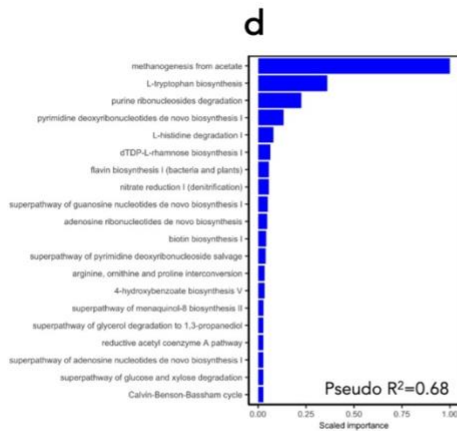
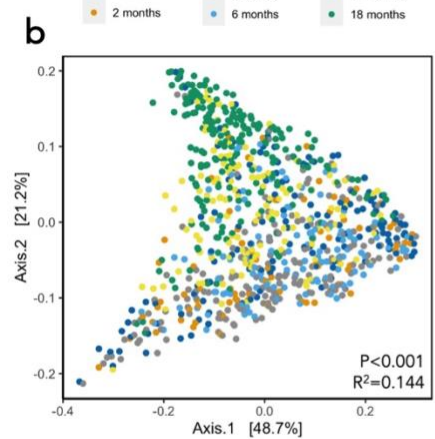
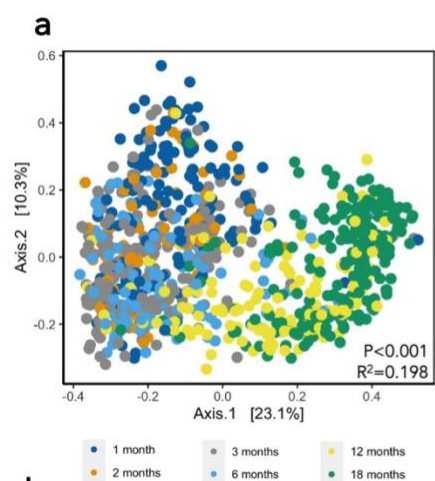
10

11

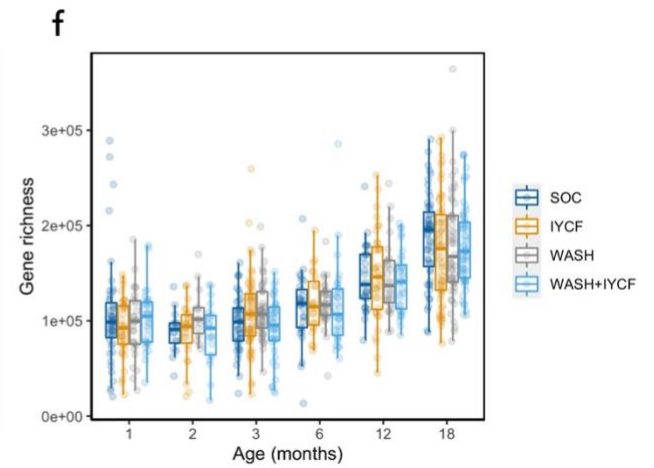
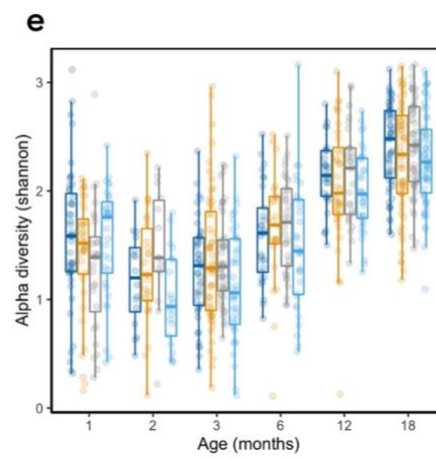
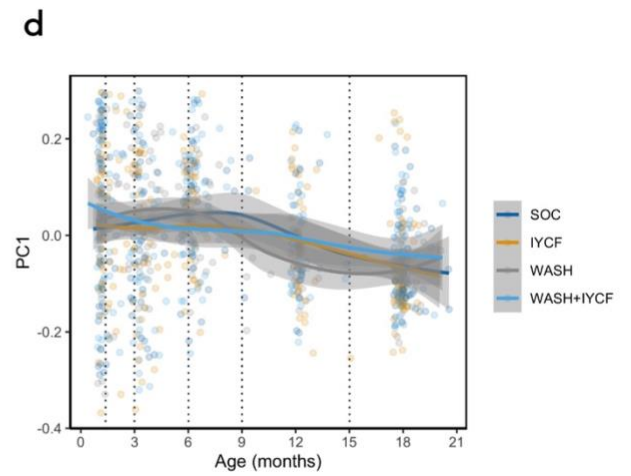
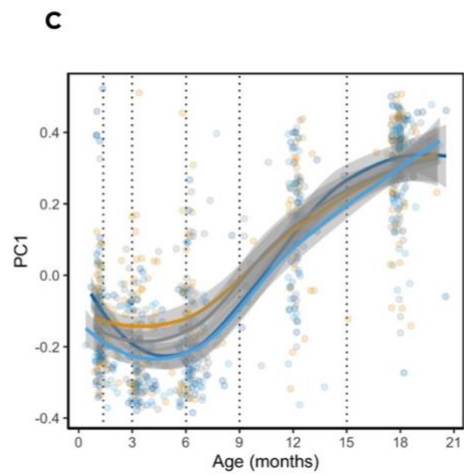
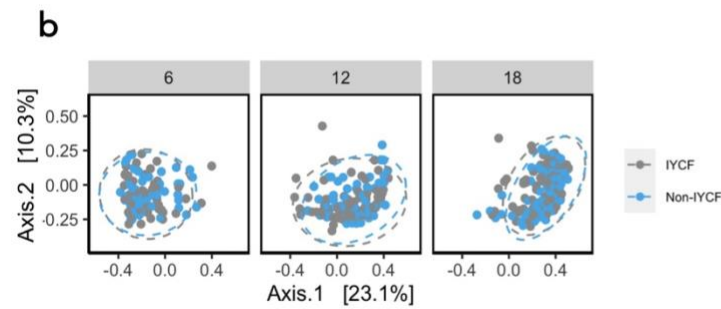
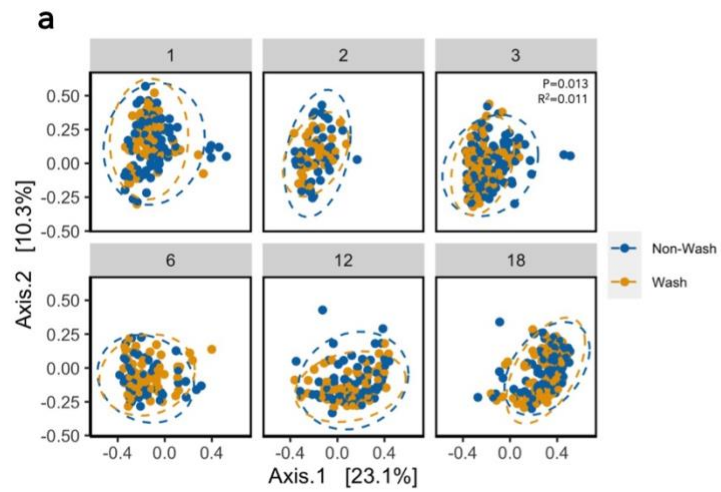
12

