1   Microbiota Assembly, Structure, and Dynamics Among Tsimane Horticulturalists

2   of the Bolivian Amazon

3

4   Daniel D. Sprockett[1], Melanie Martin[2,3], Elizabeth K. Costello[4], Adam Burns[4], Susan P. Holmes[5],

5   Michael Gurven[3,6], and David A. Relman[1,4,7,*]

6

7   [1]Department of Microbiology & Immunology, Stanford University School of Medicine, Stanford,

8   CA 94305, USA

9   [2]Department of Anthropology, University of Washington, Seattle, WA 98195, USA

10   [3]Department of Anthropology, University of California Santa Barbara, Santa Barbara, CA 93106,

11   USA

12   [4]Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

13   [5]Department of Statistics, Stanford University, Stanford, CA 94305, USA

14   [6]Broom Center for Demography, University of California Santa Barbara, Santa Barbara, CA

15   93106, USA

16   [7]Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304, USA

17

18   * Corresponding Author. Email: relman@stanford.edu

19

20   ABSTRACT

21   Little is known about the relative contributions of selective and neutral forces on human-

22   associated microbiota assembly. Here, we characterize microbial community assembly in 52

23   Tsimane infant-mother pairs, using longitudinally collected stool and tongue swab samples

24   profiled with 16S rRNA gene amplicon sequencing. The Tsimane are an indigenous Bolivian

25   population who practice infant care associated behaviors expected to increase mother-infant

26   dispersal. Infant consumption of dairy products, vegetables, and chicha (a fermented drink

27   inoculated with oral microbes) was significantly associated with gut microbiota composition. At

28   both body sites, maternal microbes at higher relative abundance were more likely to be shared.

29   Shared microbes were also higher in abundance in infants at both body sites, but decreased in

30   average relative abundance with age and were not significantly higher by 12 months of age.

31   Infant microbiotas were modeled using a neutral community model of assembly, which showed

32   that the prevalence of more than two thirds of infant-colonizing microbes could be explained

33   using neutral processes alone. The same method was applied to datasets from Finnish and

34   Bangladeshi infants, confirming that the majority of microbes colonizing infants from different

35   countries were neutrally distributed. Among the Tsimane infant and adult gut microbiota

36   samples, neutral processes were less prominent in villages with more market access. These

37   results underscore the importance of neutral processes during infant microbiota assembly, and

38   suggest that cultural changes associated with market integration may be affecting traditional

39   modes of microbiota assembly by decreasing the role of these neutral processes, perhaps

40   through changes in diet, sanitation, or access to medical care.

41   The human body becomes colonized in early life by microbial communities that are

42   involved in several key metabolic and immunologic processes, including nutrient acquisition,

43   immune programming, and pathogen exclusion. While these communities are relatively simple

44   at birth, they grow in complexity until reaching an adult-like configuration within the first few

45   years of life[1,2]. This rapid and dynamic period of community assembly is characterized by broad

46   shifts in community structure[3] that arise from both deterministic processes like host-driven

47   ecological selection[4], as well as neutral processes like dispersal and demographic stochasticity[5-

48   7]. Yet, the relative contributions of deterministic and neutral processes to early life microbiota

49   assembly—and the extent to which different environmental factors may moderate these

50   effects—remain largely unexplored.

51   In order to assess how transmission dynamics might affect community assembly, we

52   focused on the Tsimane people, an indigenous forager-horticulturalist population inhabiting the

53   Bolivian Amazon basin[8]. The Tsimane practice several infant care-associated behaviors

54   (ICABs) that potentially increase the likelihood that maternally-derived microbes disperse to

55   their offspring. For example, all Tsimane infants are vaginally birthed at home, and mothers

56   carry infants in slings during the day and share their beds with their infants at night. Infants are

57   breastfed 'on demand' 24 hours a day, and the period of exclusive breastfeeding lasts about 4

58   months[9,10], although typically, weaning is not complete until around 27 months of age[9]. During

59   the introduction of complementary foods, mothers frequently premasticate, or pre-chew, foods

60   such as rice, plantain, meat, or fish before depositing them into the mouths of their children. The

61   most commonly consumed liquids are water (usually sourced from a nearby river), water mixed

62   with sugar or fruit juice[11], and chicha, a fermented drink made from manioc, corn, or plantain.

63   Chicha made from manioc and corn is inoculated with saliva; women chew and expectorate

64   pieces of manioc during preparation, and serve the drink without cooking. Similar chicha

65   preparations have been shown to contain high loads of diverse lactic acid bacteria and yeast

66   species[12-16].

67   In comparison with industrialized societies, microbial exposure is high among the

68   Tsimane. They lack access to improved water sources and have frequent exposure to parasites

69   with high rates of early life morbidity and mortality owing to infectious disease[17-19]. Previous

70   studies have reported that Tsimane children have high levels of C-reactive protein and other

71   inflammatory markers, consistent with a high pathogen burden[19-22]. This confluence of high

72   exposure to maternal and environmental sources of microbes and high immune system

73   activation suggests high rates of microbial dispersal to Tsimane infants from family and

74   environmental sources.

75   In recent years, rapid lifestyle changes have taken place in the Tsimane population due

76   to increased access to market goods, cash economies, wage labor, education, and medicine[8].

77   These changes have not yet affected ICABs[10], and the Tsimane continue to subsist primarily on

78   horticulture and foraging, remain highly active, and exhibit negligible cardiometabolic health

79   risks[23]. However, secular trends have documented increasing sedentarism, BMI, LDL

80   cholesterol, and consumption of sugar, vegetable oil, and processed foods—particularly for

81   Tsimane residing closer to market towns[23-25]. Thus, the Tsimane represent an important

82   population in which to examine how changing lifestyle factors may affect patterns and

83   processes of host-associated microbiota assembly.

84   Here, we report on microbiota data that we generated as part of a long-term study of

85   Tsimane health and life history[8]. We first examined microbiota structure from paired fecal (n =

86   288) and oral (n = 119) samples collected longitudinally from 52 Tsimane mother-infant dyads

87   between 2012-2013, one of the largest sample collections of this type to date from a non-

88   industrialized setting. We then placed the patterns of microbiota assembly that we observed into

89   a global context by comparing Tsimane infants to previously studied infants from Bangladesh[26]

90   and Finland[27]. Finally, we expanded our cohort by including data we generated from fecal

91   samples collected in 2009 from 73 Tsimane individuals ranging in age from 1 to 58 years,

92   enabling us to examine temporal and regional variation in microbiota composition associated

4

93   with market integration. We found that even though neutral processes could largely explain

94   patterns of infant microbiota assembly, these patterns were also consistent with an effect of

95   ICABs. Gut bacterial taxa were associated with increasing market integration, and displayed

96   altered dispersal patterns during infant colonization.

97

98   RESULTS

99   To explore how ICABs affect patterns of microbial colonization in young children, stool

100   samples (reflecting the distal gut) and swab samples from the dorsum of the tongue were

101   collected from 52 Tsimane families living in six villages located along the Maniqui River in the

102   Bolivian lowlands of the Amazon basin. Samples were collected from 48 infants (0-2 years of

103   age) and 51 mothers (14 or more years of age) using a mixed longitudinal design

104   (**Supplementary Table 1** and **Fig. 1a**). Bacterial communities in these samples were profiled

105   using 16S rRNA gene amplicon sequencing. Anthropometric measures, as well as health and

106   diet surveys, were also collected at the time of sampling[8].

107

108   *Infant and Maternal Microbiota Dynamics Following Birth*

109   Bacterial diversity increased with age in both the distal gut and tongue dorsum of infants

110   over the first 18 months of life (**Fig. 1b,c**). In a linear mixed-effects model that included the

111   subject as a random effect, diversity was significantly positively correlated with age in both stool

112   ($p < 0.001$, conditional $R^2 = 0.481$) and tongue swabs ($p = 0.015$, conditional $R^2 = 0.38$),

113   although diversity increased at a faster rate in the stool samples (**Fig. 1b,c**; ANOVA, $p = 0.004$).

114   Maternal stool samples decreased in diversity over the study period (**Fig. 1b**, $p = 0.04$,

115   conditional $R^2 = 0.348$), which is consistent with other studies of the gut microbiota of women in

116   the early postpartum period[28]. The diversity of the maternal dorsal tongue communities was

117   stable during this time (**Fig. 1c**, $p = 0.105$).

118   Variation in bacterial community composition among samples from mothers and infants

119 was primarily explained by body site (14%, PERMANOVA, p < 0.001) and host age (11%,

120 PERMANOVA, p < 0.001). Village type (proximal or distal to the regional market) was also a

121 significant factor (PERMANOVA, p < 0.01), although it explained less than 0.5% of the variation.

122 Microbial communities colonizing the infant gut and tongue were distinct at all ages and became

123 increasingly differentiated with time (**Fig. 1d** and **Supplementary Fig. 1**), but were never as

124 distinct as those of adults, owing largely to incomplete maturation of these infant distal gut

125 microbiotas by 18 months of age. The tongue swabs of 16-18 month olds harbored microbial

126 communities similar in composition to adult tongue swabs (**Fig. 1d**), which contrasts with the

127 findings of a previous cross-sectional study showing that the oral microbiota of Tsimane infants

128 (9-24 months old) was distinct from that of their mothers[29]. The number of amplicon sequence

129 variants (ASVs) shared between body sites within infants was highest immediately after birth

130 (33.3%), and then declined with time (**Supplementary Fig. 1**). However, the ASV overlap

131 between body sites in 16-18 month olds was significantly higher than in adults (**Supplementary**

132 **Fig. 1**, Wilcoxon rank-sum test, p = 0.03).

133   Previous studies have reported a high prevalence of stunting and underweight in

134 Tsimane children older than 2 years, but low prevalence of undernutrition[30]. Within the current

135 study cohort of 48 Tsimane children, only one subject consistently exhibited height-for-age

136 (HAZ) and weight-for-age (WAZ) z-scores more than 2 SD below the mean. Five other subjects

137 had HAZ or WAZ scores that crossed from greater than to less than -2 SD after one year of age.

138 All infants in the longitudinal cohort were breastfeeding at all study time points. However, among

139 non-exclusively breastfed infants over 6 months of age, 63.8% of samples were collected at

140 times when the infant had a minimum dietary diversity (MDD) score of <4 (**Supplementary Fig.**

141 **2a**), indicating potential micronutrient inadequacy in their complementary foods[31]. MDD is based

142 on the 24-hour dietary recall of 7 key food groups, including grains, roots and tubers; legumes

143 and nuts; dairy products (milk, yogurt, cheese); flesh foods (meat, fish, poultry and liver/organ

6

144   meats); eggs; vitamin-A rich fruits and vegetables; and other fruits and vegetables. The infants'

145   MDD scores were positively correlated with infant gut diversity (**Supplementary Fig. 2a**, p =

146   0.027, conditional R^2 = 0.324), and accounted for a larger effect than age (β_{MDD} **=** 1.2 vs. β_{Age}

147   $_{(months)}$ = 0.6, p < 0.05 for both) on infant gut diversity. Chicha consumption, as well as vegetable

148   and dairy consumption, was significantly associated with gut microbiota composition, even after

149   controlling for age and village (**Fig. 1e**, partial CCA, ANOVA with 999 permutations, p < 0.05).

150   Interestingly, consuming chicha made from manioc or corn had an opposite association with the

151   infant gut microbiota than did consuming manioc or corn itself. In contrast to dietary factors,

152   fecal neopterin, a marker of interferon-activated monocytes and macrophages, and of intestinal

153   inflammation[32], was associated with significantly lower infant gut diversity, even when controlling

154   for age (**Supplementary Fig. 2b**, p = 0.002, conditional R^2 = 0.333).

155

156   *Sources of Infant-Colonizing Microbes*

157   To identify potential sources of infant-colonizing microbes, we first quantified the overlap

158   between infant and adult microbiotas. 84.3% of the ASVs found in all infant stool samples at all

159   time points were also found in at least one adult stool sample, while 79.9% of infant oral ASVs

160   were also found in at least one adult oral sample. On average, about half of the microbes

161   colonizing an infant's gut or tongue were also found in their mother's samples from the same

162   body site, and this percentage increased as the infant aged (**Supplementary Fig. 3a**, blue line

163   plus green line). However, the vast majority of these shared microbes were also found in other

164   adults living in the same village as the infant (**Supplementary Fig. 3a**, green line), reflecting a

165   high degree of microbial sharing among nearby adults, similar to that observed in other

166   indigenous populations[33]. Mothers from the same village shared significantly more gut microbes,

167   but not more oral microbes, than did mothers from different villages (**Supplementary Fig. 3b**,

168   Wilcoxon rank-sum test, p < 0.001), suggesting that village ecology might play a role in

169   structuring the gut microbiota. About half of the microbes found in an infant's gut were found in

7

170   at least one of their mother's stool samples (**Supplementary Fig. 3c**, purple line plus green

171   line), and about half were not found in any of their mother's samples (orange line), with very few

172   infant gut microbes ever observed on their mother's tongue (yellow line). The same was true for

173   infant tongue microbes, of which the majority of infant oral ASVs were observed on their

174   mother's tongue (**Supplementary Fig. 3c**, yellow line) and very few were observed in their

175   mother's gut (orange line).

176   Even though they accounted for only about half of the taxa in the infant distal gut and

177   tongue dorsum, the bacteria shared within a mother-infant dyad were found at higher relative

178   abundances in both the infants and their mothers than non-shared taxa. This difference in

179   average relative abundance was statistically significant until the infants reached about 12

180   months of age, after which the relative abundances of shared and non-shared taxa did not

181   significantly differ (**Fig. 2a,c**). At both body sites, the frequency at which a given microbe was

182   shared within a dyad was largely a function of its relative abundance in the mother. Microbes

183   with an average maternal abundance >0.1% were shared within a dyad at the same rate as its

184   prevalence, while microbes at 0.01% or lower abundance in mothers were almost never shared

185   within a dyad, even if the ASV was highly prevalent in infants (**Fig. 2b,d**). Overall, microbes that

186   were highly abundant in Tsimane mothers were more likely to be shared with their infants.

