Title: *Dietary plant diversity predicts early life microbiome maturation*

Teresa K. McDonald[1], Ammara Aqeel[1], Ben Neubert[1], Anna Bauer[1], Sharon Jiang[1], Olivia Osborne[1], Nolan Ives[1], Danting Jiang[10], Filemon Bucardo[6,8], Lester Gutiérrez[23], Luis Zambrana[7], Kirsten Jenkins[10], Jennifer Gilner[11], Javier Rodriguez[12], Amanda Lai[14], Jonathan P. Smith[16,17], Rinn Song[18], Kazi Ahsan[19,20], Sheraz Ahmed[21], Sanam Iram Soomro[21], Fayaz Umrani[21], Michael Barratt[19,20], Jeffrey I. Gordon[19,20], Asad Ali[21], Najeeha Iqbal[21], Jillian H. Hurst[10,11], Victoria Martin[2,3,4], Wayne Shreffler[2,3,5], Qian Yuan[2,3,4], Joe M. Brown[13], Neeraj K. Surana[1,22,15], Samuel Vilchez[6], Sylvia Becker-Dreps[6,7], Lawrence A David[1,15]

[1] Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, North Carolina, USA

[2]Harvard Medical School, Boston, Massachusetts, USA.

[3] Food Allergy Center, Massachusetts General Hospital, Boston, Massachusetts, USA

[4] Division of Pediatric Gastroenterology and Nutrition, Massachusetts General Hospital for Children, Boston, Massachusetts, USA.

[5] Division of Pediatric Allergy and Immunology, Massachusetts General Hospital for Children, Boston, Massachusetts, USA.

[6]Department of Family Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

[7]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

[8]Department of Microbiology and Immunology, University of North Carolina Chapel Hill, Chapel Hill, North Carolina, USA.

[9]Universidad Nacional Autónoma de Nicaragua León, León, Nicaragua

[10]Department of Pediatrics, Division of Infectious Diseases, Duke University School of Medicine, Durham, North Carolina, USA

[11]Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine, Duke University School of Medicine, Durham, North Carolina, USA

[12]Department of Pediatrics, Children's Clinical Research Unit, Duke University School of Medicine, Durham, North Carolina, USA

[13]Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

[14]The Aquaya Institute, San Anselmo, California, USA.

[15]Duke Microbiome Center, Duke University School of Medicine, Durham, North Carolina, USA

[16]Department of Health Policy and Management, Yale School of Public Health, New Haven, Connecticut, USA.

[17]Division of Global HIV and Tuberculosis, U.S. Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

[18]Oxford Vaccine Group, Department of Pediatrics, University of Oxford, Oxford, UK.

[19]The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63110 USA

[20] Newman Center for Gut Microbiome and Nutrition Research, Washington University School of Medicine, St. Louis, MO 63110 USA

[21] Aga Khan University, Medical College, Pakistan

[22]Department of Pediatrics, Duke University School of Medicine, Durham, North Carolina, USA

[23]Centro de Investigación de Enfermedades Tropicales (CIET). Facultad de Microbiología. Universidad de Costa Rica, San José, Costa Rica.

Summary

Despite established links between the infant gut microbiome and health, how complementary feeding shapes colonization remains unclear. Using FoodSeq, a DNA-based dietary assessment technique, we surveyed food intake across 729 children (0-3 y) from North and Central America, Africa, and Asia. We detected 199 unique plant food sequences, with only eight staples shared across all countries. Despite this global variation, early-life diets followed a common trajectory: weaning stage emerged as the dominant dietary signature across populations. Crucially, dietary diversity did not correlate with gut microbial diversity. Instead, dietary diversity and weaning stage specifically predicted the abundance of adult-associated bacteria. Our findings support a two-stage maturation model: an initial milk-dominated phase, followed by a diet-responsive phase that yields adult-like communities. Monitoring and promoting plant dietary diversity may thus support timely microbiome maturation worldwide.

Introduction

In the first years of life, the human gut transitions from a sterile state to a diverse microbial ecosystem. During this transition, the gut microbiome transforms from an 'immature' state -- characterized by a limited set of milk-adapted taxa -- to a 'mature' adult-like community harboring fiber-degrading taxa and a diversity of other commensal species[1–3]. Proper microbial succession is fundamental to healthy metabolism, immune development, and disease resistance. Indeed, disruptions in early life colonization and microbiome maturation increase risk of conditions such as obesity[4], allergy[5], and Type I diabetes[6]. Consequently, the first years of life represent a critical window when microbiome development can have lifelong health effects.

Despite the importance of the gut microbiome in healthy infant development, the specific factors guiding early microbiome colonization remain less clear. Well-established influences include antibiotic exposure, which reduces bacterial diversity and stability; caesarean delivery, which alters the initial microbial repertoire; and formula feeding, which lowers *Bifidobacterium* abundance yet increases overall microbial diversity[7]. Another pivotal factor is cessation of breastfeeding[3,2,8], which triggers a decline in the abundance of milk-adapted microbial taxa.

One area of microbiome development, that remains underdeveloped, however, is the precise role that complementary feeding plays in microbial succession[1]. While diet is known to reproducibly alter the gut microbiome in adults[9], few studies have found evidence directly linking specific dietary factors to microbiome maturation by concurrently measuring diet and the microbiome in young children. Exceptions have either tested narrow food interventions[10–13] or were conducted within a single cultural and dietary context[14–17]. Other studies have measured microbiome composition across more diverse dietary traditions[8,18]; still such studies have been limited in sample size or have often lacked detailed dietary data, leading comparisons to be on population-level dietary patterns rather than individual-level diet metrics.

A major barrier to understanding the relationship between diet and microbiome maturation is the challenge of accurately characterizing infant diet. Surveys, the gold standard and traditional approach for collecting dietary data, are labor-intensive[56] require baseline levels of literacy[51], incompletely capture minority or ethnic foods[52-54], and are vulnerable to social desirability bias[19,20]. Dietary assessment in children poses additional obstacles, as it relies on secondhand reporting by parents and guardians who may not observe all meals when children are under the supervision of other caregivers. In the United States, for instance, a quarter of 2-5 year olds' energy intake is consumed outside the home[21]. Such incomplete observation contributes to reporting bias exemplified by validation studies with doubly-labeled water that show infant dietary intake is overestimated by 7-29%[22–27]. Finally, the technical challenges with infant dietary assessment only multiply in multi-country settings, as assessments must be adapted for each different geographic, cultural, and linguistic setting[28] For example, there are over 100 versions of a standardized survey for comparing population-level global diets[29]. In concert, these barriers have prevented systematic investigation of the relationship between early life microbiome development and infant diet across diverse human populations.

Remarkable diversity in early-life diets worldwide

To perform a global survey of infant microbiome development and diet, we assembled a meta-cohort from existing longitudinal studies and population health surveillance surveys of 1036 fecal samples from 729 children aged 0-3 years from five countries across multiple continents: the USA (n=144), Nicaragua (n=457), Cambodia (n=178), Kenya (n=228), and Pakistan (n=29). These sites were selected to encompass wide variation in geography, socioeconomic status, and dietary traditions, as well as varying burdens of infectious diseases common in early childhood. Stool samples and dietary information were collected; further details on cohort characteristics and methodologies are provided in the methods and Table S1.

117 We applied FoodSeq, an objective DNA sequencing-based dietary assessment[30], to our
118 cohort. FoodSeq is an amplicon sequencing-based technique that sequences the leucine gene
119 (*trnL*) from plant plastids or the 12S rRNA gene from animal mitochondria. By mapping *trnL* and
120 12S sequencing reads to a curated reference database, FoodSeq identifies and quantifies foods
121 of plant and animal origins at a taxonomic level. Unlike traditional dietary surveys, FoodSeq
122 provides objective measurements that are independent of caregiver reporting and remain
123 consistent across varying income levels, health conditions, and infant feeding practices. The
124 diet data generated by FoodSeq also has the advantage of being temporally synchronized with
125 corresponding microbiome data, as both sets of genomic analyses can be generated from the
126 same fecal sample.

127 FoodSeq revealed extensive variety and breadth in early life diets across our global
128 cohort (Fig. 1). With plant FoodSeq, we detected 113 species and 86 assigned sequence
129 variants (ASVs) that matched at a higher level (e.g. genus), encompassing 389 potential dietary
130 plant species. Of the plant ASVs, 42% were detected exclusively in a single country (Fig. S1A),
131 with unique country-specific foods ranging from longan and lotus root in Cambodia, to jute
132 mallow and cassava in Kenya, to raspberry in the USA (Fig. 1A). By contrast, just eight global
133 staples[31] showed consistent prevalence in all five countries (detected in >5% of each country's
134 samples): wheat, corn, rice, banana/plantain, tomatoes and other nightshades (potato, etc.),
135 mangifera (mangos), and alliums (onion, leek, garlic, chives) (Fig. 1A, foods in bold).

136 Despite the breadth of plant species in our global survey, principal component analysis
137 (PCA) revealed that the primary axis of dietary variation, PC1, was likely driven by the overall
138 presence of plant matter in the diet rather than specific dietary composition (Fig. 2A-B). Unlike
139 subsequent PCs, which displayed mixed positive and negative loadings indicative of tradeoffs in
140 plant species abundance, PC1 exhibited exclusively positive loadings for common foods (Fig.
141 2C). This pattern suggests that PC1 captures the extent of plant consumption rather than
142 regional dietary differences. Supporting this interpretation, PC1 was strongly correlated with

143 both infant age (Fig. 2D, Spearman's ρ =0.25, p<0.001) and overall plant FoodSeq richness
144 (pFR, the number of unique sequences assigned to our human foods database in each sample;
145 Fig. 2E, ρ =0.69, p<0.001). These associations are consistent with the expected trajectory of
146 weaning as infants gradually incorporate more variety of solid foods into their diets as they
147 mature. Direct comparisons of dietary diversity and age confirmed our interpretation, showing
148 that pFR consistently increased across the first year of life until stabilizing (Fig. 2G; Spearman's
149 ρ =0.27, p<0.001). Taken together, these findings argue that FoodSeq effectively captures the
150 dietary shifts associated with weaning.

151 While weaning stage dominated the first PC, the second PC revealed country-specific
152 dietary signatures partitioned by their relative intakes of regional staple plants like rice (enriched
153 in Cambodia), banana/plantain (enriched in Nicaragua), or sorghum and millet (enriched in
154 Kenya). Cross-country differences also emerged in dietary diversity more broadly, with median
155 pFR ranging from 4 in Nicaragua to 10 in Cambodia (Fig. 2H). Such values aligned with
156 previously reported food variety scores, which found mean food variety ranges from 5.5 to 10
157 among children from 6-24 months in Africa and Central America[32–34]. The timing and rate of
158 dietary diversification also varied by country. For instance, the diets of Cambodian children
159 diversified early and rapidly until pFR plateaued at 8 months of age, whereas US infants
160 showed a more gradual rise in pFR until 13 months (Fig. 3B).

161 We note that in contrast to the diversity of plant-associated species and dietary patterns
162 observed in our cohort, we detected a more limited breadth of dietary animals. A total of 28
163 dietary animal species were detected. Observed animals included widely utilized livestock such
164 as pig, chicken, and cow, as well as more region-specific animals including water buffalo, the
165 primary source of milk in Pakistan, and various fish species in Cambodia. Overall, however,
166 41% of infant samples had no non-human animal DNA detected, and only 33% had more than
167 one animal species detected. Only the US and Cambodia had a non-zero median count of

168 animal species present (Fig. S1B). The limited detection of dietary animals was consistent with
169 the higher cost of animal-based foods and the majority of human animal-based foods originating
170 from cattle, poultry, and pigs[28] for the locations of infants sampled here. Given the limited
171 detected range of animal species, the distinct nutritional characteristics of different animal
172 products from the same species (beef, dairy), as well as the known importance of nutrients like
173 fiber for the gut microbiome, we focused our subsequent analyses on plant-based foods and gut
174 microbiome development.
175
176 **Patterns of microbiome maturation**
177 Prior to analyzing the relationship between early-life dietary and microbiome
178 development, we first confirmed that our cohort followed well-established patterns of early-life
179 gut microbiome maturation[3]. Microbiome alpha-diversity increased steadily from birth (Fig. 3A)
180 through age two, regardless of country (Fig. 3B). In terms of inter-individual variation (beta-
181 diversity), both country of origin (Fig. 3C) and age (Fig. 3D) were significant factors (country
182 R^2=0.4675, p=0.001, age R^2=0.4721, p=0.001; EnvFit on Unweighted Unifrac. Fig. S2A). In
183 cohorts where data were available on breastfeeding status (USA, Nicaragua, Cambodia) and
184 birth mode (USA, Nicaragua), both factors were associated with microbiome composition.
185 (Breastfeeding: R^2= 0.1377, p=0.001; Birth mode: R^2= 0.0088, p=0.014, Envfit on unweighted
186 Unifrac, controlling for age and country). Thus, consistent with previous literature, we observed
187 both increased diversification of the microbiome with age, as well as notable roles for
188 breastfeeding status and birth mode in shaping microbiome structure in our cohort[35].
189 The specific bacterial taxa observed in our infant cohort also aligned with prior studies[1–
190 3,8], as hierarchical clustering revealed a microbial succession pattern (Fig. S3A). We identified
191 an early-life cluster (Cluster 9), enriched in *Bifidobacterium* and *Streptococcus,* and a
192 'transitional' cluster (Cluster 1), abundant at ages 12-18 months, characterized by plant
193 degraders, *Blautia* and *Ruminococcus,* as well as the probiotic genus *Ligilactobacillus* (Fig. 4A).