13

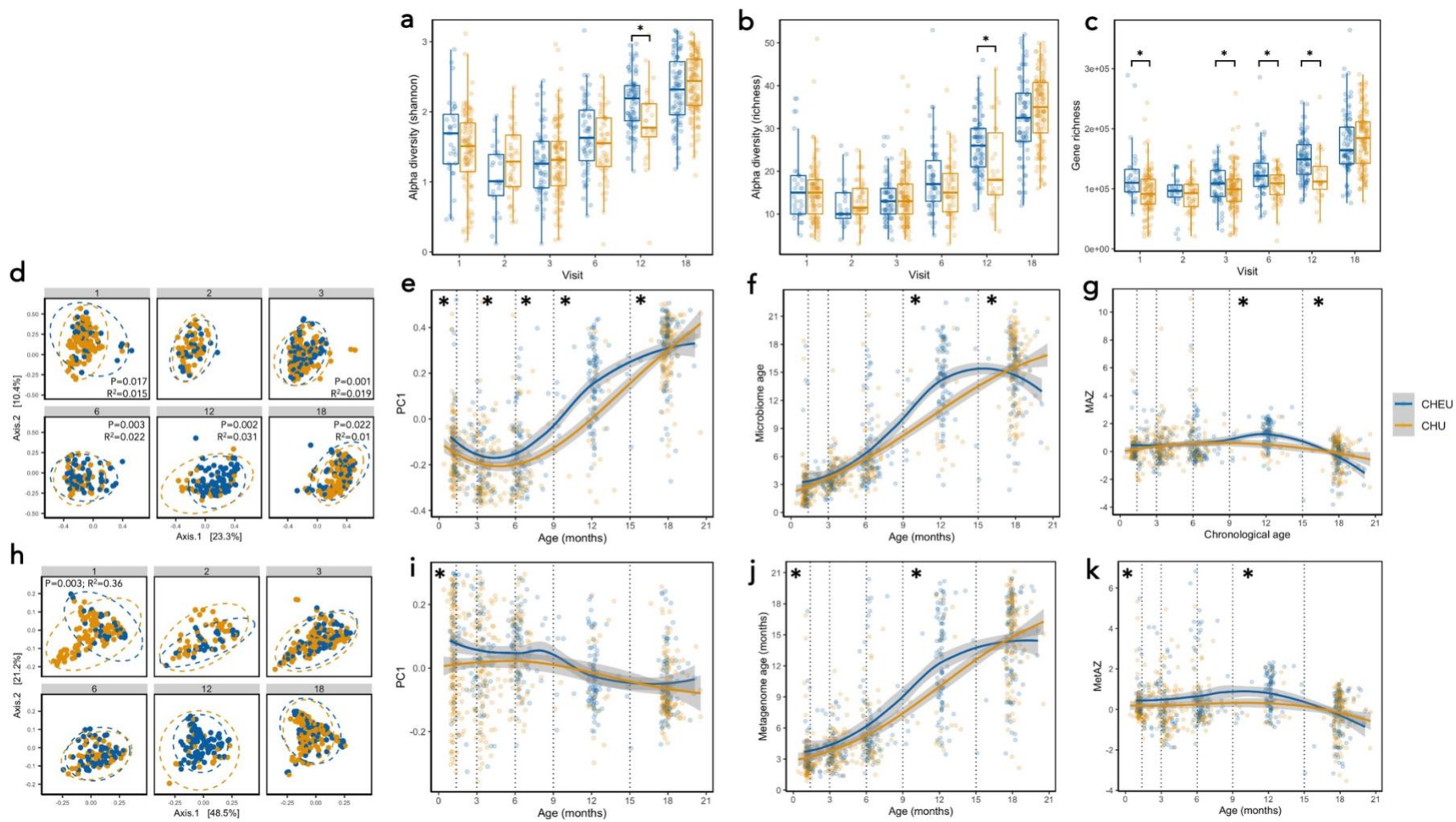
14



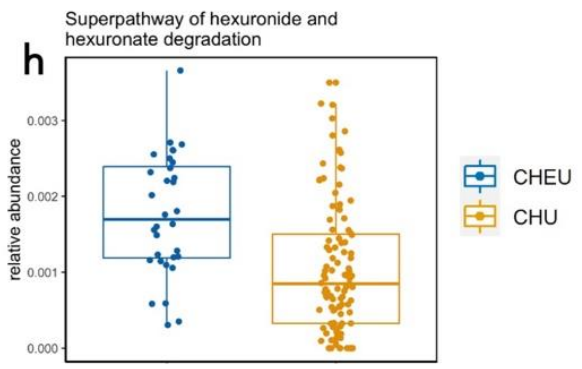
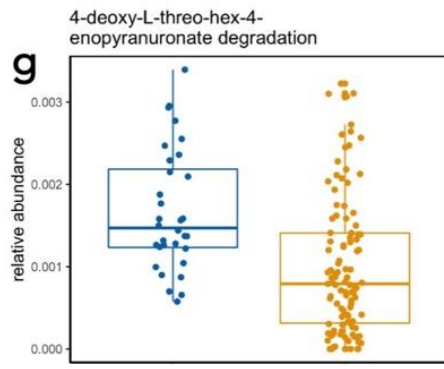
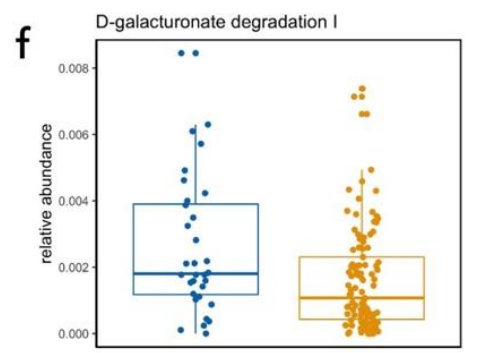
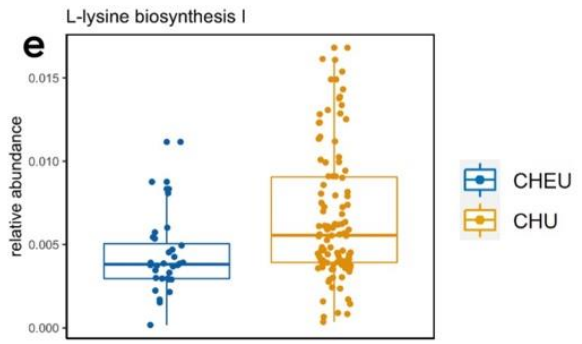
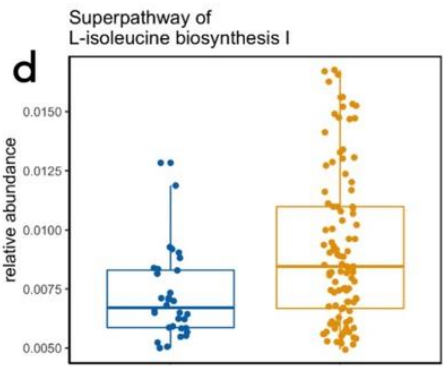
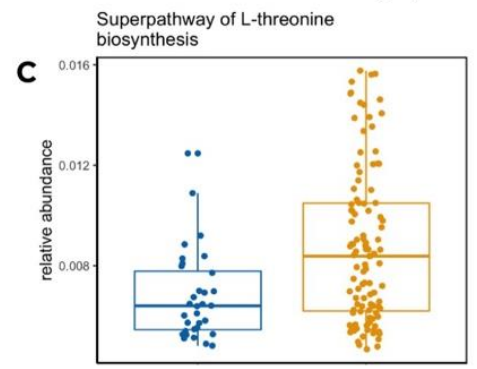
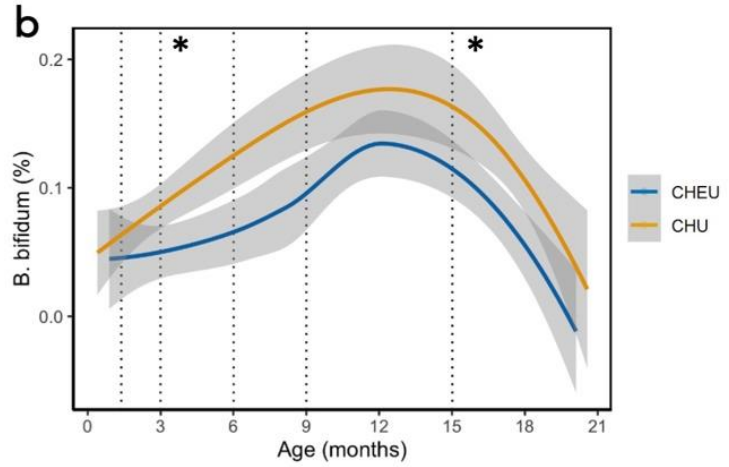
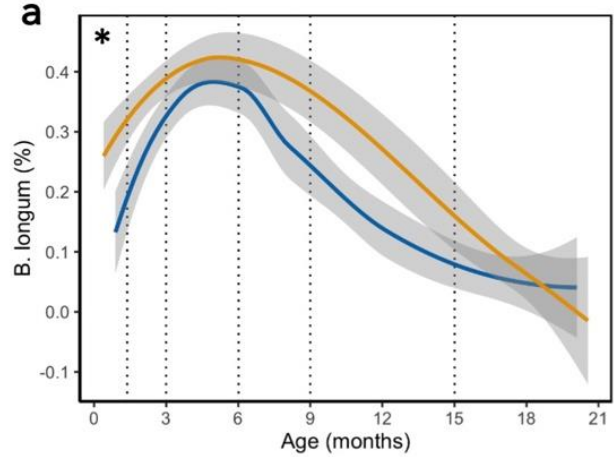
1 **Fig. 1. Compositional and functional maturation of the gut microbiome of 335**
2 **infants from rural Zimbabwe from 1-18 months of age.** PCoA of Bray-Curtis distances
3 of species (a) and metagenomic pathways (b) coloured by age category. PERMANOVA
4 model results are also plotted. The top 20 features and model pseudo-R² from XGBoost
5 models predicting age using species (c) or pathways (d) are ranked by scaled feature
6 importance and relative abundance (0-1) plotted by age (e-f) to visualize taxonomic and
7 functional microbiome succession from 1-18 months of age.