187

188   *Microbiota Assembly Rules*

189   Since many Tsimane ICABs may increase the opportunity for neutral dispersal of

190   microbes from mothers to their children, we sought to quantify the relative contributions of

191   deterministic and stochastic (*i.e.* neutral) processes during microbiota assembly. We used the

192   neutral community model (NCM)[34], which has previously been applied to the assembly of many

193   host-associated microbial communities[6,7], to address this issue. The NCM predicts the

194   prevalence of each microbe given its average relative abundance in a regional species pool.

195   Microbes that fit those predictions are inferred to be assembling neutrally, while microbes at

196   higher or lower prevalence are inferred to be under positive or negative ecological selection,

197   respectively. The NCM assumes that all microbes in a regional species pool have an equal per

198   capita ability to disperse to a local area, and once established, to have equal per capita fitness,

199   growth, and death rates[34].

200   The regional species pool was defined as containing all of the microbes able to disperse

201   to and colonize a local area, which in this case was the infant gut or tongue. We estimated the

202   composition of this pool by summing all of the local communities observed at a body site in all of

203   the infants in the cohort (*i.e.*, the 'infant pool'). However, since we were also interested in the

204   role that maternal microbes play in infant microbiota assembly, we assessed the degree to

205   which the maternal microbiota contributed to the regional species pool by calculating the NCM

206   goodness of fit (generalized R^2) when maternal samples from the same body site were used

207   instead to estimate the regional species pool (*i.e.*, the 'mother pool'). An R^2 value close to 1

208   would imply that the composition of the infant gut microbiota was consistent with neutral

209   dispersal, while a poor fit (*i.e.*, R^2 close to 0 or negative) would indicate that community

210   assembly was due to more than just neutral dispersal. In this way, the NCM served as a useful

211   mechanistic null model to help identify factors leading to a divergence from neutral dynamics or

212   specific taxa that assembled in a non-neutral manner, such as strong competitors or active

213   colonizers.

214   We found that microbes in the infant gut fit the neutral model moderately well when

215   using the infant pool (**Fig. 3a**, R^2 = 0.38), and that approximately two-thirds were neutrally

216   distributed, while 27.9% experienced positive selection. Only 5.3% experienced negative

217   selection; they included obligate anaerobes such as *Megasphaera*, *Prevotella*, *Bifidobacterium*,

218   *Bacteroides,* and *Libanicoccus*, suggesting that oxygen exposure could be an impediment to

219   transmission (**Fig. 3a**). Conversely, use of the mother pool resulted in a poorer model fit (R^2 = -

220   0.29), indicating that infant-colonizing maternal microbes were under either positive or negative

221   selection (**Fig. 3a**). Four microbes that were highly abundant and prevalent in the infant gut,

9

222   *Escherichia/Shigella* sp. (ASV 6812), *Bifidobacterium* sp. (ASV 4271), *Veillonella atypica* or *V.*

223   *dispar* (ASV 4083), and *Streptococcus* sp. (ASV 3537), were much less prevalent in the

224   maternal gut (**Fig. 3b**). Conversely, two *Prevotella* spp. ASVs (5359 and 5371) were much less

225   prevalent in the infant gut than the adult gut, especially in early life (**Fig. 3b**).

226   The NCM using the infant tongue swab pool predicted the distribution of infant oral

227   microbes well, as the majority of these taxa assembled according to neutral expectations (Fig.

228   3c, R^2 = 0.69). Interestingly, the model still fit relatively well using the mother tongue swab pool

229   (Fig. 3c, R^2 = 0.19), indicating that neutral mother-to-infant dispersal may play a more important

230   role during assembly of the infant oral microbiota than the gut microbiota. *Streptococcus* spp.

231   (ASVs 3524, 3522, and 3537), *Granulicatella adiacens* or *G. para-adiacens* (ASV 3641), and

232   *Veillonella* sp. (ASV 4083) were all highly prevalent in both infant and mother tongue swabs.

233   Two of these ASVs (4083 and 3537) were highly prevalent in the infant gut but not the adult gut,

234   suggesting that a small number of bacterial taxa may disperse from the adult mouth to the infant

235   gut and mouth.

236

237   *Infant Microbiota Assembly in a Global Context*

238   We next sought to determine whether the assembly patterns that we observed in the

239   Tsimane were conserved in infants from diverse cultures. We integrated our data with two

240   publicly available datasets generated with comparable molecular methods for profiling the distal

241   gut microbiota. The first of these datasets was created in a study of undernourished infants in

242   Bangladesh[26] (we used the well-nourished control group only), and the second in a large

243   observational study of type 2 diabetes in eastern Europe[27] (we used the Finnish infants only).

244   Since the European study did not include samples from adults, we also included gut microbiota

245   data from British adults to approximate the microbiota structure of a typical European, high-

246   income country[35]. Despite drastically different social and physical environments, access to

247   health care, and diets, the distal gut microbiota of the infants in these three studies were similar

10

248   in composition at the earliest ages based on their Jaccard distance, indicating similar initial

249   microbial colonists in societies across the globe (**Fig. 4a**). A principal curve was fit to each

250   dataset, representing the average trajectory of microbiota composition across time in infants

251   from that country. The age ranges for each infant cohort were 0.2 – 19.3 months for Bolivia

252   (Tsimane), 0 – 24.2 months for Bangladesh, and 1.2 – 36.9 months for Finland. As the children

253   became older, their microbiota trajectories diverged, and their microbiotas became more distinct

254   as they approached a population-specific adult-like state (**Fig. 4a**). By scaling the principal curve

255   position to the mean of the adult samples, and then regressing against the age of the subject we

256   found that infant microbiotas from these three countries matured, or reached an average adult-

257   like state, at significantly different rates (ANOVA, $p > 0.05$ for all pairwise comparisons)

258   (**Supplementary Fig. 4**).

259   In order to determine if neutral processes operate in similar manners among infants in

260   different cultural settings and geographies, we divided each dataset into three-month age

261   intervals up to one year of age, and then refit the NCM model using only the samples from that

262   interval as the estimate of the regional species pool, building a total of 12 additional models

263   (**Supplementary Fig. 5**). These datasets exhibited striking consistency in the role of neutral and

264   non-neutral processes across both age and geography, with the majority of microbes exhibiting

265   patterns of neutral assembly across individuals (**Fig. 4b**). Furthermore, by partitioning ASVs

266   according to country, age group, and how well they fit the neutral model, we found similar sets

267   of neutral and non-neutral microbes across geography and ontogeny (**Fig. 4c**, Jaccard

268   distance/PCoA). A heatmap of only the microbes that were positively selected revealed three,

269   *Bifidobacterium bifidum* (ASV 4264), *Bacteroides dorei or B. fragilis* (ASV 5620), *and Klebsiella*

270   sp. (ASV 6788) as common in all three cohorts (**Fig. 4e**). Other country-specific patterns were

271   also apparent, including a set of 6 *Bacteroides* ASVs (ASVs 5510, 5614, 5639, 5697, 5646, and

272   5684) that were prevalent in Finnish and Bolivian (Tsimane), but not Bangladeshi infants, as

273   well as two *Prevotella* ASVs (ASVs 5359 and 5371) that increased in prevalence with age in

11

274   both the Bolivian (Tsimane) and Bangladeshi infants, while remaining nearly absent in Finnish

275   infants (**Fig. 4e**). Interestingly, *Bacteroides* and *Prevotella* have been identified as biomarkers

276   for Western and non-Western lifestyles[36], respectively, yet both were prevalent in Tsimane

277   infants.

278   The neutral community model also predicts the probability that when an individual

279   microbe is removed from the local community, it is replaced by a microbe dispersing from the

280   regional species pool rather than by reproduction from within the local community[7,34]. This

281   migration term, *m*, increased with age in all three populations, although it increased significantly

282   more in the Tsimane infants (**Fig. 4d**). This significantly higher rate of microbial migration into

283   the Tsimane infant gut is consistent with their high exposure to maternal and environmental

284   microbes.

285

286   *Market Access Influences Tsimane Microbiota Structure*

287   In recent decades, the Tsimane have experienced profound changes to their culture and

288   life history, including increased access to processed foods in markets and greater access to

289   modern medical care. In order to understand how these lifestyle shifts may have affected the

290   assembly of the gut microbiota, we compared the gut microbiota structure of adults living in nine

291   Tsimane villages. These villages belonged to three village ecotypes, 'proximal river', 'distal

292   river', and 'forest', reflecting a gradient of cultural and lifestyle features currently in transition

293   (**Supplementary Table 1** and **Fig. 5a**), such as higher intake of processed food[24], increased

294   obesity and metabolic syndrome[37], and higher levels of urinary phthalates[38], suggesting that

295   increased economic integration and market access can have important health impacts. At the

296   time of sample collection, proximal river villages (≤ 20 km to market) were accessible by

297   road/taxi or boat, whereas distal river villages (>20 km to market) were only accessible by foot

298   or boat (**Supplementary Table 1**). Samples from the two proximal and four distal river villages

299   are those depicted in **Fig. 1a**. In addition, we included a single stool sample from each of 73

12

300   Tsimane individuals living in one of three forest villages (**Supplementary Table 1**). The set of

301   samples from the river villages was collected in September 2012 through March 2013, while that

302   from forest villages was collected in July 2009.

303   A principal coordinates analysis of these microbiota data showed clustering of stool

304   samples by ecotype of the village in which the subject lived (**Fig. 5b**, PERMANOVA, $p < 0.05$).

305   PC1, which explained 10.5% of the variation in the data, was significantly correlated with the

306   diversity of the microbiotas (**Supplementary Fig. 6**, $p < 0.001$, $R^2 = 0.289$), although there was

307   no significant difference in diversity between village types ($p < 0.05$ for all comparisons). To

308   evaluate whether these relationships might reflect an effect of market access on assembly

309   processes, we applied the NCM to samples from infants and adults living in the three different

310   village ecotypes with the regional species pool estimated from all of the samples from that age

311   group and village ecotype, and found that the goodness of fit decreased in samples from

312   villages with higher levels of market access (**Fig. 5c**). This suggested that market integration,

313   perhaps through increased access to processed foods (*i.e.* breads, pasta, refined sugar) and/or

314   medicine (*i.e.* antibiotics), may increase the role of ecological selection and diminish the role of

315   stochastic dispersal during assembly of the gut microbiota.

316   To identify how even subtle ecological and economic differences among villagers might

317   affect the gut microbiota, we used a phylogeny-based implementation of linear discriminant

318   analysis. This approach uses both the leaves of the phylogenetic tree (i.e., the ASVs), as well

319   as the nodes to identify linear combinations of features that characterize two or more sample

320   classes. The first model, built to discriminate between proximal and distal river villages, selected

321   3 predictors corresponding to 10 ASVs. Nine of those ASVs, including 2 in the family

322   Muribaculaceae (ASVs 5801 and 5805) and 7 in the family Bacteroidaceae (ASVs 5614, 5639,

323   5697, 5510, 5620, 5651, 5716) were discriminatory for the distal river samples, while only 1

324   ASV, in the family Prevotellaceae (ASV 5418), was discriminatory for the proximal river samples

325   (**Fig. 5d**). Even though diversity was not significantly lower in proximal river villages

326   (**Supplementary Fig. 6**), the small number of proximal village-discriminatory features is

327   consistent with studies suggesting that life in a high resource setting often leads to loss of gut

328   bacterial diversity[39,40]. Despite the small number of optimum features following 10-fold cross

329   validation, the model accurately predicted village type for 87.3% of samples (**Supplementary**

330   **Fig. 7**).

331   The second model, built to discriminate between river and forest villages, identified 24

332   predictors corresponding to 56 ASVs, and correctly identified the village type for 91.8% of the

333   samples (**Supplementary Fig. 7**). River villages were associated with ASVs in the bacterial

334   families Prevotellaceae (18 ASVs), Lachnospiraceae (13 ASVs), Veillonellaceae (4 ASVs),

335   Acidaminococcaceae (2 ASVs), and Ruminococcaceae (2 ASVs). Indicator taxa for Tsimane

336   forest villages included ASVs in the bacterial families Ruminococcaceae (4 ASVs), Clostridiales

337   Family XIII (2 ASVs), Lachnospiraceae (1 ASV), Muribaculaceae (1 ASV),

338   Peptostreptococcaceae (1 ASV), and Christensenellaceae (7 ASVs), as well as 1 archaeal ASV

339   in the family Methanobacteriaceae (1 ASV) (**Fig. 5d**).

340   Christensenellaceae and Methanobacteriaceae were previously identified in a study of

341   adult twins as members of a co-occurrence network of heritable gut microbiotas that were also

342   associated with body leanness[35,41]. Since Tsimane have low but increasing rates of obesity[23],

343   we sought to determine if a similar network was present in the Tsimane adults. The optimal

344   network topology constructed from adult stool microbial communities contained a subnetwork

345   involving the most highly connected node, ASV 12 (*Methanobrevibacter*), as well as 5 ASVs in

346   family Christensenellaceae (**Fig. 6a**), indicating strong associations among these taxa. This

347   subnetwork also included the second most highly connected node, *Prevotella copri* (ASV 5359).