194 Following this transition, a late-microbiome cluster (Cluster 5) emerged between 21-36 months
195 that more closely resembled the adult microbiome, featuring *Bacteroides vulgatus* and
196 *Faecalibacterium prausnitzii,* which ferment plant fiber into butyrate and other short chain fatty
197 acids, and *Prevotella copri,* a core fiber-metabolizing taxon[36] (Fig. 4A). In parallel, we
198 constructed a Random Forest model that successfully predicted infant age given input
199 microbiome data. The RF model identified *Faecalibacterium* and *Bifidobacterium* as top
200 predictor variables, mirroring both our hierarchical clustering results and previously published
201 models of infant microbiome development[1] (Fig.S4).
202
203 **Dietary diversity does not predict microbiome diversity**
204 Having found that infant microbiota in our cohort matured as expected from previous
205 studies, we next compared patterns of dietary maturation to gut microbiome maturation.
206 Surprisingly, we did not see consistent evidence that dietary diversity predicted gut microbial
207 diversity. Microbial alpha diversity continued to increase until at least 14-16 months of age (Fig.
208 3A) well after dietary diversity (as measured by pFR) had plateaued (Fig. 3B; 2G). We did
209 observe associations between dietary diversity and microbiome diversity within several
210 countries but associations were not consistent; for example, pFR correlated with microbiome
211 Shannon diversity in Kenya and the United States, whereas in Nicaragua, the relationship was
212 reversed (Fig. S2C). Further, when we constructed global models controlling for age and
213 country, pFR did not correlate wtih overall microbiome diversity measures (Shannon diversity
214 and 16S richness both non-significant (Fig. S2D, F). Thus, although children began eating
215 complex diets soon after introduction of solid foods, the gut microbiome continued to diversify
216 even after dietary complexity had been reached.
217
218 **Diet is associated with later stage microbes during gut colonization**

219 We next hypothesized that although dietary diversity may not directly correlate with
220 overall microbial diversification, plant richness and weaning stage would be associated with
221 microbiome maturity. Indeed, despite substantial geographic variation in diet, we found pFR was
222 positively correlated with the ‘Transitional Microbiome’ cluster, comprised of *Blautia* and
223 *Ruminococcus*, and the ‘Late Microbiome’ cluster, comprised of *Bacteroides*, *Faecalibacterium*,
224 and *Prevotella*, even after controlling for age and country (Fig. 4C, D). Likewise, PC1 was
225 positively correlated with 'Late Microbiome' cluster abundance (Fig. 4E).
226 Given that we sourced fecal samples across multiple study sites, we re-fitted all models
227 while (i) adjusting for diarrhea, weight-for-height z-score, RUTF intake, and sequencing depth,
228 and (ii) adding participation in the Cambodian nutrition-education intervention. None of these
229 adjustments altered the direction or significance of the associations between plant richness
230 (pFR) and the Transitional or Late microbiome clusters, nor did they create any association with
231 Shannon diversity (Fig. S5,6).
232 Ultimately, as increased dietary diversity (i.e., pFR) and weaning stage (i.e., PC1) both
233 reflect complementary feeding and the introduction of an adult-like diet, our findings support the
234 hypothesis that introduction of an adult-like diet stimulates maturation toward an adult-like
235 microbiome. Notably, the 'Early Microbiome' cluster did not correlate with dietary diversity (Fig.
236 4B) or PC1, consistent with the hypothesis that milk consumption, rather than solid food
237 introduction, primarily shapes gut microbial ecology prior to weaning. These results suggest that
238 the transition to an adult-like microbiome specifically is driven by the cumulative introduction of
239 diverse solid foods, but that other transitions within the microbiome maturation process, such as
240 overall microbiome diversification, are not driven by solid food diversity. These transitions are
241 furthermore unlikely to be driven by consumption of particular foods or specific dietary patterns,
242 as we observed conserved microbiome microbial responses to diversity across countries with
243 highly unique and non-overlapping diets. More broadly, our findings point to a developmental
244 sequence wherein milk consumption governs early microbiome composition, while the

245 subsequent introduction of solid plant foodsdrives the transition toward a mature gut microbial
246 community.
247
248

249 **Discussion**

250 In concert, our findings here refine existing models linking weaning to microbiome
251 development in the infant gut. Our data do not reveal simple linear relationships between dietary
252 and microbiome diversity across early life. This negative finding, however, is consistent with our
253 increasing understanding that microbial ecosystem diversity does not necessarily predict
254 community function[37,38]. Instead, our data support a two-stage developmental model: an early
255 phase governed by milk consumption, followed by a maturation phase shaped by dietary
256 diversity. In this second stage, plant dietary diversity and the child's physiologic age together
257 predict the colonization by select groups of gut bacterial taxa associated with adult-like
258 microbiome function. Notably, despite the broad range of complementary feeding patterns
259 across our geographically diverse cohorts, we observed similar successional trends,
260 underscoring the functional redundancy[39] that gut microbes can share, regardless of
261 environmental differences.

262 Early life nutrition remains one of the most critical challenges in global health, with
263 lasting consequences for child development and lifelong health outcomes[40]. Our findings
264 confirm that adequate and diverse plant consumption during the complementary feeding period
265 foster maturation of the gut microbiome toward an adult-like state rich in fiber-degrading
266 bacteria, extending the known benefits of dietary diversity beyond micronutrient adequacy[41,42]
267 and food allergy prevention[43]. Importantly, we show that simply counting the number of distinct
268 plant foods effectively measures this microbial transition, potentially offering a more accessible
269 and practical complement to widely used dietary indices such as the WHO minimum dietary

diversity score (MDD), which assesses diversity as consumption across at least five of eight food groups.[44] Implementing straightforward, 'pro-gut' complementary feeding guidance based on overall plant dietary richness could enable easier and more precise monitoring by caregivers and health workers across diverse settings, ultimately supporting appropriate microbiome maturation among infants worldwide.

Materials and Methods

Study populations and stool sampling

Cambodia

Cambodian samples were obtained as extracted DNA from stool samples collected as part of the Integrated Nutrition, Hygiene and Sanitation Project (NOURISH), conducted across three rural provinces[45]. The NOURISH study involved both nutrition and sanitation interventions. The nutrition intervention specifically consisted of caregiver nutrition education, direct cash payments contingent upon regular health visit attendance, and provision of two $5 food vouchers. Of the Cambodian samples in our study, 101 (56.7%) originated from children enrolled in either of the nutrition intervention arms (arm 1: nutrition intervention only; arm 2: nutrition intervention+sanitation). Our selection prioritized samples from children who had not experienced diarrhea or fever within the week preceding sample collection; however, 14 samples came from children who had recently experienced diarrhea, 156 from children without recent diarrhea, and five lacking diarrhea data. Additionally, seven samples came from children who had received RUTF or a local alternative, independent of the intervention protocol. The cohort overall exhibited a high burden of enteric pathogens, although most pathogen-positive children were asymptomatic at the time of sample collection.

Kenya

295 Kenyan samples were obtained as extracted DNA from stool samples collected in Kisumu,
296 Kenya as part of the TOTO study[46]. The TOTO cohort comprised children generally
297 representative of the broader Kenyan pediatric population. Children were enrolled in the TOTO
298 study when presenting for clinical care with symptoms of respiratory illness raising clinical
299 suspicion for tuberculosis infection. Full study enrollment criteria and clinical characteristics of
300 the original TOTO study have been previously published[46]. Of the 228 Kenyan samples included
301 in our analysis, 70 (30.8%) reported receiving nutritional supplmentation with RUTF or a local
302 alternative as part of clinical care, independent of microbiome sampling protocol.
303
304 *Nicaragua*
305 Nicaraguan samples were obtained as stool slurries (stored 1:1 in phosphate-buffered saline,
306 PBS) from the *Sapovirus*-Associated Gastro-Enteritis (SAGE) study[47], which enrolled children at
307 risk of diarrheal disease in León, Nicarauga. Full enrollment criteria and clinical characteristics
308 of the original SAGE study have been previously published[47]. In addition to the previously
309 described control (n=33) and diarrheal (n=32) samples, we obtained n=391 additional stool
310 samples collected for monthly surveillance from study participants. Non-diarrheal surveillance
311 samples were specifically selected from children who had not experienced a diarrhea episode
312 within the preceeding month. Only surveillance samples from children aged 2-12 months were
313 included in our 16S microbiome analysis.
314
315 *Pakistan*
316 The Pakistan samples were obtained as extracted DNA from fecal samples collected as part of
317 the Study of Environmental Enteropathy (SEEM), a prospective longitudinal cohort study,
318 conducted in Matiari, a rural area in Sindh, Pakistan, between March 2016 and March 2019.
319 The SEEM study collected samples from 50 healthy control children and longitudinal samples
320 from malnourished children (WHZ< -2), who were part of a malnourishment intervention arm.

321 For our analysis, we included stool samples (n=29) exclusively from the healthy control group
322 (WHZ > 0), none of whom received nutritional interventions. Approximately 1g of stool was
323 collected from each participant by community health workers within 10 minutes of defecation.
324 Samples were immediately transported from field sites in 4°C cooler boxes, stored at -80°C
325 upon arrival at the laboratory, and shipped on dry ice for subsequent analysis.
326
327 *United States*
328 US samples consisted of frozen solid stool or stool slurry obtained from participants enrolled in
329 the Project HOPE 1000 study[48] (Durham, NC, USA) and Gastrointestinal Microbiome & Allergic
330 Proctocolitis Study (GMAP) study (Boston, MA, USA)[49]. Project HOPE 1000 aims to develop a
331 biorepository of longitudinal samples collected from healthy infants over the first 1000 days of
332 life, with the goal of identifying early life determinants of health. The GMAP study prospectively
333 collected stool samples from children, some of whom subsequently developed allergies and/or
334 proctocolitis; however, GMAP samples analyzed in this study were exclusively from infants who
335 did not develop allergic disease. Breastfeeding status was documented at the time of each
336 sample collection for Project HOPE 1000. For GMAP, breastfeeding status was recorded at
337 birth and 4 months of age, but not at the exact time of stool sample collection. Therefore, infants
338 exclusively receiving formula at 4 months of age in GMAP were presumed to not be breastfed at
339 the time of stool collection.
340
341 **DNA Extraction**
342 *Cambodia:* Stool samples were preserved 1:1 in Zymo DNA/RNA Shield buffer (Zymo
343 Research, Irvine, CA) and then stored at -20°C. They were then extracted using the QIAamp 96
344 Virus QIAcube HT (Qiagen, Germany) protocol. 200mg of solid stool or 200uL of liquid stool
345 were combined with 1mL of stool lysis buffer (ASL, Qiagen, Germany) and vortexed in an SK38

346 soil grinding tube (Bertin Corp., Rockville, MD) for five minutes (Vortex Genie 2, Scientific
347 Industries, Bohemia, NY). Following a ten minute incubation at room temperature, the lysate
348 was centrifuged at 12,000 g for two minutes (Thermo Fisher Scientific, Waltham, MA). 200uL of
349 this lysate was then used for DNA/RNA extraction with the QIAamp 96 Virus QIAcube HT
350 protocol.
351
352 *Kenya:* 1g of stool was diluted in 10mL of sterile water and emulsified with sterile glass beads.
353 The stool slurry was then filtered. DNA was extracted using the MagAttract PowerSoil DNA EP
354 kit (Qiagen, Germany) according to the manufacturer's instructions.
355
356 *Pakistan*: 50mg of frozen fecal sample was extracted using phenol-chloroform extraction.
357 Briefly, samples were ground in liquid nitrogen using a mortar and pestle. 50mg of this was then
358 transferred to a 2mL vial. 500uL of 0.1mm diameter zirconia/silica beads were added to each
359 sample, along with 500 µL of phenol: chloroform: isoamyl alcohol (25:24:1), 210 µL of 20% SDS
360 and 500 µL of buffer (200 mM NaCl, 200 mM Trizma base, 20 mM EDTA). Samples underwent
361 bead beating (Mini-Beadbeater-8; Biospec) to extract DNA. Extracted DNA was purified,
362 (Qiaquick columns, Qiagen), then eluted in 70 µL of Tris-EDTA (TE) buffer, and then quantified
363 (Quant-iT dsDNA broad range kit; Invitrogen). Each DNA sample was adjusted to a
364 concentration of 1 ng/uL.
365
366 *Nicaragua and USA:*
367 The Nicaraguan and USA (Durham, Boston) stool samples were diluted 1:10 in filtered
368 phosphate buffered saline (0.2uM filter) and then stored at -80°C. Samples were thawed on ice
369 in an anaerobic chamber. Samples were vortexed 30s and then 250uL aliquots were taken for
370 DNA extraction. Slurry aliquots were then randomized and extracted using the Power Soil Pro
371 MagAttract kit (Qiagen, Germany) according to manufacturer's instructions.