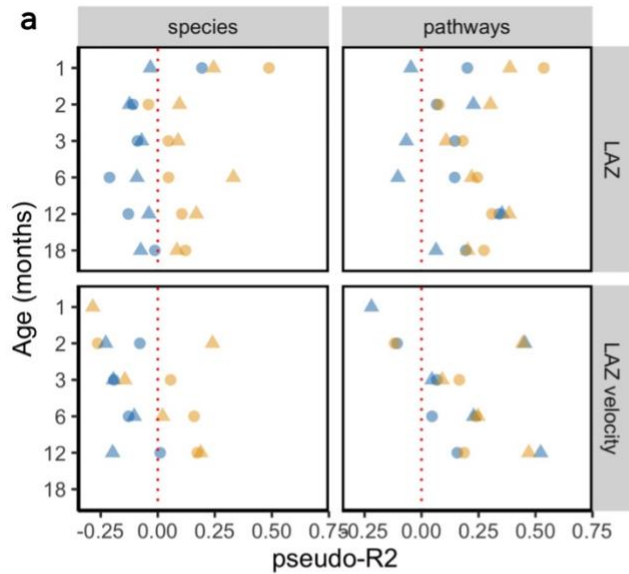
1 **Fig. 2. Impact of randomized WASH and IYCF interventions on infant gut**
2 **microbiome.** PCoA of Bray-Curtis distances species coloured by WASH vs non-WASH
3 arms (a), including PERMANOVA model results, and IYCF vs non-IYCF arms (b) are
4 plotted in addition to the first component (PC1) from PCoA of species (c) and pathways
5 (d). The IYCF intervention was introduced after 6 months of age, therefore direct
6 comparisons of IYCF vs non-IYCF arms are not shown in the 1-, 2- and 3-month age
7 categories. No significant differences were observed in Shannon alpha diversity (e) and
8 gene richness (f) according to trial arm.



1 **Fig. 3. Maternal HIV infection comprehensively alters infant gut microbiome**
2 **diversity and maturity.** Shannon alpha diversity (a), species richness (b) and gene
3 richness (c) shows significant over-diversification in CHEU vs CHU (Wilcoxon rank-sum
4 test; * $p < 0.05$). PCoA of Bray-Curtis distances (d) and PC1 (e) of species composition in
5 CHEU vs CHU show significant differences throughout 18 months of life (PERMANOVA).
6 Microbiome age (f) and microbiome-for-age Z score (MAZ; g) shows significant
7 differences in gut microbiome maturity in CHEU vs CHU (linear regression analyses).
8 PCoA of microbiome gene pathways shows differences in CHEU vs CHU at 1 month of
9 age (h-i) in addition to differences in metagenomic maturity (j-k).



1 **Fig. 4. Maternal HIV infection is associated with reduced abundance of**
2 **Bifidobacteria abundance and amino acid biosynthesis genes.** Relative abundance
3 (0-1) of *Bifidobacterium longum* (a) and *B. bifidum* (b) in the gut microbiome of CHEU and
4 CHU at each age category via multivariate regression analyses. Multivariate regression
5 of gene pathways demonstrates reduced abundance of amino acid biosynthetic pathways
6 (c-e) and increase in abundance of pathways involved in degradation of sugar derivatives
7 (f-h). Multivariate regression using MaAsLin2 using default $q < 0.25$ cut-off. * $q < 0.1$.

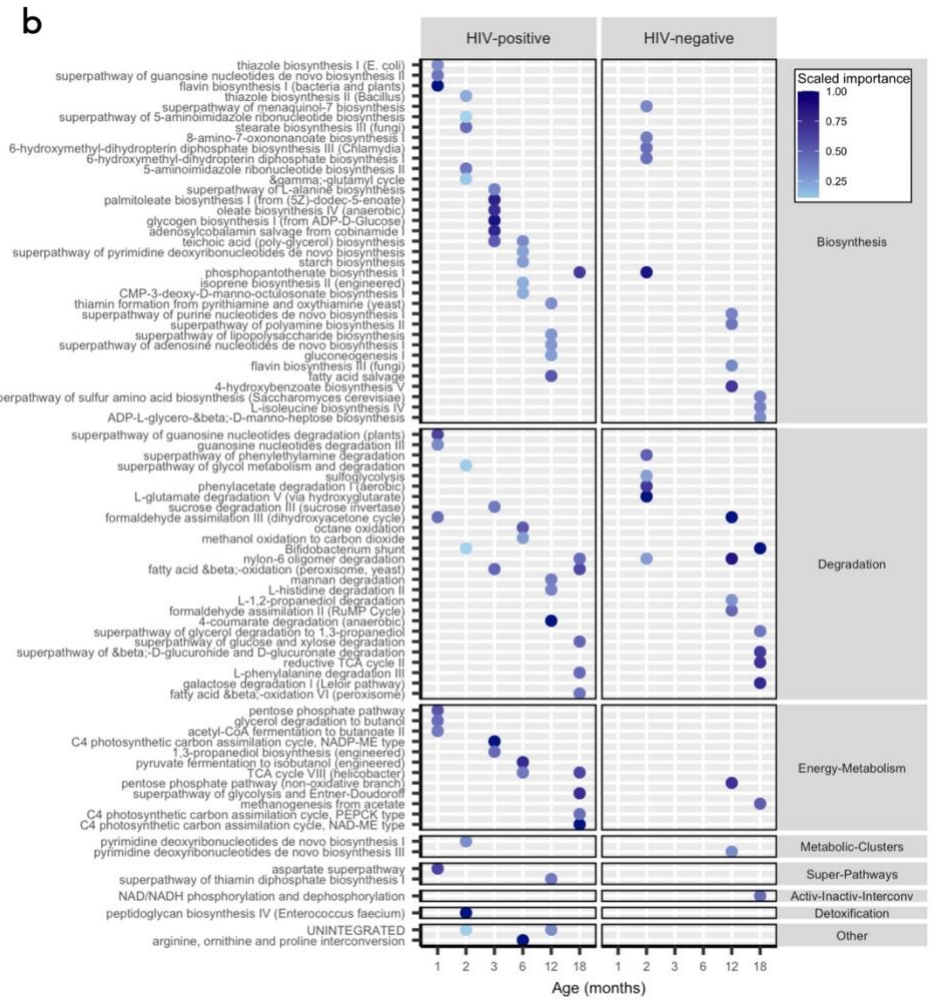
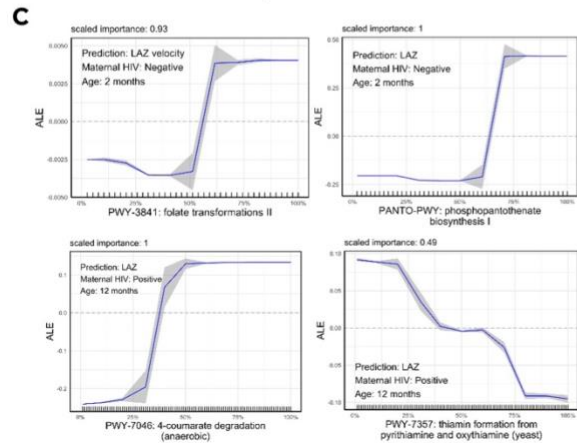


Features

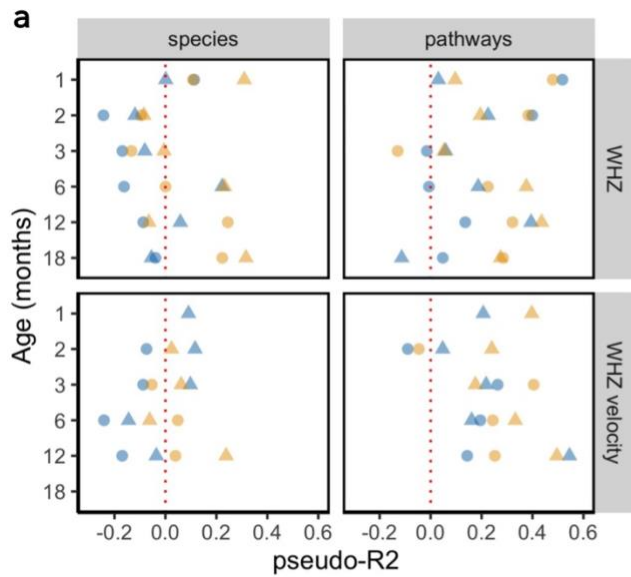
- Microbiome
- Microbiome + epi

Mother HIV status

- Positive
- ▲ Negative



1 **Fig. 5. Prediction of attained LAZ and LAZ velocity using XGBoost models.**
2 Performance of XGBoost models as assessed by pseudo-R² values for prediction of
3 attained LAZ and LAZ velocity (LAZ increase per day to next study visit) using species or
4 metagenomic pathways, stratified by age category and maternal HIV status (a). Models
5 were run using microbiome features alone (species or metagenomic pathways; blue
6 points) and in combination with epidemiological variables (yellow points). The top ranked
7 pathways predicting LAZ at each age category are plotted (b), stratified by maternal HIV
8 status and coloured by scaled importance in the XGBoost model. Accumulated effect
9 plots (ALE) of representative pathways ranking highly in XGBoost model predictions
10 display change in predicted linear growth (LAZ or LAZ velocity) by percentile of the feature
11 abundance distribution. Tick marks on the x-axis are a rug plot of individual feature
12 abundance percentiles. ALEs were generated using the *ALEplot* package and were
13 plotted using *ggplot2*. Standard deviations (sd) were calculated per increment in
14 microbiome feature and were used to calculate and plot increment-wise 95% confidence
15 intervals as the average change in the outcome $\pm 1.96(sd/\sqrt{n})$, where n is the number
16 of observed feature values, and sd is the standard deviation of the change in the outcome
17 variable in an interval.