348   Although these two taxa frequently co-occurred, their abundances were inversely related (**Fig.**

349   **6b**, $p < 0.001$, marginal $R^2 = 0.205$). *Methanobrevibacter and Prevotella* have been shown to be

350   positively correlated in the guts of humans[42] and western lowland gorillas[43], especially in the

351   setting of seasonal, carbohydrate-rich diets. These taxa also had distinct colonization dynamics

14

352   in the gut of infants from Bolivia (Tsimane), Bangladesh, and Finland. *Prevotella copri*, the

353   same ASV that was abundant in Bolivian and Bangladeshi infants but essentially absent in

354   Finnish infants, increased in abundance beginning at approximately 6 months of age, and

355   remained at high levels throughout adulthood (**Fig. 6c**). *Methanobrevibacter*, on the other hand,

356   was not detected in these gut microbiotas until later in adolescence (**Fig. 6c**). However, since

357   *Methanobrevibacter* was more common in forest villages, and these villages did not include

358   many samples from young children, this pattern may also have been a result of ecological,

359   seasonal, or societal differences between villages.

360

361   DISCUSSION

362   Here we demonstrate that the Tsimane, an indigenous forager-horticulturalist population,

363   display a distinct microbiota maturation trajectory among their infants. These colonization

364   dynamics may be the result of intensive ICABs. Strain-resolved investigations of mother-infant

365   transmission have found that most vertically transmitted microbes likely originate from the same

366   body site in the mother, and that colonization of strains originating from other maternal body

367   sites is relatively rare[44]. However, we observed evidence of maternal oral-to-infant gut

368   transmission among Tsimane participants, which could be a reflection of traditional intensive

369   ICABs like frequent premastication, and consumption of chicha inoculated with maternal oral

370   microbes. A study of Italian 4-month olds found that 9.5% of bacterial strains in their guts were

371   also found on their mother's tongue[44], whereas we observed an average of 23.7% of 3-5 month

372   old Tsimane infant gut ASVs on their mother's tongue. While metagenomic sequencing allows

373   investigators to track the transmission of specific strains, our 16S rRNA profiling approach,

374   paired with the inference of exact sequence variants, enabled us to perform broad surveys of

375   assembly patterns that can serve as an "upper bound" for estimating transmission rates. This

376   approach also allowed us to apply statistical models to quantify underlying ecological processes

377   that drive host-microbe biology. One important caveat is our assumption that microbes that were

15

378   shared between mothers and infants imply mother-to-infant transmission. Importantly, we

379   cannot directly observe the direction of transmission, nor can we exclude the possibility that

380   both mothers and children were colonized by microbes from other, unobserved sources.

381   The patterns of microbial prevalence and abundance fit the NCM less well in both adults

382   and infants living in proximal river villages than did those from villages with less market

383   integration. This suggests that the structure of the Tsimane microbiota may be changing in the

384   face of rapid socioeconomic shifts in ways that decrease the contribution of neutral processes

385   during childhood microbiota assembly, potentially altering interactions between highly abundant

386   microbes like *Methanobrevibacter* and *Prevotella copri.* Perhaps it is not surprising that

387   differences in subsistence strategies have been associated with differences in the gut

388   microbiota of indigenous populations[45], since even relatively modest differences in lifestyle can

389   also lead to differences in the gut microbiota[46]. A recent study identified microbes that were

390   differentially abundant in indigenous Bassa infants living in rural Nigeria versus infants living in

391   nearby urban centers[47], hinting that the observed microbiota differences in adults from

392   traditional societies may be influenced by early life factors[40]. Here, we suggest that the

393   differences observed between infants living in traditional societies and those living in or near

394   industrialized cities might be the result of reduced neutral dispersal of microbes that

395   accompanies the hygienic, medical, or societal changes associated with increased market

396   integration. Alternatively, these changes may be the result of maternal microbiotas that have a

397   decreased ability to stably colonize infants. Our findings lend support to the hypothesis that

398   familial microbial transmission in western societies deviates from traditional modes of

399   transmission[48].

400   While there are substantial differences in bacterial taxa between adult populations from

401   different regions and cultures, the prominent role of neutral assembly in infant guts is strikingly

402   consistent. This is especially important in the transmission of commensal microbes in

403   indigenous populations[49], since these relatively isolated societies experience low levels of

16

404   microbial dispersal from outside their local communities, and exhibit a high degree of

405   homogeneity in their microbiota composition[33]. Collectively, these observations of the microbiota

406   in traditional populations have broad implications for understanding how ecological interactions

407   operate during microbiota assembly, as well as how those patterns are altered due to broad

408   societal changes that have arisen in concert with increased participation in market economies.

409

410   METHODS

411   *Ethics Approval*

412   All study protocols were approved by the University of California Santa Barbara

413   Institutional Review Board on Human Subjects. Permission to conduct research was granted to

414   the Tsimane Life History Project (THLHP) and their research affiliates. The THLHP maintains

415   formal agreements with the local municipal government of San Borja and the Tsimane

416   governing body. Consent was obtained from village leaders and individual participants upon

417   initiating research activities in each village.

418

419   *Sample collection and processing*

420   The two sets of samples (2012-2013 and 2009) were collected as part of a long-term

421   study of Tsimane health and life history, which also included detailed surveys of the subject's

422   health, diet, and lifestyle. The first set of samples presented in this manuscript was collected

423   from September 2012 through March 2013, and focused on 52 mother-child dyads and used a

424   mixed longitudinal design. These individuals resided in six villages along the Maniqui River.

425   These river villages were grouped into 'proximal' or 'distal' villages, based on their proximity to

426   the same, nearest market town. Samples were collected together with data on breastfeeding

427   status, 24-hour dietary recall, anthropometrics, and symptoms of infectious disease, as part of a

428   study of changes in infant feeding and related maternal and infant health outcomes[9]. The

429   second analysis presented in this manuscript included samples that were collected in July 2009;

17

430   a cross-sectional design specified collection of single stool samples from each of 73 Tsimane

431   individuals ranging in age from 1 to 58 years old. These individuals resided in one of three forest

432   villages between 31 and 68 kilometers from the nearest regional market.

433   Adult and infant fecal samples were collected in plastic containers by the adult subjects,

434   which were then given to investigators typically within 1-2 hours. A portion of each sample was

435   transferred to a cryotube and immediately frozen in liquid nitrogen. Oral samples were collected

436   with a buccal cell collection swab (Catch-All[TM] Sample CollectionSwab; Epicentre[R]), transferred

437   to cryotubes and stored in liquid nitrogen. All samples were shipped on dry ice to a laboratory in

438   the United States, where they remained at -80°C until they were processed. DNA was extracted

439   from the samples using the MOBIO PowerSoil-htp Kit (MOBIO, Carlsbad, CA) following the

440   manufacturer's instructions, including a 2 x 10 minute bead-beating step using the Retsch 96

441   Well Plate Shaker at speed 20. The V4 hypervariable region of the 16S rRNA gene was then

442   PCR-amplified in triplicate using bacterial specific primers (515F and 806R) that include error-

443   correcting barcodes and Illumina adapters[50]. Amplicons were then pooled in equimolar ratios

444   before being sequenced on two runs of an Illumina MiSeq 2x250 PE, generating a total of 35.5

445   million raw reads.

446

447   *16S rRNA gene sequencing and data processing*

448   Raw reads were denoised into Amplicon Sequence Variants (ASVs) using the DADA2

449   pipeline (`dada2`, Version 1.9). Following generation of an ASV table, sequences were chimera-

450   checked, and those remaining that were not 230-235 bps in length were removed. In addition to

451   these data, publicly available 16S rRNA datasets from Bangladesh[26], Finland[27], and England[35]

452   were downloaded, run through the DADA2 pipeline, and combined at the level of ASVs with the

453   Tsimane dataset.

454   Taxonomic assignments were made using the RDP classifier and the SILVA nr database

455   v132. ASVs that were not identified as Domain *Bacteria* or Domain *Archaea* were excluded from

18

456   further analyses. A phylogenetic tree of the total set of ASVs was inferred using the fragment

457   insertion function (SEPP and `pplacer`) in QIIME2 and the full SILVA v132 tree. An aggregated

458   mapping file from all 4 of the datasets, the ASV table, the taxonomy table, and the phylogenetic

459   tree were imported into R and combined into a single phyloseq object (`phyloseq`, version

460   1.24.2).

461

462   *Data analysis*

463   Both alpha-diversity (Shannon Diversity Index) and beta-diversity (Jaccard Index, Bray-

464   Curtis Distance) analyses were performed in R. Statistical analyses including Wilcoxon tests

465   and PERMANOVA tests were also performed in R. PCoA and canonical correspondence

466   analysis (CCA) ordinations, boxplots, and scatter plots, were generated using the `ggplot2`

467   package in R (version 3.2.0). Heatmaps were generated using the `pheatmap` R package

468   (version 1.0.12). Amplicon sequence variant (ASV) abundances were centered and log ratio-

469   transformed before performing a PCoA on the Jaccard pairwise distances. The partial CCA was

470   performed on centered and log transformed ASV abundances constrained against a matrix of

471   diet survey data, including the 7 MDD food groups, as well as information on meal frequency,

472   frequency of consuming premasticated foods and liquids, frequency of consuming chicha, as

473   well as chicha type. The effects of infant age and village were controlled using a conditioning

474   matrix. Significance was assessed using an ANOVA-like permutation test with 999

475   permutations. The principal curve was fit to the PCoA in **Fig. 4a** using the `princurve` R

476   package (version 2.1.3).

477   For the treeDA analysis in **Fig. 5**, ASVs were removed if they were not observed at least

478   10 times in at least 20% of Tsimane adult stool samples. ASVs were inverse hyperbolic sine-

479   transformed before running the tree-based LDA algorithm in the treeDA R package (version

19

480   0.0.3). The LDA score of each adult stool sample was calculated from a phylogeny-based form

481   of linear discriminant analysis optimized with 10-fold cross validation.

482   Network analysis was performed using the SPIEC-EASI function in the `SpiecEasi` R

483   package (version 1.0.6) to infer an underlying graphical network model using both sparse

484   neighborhood and inverse covariance selection[51]. ASVs were removed that were not observed

485   at least 10 times in at least 20% of Tsimane adult stool samples. The network was constructed

486   with the SPIEC-EASI network inference algorithm using the Meinshausen-Buhlmann

487   Neighborhood Selection method. The size of the nodes were scaled to represent their degree of

488   connectedness.

489

490   *Modeling neutral community assembly*

491   As applied by Sloan *et al.* (2006), neutral assembly theory assumes that all microbes in

492   a regional species pool have an equal ability to disperse to a local area, and once established,

493   all have equal fitness, growth, and death rates[34]. These assumptions can be tested using a

494   nonlinear least squares model available in the `minpack.lm` R package (version 1.2-1) to

495   predict the prevalence of a microbe in a local community based on its average relative

496   abundance in the regional species pool. Microbial distributions that are consistent with the

497   model's predictions are well explained by neutral dispersal, while taxa that deviate from the

498   model likely experience either positive or negative selection. Our approach was a modified

499   version of one used in Burns *et al.*[7]; modeling and plotting functions are available in the `tyRa` R

500   package (version 0.1.0) available at https://danielsprockett.github.io/tyRa/.

501

502   *Data and Workflow Availability*

503   Raw sequencing data will be made available at NCBI Sequence Read Archive (SRA

504   Project ID pending). A detailed description of these analyses, along with all analysis code and

505   raw data, are available at the Stanford Digital Repository (https://purl.stanford.edu/tv993xn7633)

506   and as Supplementary Material.

507

508   ACKNOWLEDGEMENTS

519

REFERENCES

1. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).

2. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5,** e177 (2007).

3. Sharon, I. *et al.* Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res.* **23,** 111–120 (2013).

4. Yassour, M., Xavier, R. J. & Huttenhower, C. Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host Microbe* **24,** 146–154.e4 (2018).

5. Woodcock, S. *et al.* Neutral assembly of bacterial communities. *FEMS Microbiology Ecology* **62,** 171–180 (2007).

6. Venkataraman, A. *et al.* Application of a neutral community model to assess structuring of the human lung microbiome. *MBio* **6,** e02284-14 (2015).

7. Burns, A. R. *et al.* Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J.* **10**(3), 665-664 (2015).

8. Gurven, M. *et al.* The Tsimane Health and Life History Project: Integrating anthropology and biomedicine. *Evolutionary Anthropology: Issues, News, and Reviews* **26,** 54–73 (2017).

9. Martin, M. A., Garcia, G., Kaplan, H. S. & Gurven, M. D. Conflict or congruence? Maternal and infant-centric factors associated with shorter exclusive breastfeeding durations among the Tsimane. *Soc. Sci. Med.* **170,** 9–17 (2016).

10. Veile, A., Martin, M., McAllister, L. & Gurven, M. Modernization is associated with

545   intensive breastfeeding patterns in the Bolivian Amazon. *Soc. Sci. Med.* **100,** 148–

546   158 (2014).

547 11.   Rosinger, A. & Tanner, S. Water from fruit or the river? Examining hydration

548   strategies and gastrointestinal illness among Tsimane' adults in the Bolivian

549   Amazon. *Public Health Nutrition* **18,** 1098–1108 (2015).

550 12.   Freire, A. L., Zapata, S., Mosquera, J., Mejia, M. L. & Trueba, G. Bacteria

551   associated to human saliva are major microbial components of Ecuadorian

552   indigenous beers (chicha). **4**, e1962 (2016). doi:10.7287/peerj.preprints.1520v2

553 13.   Elizaquível, P. *et al.* Pyrosequencing vs. culture-dependent approaches to analyze

554   lactic acid bacteria associated to chicha, a traditional maize-based fermented

555   beverage from Northwestern Argentina. *Int. J. Food. Microbiol.* **198,** 9–18 (2015).

556 14.   Colehour, A. M. *et al.* Local domestication of lactic acid bacteria via cassava beer

557   fermentation. *PeerJ* **2,** e479 (2014).

558 15.   Resende, L. V. *et al.* Microbial community and physicochemical dynamics during

559   the production of 'Chicha', a traditional beverage of Indigenous people of Brazil.