372

373 **DNA Sequencing**

374 *16S rRNA Sequencing*

375 All samples were PCR-amplified using primers targeted to the V4 region of the 16S rRNA gene
376 using primers 515F and 806R as previously described.[50] Kenyan samples were quantified using
377 a Qubit dsDNA HS assay kit (ThermoFisher), pooled in equimolar concentration, and
378 sequenced on a MiSeq sequencer (Illumina; 2x250 base pair). Cambodian, Nicaraguan,
379 Pakistan, and USA samples were processed as previously described[50–52], diluted to a 10nM
380 concentration, and sequenced using an Illumina MiniSeq and a MiniSeq High Output kit
381 (FC420-1004) with paired-end 150-bp reads.

382

383 *FoodSeq*

384 FoodSeq was carried out as described in previously published protocols.[53] For plant FoodSeq,
385 samples were amplified in a 10uL volume with 3 µL of extracted DNA template and either a) 0.5
386 µL each of 10 µM forward and reverse primers (IDT), 2 µL of 5X KAPA HiFi buffer, 0.3 µL of 10
387 mM dNTPs, 0.1 µL of 100X SYBR Green I (Life Technologies), 0.1 µL KAPA HiFi polymerase,
388 3.5 µL nuclease-free water, or b) 5uL of KAPA HiFi 2X, 0.3uL each of 10 µM forward and
389 reverse primers (IDT), 0.1uL of 100X SYBR Green I (Life Technologies) and 1.3uL of nuclease-
390 free water. Method (b) uses the same components as method (a), but in a premixed format with
391 a slightly higher concentration of polymerase. Cycling conditions were 95°C for 3 minutes,
392 followed by 35 cycles of 98°C for 20 seconds, 63°C for 15 seconds, and 72°C for 15 seconds.

393

394 For animal FoodSeq, samples were amplified in a 10uL volume with 1uL of extracted DNA
395 template and 0.5uL of 10uM forward and reverse primer and 1uL of 100uM human blocking
396 primer (IDT), 0.1 uL of 100X SYBR Green, 0.25uL of 20mg/mL bovine serum albumin, and 5uL

397 of 2X Accustart. The amplification primers were 12SV5F and 12SV5R[54] with Illumina overhang
398 adapter sequences added at the 5′ end, and the human blocking primer was DeBarba14
399 HomoB[55]. Cycling conditions were an initial denaturation at 94°C for 3 minutes, followed by 35
400 cycles of 94°C for 20 seconds, 57°C for 15 seconds, and 72°C for 1 minute.
401
402 For both plant and animal FoodSeq, primary amplification was followed by secondary PCR
403 amplification to add Illumina adapters and dual 8-bp indices for sample multiplexing. Samples
404 were amplified in a 50uL volume using 5uL primary PCR product diluted 1:100 in nuclease free
405 water, 25uL KAPA HiFi 2X Buffer, 0.5uL 100X SYBR Green I, 9.5uL nuclease free water, 5uL
406 each of 2.5uM forward and reverse indexing primers.
407
408 Amplicons were cleaned (Ampure XP, Beckman Coulter, Brea, CA), quantified (QuantIT dsDNA
409 assay kit, Invitrogen, Waltham, MA), and pooled in equimolar ratios to create a sequencing pool.
410 If samples were too low in concentration to be pooled at a equimolar ratio, they were added up
411 to a set volume of 15-20uL. The pooled library was then concentrated, gel purified, quantified by
412 both fluorimeter and either qPCR or BioAnalyzer (Agilent, Santa Clara, CA). If trnL and 12SV5
413 amplicons were to be sequenced on the same run, they were combined at this point in
414 equimolar amounts. Final sequencing library was spiked with 30% PhiX (Illumina, San Diego,
415 CA) to mitigate low nucleotide diversity. Libraries were sequenced on an Illumina MiniSeq
416 according to manufacturer's instructions using a 150-cycle Mid or High kit, (Illumina, San Diego,
417 CA, USA), depending on the number of samples in each pool.
418
419 **Data analysis**
420 All packages used are available as a Singularity container hosted by the Duke Gitlabs
421 (https://gitlab.oit.duke.edu/lad-lab).
422 *16S rRNA*

423 All code used in bioinformatic pipeline to process 16S rRNA amplicon data can be found on
424 Github at tjk30/16S-pipeline. To summarize, following conversion to fastq format using Illumina's
425 bcl2fastq package, DNA sequencing results were demultiplexed using qiime2 with the --p-no-
426 golay-error-correction flag. Trimmomatic was used to remove primer sequences. Dada2
427 (v1.22.0) was then used to filter, trim, merge, and assign species to reads using the SILVA
428 database v138.1. Samples with <5000 total reads were excluded from analysis.
429
430 Richness was calculated as the sum of unique ASVs that had reads>0. Shannon diversity was
431 calculated using estimate_richness() from the *phyloseq* package (v.1.48.0) in R (v.4.1).
432
433 Unweighted Unifrac was performed using the SILVA v138.1 SSU phylogenetic tree and the
434 Unifrac() function from the *phyloseq* package. Ordination of sample distances was then
435 performed using cmdscale() from the R package *stats* (v4.4.1) with the flag eig=TRUE. Envfit
436 was run using the envfit() function from the R package *vegan (*v2.6.6.1). For calculation of the
437 variation captured by age and country, both variables were included in the model. For the
438 breastfeeding and birth mode models, each model was run controlling for country and age.
439
440 For microbiome clustering analysis, we excluded samples from children known to be
441 experiencing diarrhea at the time of stool sample collection (n=33/934) or acute malnutrition
442 (WHZ <-2), as these conditions are known to significantly impact gut microbiome
443 composition[56],[1]. Diarrhea data were available only from the Cambodian and Nicaraguan
444 cohorts. Age was binned as follows: 0-2mo, 3-5mo, 6-8mo, 8-10mo, 11-13mo, 14-16mo,17-
445 19mo, 20-22mo, 23-25mo, and 26+mo (binned together due to limited sampling >24mo in
446 Cambodia and USA). Clustering was limited to 0-36 month old children to ensure that each age
447 bin consisted of children from multiple countries, as only Kenya provided samples from children
448 older than 36 months. Taxa were filtered to those with at least 0.5% relative abundance in at

449 least one age bin. Rare taxa, defined as taxa detected in two or fewer samples, were removed
450 from analyses.
451
452 The samples that passed filtering criteria (n=818) were then used to calculate the
453 average relative abundance of each ASV within each age bin. In order to normalize across taxa
454 with different relative abundance ranges across age, relative abundances for each taxon were
455 converted to z-scores. Clustering analysis was performed on the Euclidean distances of the
456 taxa z-scores using Unweighted Pair Group Method with Arithmetic Mean (UPGMA;
457 implemented with the hclust() function from the R package *stats*, method='average'). Clustering
458 analysis was then used to generate an k=10 clusters using the *stats* function cutree().
459
460 After identifying microbiome clusters in the subset of healthy children, we reincorporated the
461 excluded "sick" samples back into the dataset for subsequent analyses, as we wanted to
462 comprehensively evaluate the identified microbial clusters within the context of broader
463 childhood conditions and maximize statistical power. We calculated the cluster abundances for
464 Clusters 1-10 for all samples in the full dataset (n=934) by merging the taxa identified within
465 each cluster and then performing centered log ratio (clr) normalization using the R package
466 microbiome (v1.26.0).
467
468 *FoodSeq*
469 All code and databases used in the bioinformatic pipeline to process FoodSeq amplicon data
470 can be found on Github at LAD-lab/mb-pipeline. Bioinformatic processing was conducted as
471 previously described[30]. To summarize, following conversion to fastq format and demultiplexing
472 using Illumina's bcl2fastq package, primer sequences were trimmed using the qiime2 cutadapt
473 wrapper with the following flags: --p-error-rate 0.15, --p-minimum-length 1 , --p-overlap 5, p-
474 discard-untrimmed. Paired reads were then merged using the qiime2 dada2 wrapper with the

475 following flags: --p-trunc-len-f 0, --p-trunc-len-r 0, --p-max-ee-f 2, --p-max-ee-r 2, --p-trunc-q 2,
476 p-min-overlap 12, --p-pooling-method 'independent'.
477
478 As we used a non-error-correcting polymerase for the 12SV5 PCR amplification reactions to
479 maximize human blocking primer efficiency (resulting in more error-prone amplification), ASVs
480 mapping to the same species were glommed using tax_glom() in the R package *phyloseq* prior
481 to analysis. This meant that ASVs were assigned to the most similar species in the database
482 used from Schneider et al.[57]
483
484 pFR was calculated as the sum of unique assigned ASVs that had any detectable reads.
485 Unassigned ASVs that did not match a human food in our database were excluded from pFR.
486 Samples with a pFR >5 standard deviations outside the median were excluded from further
487 analysis.
488
489 For food clustering analysis, foods that did not have at least one read in one sample were
490 removed and a matrix was generated of the percent of samples in each country where each
491 assigned ASV was detected. Clustering was performed on the Euclidean distance between
492 each food based on its prevalence in the 5 countries using the “complete” method in the
493 function hclust() from the R package *stats*. The R package *ggtree* was then used for
494 dendrogram visualization.
495
496 **Statistical Analysis**
497 *Data Ordinations*
498 To account for the compositional nature of the data, we calculated the centered-log ratio (CLR)
499 transformation of DNA sequencing read counts using the transform() function of *microbiome*

v1.26.0. Principal component analysis (PCA) was then performed on CLR-transformed read counts using the prcomp() function in the *stats* R package (v4.4.1).

Linear model design

Generalized linear models were used to model the relationship between dietary factors and microbiome outcomes. We included age as a covariate in our models due to the strong relationships between age and our microbiome outcomes of interest established in previous literature.[1] We further controlled for country due to differences in age distribution, malnutrition, enteric pathogen exposure, and breastfeeding duration between the countries assessed. Our choice to include interaction terms in our models was determined by choosing the model with the lowest AIC and significant chi-squared p-value in a likelihood ratio test between the reduced model and saturated model.

The relationship between dietary diversity and microbial diversity was modeled using the call *glm(microbiome_diversity ~ pFR + country + age_months + country:age_months + pFR:country, family = 'gaussian')*. The relationship between dietary PC1 and microbiome clusters was modeled using the call *glm(cluster ~ age_months + country + country:age_months + PC1 + country:PC1 + age_months:PC1, family = 'gaussian')*. Cluster associations with subsequent PCs were not tested, as country was highly collinear with PC2, accounting for half of its variance (ANOVA, $p < 2e{-}16$; $R^2 \approx 0.50$ for country), which would render PC2 effects indistinguishable from country-level associations.

When fitting within-country models, the following models were used, depending on available covariate data:

USA and Kenya (breastfeeding data not available):

glm(microbiome_outcome ~ diet_variable + age_months, family = 'gaussian')

Pakistan: (lacked breastfeeding data; all samples came from one timepoint (12-15 months):

glm(microbiome_outcome ~ diet_variable, family = 'gaussian')

Cambodia and Nicaragua (breastfeeding data available):

glm(microbiome_outcome ~ diet_variable + currently_breastfed + age_months + currently_breastfed:age_months + diet_variable:currently_breastfed, family= 'gaussian')

When accounting for the effects of WHZ on microbiome richness, the following models were used:

Kenya:

glm(microbiome_diversity ~ WHZ + pFR + age_months + sex)

Cambodia:

glm(microbiome_diversity ~ WHZ + pFR + age_months + sex + currently_breastfed + pFR:currently_breastfed + currently_breastfed:age_months + WHZ:currently_breastfed)

Institutional and Ethical Review

Stool and extracted stool DNA samples were obtained through collaborations with existing studies via an IRB exemption (Pro00100567) from the Instutitional Review Board at Duke University.