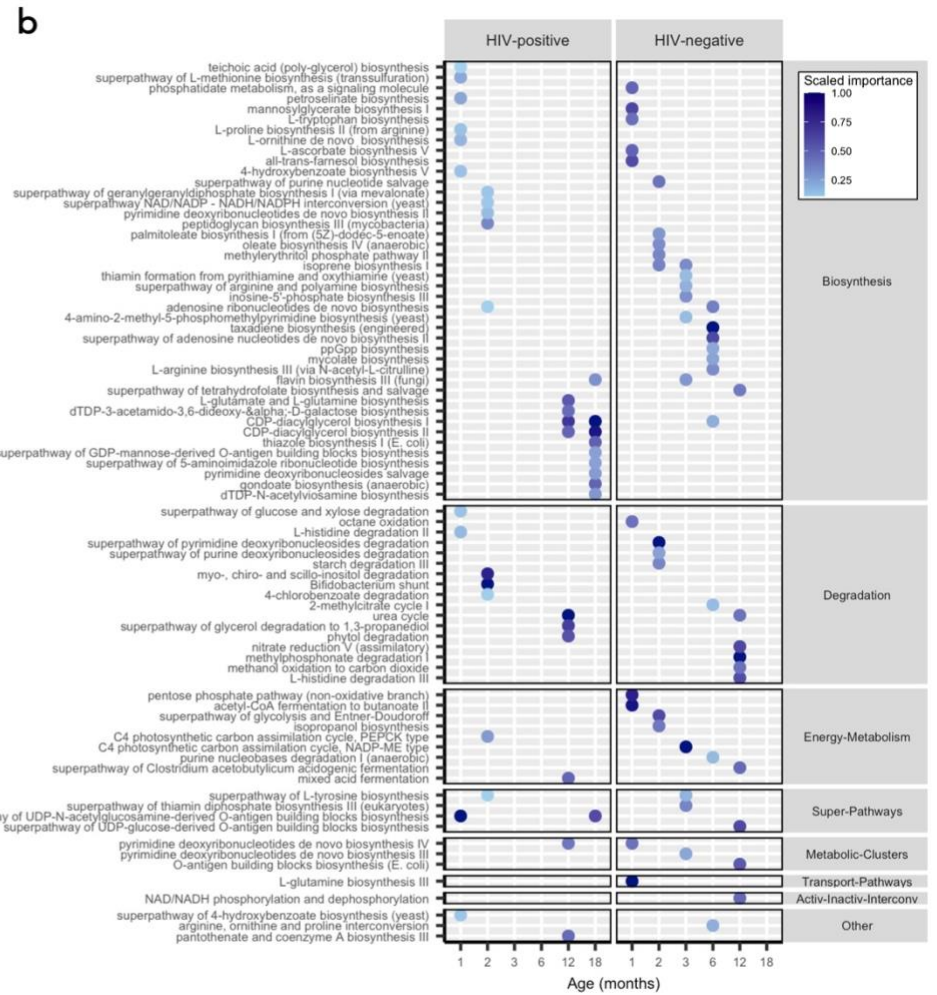
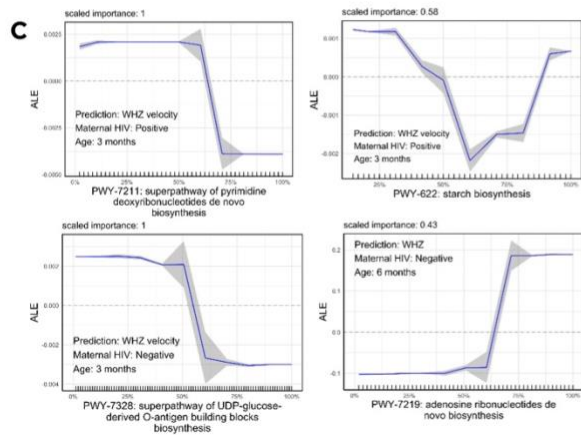


Features

- Microbiome
- Microbiome + epi

Mother HIV status

- Positive
- Negative



1 **Fig. 6. Prediction of attained LAZ and LAZ velocity using XGBoost models.**
2 Performance of XGBoost models as assessed by pseudo-R² values for prediction of
3 attained WHZ and WHZ velocity (WHZ increase per day to next study visit) using species
4 or metagenomic pathways, stratified by age category and maternal HIV status (a). Models
5 were run using microbiome features alone (species or metagenomic pathways; blue
6 points) and in combination with epidemiological variables (yellow points). The top ranked
7 pathways predicting WHZ at each age category are plotted (b), stratified by maternal HIV
8 status and coloured by scaled importance in the XGBoost model. Accumulated effect
9 plots (ALE) of representative pathways ranking highly in XGBoost model predictions
10 display change in predicted linear growth (LAZ or LAZ velocity) by percentile of the feature
11 abundance distribution. Tick marks on the x-axis are a rug plot of individual feature
12 abundance percentiles. ALEs were generated using the *ALEplot* package and were
13 plotted using *ggplot2*. Standard deviations (sd) were calculated per increment in
14 microbiome feature and were used to calculate and plot increment-wise 95% confidence
15 intervals as the average change in the outcome $\pm 1.96(sd/\sqrt{n})$, where n is the number
16 of observed feature values, and sd is the standard deviation of the change in the outcome
17 variable in an interval.

Table S1. Baseline characteristics of infants in SHINE trial and microbiome substudy

	SHINE	Microbiome substudy
	N=4727	N=335
Mothers HIV+	738 (15.6%)	99 (29.6%)
Female	2329 (49.5%)	147 (43.9%)
Birthweight, kg (sd)	3.07 (0.5)	3.09 (0.49)
Low birthweight	410 (9.7%)	27 (8.2%)
Gestational age, weeks (sd)	38.4 (4.03)	39.2 (3.43)
Standard vaginal delivery	4000 (92.5%)	308 (93.9%)
Institutional delivery	3752 (88.2%)	294 (89.9%)
Exclusive breastfeeding	2397 (87.9%)	292 (91.0%)
LAZ - 1 month, (sd)	-0.9 (1.35)	-0.96 (1.25)
LAZ - 3 months, (sd)	-0.91 (1.33)	-0.94 (1.18)
LAZ - 6 months, (sd)	-0.95 (1.25)	-0.91 (1.15)
LAZ - 12 months, (sd)	-1.28 (1.22)	-1.21 (1.05)
LAZ - 18 months, (sd)	-1.58 (1.13)	-1.48 (1.07)
Trial arm		
SOC	1126 (23.8%)	88 (26.3%)
IYCF	1121 (23.7%)	93 (27.8%)
WASH	1201 (25.4%)	71 (21.2%)
WASH+IYCF	1279 (27.1%)	83 (24.8%)
Household size, [IQR]	5.00 [3.00;6.00]	5.00 [4.00;6.00]
Mother age, years (sd)	26.2 (6.66)	28.3 (6.35)
Mother height, cm (sd)	160.1 (5.94)	161 (5.78)
Parity, [IQR]	2.00 [1.00;3.00]	2.00 [1.00;3.00]
Maternal MUAC, cm (sd)	26.4 (3.08)	27.1 (3.23)
Maternal education, years (sd)	9.52 (1.85)	9.39 (1.90)
Mothers employed	383 (8.7%)	30 (9.17%)
Open defecation in household	1952 (44.8%)	154 (47.1%)
Any latrine in household	1743 (39.8%)	128 (40.3%)
Improved latrine in household	1535 (35.1%)	119 (37.4%)
Improved floor in household	2370 (54.2%)	183 (56.5%)
Household owns chickens	3512 (79.0%)	264 (80.5%)
Livestock in household	1691 (36.6%)	146 (44.0%)