560   *World J Microbiol Biotechnol* **34,** 46

561 16.   Puerari & Schwan, R. F. Physicochemical and microbiological characterization of

562   chicha, a rice-based fermented beverage produced by Umutina Brazilian

563   Amerindians. *Food Microbiology* **46,** 210–217 (2015).

564 17.   McAllister, L., Gurven, M., Kaplan, H. & Stieglitz, J. Why do women have more

565   children than they want? Understanding differences in women's ideal and actual

566   family size in a natural fertility population. *Am. J. Hum. Biol.* **24,** 786–799 (2012).

567 18.   Gurven, M. Infant and fetal mortality among a high fertility and mortality population

568   in the Bolivian Amazon. *Soc. Sci. Med.* (2012).

569 19.   Blackwell, A. D. *et al.* Evidence for a peak shift in a humoral response to helminths:

570   age profiles of IgE in the Shuar of Ecuador, the Tsimane of Bolivia, and the U.S.

571   NHANES. *PLoS Negl Trop Dis* **5,** e1218 (2011).

572   20.   McDade, T. W. *et al.* Predictors of C☐reactive protein in Tsimane' 2 to 15

573   year☐olds in lowland Bolivia. *Am. J. Phys. Anthropol.* **128,** 906–913 (2005).

574   21.   Vasunilashorn, S. *et al.* Blood lipids, infection, and inflammatory markers in the

575   Tsimane of Bolivia. *Am. J. Hum. Biol.* **22,** 731–740 (2010).

576   22.   Vasunilashorn, S. *et al.* Inflammatory Gene Variants in the Tsimane, an Indigenous

577   Bolivian Population with a High Infectious Load. *Biodemography & Social Biol.* **57,**

578   33–52 (2011).

579   23.   Kaplan, H. *et al.* Coronary atherosclerosis in indigenous South American Tsimane:

580   a cross-sectional cohort study. *Lancet* **389,** 1730–1739 (2017).

581   24.   Kraft, T. S. *et al.* Nutrition transition in 2 lowland Bolivian subsistence populations.

582   *Am. J. Clin. Nutr.* **108,** 1183–1195 (2018).

583   25.   Bethancourt, H. J., Leonard, W. R., Tanner, S., Schultz, A. F. Longitudinal changes

584   in measures of body fat and diet among adult Tsimane' forager-horticulturalists of

585   Bolivia, 2002-2010. *Obesity* **27**(8), 1347-1359 (2019).

586   26.   Subramanian, S. *et al.* Persistent gut microbiota immaturity in malnourished

587   Bangladeshi children. *Nature* **510,** 417–421 (2014).

588   27.   Vatanen, T., Kostic, A. D., d'Hennezel, E. & Siljander, H. Variation in Microbiome

589   LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell* **165**(4), 842-853

590   (2016).

591   28.   DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during

592   pregnancy. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 11060–11065 (2015).

593   29.   Han, C. S. *et al.* Salivary microbiomes of indigenous Tsimane mothers and infants

594   are distinct despite frequent premastication. *PeerJ* **4,** e2660 (2016).

595   30.   Blackwell, A. D. *et al.* Growth references for Tsimane forager-horticulturalists of the

596   Bolivian Amazon. *Am. J. Phys. Anthropol.* **162,** 441–461 (2017).

597 31. World Health Organization. Indicators for Assessing Infant and Young Child

598 Feeding Practices. 1–26 (WHO Press, 2008).

599 32. Guerrant, R. L. *et al.* Biomarkers of environmental enteropathy, inflammation,

600 stunting, and impaired growth in children in northeast Brazil. *PLoS ONE* **11,**

601 e0158772 (2016).

602 33. Clemente, J. C., Pehrsson, E. C. & Blaser, M. J. The microbiome of uncontacted

603 Amerindians. *Science Advances* **1**(3) e1500183 (2015).

604 34. Sloan, W. T. *et al.* Quantifying the roles of immigration and chance in shaping

605 prokaryote community structure. *Environmental Microbiology* **8,** 732–740 (2006).

606 35. Goodrich, J. K. *et al.* Genetic Determinants of the Gut Microbiome in UK Twins. *Cell*

607 *Host Microbe* **19,** 731–743 (2016).

608 36. Gorvitovskaia, A., Holmes, S. P. & Huse, S. M. Interpreting Prevotella and

609 Bacteroides as biomarkers of diet and lifestyle. *Microbiome* **4,** 15 (2016).

610 37. Rosinger, A. Precursors to overnutrition: The effects of household market food

611 expenditures on measures of body composition among Tsimane adults in lowland

612 Bolivia. *Soc. Sci. Med.* **92,** 53–60 (2013).

613 38. Sobolewski, Weiss & Martin, M. Toxicoanthropology: Phthalate exposure in relation

614 to market access in a remote forager-horticulturalist population. *International*

615 *Journal of Hygiene and Environmental Health* **220,** 799–809 (2017).

616 39. Smits, S. A. *et al.* Seasonal cycling in the gut microbiome of the Hadza hunter-

617 gatherers of Tanzania. *Science* **357,** 802–806 (2017).

618 40. Mancabelli, L. *et al.* Meta-analysis of the human gut microbiome from urbanized

619 and pre-agricultural populations. *Environmental Microbiology* **19,** 1379–1390

620 (2017).

621 41. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159,** 789–799

622 (2014).

623 42.   Hoffmann, C. *et al.* Archaea and Fungi of the Human Gut Microbiome: Correlations

624   with Diet and Bacterial Residents. *PLoS ONE* **8,** e66019 (2013).

625 43.   Hicks, A. L. *et al.* Gut microbiomes of wild great apes fluctuate seasonally in

626   response to diet. *Nat. Commun.* **9,** 1786 (2018).

627 44.   Ferretti, P. *et al.* Mother-to-Infant Microbial Transmission from Different Body Sites

628   Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe* **24,** 133–145.e5

629   (2018).

630 45.   Obregon-Tito, A. J. *et al.* Subsistence strategies in traditional societies distinguish

631   gut microbiomes. *Nat. Commun.* **6,** 1–9 (2015).

632 46.   Stagaman, K. *et al.* Market Integration Predicts Human Gut Microbiome Attributes

633   across a Gradient of Economic Development. *mSystems* **3,** e00122–17 (2018).

634 47.   Ayeni, F. A. *et al.* Infant and Adult Gut Microbiome and Metabolome in Rural Bassa

635   and Urban Settlers from Nigeria. *Cell Rep.* **23,** 3056–3067 (2018).

636 48.   Blaser, Martin The theory of disappearing microbiota and the epidemics of chronic

637   diseases. *Nat. Rev. Immunol.* **17**, 461–463 (2017).

638 49.   Brito, I. L. *et al.* Transmission of human-associated microbiota along family and

639   social networks. *Nat. Microbiol.* **4**, 964–971 (2019).

640 50.   Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the

641   Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–1624 (2012).

642 51.   Kurtz, Z. D. *et al.* Sparse and compositionally robust inference of microbial

643   ecological networks. *PLoS Comput. Biol.* **11,** e1004226 (2015).

644

645

646   FIGURE LEGENDS

647   Figure 1. **Infant microbiota dynamics and dietary factors associated with maturation.** (a)

648   Timeline denoting fecal sample and tongue swab collections for each mother-child dyad relative

649   to the infant's birth date. For maternal samples, time refers to the immediate postpartum period.

650   Infant stool samples are circles, adult stool samples are squares, infant and adult tongue swabs

651   are triangles that point up and down, respectively. Shapes are colored by sample type (red =

652   stool, blue = tongue swabs), and the color darkens as the subject's age increases. Shapes and

653   colors are consistent across Figs. 1a–e. (b) Shannon Diversity Index regressed against the time

654   since the infant's birth for stool samples. Lines indicate the linear mixed-effects regression of

655   diversity on time since delivery, while treating subject as a random effect. The shading indicates

656   the 95% confidence interval. The conditional R^2 describes the proportion of variation explained

657   by both the fixed and random factors. (c) Shannon Diversity Index regressed against the time

658   since the infant's birth for tongue swab samples. Figure details are the same as in Fig. 1b. (d)

659   Principal coordinate analysis (PCoA) of microbiota taxa composition in stool samples and

660   tongue swabs from Tsimane dyads. (e) A partial canonical correspondence analysis (CCA) of

661   ASV abundances, constrained against a matrix of diet survey data. The effects of infant age and

662   village were controlled using a conditioning matrix. Significance was assessed using an

663   ANOVA-like permutation test with 999 permutations, and only significant (p < 0.05) factors are

664   displayed.

665

666   Figure 2. **Shared microbes are highly abundant within mother-infant dyads.** (a) Average

667   relative abundance of ASVs that are shared or not shared within the stool samples of an infant-

668   mother dyad plotted against the time since infant birth. Lines represent the best fit of a

669   generalized additive model to either shared or not shared ASVs in the mother's (left) or infant's

670   (right) stool samples. (b) Relationship between ASV prevalence in infant gut samples and the

27

671   rate at which the ASV is shared within infant-mother dyads. The points are colored according to

672   their average relative abundance in maternal gut samples. (c) Average relative abundance of

673   ASVs that are shared or not shared within the tongue swabs of an infant-mother dyad plotted

674   against the time since infant birth. Figure details are the same as in Fig. 2a. (d) Relationship

675   between ASV prevalence in infant tongue swabs and the rate at which the ASV is shared within

676   infant-mother dyads. Figure details are the same as in Fig. 2b.

677

678   Figure 3. **Infant microbiotas assemble according to neutral rules.** (a) Neutral community

679   model fit to each ASV observed in the stool of Tsimane infants, using either infant stool pool

680   (left) or adult stool pool (right), as estimates of the regional species pool. Points are colored

681   according to whether the taxon prevalence in infant samples is above (green), at (yellow), or

682   below (red) the predicted prevalence according to the NCM. (b) Heatmap of the log normalized

683   abundances of the ASVs from infant (left) and mother (right) stool samples that were observed

684   in at least 25% of infant stool samples (n = 32). Rows were clustered on the y-axis using Ward's

685   minimum variance method. Row label colors correspond to the point colors in Fig. 3a. Columns

686   were sorted by age for infants, and time since birth for mothers. (c) Neutral community model fit

687   to each ASV observed from the tongue swabs of Tsimane infants, using either the infant tongue

688   swab pool (left) or adult tongue swab pool (right), as estimates of the regional species pool.

689   Figure details are the same as in Fig. 3a. (d) Heatmap of the log normalized abundances of

690   ASVs from infant (left) and mother (right) stool samples that were observed in at least 25% of

691   infant samples (n = 27). Figure details are the same as in Fig. 3b.

692

693   Figure 4. **Patterns of selection and neutral assembly are consistent across ontogeny and**

694   **geography.** (a) Principal coordinate analysis (PCoA) of stool samples from Bolivian (Tsimane)

695   dyads, as well as previously published 16S rRNA datasets of Bangladeshi and Finnish infants,

28

696   as well as English adults. Points are shaped and colored according to country, and the color

697   darkens with increased age of the subject (Finland - blue squares, England – green squares,

698   Bolivia (Tsimane) – red circles, Bangladesh – yellow triangles). A principal curve was fit for each

699   of the three populations (Bolivian (Tsimane), Bangladeshi, and European (Finnish/English)). (b)

700   A summary of the proportion of ASVs present in each country and age group that fit the model

701   (yellow), or was above (green) or below (red) the model's prediction. See Supplementary Fig. 5

702   for details. (c) Principal coordinate analysis using the Jaccard distance for each of the 3 data

703   partitions of four age groups from 3 countries (36 measurements in total). The color indicates

704   the predicted fit, the shape indicates the country whence the samples were collected, and the

705   size of each shape is scaled to indicate the age group of the subject. (d) The estimated

706   migration rate (m) given by the neutral community model for each country and age group.

707   Vertical lines represent 95% confidence intervals. Colors and shapes are the same as in Fig.

708   4a. (e) A heatmap of the log normalized abundances of ASVs that were estimated to be above

709   their predicted prevalence given their average relative abundance in all 12 green partitions in

710   Fig. 4c. Rows were clustered on the y-axis using Ward's minimum variance method. Columns

711   were sorted by infant age (0-12 months).

712

713   Figure 5. **Market access is associated with Tsimane gut microbiota structure.** (a) A map

714   showing the location of the 9 Tsimane villages whence stool samples were collected. The size

715   of each circle was scaled to the number of samples collected from that village, with the

716   exception of the market town, where no samples were collected. The map insert denotes the

717   region of Bolivia being displayed. (b) Principal coordinate analysis (PCoA) on the Bray-Curtis

718   distance of 16S rRNA profiles from adult stool samples. Village types ('proximal' river, 'distal'

719   river, and 'forest') have significantly different distributions (PERMANOVA with 1000

720   permutations, p = 0.001). Samples are colored by village and their shapes denote the village

29

721   type (proximal river villages – circles, distal river villages – triangles, forest villages – squares).

722   (c) Neutral community model fit to each ASV observed in the stool of Tsimane adults (top row)

723   or infants (bottom row) from each of three village types (proximal river – left, distal river – right,

724   and forest – left). In each case, the composition of the regional species pool was estimated

725   using the samples from the age group and village type. The gray line is the predicted distribution

726   (shaded area is 95% CI), and points are colored according to whether their prevalence in infant

727   samples was above (green), at (yellow), or below (red) their predicted prevalence according to

728   the NCM. (d) Phylogenetic tree with ASVs found at least 10 times in at least 20% of Tsimane

729   adult stool samples. Discriminatory ASVs were identified using the tree-based LDA algorithm in

730   the `treeDA` R package (version 0.0.3). ASVs are denoted by shapes based on two sets of

731   comparisons (circles for river vs. forest and squares for proximal river vs. distal river), and are

732   colored based on their discriminatory strength. ASVs are colored by taxonomic family and

733   significantly discriminatory ASVs are labeled with their ASV number.