Use of Kenyan pediatric stool samples in this analysis was reviewed and approved by the institutional review boards of the US Centers for Disease Control and Prevention, the Kenya Medical Research Institute, and the Jaramogi Oginga Odinga Hospital. Written informed

consent for the use of these stool samples was given by parents or legal guardians of participants.

U.S. CDC Disclaimer: This work has been supported in part by the President's Emergency Plan for AIDS Relief through the U.S. Centers for Disease Control and Prevention, The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the funding agencies.

Acknowledgements

Nicaraguan samples: The SAGE study was made possible with the NIH grants R01AI127845 and: K24AI141744.

Kenyan samples: We acknowledge the Duke Microbiome Core Facility for performing the DNA extractions, quality control, and preparing the 16S rRNA gene sequencing library; the Duke Sequencing and Genomic Technologies Shared Resource for sequencing the 16S rRNA gene library. We also thank the Duke Microbiome Core Facility for performing DNA BioAnalyzer quality assessments.

Pakistan samples: SEEM study was a multi-national collaboration between Aga Khan University (Pakistan), Washington University in St. Louis, the University of Virginia, and Cincinnati Children's Hospital Medical Center, funded by the Bill and Melinda Gates Foundation (2016-2019). Marty Meier provided technical assistance in stool DNA purification.

Linda Adair, Jun Zeng, Brianna Petrone, Michelle Kirtley, provided valuable scientific guidance to this study.

Finally, we thank all of the children who were participants in this study, as well as their parents. Without their time, effort, and participation, this study would not have occurred.

Figures

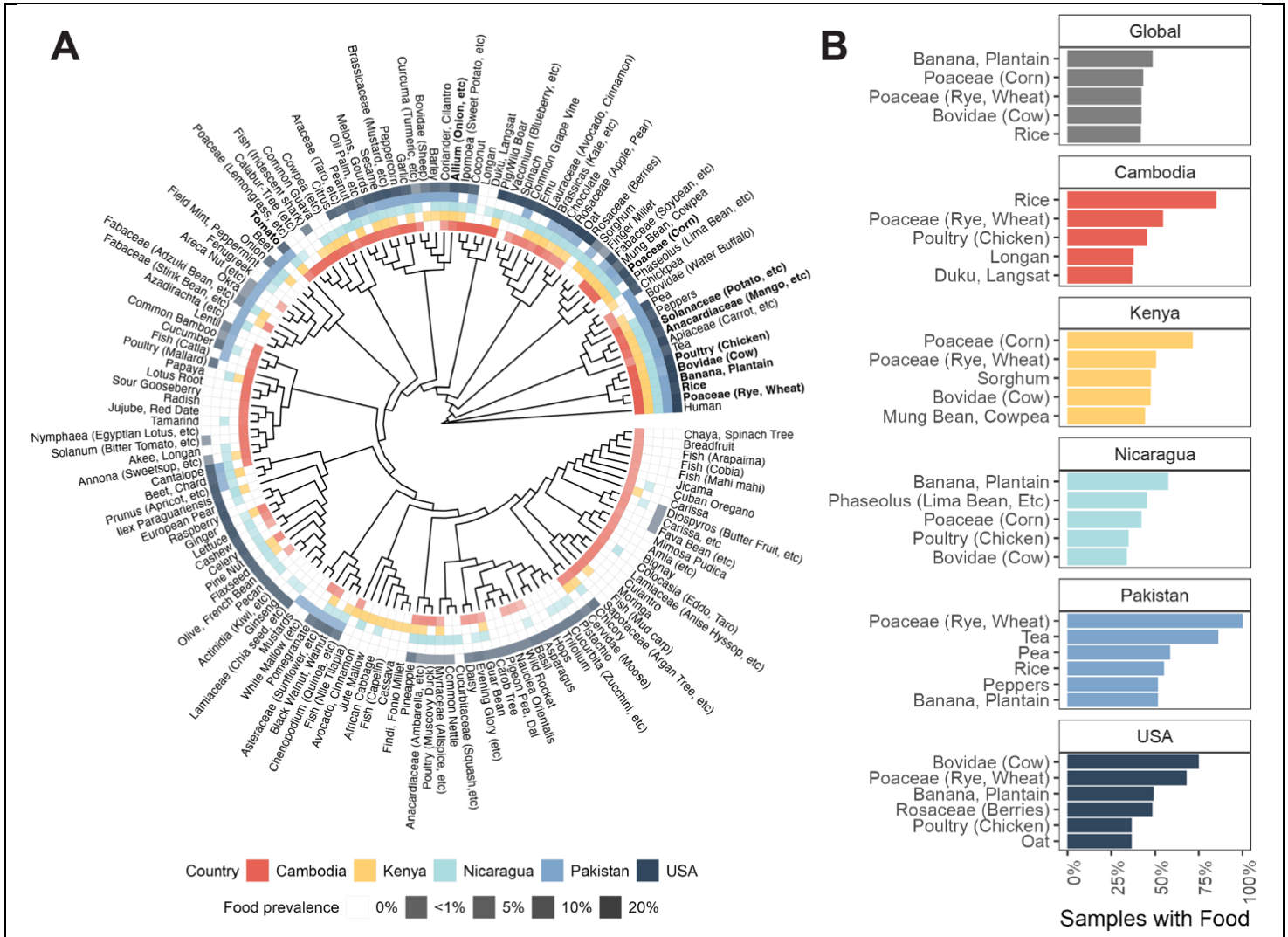

Fig. 1: Early life diet is diverse and varies globally. (A) Dendrogram of food prevalences. Rare foods (<1% frequency in at least one country) were filtered out prior to plotting. Sequences that matched to a higher taxonomic level than species are labeled with the lowest taxonomic level assigned (plants) or a representative taxonomic level (animals), and then a common name of a food that sequence matches. Sequences with multiple foods or food names associated with them are labeled with 'etc'. The 8 plant and 2 animal foods detected in at least 5% of samples in all 5 countries are labeled in bold. (B) Top 5 most prevalent foods in each country.

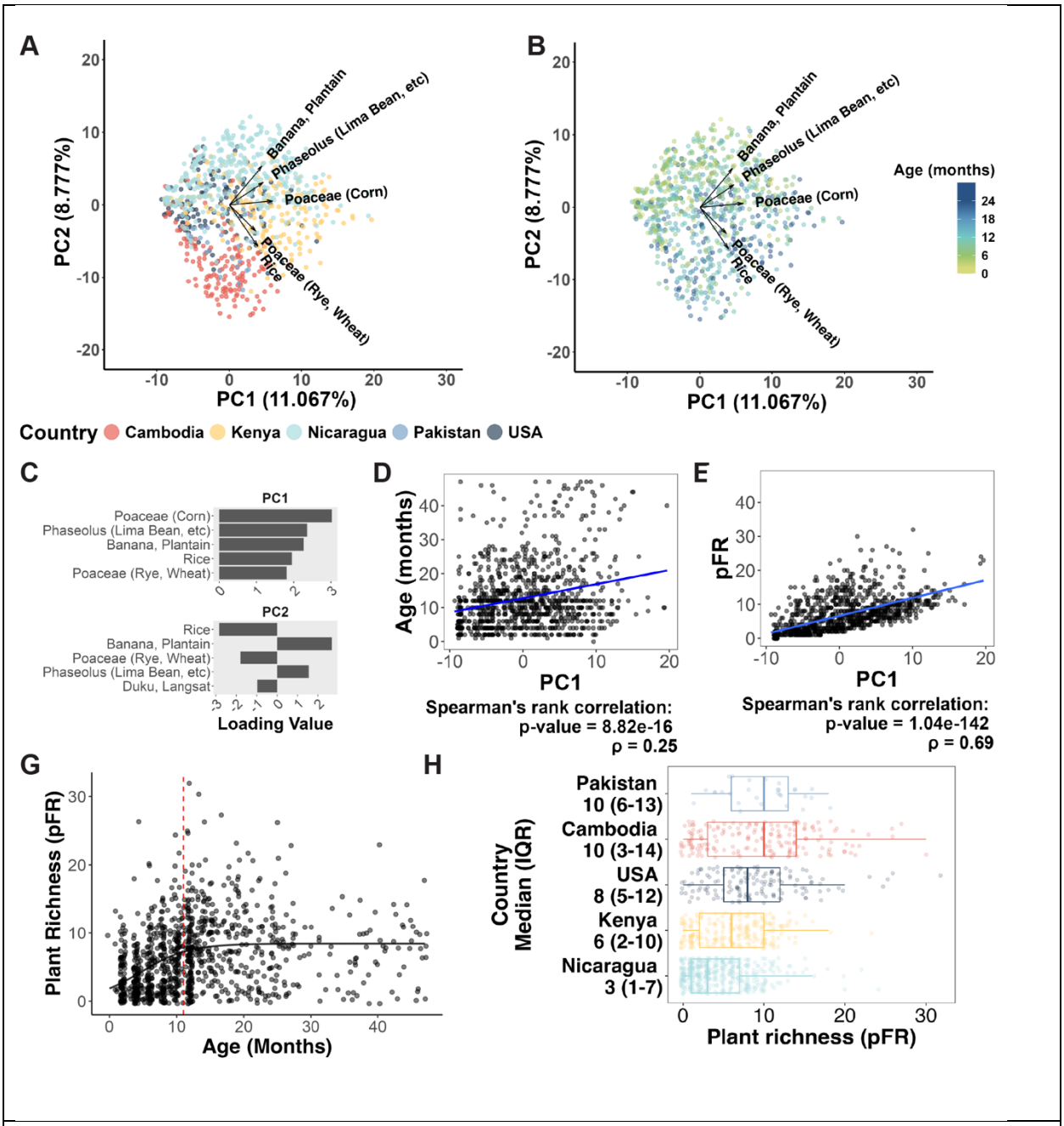

Fig. 2: Dietary variation across age and country. **(A)** Plant FoodSeq PCA colored by country. Country accounted for 15% of variation, age accounted for 0.6% variation, and dietary diversity accounted for 5% of variation. (PERMANOVA, p-value <0.001 for all). **(B)** PCA colored by age **(C)** PCA loadings. **(D)** Correlation between age and PC1. (E) Correlation between pFR and PC1. **(G)** pFR over time. A logistic line (black) was fit to the data. The red dashed line represents the plateau point at which the increase in pFR decelerates, calculated as the x value at which the first derivative of the line reaches 50% of its maximum. **(H)** pFR by country.

581

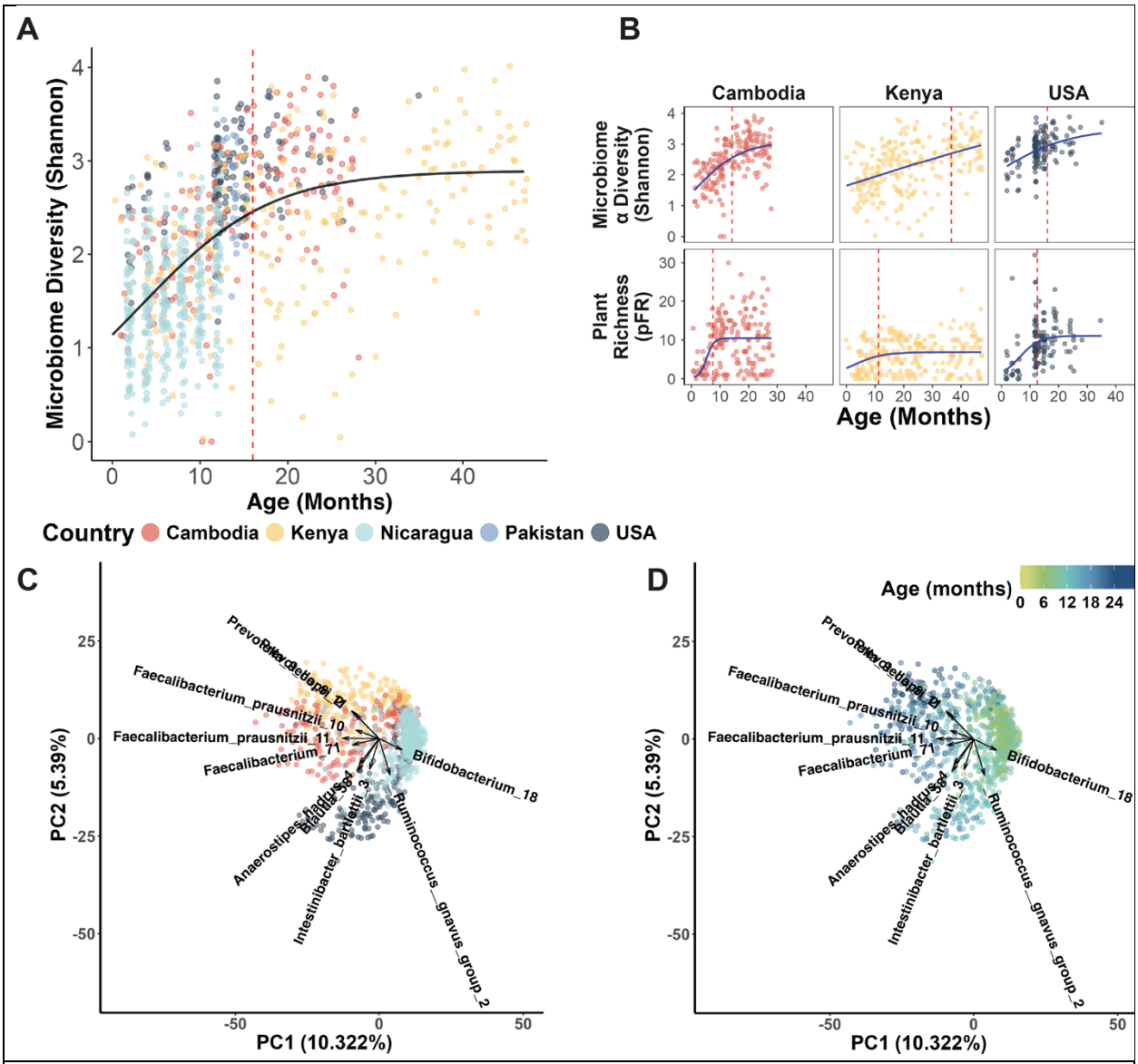

Fig. 3 The early life microbiome varies primarily by geography and age. **(A)** Microbiome diversity increases with age. A logistic line (black) was fit to the data. The red dashed line represents the point at which diversity plateaus with respect to age, defined as the point at which the first derivative of the line reached 50% of its maximum. **(B)** Dietary diversity plateaus prior to microbiome diversity. A logistic line (black) was fit to the data and the plateau point (red dashed line) calculated as in (A). **(C)** PCA of 16S rRNA sequencing data colored by country. **(D)** PCA of 16S data colored by age in months.