Table S2. Detected Eukaryota and Archaea species prior to prevalence filtering

Eukaryota	Archaea
<i>Blastocystis sp subtype 1</i>	<i>Methanobrevibacter oralis</i>
<i>Candida albicans</i>	<i>Methanobrevibacter smithii</i>
<i>Cryptosporidium hominis</i>	<i>Methanobrevibacter woesei</i>
<i>Cryptosporidium meleagridis</i>	<i>Methanosphaera stadtmanae</i>
<i>Cryptosporidium parvum</i>	
<i>Giardia intestinalis</i>	
<i>Saccharomyces cerevisiae</i>	

Table S3. (Auxiliary supplementary file) Multivariate regression analysis

examining the effect of maternal HIV infection, age (days), exclusive breastfeeding status, delivery mode and trial arm on taxonomic microbiome composition

Table S4. (Auxiliary supplementary file) Multivariate regression analysis examining

the effect of maternal HIV infection, age (days), exclusive breastfeeding status, delivery mode and trial arm on gene pathway microbiome composition

Table S5. (Auxiliary supplementary file) Epidemiological variables included in

XGBoost models

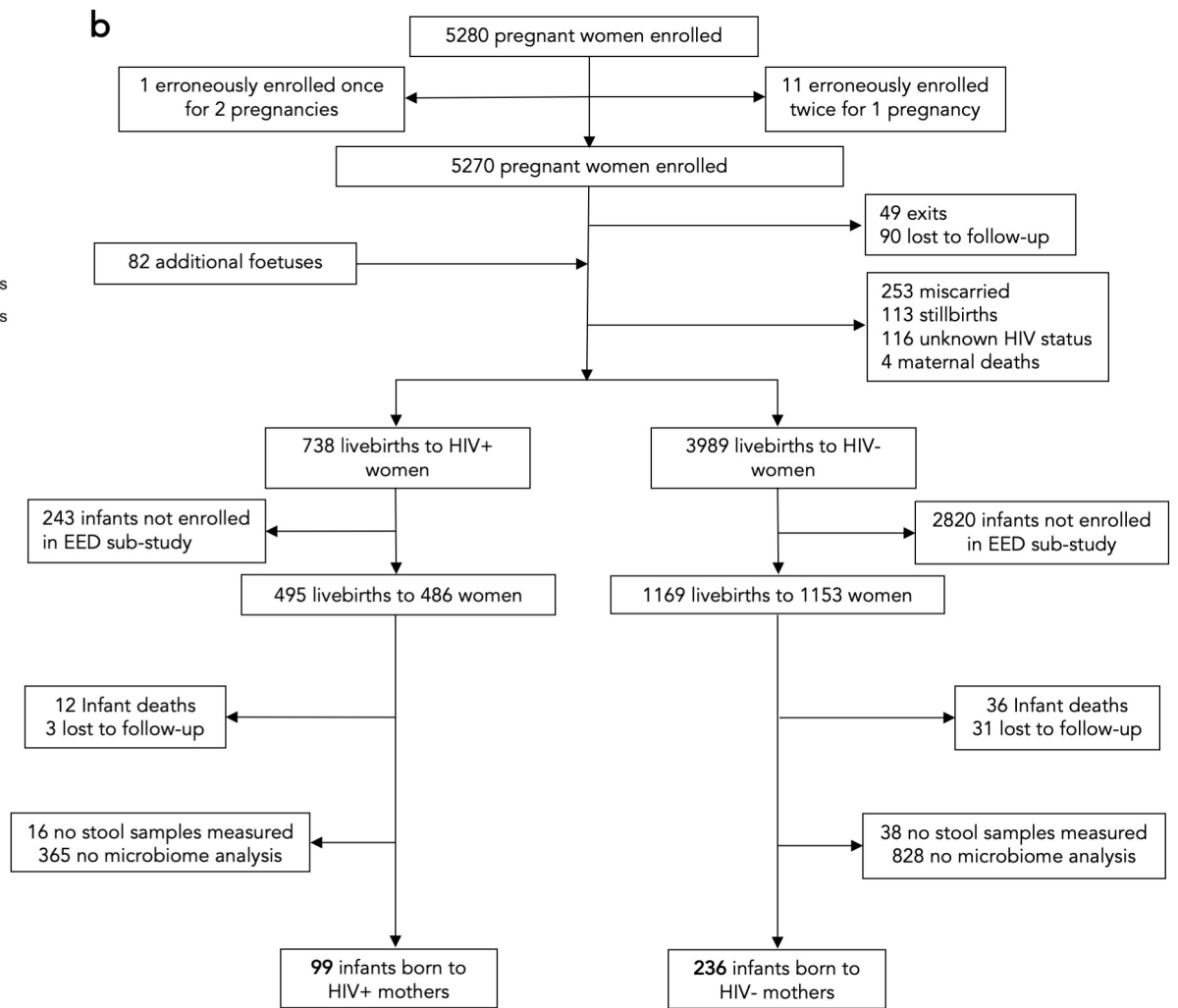
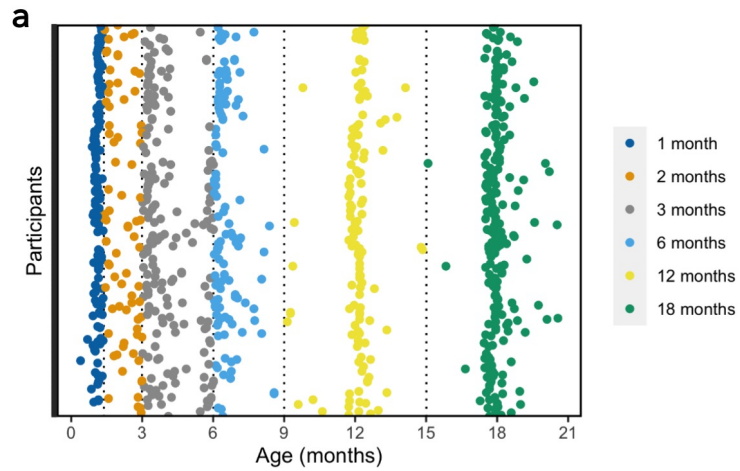


Figure S1. Participants and samples included in the study. 875 stool samples

collected from 335 unique infants underwent whole metagenome shotgun

sequencing and were categorized into 6 age groups (a). CONSORT diagram of the

participants from the SHINE trial included in the infant microbiome sub-study (b).