734

735   Figure 6. **Network analysis reveals an inverse relationship between highly-connected**

736   **taxa.** (a) A network diagram of ASVs observed at least 10 times in at least 20% of Tsimane

737   adult stool samples. The size of the node was scaled to represent its degree of connectedness.

738   The insert shows the complete network colored by network degree, while the main figure shows

739   the highly-connected subgraph colored by each ASV's taxonomic family. Positive associations

740   between ASV nodes are colored gray, while negative associations are colored red. (b) A scatter

741   plot of the log transformed relative abundances of the two most highly connected ASVs in the

742   network, *Methanobrevibacter* sp. (ASV 12) and *Prevotella copri* (ASV 5359). The red line

743   indicates the linear mixed-effects regression while treating subject as a random effect, and

744   shading indicates the 95% confidence interval. Samples are colored by village, and their shapes

745   represent the village type (proximal river – circles, distal river – triangles, forest – squares). (c)

746 Scatter plot of the relative abundances of *Methanobrevibacter* sp. (ASV 12) and *Prevotella copri*

747 (ASV 5359) and the subject's age in months at the time of collection. The x-axis was log

748 transformed for clarity when plotting both infant and adult samples. Ages 1, 2, 3, 4, 5, 10, 20,

749 30, 40, 50, and 60 years are denoted by horizontal dashed lines.

750

751 Supplementary Figure 1. **Gut and oral microbiotas diverge as infants age.** Average number

752 of ASVs shared between body sites within a subject based on samples collected within 14 days

753 of each other. Boxplot color darkens with increasing age. Adults and 16-18 month olds were the

754 only sequential age groups whose microbiotas were significantly different from one another

755 (Wilcoxon rank-sum test, p = 0.03).

756

757 Supplementary Figure 2. **Diet and inflammation impact the infant gut microbiota.** (a) The

758 World Health Organization's minimum diet diversity (MDD) score regressed against the

759 Shannon Diversity Index from the infant stool samples. The MDD was validated to assess the

760 diversity of complementary foods of 6-23 month old children in a diverse range of cultural

761 contexts. It should only be interpreted as an estimate of micronutrient sufficiency in children

762 aged 6-23 months, which are denoted by a diamond shape, and the color darkens as the

763 subject's age increases. MDD of 4 is considered to be sufficiently diverse, as indicated by the

764 vertical dashed line. Lines indicate the linear mixed-effects regression of diversity on MDD,

765 while treating subject as a random effect. The shading indicates the 95% confidence interval.

766 The conditional R^2 describes the proportion of variation explained by both fixed and random

767 factors. (B) Levels of fecal neopterin, an inflammatory marker normally produced by activated

768 macrophages, regressed against the Shannon Diversity Index in infant stool samples. Lines

769 indicate the linear mixed-effects regression of diversity on neopterin, while treating subject as a

770 random effect. The shading indicates the 95% confidence interval. The conditional R^2 describes

771   the proportion of variation explained by both fixed and random factors. A neopterin level above

772   10, as indicated by the vertical dashed line, indicates clinically relevant inflammation.

773

774   Supplementary Figure 3. **Sources of infant-colonizing microbes.** (a) Scatter plot of the

775   average number of ASVs found in each infant stool sample that were also found in that infant's

776   mother but not in other adults from their village (blue), in both their mother and other mothers

777   from their village (green), in other mothers from the infant's village but not their mother (yellow),

778   or in neither their mother or other adults (orange). Non-parent adults from each village were

779   down-sampled to control for differences in the number of samples collected from each village.

780   Left panel – infant stool samples, right panel – infant tongue swabs. (b) Average number of

781   ASVs shared between multiple samples from the same mother, between mothers living in the

782   same village, or mothers living in different villages. Wilcoxon Rank Sum Test. ∗∗∗∗: $p \leq 0.0001$,

783   ∗∗: $p \leq 0.01$, ∗: $p \leq 0.05$, ns: $p > 0.05$. Left panel – adult stool samples, right panel – adult

784   tongue swabs. (c) Scatter plot of the average number of ASVs in each infant stool sample that

785   was also found in their mother's stool but not in her tongue swab (purple), in both their mother's

786   stool and tongue swab (green), in their mother's tongue swab but not their mother's stool

787   (yellow), or not found in any of their mother's samples (orange).

788

789   Supplementary Figure 4. **Infants from different countries reach a population-specific adult-**

790   **like state at different rates.** (a) In Fig. 4a, a PCoA was performed on 16S rRNA datasets from

791   three populations (Bolivian (Tsimane), Bangladeshi, and European (Finnish/British)), and a

792   principal curve was fit to each set of points using the `princurve` function in the `princurve` R

793   package (version 2.1.3). The distance along each curve was scaled so that the mean location of

794   the adult samples was equal to one (as indicated by the horizontal gray line), making an infant's

795   location along the curve analogous to their position along the average maturational trajectory

796   towards the average adult in their society. Here, we show the linear regression of the age of the

797   infant from whom each sample was collected against its position along its principal curve. Lines

798   indicate the linear mixed-effects regression of principal curve position on the infant's age in

799   months, while treating subject as a random effect. The shading indicates the 95% confidence

800   interval. The vertical dashed lines indicate the age at which the infant line from each country

801   crosses the adult average position for that country.

802

803   Supplementary Figure 5. **Neutral processes explain microbial community assembly across**

804   **age groups and geography.** Scatterplot of the prevalence of each ASV in the infant gut versus

805   its mean relative abundance in the regional species pool for each age group and country. The

806   gray line is their predicted distribution (shaded area is 95% confidence interval) based on the

807   neutral community model. Points are colored by the ASV's fit to the model: above prediction–

808   green, below prediction– red, at prediction – yellow.

809

810   Supplementary Figure 6. **Gut microbiota diversity among Tsimane villages.** (a) Shannon

811   Diversity Index plotted against the PC1 value from Fig. 5b (R^2 = 0.289, p < 0.001). The red line

812   indicates the linear mixed-effects regression while treating subject as a random effect, and

813   shading indicates the 95% confidence interval. The conditional R^2 describes the proportion of

814   the variation explained by both fixed and random factors. (b) Boxplot comparing the average

815   diversity of adult stool microbiota of samples collected from proximal river, distal river, and forest

816   Tsimane villages (p > 0.05 for all comparisons).

817

818   Supplementary Figure 7. **Bacterial taxa discriminate between village types.** (a) The LDA

819   score of each adult stool sample was calculated from a phylogeny-based form of linear

820   discriminant analysis (treeDA, version 0.0.3). Higher scores indicate samples more likely to

821   have been collected from a proximal river village, while lower scores indicate samples more

822   likely to have been collected from a distal river village. The dashed gray line represents the

823   optimal LDA score cut-off between groups, where it accurately classifies 87.3% of samples.

824   Samples are colored by village, and their shapes represent the village type (proximal river –

825   circles, distal river – triangles, forest – squares). (b) The LDA score of each adult stool sample

826   was calculated from a phylogeny-based form of linear discriminant analysis (treeDA, version

827   0.0.3). Higher scores indicate samples more likely from a forest village, while lower scores

828   indicate samples more likely from a river village. The dashed gray line represents the optimal

829   LDA score cut-off between groups, where it accurately classifies 91.8% of samples. The colors

830   and shapes are the same as in panel a.

831

832   **Supplementary Table 1. Demographics of Tsimane villages and study subjects**

833   Villages are listed in order of increasing distance from the closest market town. Age classes are

834   defined as Infant – under 2 years, Child – 2 to 13 years old, Adult – 14 years old and older. The

835   73 fecal samples collected from forest communities were collected in July 2009, while the

836   remaining fecal and oral samples from proximal and distal river communities were collected

837   between September 2012 and March 2013. *Population sizes for forest villages were collected

838   during a 2009 census, while the remaining were collected in a 2012 census.
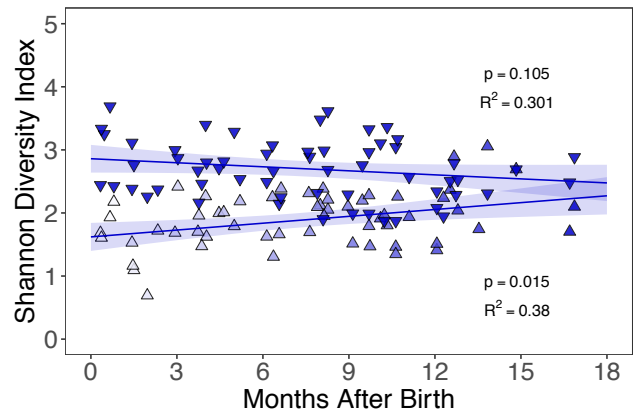
Figure 1
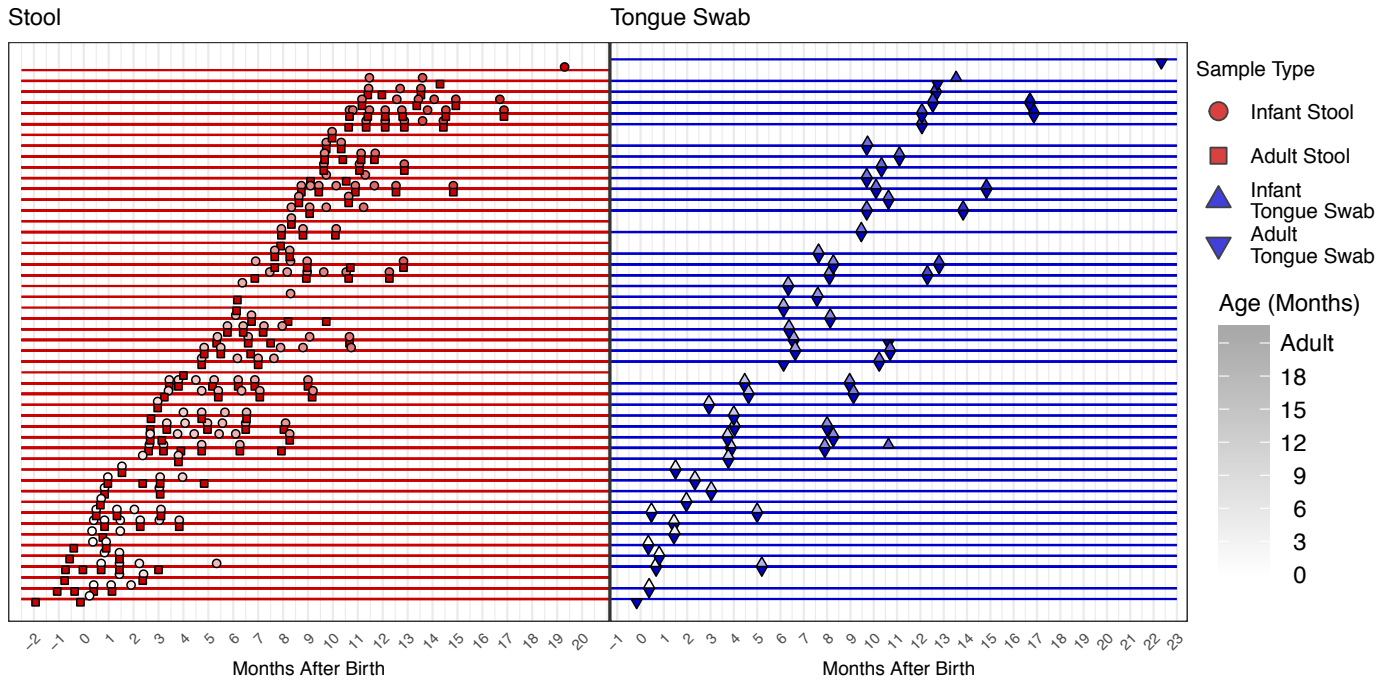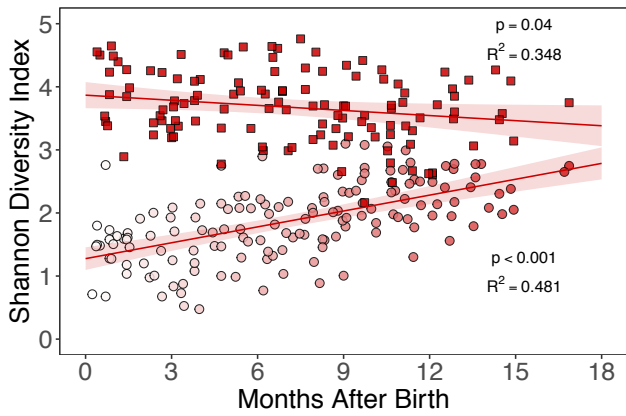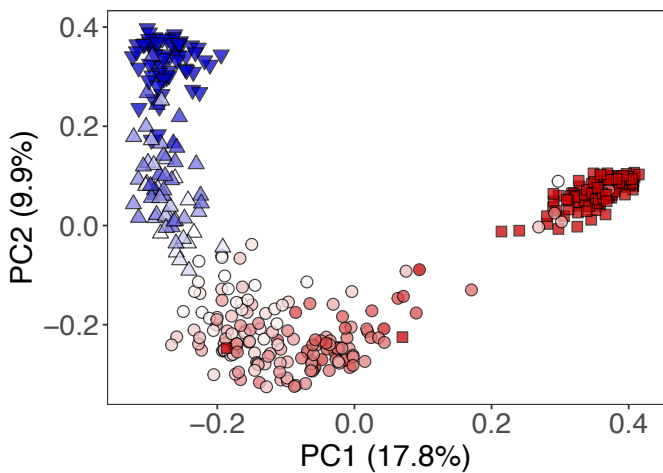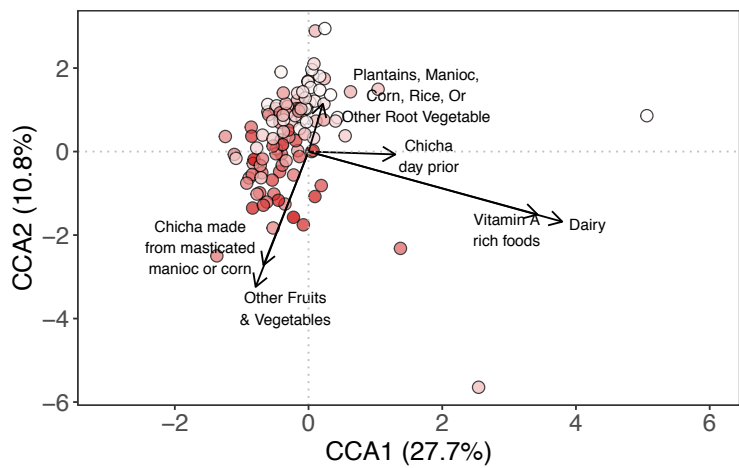
Figure 1. **Infant microbiota dynamics and dietary factors associated with maturation.**
(a) Timeline denoting fecal sample and tongue swab collections for each mother-child dyad relative to the infant's birth date. For maternal samples, time refers to the immediate postpartum period. Infant stool samples are circles, adult stool samples are squares, infant and adult tongue swabs are triangles that point up and down, respectively. Shapes are colored by sample type (red = stool, blue = tongue swabs), and the color darkens as the subject's age increases. Shapes and colors are consistent across Figs. 1a–e. (b) Shannon Diversity Index regressed against the time since the infant's birth for stool samples. Lines indicate the linear mixed-effects regression of diversity on time since delivery, while treating subject as a random effect. The shading indicates the 95% confidence interval. The conditional R^2 describes the proportion of variation explained by both the fixed and random factors. (c) Shannon Diversity Index regressed against the time since the infant's birth for tongue swab samples. Figure details are the same as in Fig. 1b. (d) Principal coordinate analysis (PCoA) of microbiota taxa composition in stool samples and tongue swabs from Tsimane dyads. (e) A partial canonical correspondence analysis (CCA) of ASV abundances, constrained against a matrix of diet survey data. The effects of infant age and village were controlled using a conditioning matrix. Significance was assessed using an ANOVA-like permutation test with 999 permutations, and only significant ($p < 0.05$) factors are displayed.
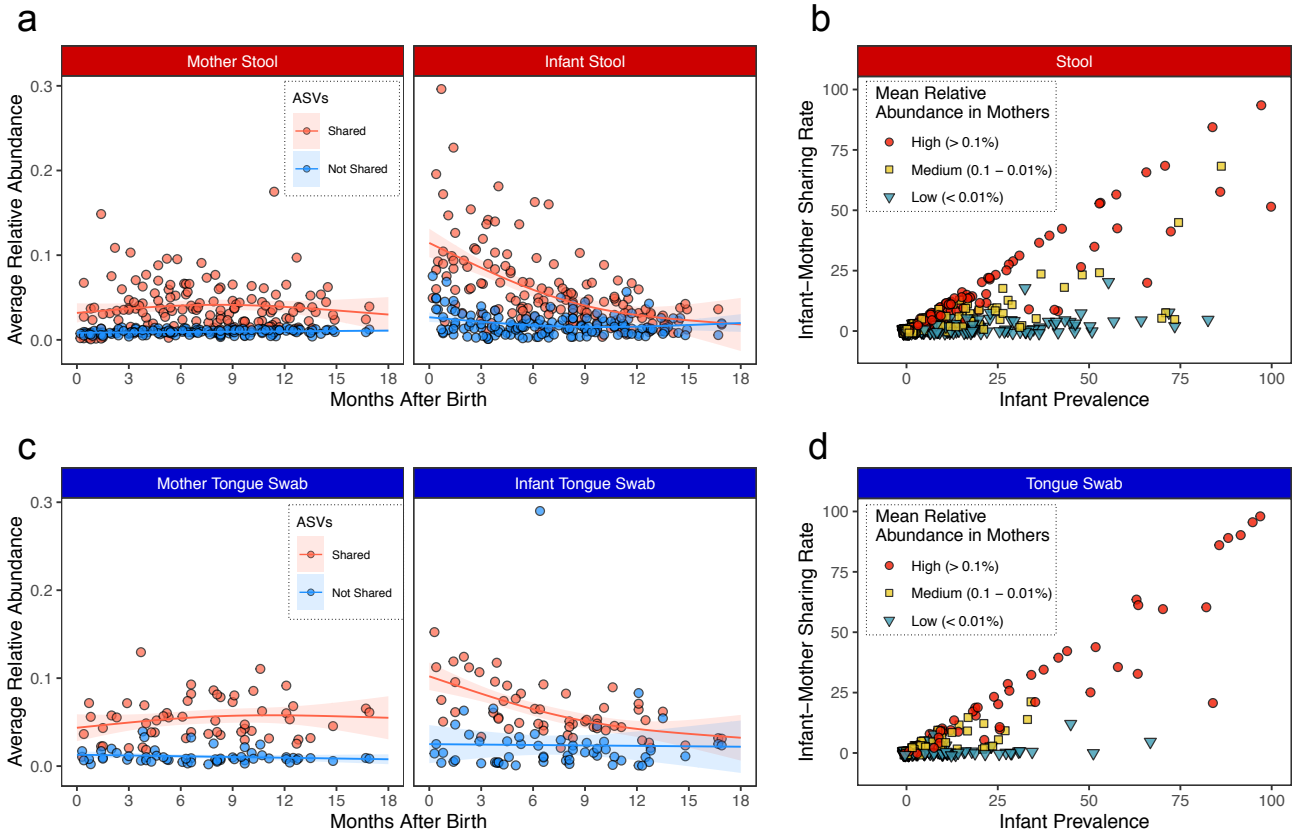
Figure 2