582

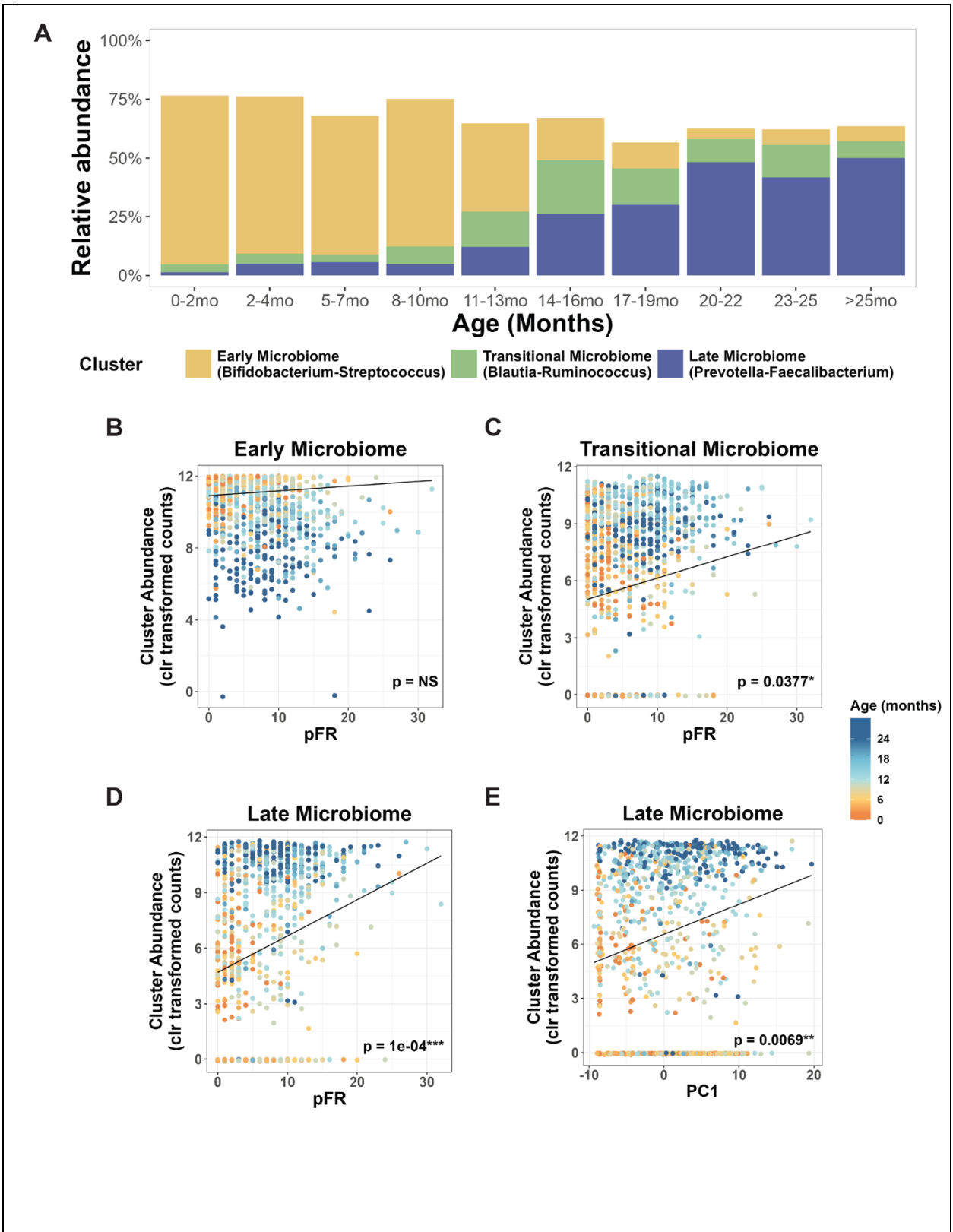

Fig. 4 Plant dietary diversity is correlated with the abundance of mature-microbiome marker bacteria. (A) Cluster abundance over time. **(B)** The abundance of the 'Early Microbiome' cluster, plotted against plant richness, **(C)** The abundance of the 'Transitional

Microbiome' cluster, plotted against plant richness. (**D-E)** The abundance of the 'Late Microbiome' cluster, plotted against plant richness **(D)** and PC1 **(E).** For B-E, lines plot the coefficient derived from a generalized linear model (GLM) controlling for age and country.

Table S1: Cohort Characteristics	Cambodia	Kenya	Nicaragua	Pakistan	USA
Samples	178	228	457	29	144
Participants	178	228	141	29	132
Age range of cohort (months)	1-28	0-47	2-25	12-15	2-35
Age, Months (mean, SD)	14.31 ,7.47	22.44, 13.28	7.33, 4.09	13.86, 0.58	13.42, 5.84
Male (%)	95 (53.4)	111 (48.7)	227 (49.7)	17 (58.6)	76 (52.8)
WHZ (mean (SD))	-0.69 (0.98)	-0.92 (1.38)	0.64 (1.19)	-0.15(0.66)	NaN (NA)
HAZ (mean (SD))	-1.09 (2.35)	-0.42(2.41)	0.18 (1.20)	-0.99(1.21)	NaN (NA)
WAZ (mean (SD))	-1.11 (1.47)	-0.86 (1.82)	0.55 (1.02)	-0.59(0.91)	NaN (NA)
Received nutrition (RUTF or local alternative)	7 (4%)	70 (30.8%)	NA	NA	NA
National Human Development Index [58]	Low	Medium	Medium	Low	Very High
Urbanization	Rural	Urban	Urban	Rural	Urban/Suburban
Currently Breastfed (%)					
Don't Know	9 (5.1)	228 (100.0)	0 (0.0)	29 (100.0)	71 (49.3)
No	69 (38.8)	0 (0.0)	133 (29.1)	0 (0.0)	54 (37.5)
Yes	100 (56.2)	0 (0.0)	324 (70.9)	0 (0.0)	19 (13.2)
pFR (mean (SD))	9.37 (6.69)	6.20 (4.30)	4.48 (3.88)	9.90 (4.44)	8.90 (5.71)
Diarrhea (%)					
Don't Know	5 (2.8)	228 (100.0)	392 (85.8)	29 (100.0)	144 (100.0)
No	159 (89.3)	0 (0.0)	33 (7.2)	0 (0.0)	0 (0.0)
Yes	14 (7.9)	0 (0.0)	32 (7.0)	0 (0.0)	0 (0.0)

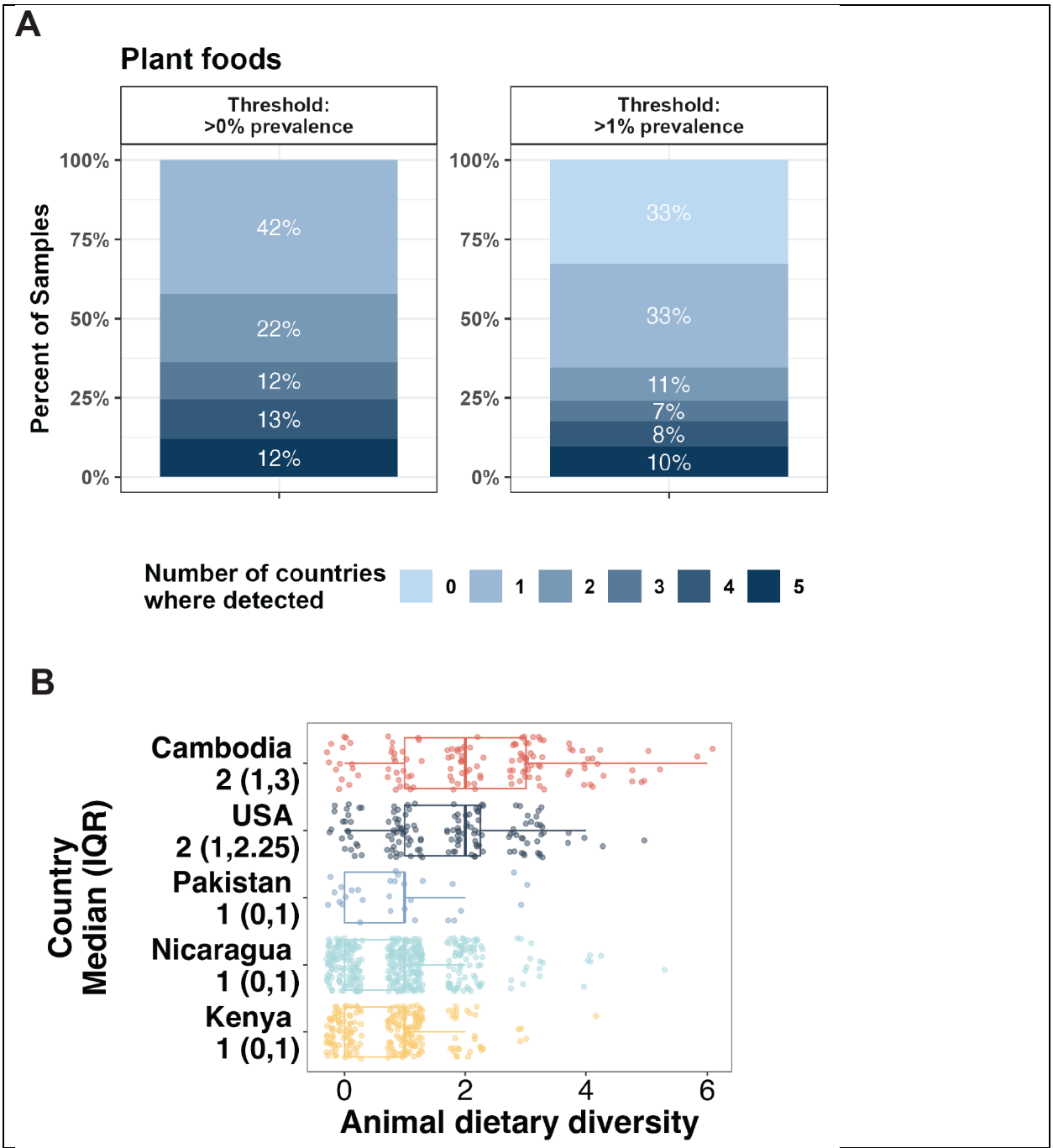

Fig. S1 (A) Distribution of foods by how many countries they were each detected in. A food was counted as 'detected' if it was detected in at least 1 sample in that country (>0% threshold), or at least 1% of samples in that country (>1% threshold). (B) Animal FoodSeq diversity by country.