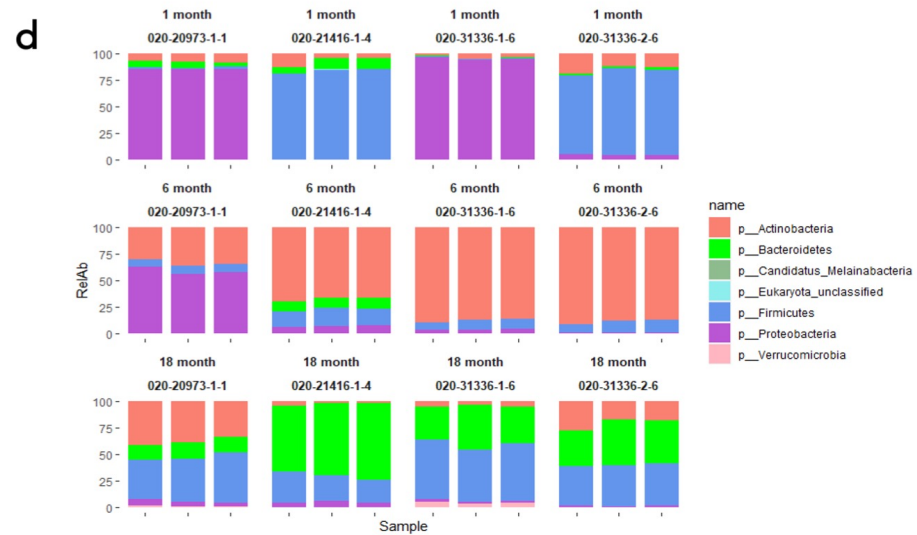
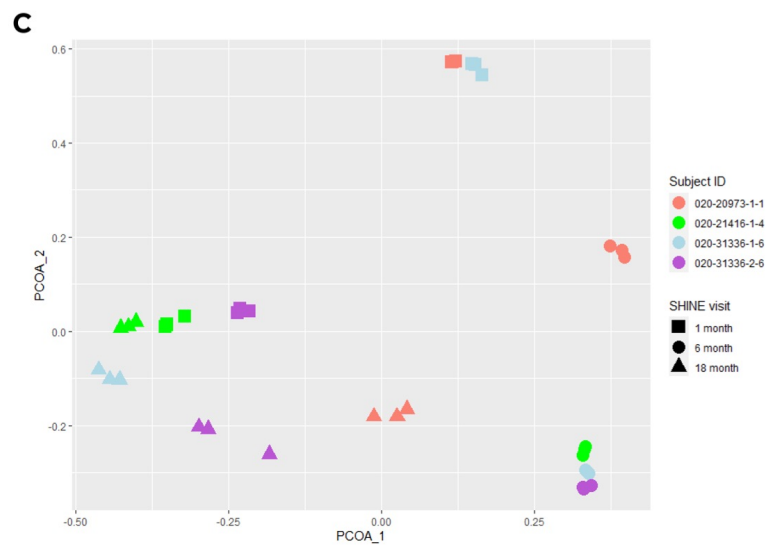
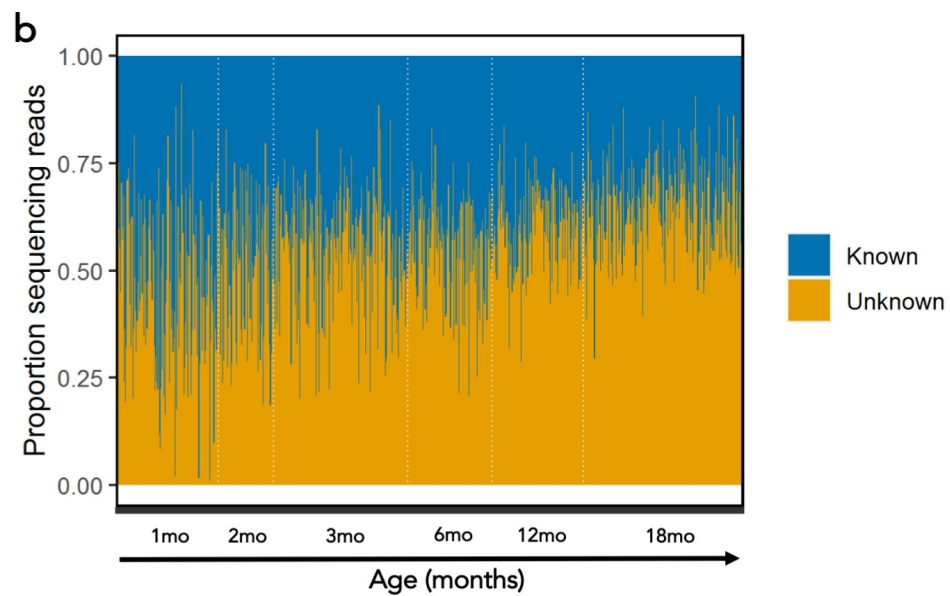
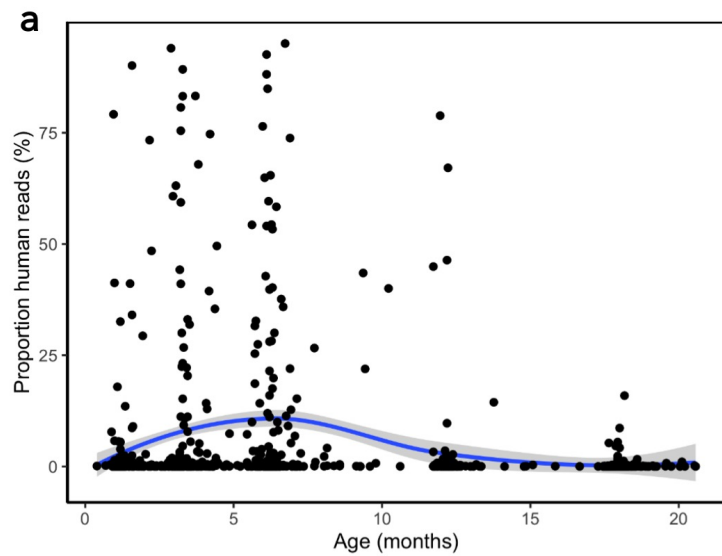


Figure S2. Whole metagenome sequencing performance. A median of 0.05%

sequencing reads were assigned to the human genome in each sample (a), which varied by age at stool sample collection. The percentage of sequencing reads that could be aligned to known sequences using the MetaPhlAn3 and HUMAnN3 pipelines decreased in stool samples collected at older ages (b). PCoA (c) and phylum relative abundances (d) of sequencing sample replicates showed high reproducibility and little technical variation.