Figure 2. **Shared microbes are highly abundant within mother-infant dyads.** (a) Average relative abundance of ASVs that are shared or not shared within the stool samples of an infant-mother dyad plotted against the time since infant birth. Lines represent the best fit of a generalized additive model to either shared or not shared ASVs in the mother's (left) or infant's (right) stool samples. (b) Relationship between ASV prevalence in infant gut samples and the rate at which the ASV is shared within infant-mother dyads. The points are colored according to their average relative abundance in maternal gut samples. (c) Average relative abundance of ASVs that are shared or not shared within the tongue swabs of an infant-mother dyad plotted against the time since infant birth. Figure details are the same as in Fig. 2a. (d) Relationship between ASV prevalence in infant tongue swabs and the rate at which the ASV is shared within infant-mother dyads. Figure details are the same as in Fig. 2b.
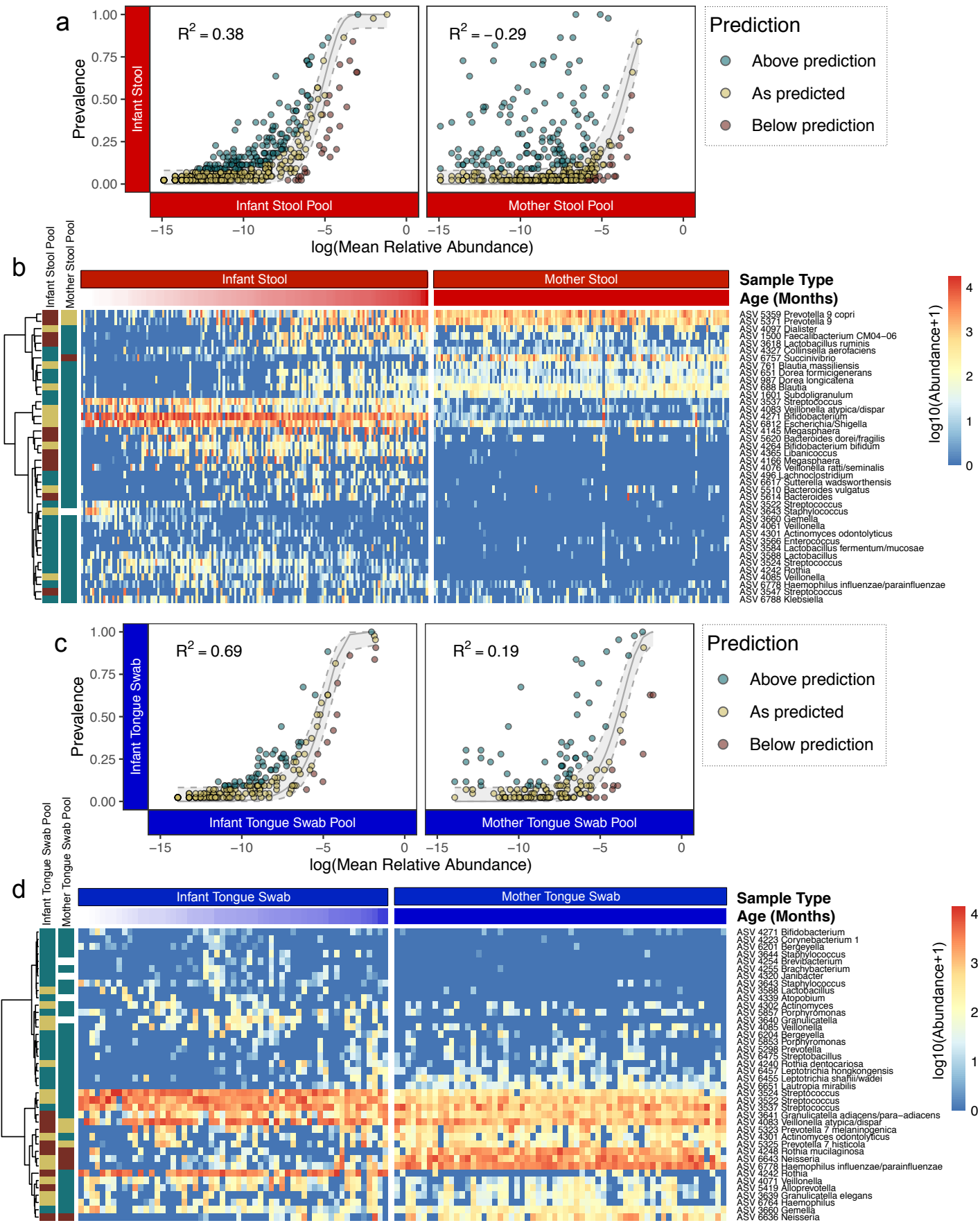
Figure 3

Figure 3. **Infant microbiotas assemble according to neutral rules.** (a) Neutral community model fit to each ASV observed in the stool of Tsimane infants, using either infant stool pool (left) or adult stool pool (right), as estimates of the regional species pool. Points are colored according to whether the taxon prevalence in infant samples is above (green), at (yellow), or below (red) the predicted prevalence according to the NCM. (b) Heatmap of the log normalized abundances of the ASVs from infant (left) and mother (right) stool samples that were observed in at least 25% of infant stool samples (n = 32). Rows were clustered on the y-axis using Ward's minimum variance method. Row label colors correspond to the point colors in Fig. 3a. Columns were sorted by age for infants, and time since birth for mothers. (c) Neutral community model fit to each ASV observed from the tongue swabs of Tsimane infants, using either the infant tongue swab pool (left) or adult tongue swab pool (right), as estimates of the regional species pool. Figure details are the same as in Fig. 3a. (d) Heatmap of the log normalized abundances of ASVs from infant (left) and mother (right) stool samples that were observed in at least 25% of infant samples (n = 27). Figure details are the same as in Fig. 3b.