585
586
587

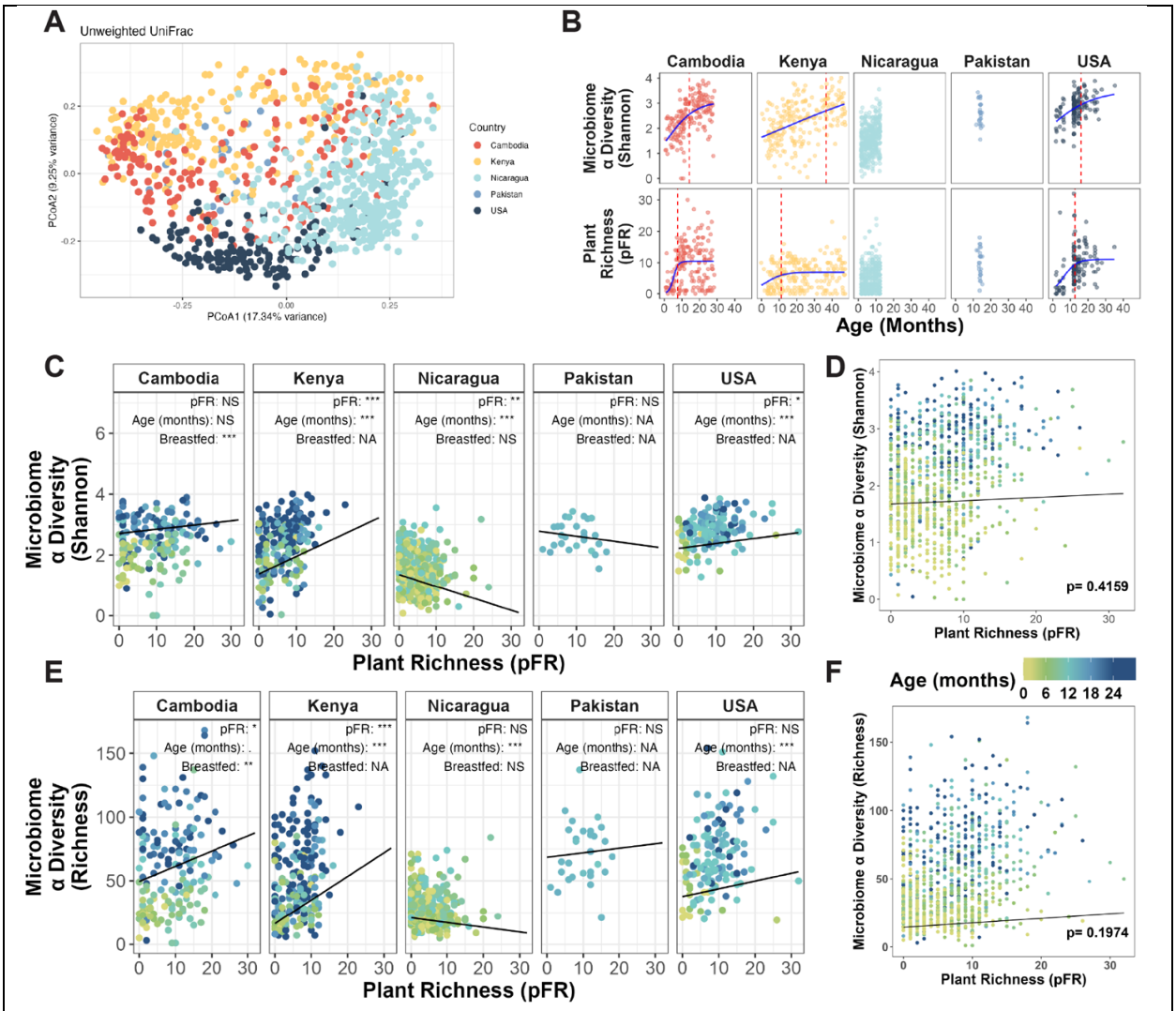

Fig. S2 (A) Unweighted Unifrac ordination. **(B)** Plant and microbiome diversity versus age, with Pakistan and Nicaragua included (see: Fig. 3B). **(C)** Scatterplots showing relationships between microbial Shannon diversity and pFR within country. (D) Scatterplot showing relationship between microbial Shannon diversity and pFR. (E) Scatterplot showing relationships between microbial richness and pFR. For C-E, line plots the coefficient of pFR derived from a generalized linear model controlling for age (C-E), and, where applicable, country (D-E) and breastfeeding status (C, Cambodia and Nicaragua).

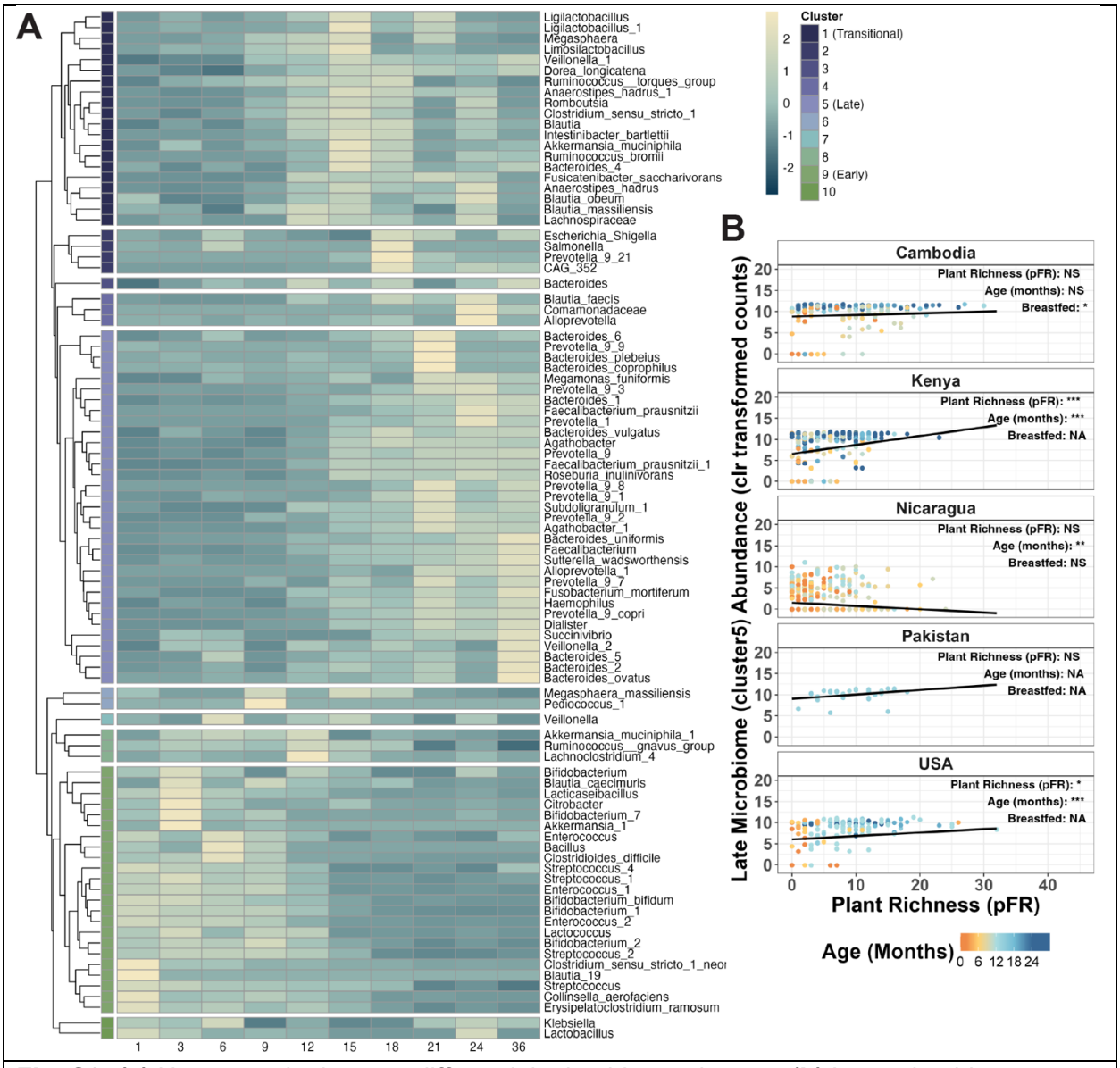

Fig. S3: (a) Heatmap plotting age-differential microbiome clusters. **(b)** Late microbiome abundance correlated with pFR within country while controlling for breastfeeding status, where available, and age. Similar to microbiome diversity, late microbiome abundance is positively correlated with plant richness in Kenya and the USA.

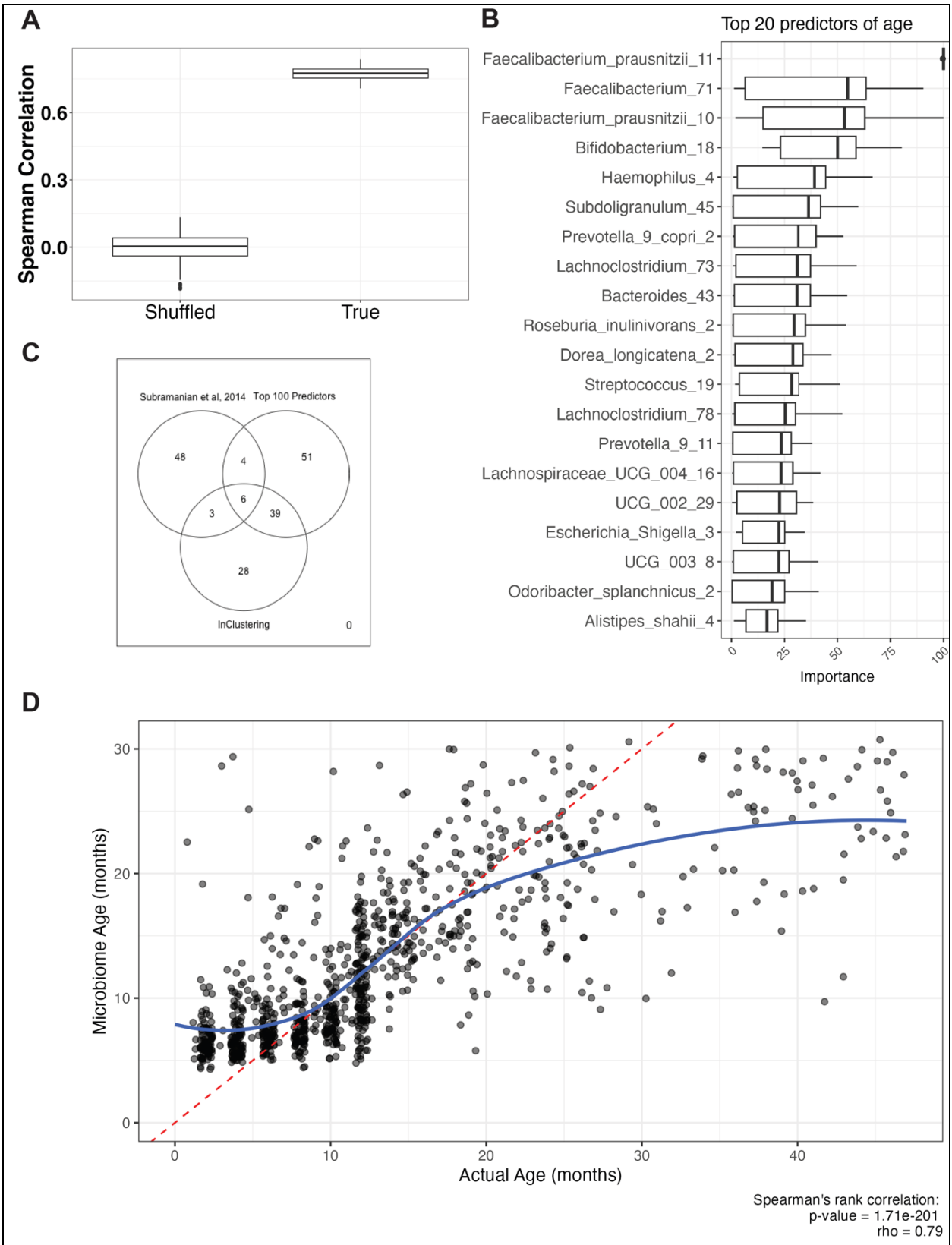

Fig. S4 (A) Model fit for classifier predicting age from shuffled or true 16S rRNA relative abundance data. (B) Top predictors from Random Forest model. (C) Exact sequence

matches between the top 100 predictors in the Random Forest model, the taxa identified in the Early, Transitional, and Late clusters, and a previous study that first employed this analysis. (D) Plot of microbiome age versus actual age. Red dashed line is a theoretical perfect concordance between the two.