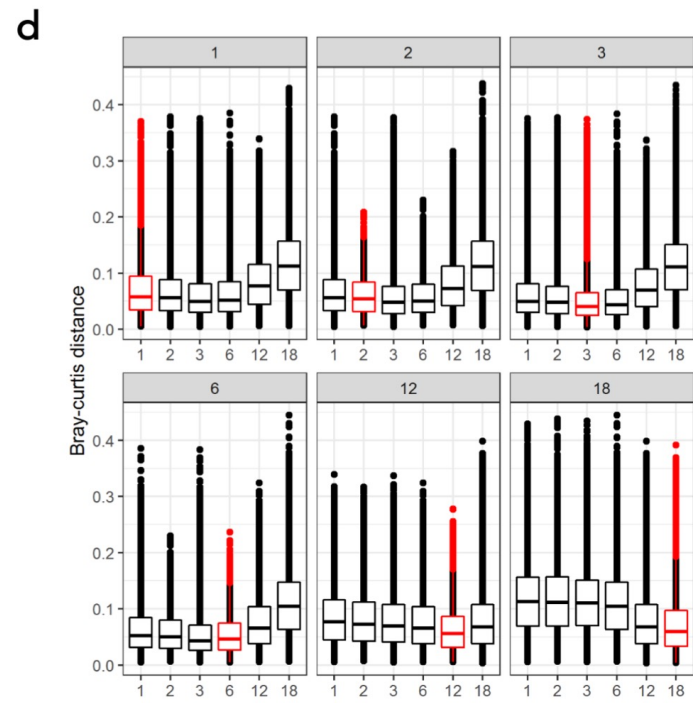
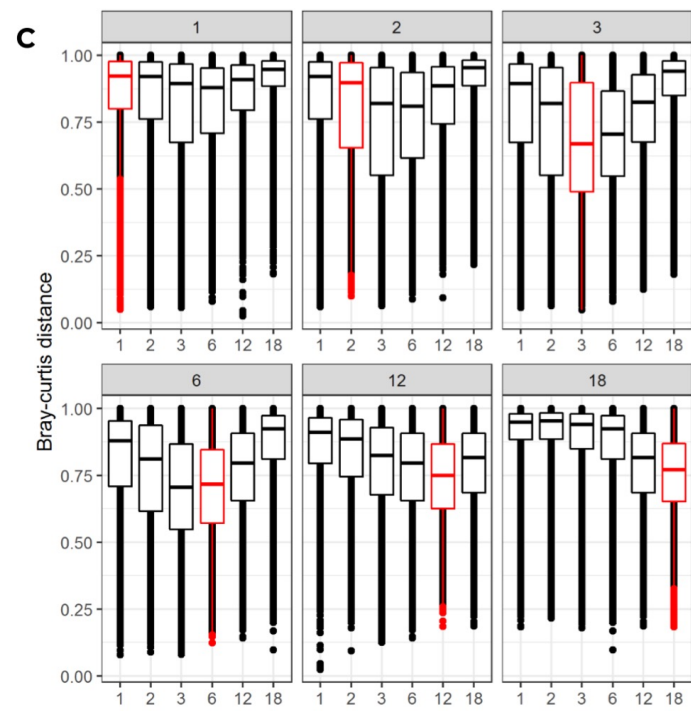
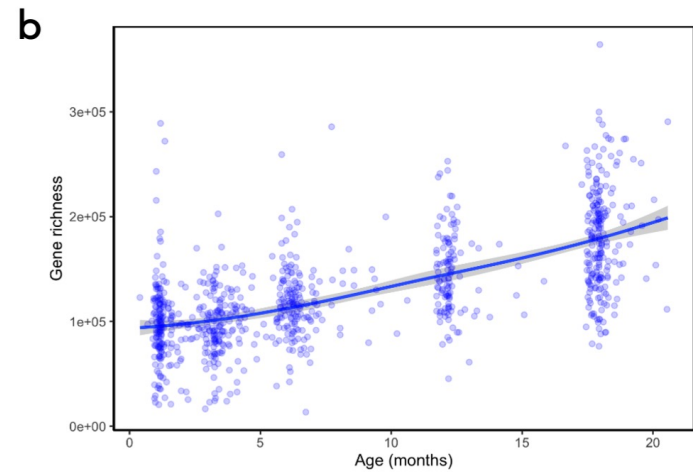
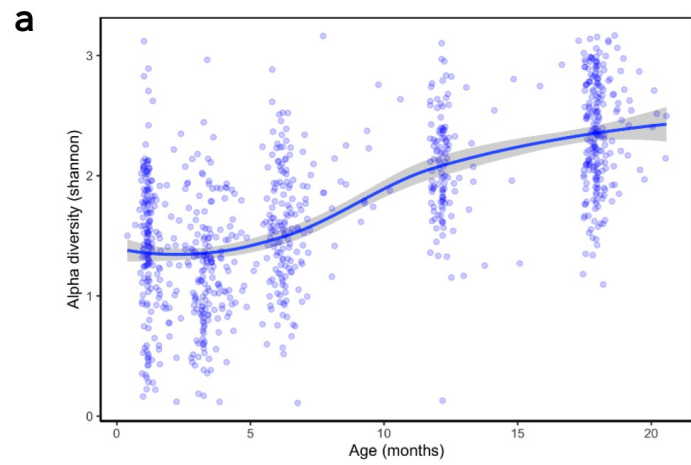
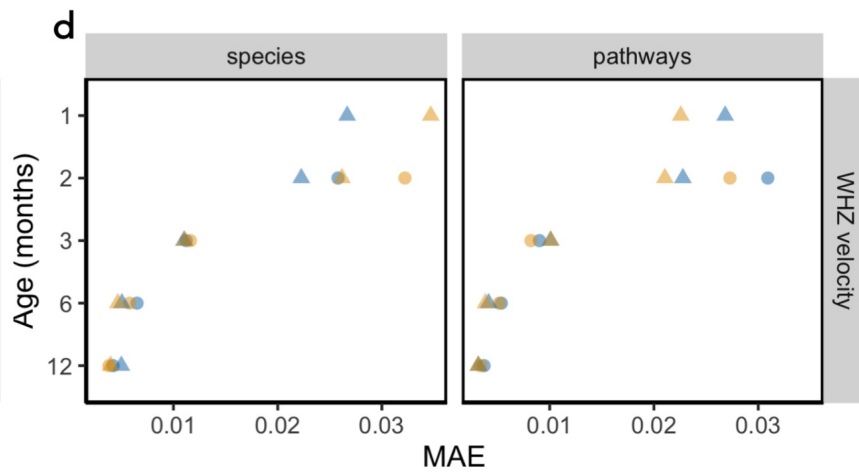
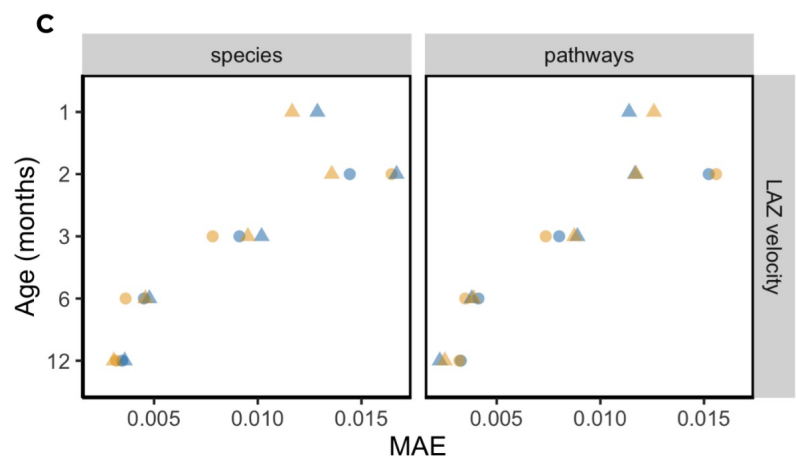
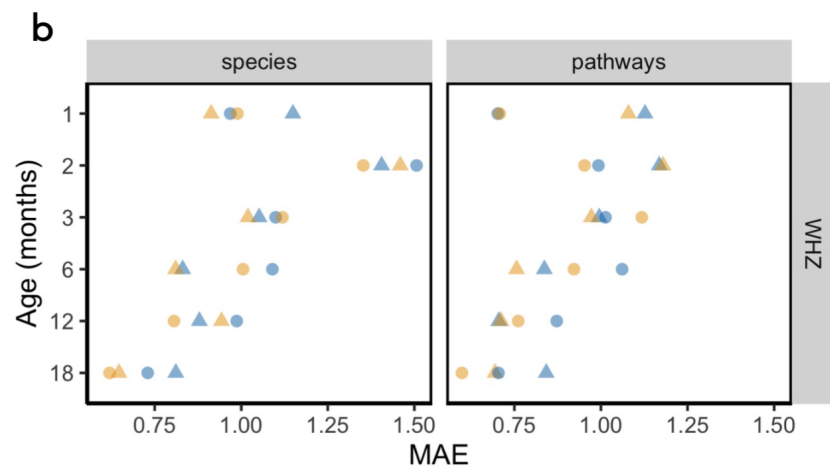
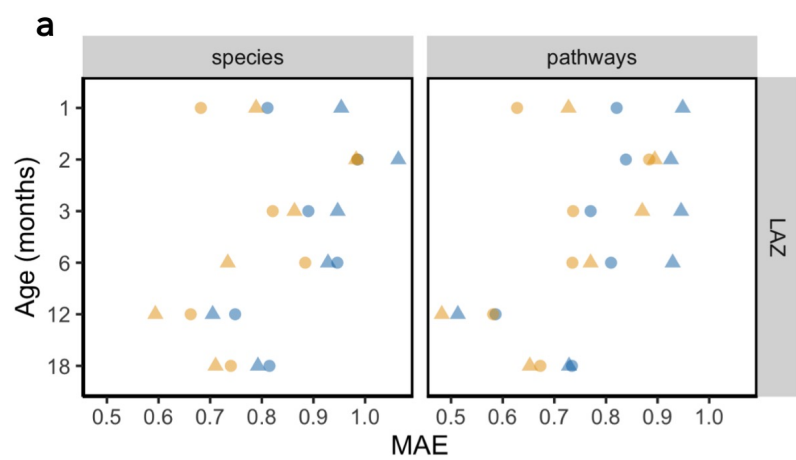


Figure S3. Diversity metrics in entire dataset. Shannon alpha diversity (a) and

gene richness (b) across the entire dataset revealed stable diversity up to 4-5 months of age followed by rapid taxonomic and functional diversification. Bray-curtis distances between samples at within and across each age visit showed low inter-individual variability in species composition which increase with age (c), and high inter-individual variation in metagenome pathways (d).



Features

- Microbiome
- Microbiome + epi

Mother HIV status

- Positive
- ▲ Negative

Figure S4. XGBoost model performance metrics as assessed by mean

absolute error. Mean absolute error (MAE) in XGBoost model performances in models predicting LAZ (a), WHZ (b), LAZ velocity (c) and WHZ velocity (d) stratified by maternal HIV status and age categories. Models were run using microbiome features alone (species or metagenomic pathways) and in combination with epidemiological variables.