Figure 4

Figure 4. **Patterns of selection and neutral assembly are consistent across ontogeny and geography.** (a) Principal coordinate analysis (PCoA) of stool samples from Bolivian (Tsimane) dyads, as well as previously published 16S rRNA datasets of Bangladeshi and Finnish infants, as well as English adults. Points are shaped and colored according to country, and the color darkens with increased age of the subject (Finland - blue squares, England – green squares, Bolivia (Tsimane) – red circles, Bangladesh – yellow triangles). A principal curve was fit for each of the three populations (Bolivian (Tsimane), Bangladeshi, and European (Finnish/English)). (b) A summary of the proportion of ASVs present in each country and age group that fit the model (yellow), or was above (green) or below (red) the model's prediction. See Supplementary Fig. 5 for details. (c) Principal coordinate analysis using the Jaccard distance for each of the 3 data partitions of four age groups from 3 countries (36 measurements in total). The color indicates the predicted fit, the shape indicates the country whence the samples were collected, and the size of each shape is scaled to indicate the age group of the subject. (d) The estimated migration rate (m) given by the neutral community model for each country and age group. Vertical lines represent 95% confidence intervals. Colors and shapes are the same as in Fig. 4a. (e) A heatmap of the log normalized abundances of ASVs that were estimated to be above their predicted prevalence given their average relative abundance in all 12 green partitions in Fig. 4c. Rows were clustered on the y-axis using Ward's minimum variance method. Columns were sorted by infant age (0-12 months).
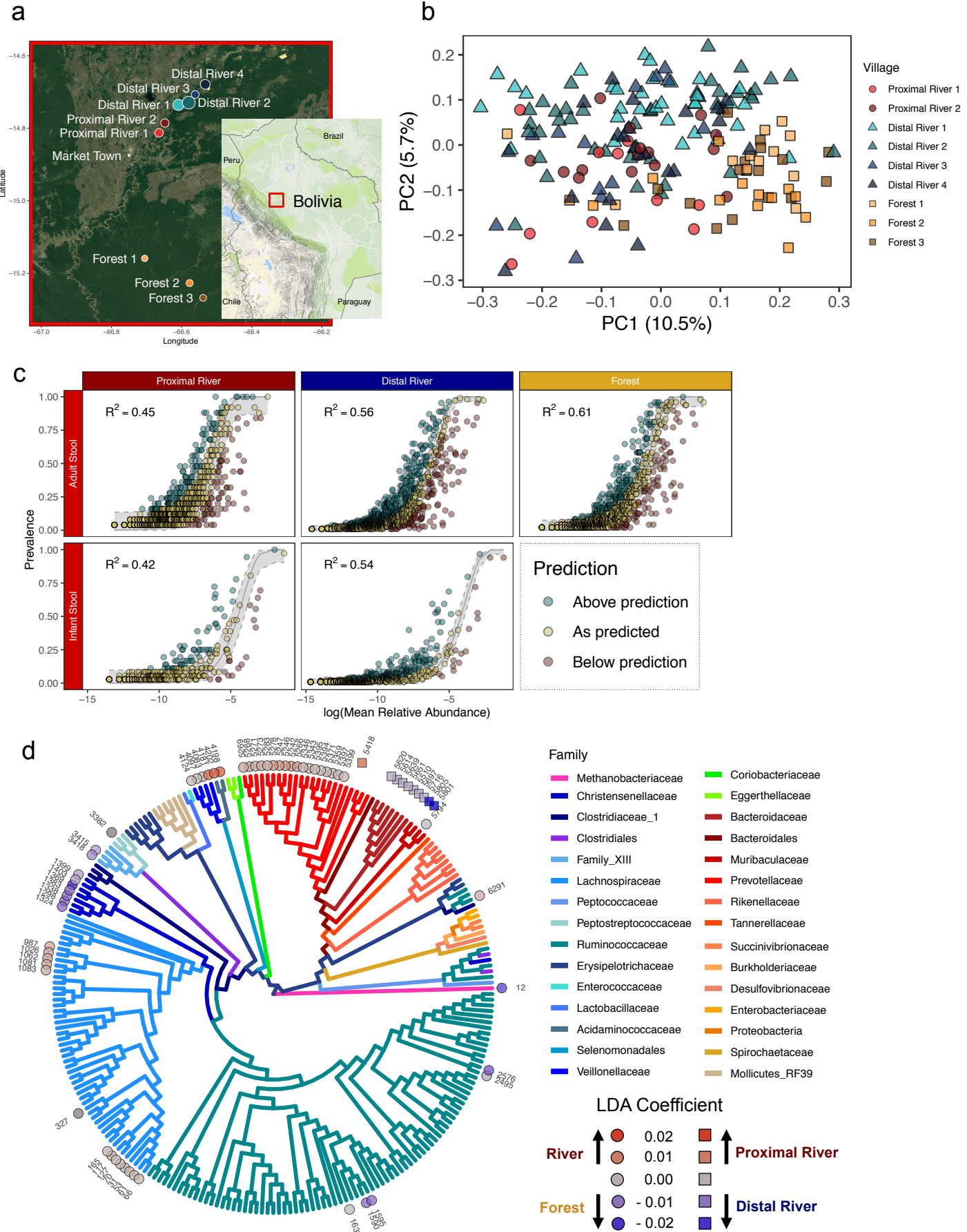
Figure 5

a

Distal River 4
Distal River 3
Distal River 1
Distal River 2
Proximal River 2
Proximal River 1
Market Town

Brazil
Peru
Bolivia
Chile
Paraguay

Forest 1
Forest 2
Forest 3

Latitude
Longitude

b

PC2 (5.7%)
PC1 (10.5%)

Village
- Proximal River 1
- Proximal River 2
- Distal River 1
- Distal River 2
- Distal River 3
- Distal River 4
- Forest 1
- Forest 2
- Forest 3

c

Proximal River | Distal River | Forest

Adult Stool
$R^2 = 0.45$ | $R^2 = 0.56$ | $R^2 = 0.61$

Infant Stool
$R^2 = 0.42$ | $R^2 = 0.54$

Prevalence
log(Mean Relative Abundance)

Prediction
- Above prediction
- As predicted
- Below prediction

d

Family
- Methanobacteriaceae
- Christensenellaceae
- Clostridiaceae_1
- Clostridiales
- Family_XIII
- Lachnospiraceae
- Peptococcaceae
- Peptostreptococcaceae
- Ruminococcaceae
- Erysipelotrichaceae
- Enterococcaceae
- Lactobacillaceae
- Acidaminococcaceae
- Selenomonadales
- Veillonellaceae
- Coriobacteriaceae
- Eggerthellaceae
- Bacteroidaceae
- Bacteroidales
- Muribaculaceae
- Prevotellaceae
- Rikenellaceae
- Tannerellaceae
- Succinivibrionaceae
- Burkholderiaceae
- Desulfovibrionaceae
- Enterobacteriaceae
- Proteobacteria
- Spirochaetaceae
- Mollicutes_RF39

LDA Coefficient

River ↑
- 0.02
- 0.01
- 0.00
Forest ↓
- -0.01
- -0.02

Proximal River ↑
Distal River ↓

Figure 5. **Market access is associated with Tsimane gut microbiota structure.** (a) A map showing the location of the 9 Tsimane villages whence stool samples were collected. The size of each circle was scaled to the number of samples collected from that village, with the exception of the market town, where no samples were collected. The map insert denotes the region of Bolivia being displayed. (b) Principal coordinate analysis (PCoA) on the Bray-Curtis distance of 16S rRNA profiles from adult stool samples. Village types ('proximal' river, 'distal' river, and 'forest') have significantly different distributions (PERMANOVA with 1000 permutations, p = 0.001). Samples are colored by village and their shapes denote the village type (proximal river villages – circles, distal river villages – triangles, forest villages – squares). (c) Neutral community model fit to each ASV observed in the stool of Tsimane adults (top row) or infants (bottom row) from each of three village types (proximal river – left, distal river – right, and forest – left). In each case, the composition of the regional species pool was estimated using the samples from the age group and village type. The gray line is the predicted distribution (shaded area is 95% CI), and points are colored according to whether their prevalence in infant samples was above (green), at (yellow), or below (red) their predicted prevalence according to the NCM. (d) Phylogenetic tree with ASVs found at least 10 times in at least 20% of Tsimane adult stool samples. Discriminatory ASVs were identified using the tree-based LDA algorithm in the treeDA R package (version 0.0.3). ASVs are denoted by shapes based on two sets of comparisons (circles for river vs. forest and squares for proximal river vs. distal river), and are colored based on their discriminatory strength. ASVs are colored by the taxonomic family and significantly discriminatory ASVs are labeled with their ASV number.
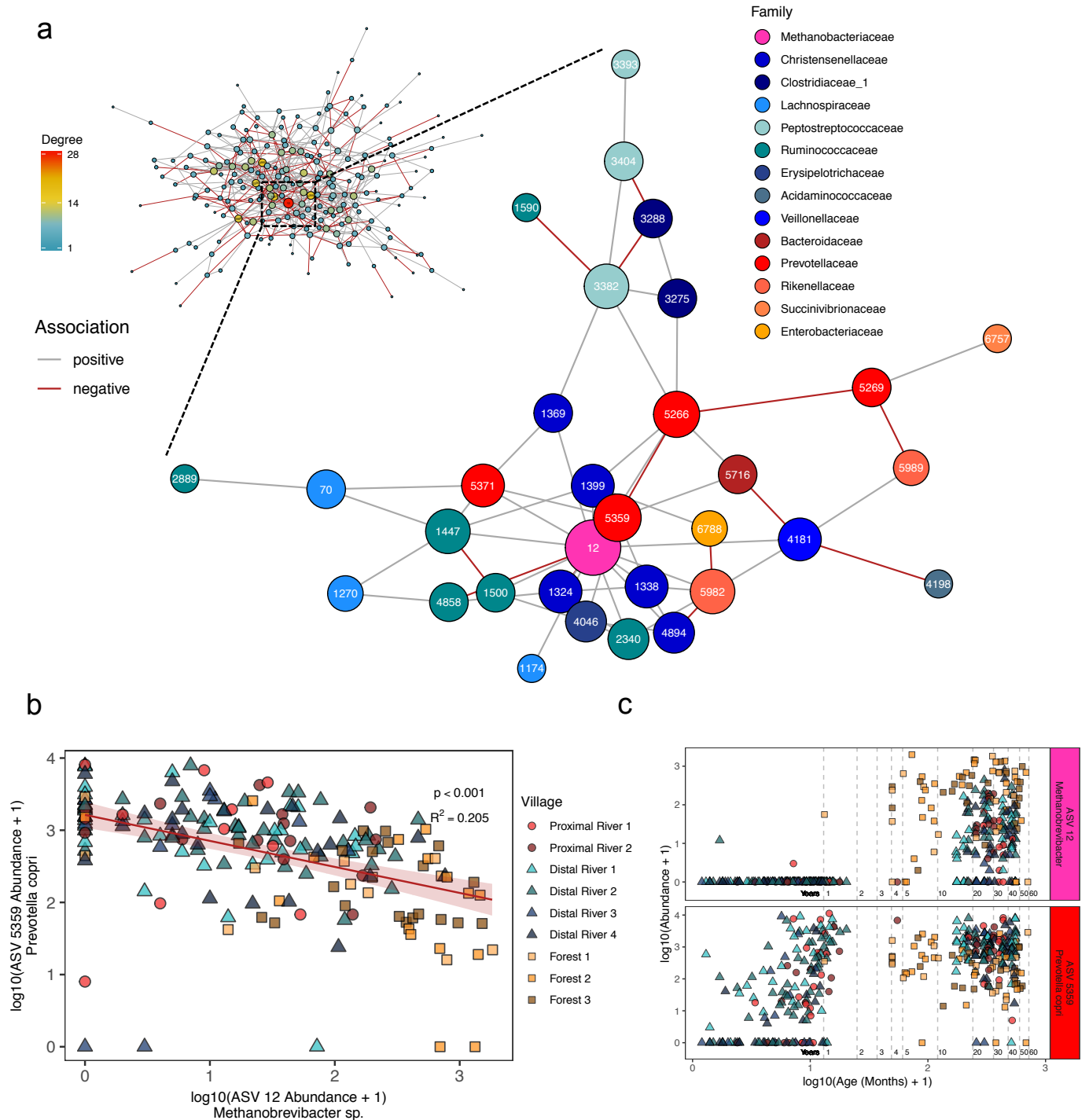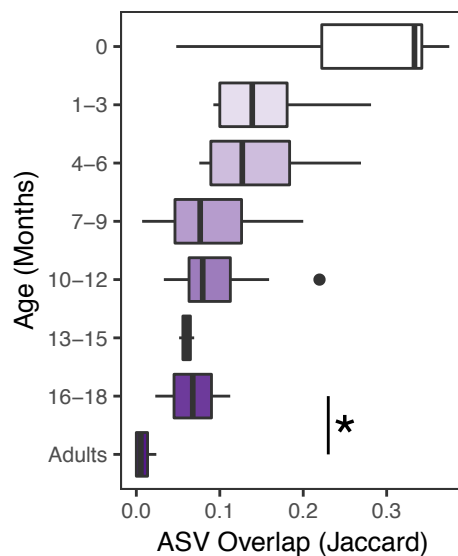
Figure 6

Figure 6. **Network analysis reveals an inverse relationship between highly-connected taxa.** (a) A network diagram of ASVs observed at least 10 times in at least 20% of Tsimane adult stool samples. The size of the node was scaled to represent its degree of connectedness. The insert shows the complete network colored by network degree, while the main figure shows the highly-connected subgraph colored by each ASV's taxonomic family. Positive associations between ASV nodes are colored gray, while negative associations are colored red. (b) A scatter plot of the log transformed relative abundances of the two most