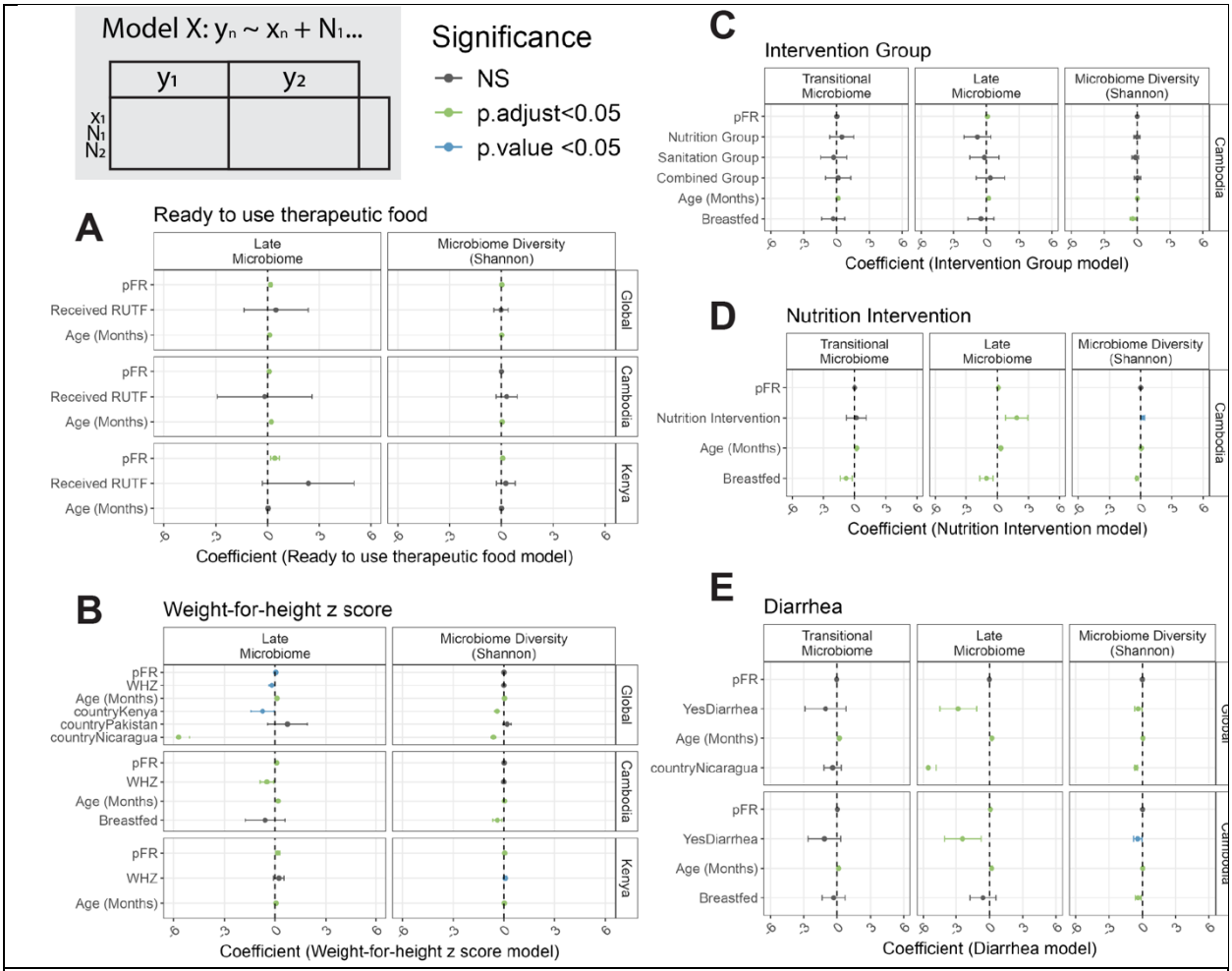

Fig. S5 Generalized linear models controlling for country-specific variables. **(A)** Models controlling for RUTF consumption. **(B)** Models controlling for weight-for-height z score (WHZ). **(C)** Models controlling for study intervention group within the Cambodian cohort. **(D)** Models controlling for whether or not Cambodian children were in an intervention group that included a nutrition intervention. **(E)** Models controlling for diarrhea. Each column represents an outcome variable, and each row represents a predictor variable. Results are grouped by whether the test was run on all samples with non-NA values for that variable ('Global') or within a specific country. Missing countries did not collect data for that variable.

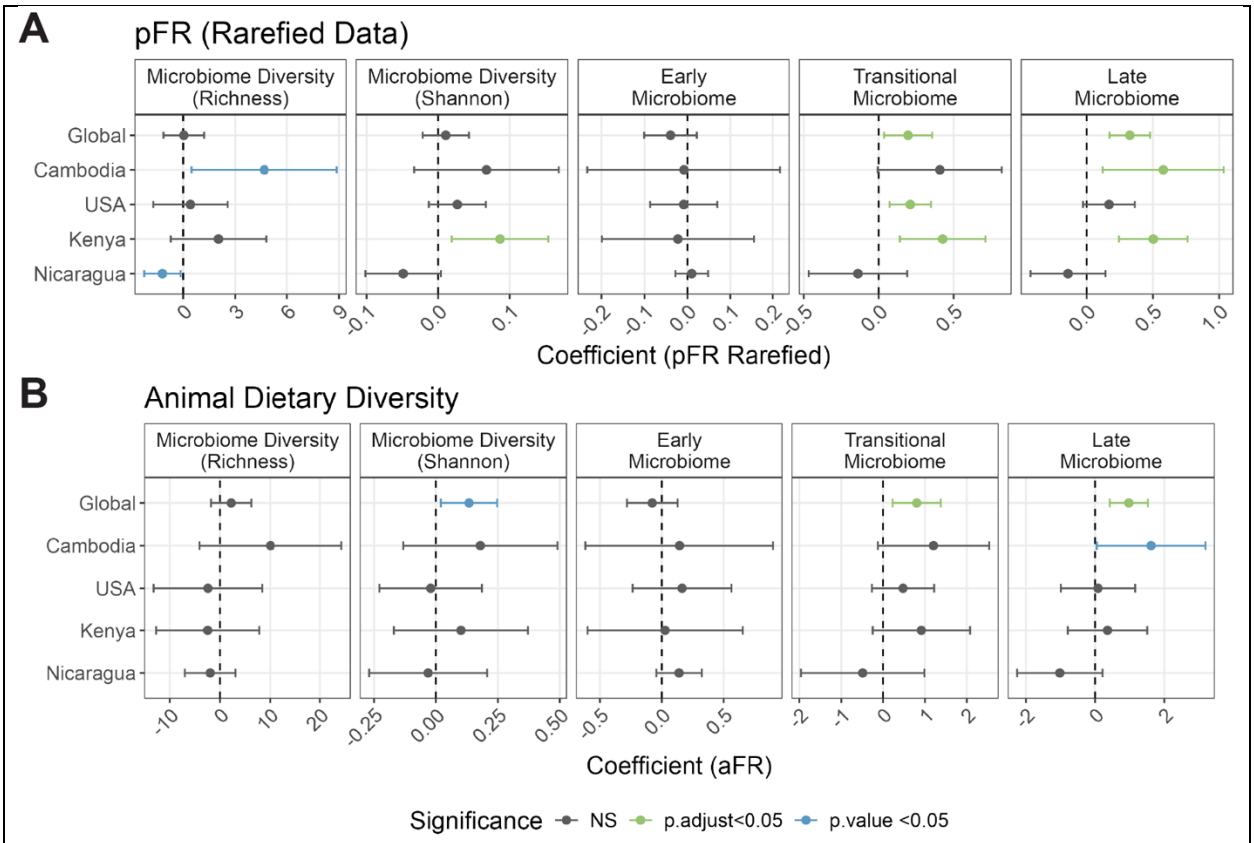

Fig. S6. Summary of generalized linear model results for other diet metrics. **(A)** Model results for pFR calculated from data rarefied to 500 reads. **(B)** Model results for animal dietary diversity (aFR). Y axis is the dataset the test was run on ('global' for all samples, or within a specific country.) Pakistan was omitted due to low sample size (n=29). X axis represents the coefficient, indicative of effect size, of that variable in the model.

Works Cited

1. Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M.A., Benezra, A., DeStefano, J., Meier, M.F., Muegge, B.D., et al. (2014). Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature *510*, 417–421. https://doi.org/10.1038/nature13421.

2. Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., et al. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe *17*, 690–703. https://doi.org/10.1016/j.chom.2015.04.004.

3. Fahur Bottino, G., Bonham, K.S., Patel, F., McCann, S., Zieff, M., Naspolini, N., Ho, D., Portlock, T., Joos, R., Midani, F.S., et al. (2025). Early life microbial succession in the gut follows common patterns in humans across the globe. Nat. Commun. *16*, 660. https://doi.org/10.1038/s41467-025-56072-w.

4. Ajslev, T.A., Andersen, C.S., Gamborg, M., Sørensen, T.I.A., and Jess, T. (2011). Childhood overweight after establishment of the gut microbiota: the role of delivery mode,

612 pre-pregnancy weight and early administration of antibiotics. Int. J. Obes. *35*, 522–529.
613 https://doi.org/10.1038/ijo.2011.27.

614 5. Bager, P., Wohlfahrt, J., and Westergaard, T. (2008). Caesarean delivery and risk of atopy
615 and allergic disease: meta-analyses. Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol. *38*,
616 634–642. https://doi.org/10.1111/j.1365-2222.2008.02939.x.

617 6. Clausen, T.D., Bergholt, T., Bouaziz, O., Arpi, M., Eriksson, F., Rasmussen, S., Keiding, N.,
618 and Løkkegaard, E.C. (2016). Broad-Spectrum Antibiotic Treatment and Subsequent
619 Childhood Type 1 Diabetes: A Nationwide Danish Cohort Study. PloS One *11*, e0161654.
620 https://doi.org/10.1371/journal.pone.0161654.

621 7. Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., D. Lieber, A., Wu, F.,
622 Perez-Perez, G.I., Chen, Y., et al. (2016). Antibiotics, birth mode, and diet shape
623 microbiome maturation during early life. Sci. Transl. Med. *8*, 343ra82.
624 https://doi.org/10.1126/scitranslmed.aad7121.

625 8. Stewart, C.J., Ajami, N.J., O'Brien, J.L., Hutchinson, D.S., Smith, D.P., Wong, M.C., Ross,
626 M.C., Lloyd, R.E., Doddapaneni, H., Metcalf, G.A., et al. (2018). Temporal development of
627 the gut microbiome in early childhood from the TEDDY study. Nature *562*, 583–588.
628 https://doi.org/10.1038/s41586-018-0617-x.

629 9. David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E.,
630 Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and
631 reproducibly alters the human gut microbiome. Nature *505*, 559–563.
632 https://doi.org/10.1038/nature12820.

633 10. Krebs, N.F., Sherlock, L.G., Westcott, J., Culbertson, D., Hambidge, K.M., Feazel, L.M.,
634 Robertson, C.E., and Frank, D.N. (2013). Effects of Different Complementary Feeding
635 Regimens on Iron Status and Enteric Microbiota in Breastfed Infants. J. Pediatr. *163*, 416-
636 423.e4. https://doi.org/10.1016/j.jpeds.2013.01.024.

637 11. Qasem, W., Azad, M.B., Hossain, Z., Azad, E., Jorgensen, S., Castillo San Juan, S., Cai,
638 C., Khafipour, E., Beta, T., Roberts, L.J., et al. (2017). Assessment of complementary
639 feeding of Canadian infants: effects on microbiome & oxidative stress, a randomized
640 controlled trial. BMC Pediatr. *17*, 54. https://doi.org/10.1186/s12887-017-0805-0.

641 12. Tang, M., Frank, D., Hendricks, A., Ir, D., and Krebs, N. (2019). Protein Intake During Early
642 Complementary Feeding Affects the Gut Microbiota in U.S. Formula-fed Infants (FS04-03-
643 19). Nutr. 2019 Abstr. *3*, nzz048.FS04-03-19. https://doi.org/10.1093/cdn/nzz048.FS04-03-
644 19.

645 13. Plaza-Diaz, J., Bernal, M.J., Schutte, S., Chenoll, E., Genovés, S., Codoñer, F.M., Gil, A.,
646 and Sanchez-Siles, L.M. (2021). Effects of Whole-Grain and Sugar Content in Infant
647 Cereals on Gut Microbiota at Weaning: A Randomized Trial. Nutrients *13*, 1496.
648 https://doi.org/10.3390/nu13051496.

649 14. Smith-Brown, P., Morrison, M., Krause, L., and Davies, P.S.W. (2016). Dairy and plant
650 based food intakes are associated with altered faecal microbiota in 2 to 3 year old
651 Australian children. Sci. Rep. *6*, 32385. https://doi.org/10.1038/srep32385.

652 15. Smith-Brown, P., Morrison, M., Krause, L., and Davies, P.S.W. (2019). Microbiota and Body
653 Composition During the Period of Complementary Feeding. J. Pediatr. Gastroenterol. Nutr.
654 *69*, 726–732. https://doi.org/10.1097/MPG.0000000000002490.