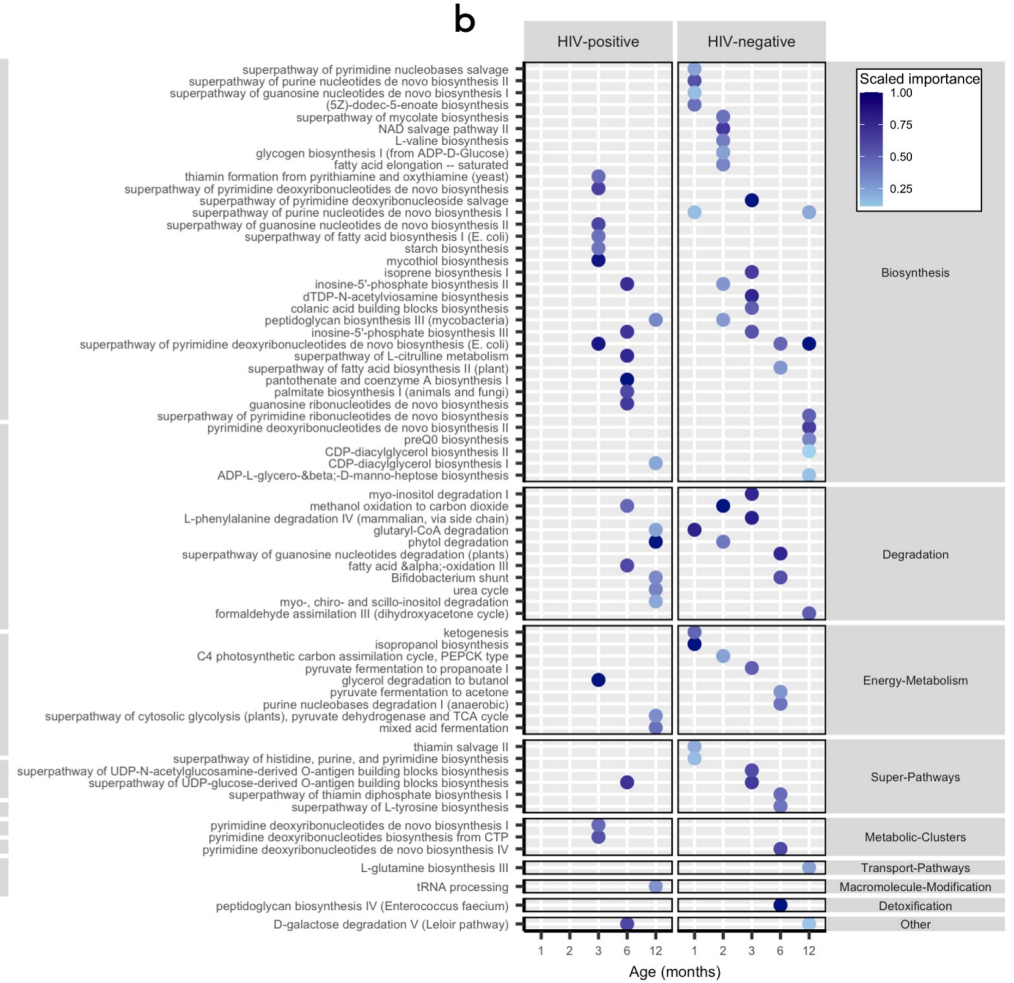
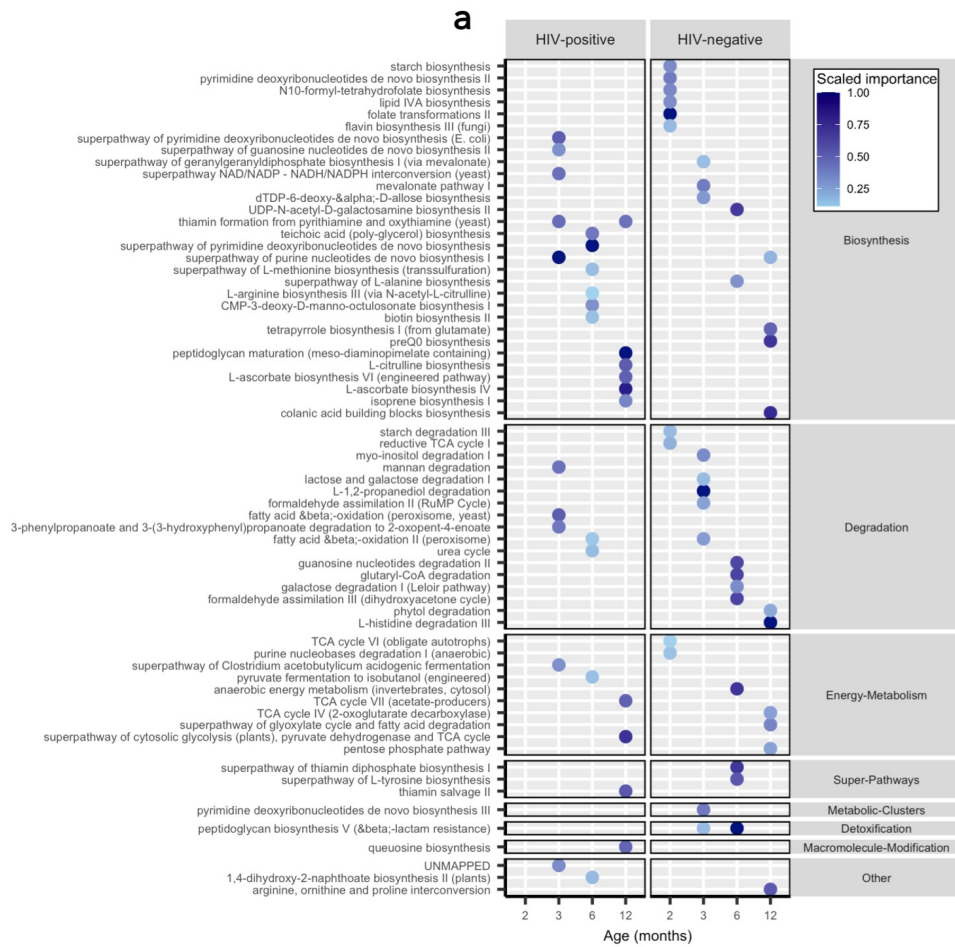


Figure S5. Top ranked pathways in XGBoost models predicting growth

velocity. Top ranked features in XGBoost model predictions of LAZ velocity (a) and WHZ velocity (b) stratified by maternal HIV status. Only features from XGBoost models with pseudo- $R^2 > 0$ are plotted.

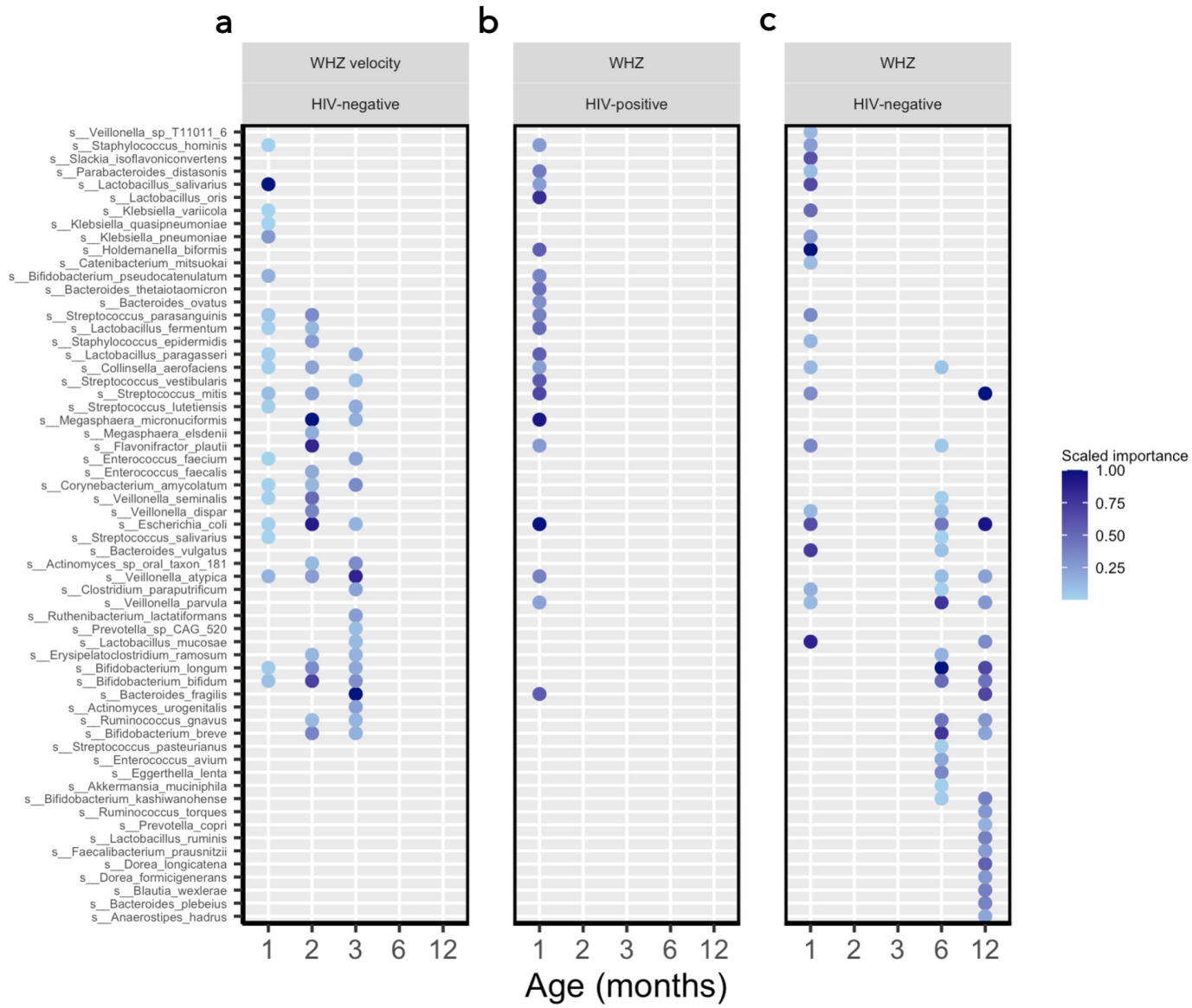


Figure S6. Figure S5. Top ranked species in XGBoost models predicting

growth. Top ranked features in XGBoost model predictions of WHZ velocity (a) and WHZ in children born to HIV+ (b) and HIV- mothers (c). Only features from XGBoost models with pseudo- $R^2 > 0$ are plotted.