655 16. Thompson, A.L., Monteagudo-Mera, A., Cadenas, M.B., Lampl, M.L., and Azcarate-Peril,
656 M.A. (2015). Milk- and solid-feeding practices and daycare attendance are associated with
657 differences in bacterial diversity, predominant communities, and metabolic and immune
658 function of the infant gut microbiome. Front. Cell. Infect. Microbiol. *5*.
659 https://doi.org/10.3389/fcimb.2015.00003.

660 17. Laursen Martin Frederik, Andersen Louise B. B., Michaelsen Kim F., Mølgaard Christian,
661 Trolle Ellen, Bahl Martin Iain, Licht Tine Rask, and Suen Garret (2016). Infant Gut
662 Microbiota Development Is Driven by Transition to Family Foods Independent of Maternal
663 Obesity. mSphere *1*, e00069-15. https://doi.org/10.1128/mSphere.00069-15.

664 18. Olm, M.R., Dahan, D., Carter, M.M., Merrill, B.D., Yu, F.B., Jain, S., Meng, X., Tripathi, S.,
665 Wastyk, H., Neff, N., et al. (2022). Robust variation in infant gut microbiome assembly
666 across a spectrum of lifestyles. Science *376*, 1220–1223.
667 https://doi.org/10.1126/science.abj2972.

668 19. Kipnis, V., Subar, A.F., Midthune, D., Freedman, L.S., Ballard-Barbash, R., Troiano, R.P.,
669 Bingham, S., Schoeller, D.A., Schatzkin, A., and Carroll, R.J. (2003). Structure of Dietary
670 Measurement Error: Results of the OPEN Biomarker Study. Am. J. Epidemiol. *158*, 14–21.
671 https://doi.org/10.1093/aje/kwg091.

672 20. Subar, A.F., Freedman, L.S., Tooze, J.A., Kirkpatrick, S.I., Boushey, C., Neuhouser, M.L.,
673 Thompson, F.E., Potischman, N., Guenther, P.M., Tarasuk, V., et al. (2015). Addressing
674 Current Criticism Regarding the Value of Self-Report Dietary Data12. J. Nutr. *145*, 2639–
675 2645. https://doi.org/10.3945/jn.115.219634.

676 21. What We Eat in America, NHANES 2017-March 2020 Prepandemic, individuals 2 years and
677 over (excluding breast-fed children), day 1.

678 22. Burrows, T.L., Martin, R.J., and Collins, C.E. (2010). A Systematic Review of the Validity of
679 Dietary Assessment Methods in Children when Compared with the Method of Doubly
680 Labeled Water. J. Am. Diet. Assoc. *110*, 1501–1510.
681 https://doi.org/10.1016/j.jada.2010.07.008.

682 23. Davies, P.S.W., Coward, W.A., Gregory, J., White, A., and Mills, A. (1994). Total energy
683 expenditure and energy intake in the pre-school child: a comparison. Br. J. Nutr. *72*, 13–20.
684 https://doi.org/10.1079/BJN19940005.

685 24. Fisher, J.O., Butte, N.F., Mendoza, P.M., Wilson, T.A., Hodges, E.A., Reidy, K.C., and
686 Deming, D. (2008). Overestimation of infant and toddler energy intake by 24-h recall
687 compared with weighed food records. Am. J. Clin. Nutr. *88*, 407–415.
688 https://doi.org/10.1093/ajcn/88.2.407.

689 25. Lanigan, J.A., Wells, J.C.K., Lawson, M.S., and Lucas, A. (2001). Validation of food diary
690 method for assessment of dietary energy and macronutrient intake in infants and children
691 aged 6–24 months. Eur. J. Clin. Nutr. *55*, 124–129. https://doi.org/10.1038/sj.ejcn.1601128.

26. Ravelli, M.N., and Schoeller, D.A. (2020). Traditional Self-Reported Dietary Instruments Are Prone to Inaccuracies and New Approaches Are Needed. Front. Nutr. *7*. https://doi.org/10.3389/fnut.2020.00090.

27. Wallace, A., Kirkpatrick, S.I., Darlington, G., and Haines, J. (2018). Accuracy of Parental Reporting of Preschoolers' Dietary Intake Using an Online Self-Administered 24-h Recall. Nutrients *10*, 987. https://doi.org/10.3390/nu10080987.

28. Data FAO/WHO GIFT | Global Individual Food Consumption Data Tool GIFT. https://www.fao.org/gift-individual-food-consumption/data/en.

29. Global Diet Quality Project (2022). Measuring what the world eats: Insights from a new approach. (Geneva: Global Alliance for Improved Nutrition (GAIN); Boston, MA: Harvard T.H. Chan School of Public Health, Department of Global Health and Population.).

30. Petrone, B.L., Aqeel, A., Jiang, S., Durand, H.K., Dallow, E.P., McCann, J.R., Dressman, H.K., Hu, Z., Tenekjian, C.B., Yancy, W.S., et al. (2023). Diversity of plant DNA in stool is linked to dietary quality, age, and household income. Proc. Natl. Acad. Sci. *120*, e2304441120. https://doi.org/10.1073/pnas.2304441120.

31. Loftus, T. (1995). Staple foods: what do people eat? In Dimensions of Need (Food and Agriculture organization of the United Nations).

32. Mank, I., Vandormael, A., Traoré, I., Ouédraogo, W.A., Sauerborn, R., and Danquah, I. (2020). Dietary habits associated with growth development of children aged < 5 years in the Nouna Health and Demographic Surveillance System, Burkina Faso. Nutr. J. *19*, 81. https://doi.org/10.1186/s12937-020-00591-3.

33. Steyn, N.P., Nel, J.H., Nantel, G., Kennedy, G., and Labadarios, D. (2006). Food variety and dietary diversity scores in children: are they good indicators of dietary adequacy? Public Health Nutr. *9*, 644–650. https://doi.org/10.1079/phn2005912.

34. Mezgebo, K., Lema, T.B., and Neela, S. (2021). Food variety, dietary diversity scores and dietary quality of complementary foods consuming by 6-24 months aged children in Jimma town, southwest Ethiopia. Nutr. Food Sci. *51*, 323–344. https://doi.org/10.1108/NFS-02-2020-0033.

35. Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. *107*, 11971. https://doi.org/10.1073/pnas.1002601107.

36. Abdelsalam, N.A., Hegazy, S.M., and Aziz, R.K. The curious case of Prevotella copri. Gut Microbes *15*, 2249152. https://doi.org/10.1080/19490976.2023.2249152.

37. Johnson, K.V.-A., and Burnet, P.W.J. (2016). Microbiome: Should we diversify from diversity? Gut Microbes *7*, 455–458. https://doi.org/10.1080/19490976.2016.1241933.

38. Starke, R., Capek, P., Morais, D., Callister, S.J., and Jehmlich, N. (2020). The total microbiome functions in bacteria and fungi. J. Proteomics *213*, 103623. https://doi.org/10.1016/j.jprot.2019.103623.

731 39. Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T.,
732 Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., et al. (2012). Structure, function
733 and diversity of the healthy human microbiome. Nature *486*, 207–214.
734 https://doi.org/10.1038/nature11234.

735 40. Branca, F., Lartey, A., Oenema, S., Aguayo, V., Stordalen, G.A., Richardson, R., Arvelo, M.,
736 and Afshin, A. (2019). Transforming the food system to fight non-communicable diseases.
737 BMJ *364*, l296. https://doi.org/10.1136/bmj.l296.

738 41. Arimond, M., and Ruel, M.T. (2004). Dietary Diversity Is Associated with Child Nutritional
739 Status: Evidence from 11 Demographic and Health Surveys. J. Nutr. *134*, 2579–2585.
740 https://doi.org/10.1093/jn/134.10.2579.

741 42. Kennedy, G.L., Pedro, M.R., Seghieri, C., Nantel, G., and Brouwer, I. (2007). Dietary
742 diversity score is a useful indicator of micronutrient intake in non-breast-feeding Filipino
743 children. J. Nutr. *137*, 472–477. https://doi.org/10.1093/jn/137.2.472.

744 43. Roduit, C., Frei, R., Depner, M., Schaub, B., Loss, G., Genuneit, J., Pfefferle, P., Hyvärinen,
745 A., Karvonen, A.M., Riedler, J., et al. (2014). Increased food diversity in the first year of life
746 is inversely associated with allergic diseases. J. Allergy Clin. Immunol. *133*, 1056-1064.e7.
747 https://doi.org/10.1016/j.jaci.2013.12.1044.

748 44. Indicators for assessing infant and young child feeding practices: definitions and
749 measurement methods (2021).

750 45. Lai, A., Velez, I., Ambikapathi, R., Seng, K., Levy, K., Kowalsky, E., Holcomb, D.,
751 Konstantinidis, K.T., Cumming, O., and Brown, J. (2022). Independent and combined effects
752 of nutrition and sanitation interventions on enteric pathogen carriage and child growth in
753 rural Cambodia: a factorial cluster-randomised controlled trial. Preprint at medRxiv,
754 https://doi.org/10.1101/2021.05.21.21257546 https://doi.org/10.1101/2021.05.21.21257546.

755 46. Song, R., Click, E.S., McCarthy, K.D., Heilig, C.M., Mchembere, W., Smith, J.P., Fajans, M.,
756 Musau, S.K., Okeyo, E., Okumu, A., et al. (2021). Sensitive and Feasible Specimen
757 Collection and Testing Strategies for Diagnosing Tuberculosis in Young Children. JAMA
758 Pediatr. *175*, e206069. https://doi.org/10.1001/jamapediatrics.2020.6069.

759 47. Becker-Dreps, S., Allali, I., Monteagudo, A., Vilchez, S., Hudgens, M.G., Rogawski, E.T.,
760 Carroll, I.M., Zambrana, L.E., Espinoza, F., and Azcarate-Peril, M.A. (2015). Gut
761 Microbiome Composition in Young Nicaraguan Children During Diarrhea Episodes and
762 Recovery. Am. J. Trop. Med. Hyg. *93*, 1187–1193. https://doi.org/10.4269/ajtmh.15-0322.

763 48. Study Details | Project HOPE 1000: Health Outcomes Pregnancy Exposures |
764 ClinicalTrials.gov https://clinicaltrials.gov/study/NCT04417283.

765 49. Martin, V.M., Virkud, Y.V., Dahan, E., Seay, H.L., Itzkovits, D., Vlamakis, H., Xavier, R.,
766 Shreffler, W.G., Yuan, Q., and Yassour, M. (2022). Longitudinal disease-associated gut
767 microbiome differences in infants with food protein-induced allergic proctocolitis. Microbiome
768 *10*, 154. https://doi.org/10.1186/s40168-022-01322-y.

769 50. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens,
770 S.M., Betley, J., Fraser, L., Bauer, M., et al. (2012). Ultra-high-throughput microbial

community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. *6*, 1621–1624. https://doi.org/10.1038/ismej.2012.8.

51. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. *108*, 4516–4522. https://doi.org/10.1073/pnas.1000080107.

52. Maurice, C.F., Haiser, H.J., and Turnbaugh, P.J. (2013). Xenobiotics Shape the Physiology and Gene Expression of the Active Human Gut Microbiome. Cell *152*, 39–50. https://doi.org/10.1016/j.cell.2012.10.052.

53. Petrone, B.L., Aqeel, A., Jiang, S., Durand, H.K., Dallow, E.P., McCann, J.R., Dressman, H.K., Hu, Z., Tenekjian, C.B., Yancy, W.S., et al. (2022). DNA metabarcoding captures dietary plant diversity in individuals and cohorts. Preprint at medRxiv, https://doi.org/10.1101/2022.06.13.22276343 https://doi.org/10.1101/2022.06.13.22276343.

54. Shehzad, W., Riaz, T., Nawaz, M.A., Miquel, C., Poillot, C., Shah, S.A., Pompanon, F., Coissac, E., and Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (Prionailurus bengalensis) in Pakistan. Mol. Ecol. *21*, 1951–1965. https://doi.org/10.1111/j.1365-294X.2011.05424.x.

55. De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., and Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol. Ecol. Resour. *14*, 306–323. https://doi.org/10.1111/1755-0998.12188.

56. David, L.A., Weil, A., Ryan, E.T., Calderwood, S.B., Harris, J.B., Chowdhury, F., Begum, Y., Qadri, F., LaRocque, R.C., and Turnbaugh, P.J. (2015). Gut Microbial Succession Follows Acute Secretory Diarrhea in Humans. mBio *6*, 10.1128/mbio.00381-15. https://doi.org/10.1128/mbio.00381-15.

57. Schneider, J. (2021). Comprehensive coverage of human last meal components revealed by a forensic DNA metabarcoding approach | Scientific Reports. Sci. Rep. *11*.

58. Nations, U. Human Development Index (United Nations).