highly connected ASVs in the network, *Methanobrevibacter* sp. (ASV 12) and *Prevotella copri* (ASV 5359). The red line indicates the linear mixed-effects regression while treating subject as a random effect, and shading indicates the 95% confidence interval. The conditional R^2 describes the proportion of variation explained by both the fixed and random factors. Samples are colored by village, and their shapes represent the village type (proximal river – circles, distal river – triangles, forest – squares). (c) Scatter plot of the relative abundances of *Methanobrevibacter* sp. (ASV 12) and *Prevotella copri* (ASV 5359) and the subject's age in months at the time of collection. The x-axis was log transformed for clarity when plotting both infant and adult samples. Ages 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, and 60 years are denoted by horizontal dashed lines.
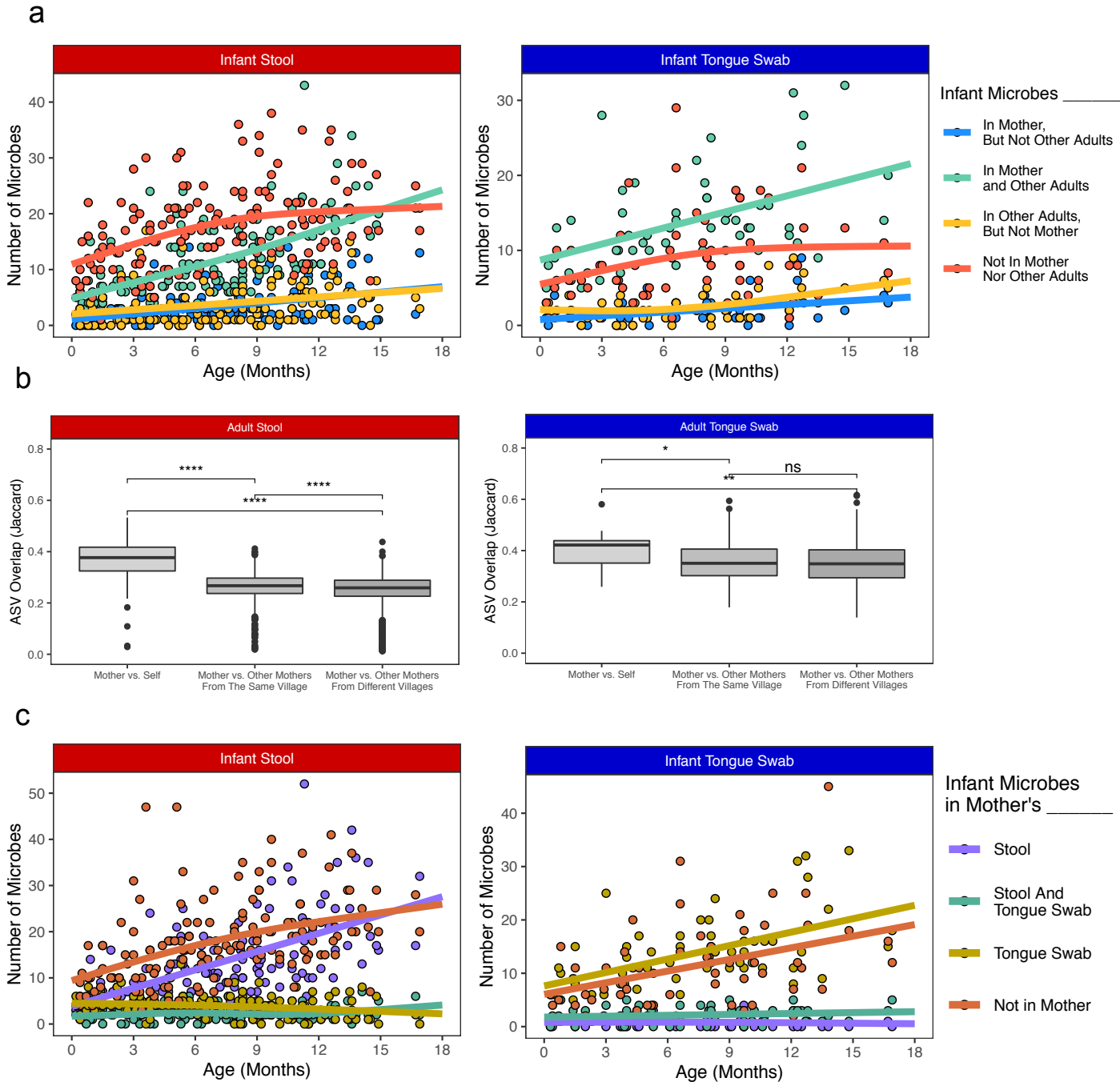
Supplementary Figure 1

Supplementary Figure 1. **Gut and oral microbiotas diverge as infants age.** Average number of ASVs shared between body sites within a subject based on samples collected within 14 days of each other. Boxplot color darkens with increasing age. Adults and 16-18 month olds were the only sequential age groups whose microbiotas were significantly different from one another (Wilcoxon rank-sum test, p = 0.03).

Supplementary Figure 2
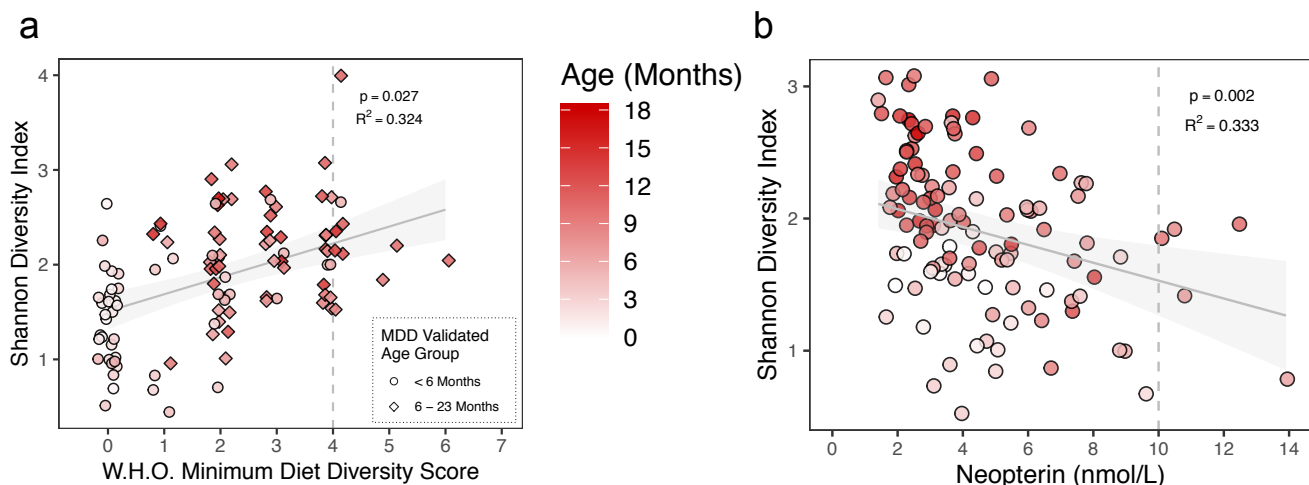
Supplementary Figure 2. **Diet and inflammation impact the infant gut microbiota.** (a) The World Health Organization's minimum diet diversity (MDD) score regressed against the Shannon Diversity Index from the infant stool samples. The MDD was validated to assess the diversity of complementary foods of 6-23 month old children in a diverse range of cultural contexts. It should only be interpreted as an estimate of micronutrient sufficiency in children aged 6-23 months, which are denoted by a diamond shape, and the color darkens as the subject's age increases. MDD of 4 is considered to be sufficiently diverse, as indicated by the vertical dashed line. Lines indicate the linear mixed-effects regression of diversity on MDD, while treating subject as a random effect. The shading indicates the 95% confidence interval. The conditional R^2 describes the proportion of variation explained by both fixed and random factors. (B) Levels of fecal neopterin, an inflammatory marker normally produced by activated macrophages, regressed against the Shannon Diversity Index in infant stool samples. Lines indicate the linear mixed-effects regression of diversity on neopterin, while treating subject as a random effect. The shading indicates the 95% confidence interval. The conditional R^2 describes the proportion of variation explained by both fixed and random factors. A neopterin level above 10, as indicated by the vertical dashed line, indicates clinically relevant inflammation.
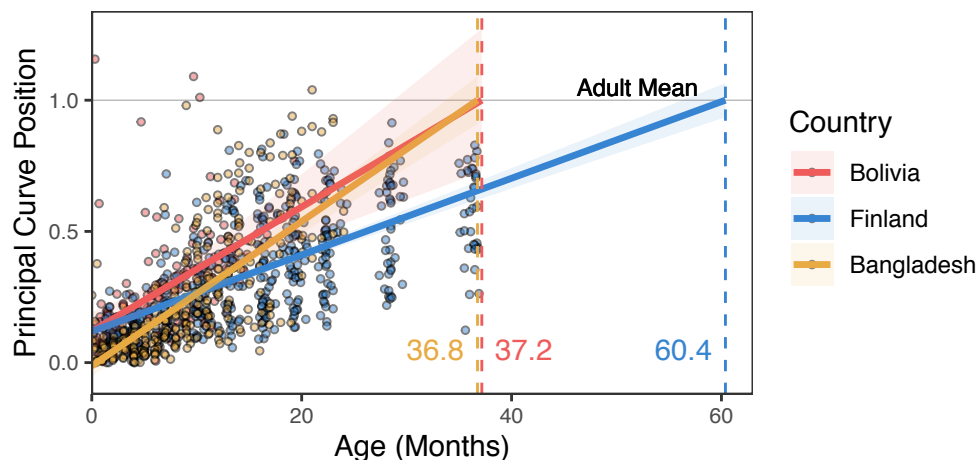
Supplementary Figure 3

a

b

c

Supplementary Figure 3. **Sources of infant-colonizing microbes.** (a) Scatter plot of the average number of ASVs found in each infant stool sample that were also found in that infant's mother but not in other adults from their village (blue), in both their mother and other mothers from their village (green), in other mothers from the infant's village but not their mother (yellow), or in neither their mother or other adults (orange). Non-parent adults from each village were down-sampled to control for differences in the number of samples collected from each village. Left panel – infant stool samples, right panel – infant tongue swabs. (b) Average number of ASVs shared between multiple samples from the same mother, between

mothers living in the same village, or mothers living in different villages. Wilcoxon Rank Sum Test. ****: p ≤ 0.0001, **: p ≤ 0.01, *: p ≤ 0.05, ns: p > 0.05. Left panel – adult stool samples, right panel – adult tongue swabs. (c) Scatter plot of the average number of ASVs in each infant stool sample that was also found in their mother's stool but not in her tongue swab (purple), in both their mother's stool and tongue swab (green), in their mother's tongue swab but not their mother's stool (yellow), or not found in any of their mother's samples (orange).
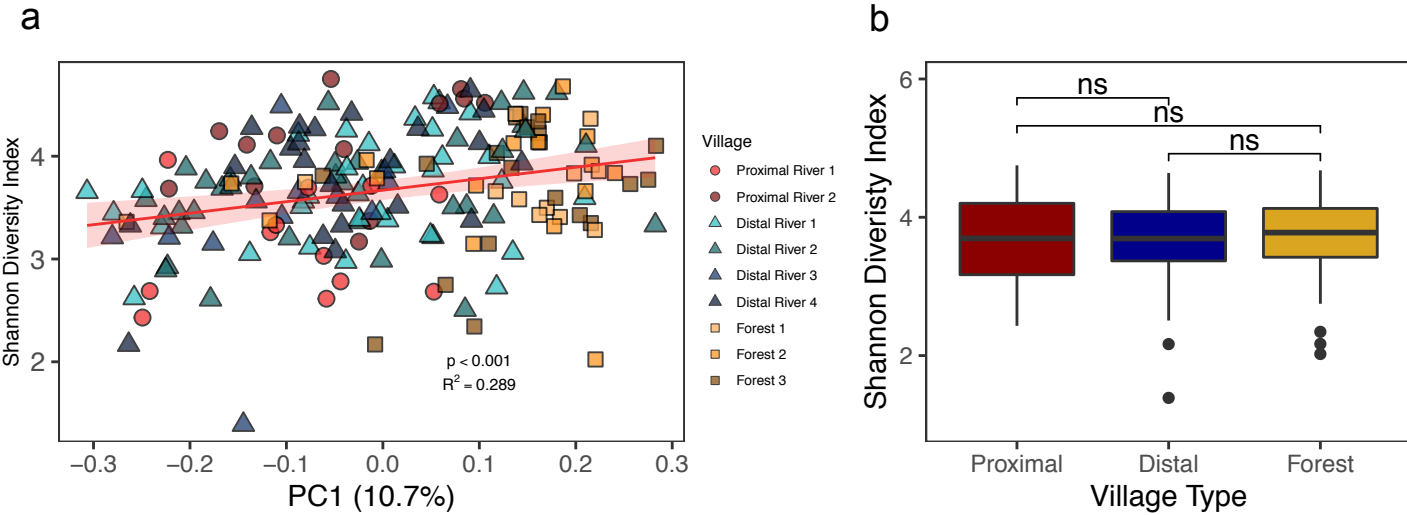
Supplementary Figure 4

Supplementary Figure 4. **Infants from different countries reach a population-specific adult-like state at different rates.** (a) In Fig. 4a, a PCoA was performed on 16S rRNA datasets from three populations (Bolivian (Tsimane), Bangladeshi, and European (Finnish/British)), and a principal curve was fit to each set of points using the princurve function in the princurve R package (version 2.1.3). The distance along each curve was scaled so that the mean location of the adult samples was equal to one (as indicated by the horizontal gray line), making an infant's location along the curve analogous to their position along the average maturational trajectory towards the average adult in their society. Here, we show the linear regression of the age of the infant from whom each sample was collected against its position along its principal curve. Lines indicate the linear mixed-effects regression of principal curve position on the infant's age in months, while treating subject as a random effect. The shading indicates the 95% confidence interval. The vertical dashed lines indicate the age at which the infant line from each country crosses the adult average position for that country.
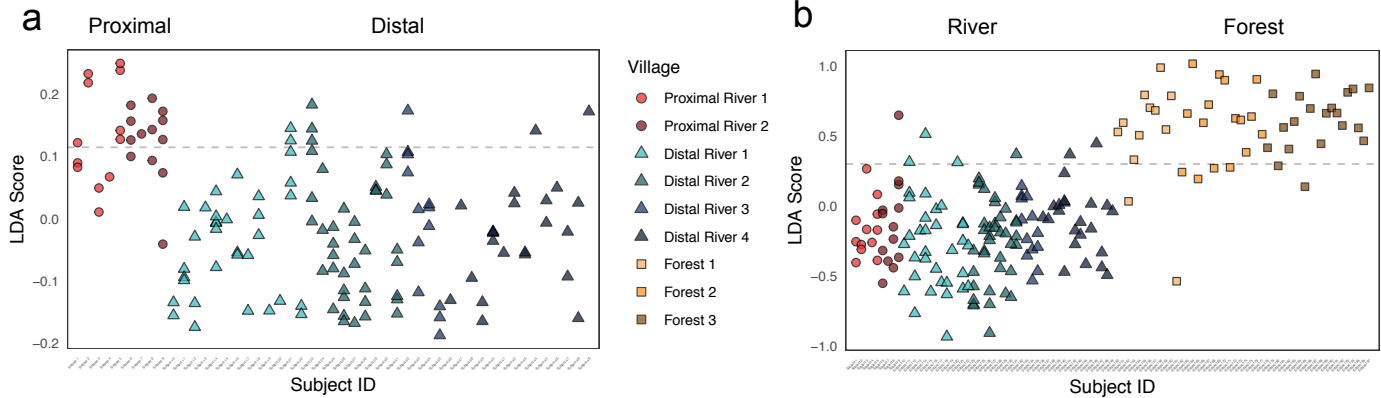
Supplementary Figure 5
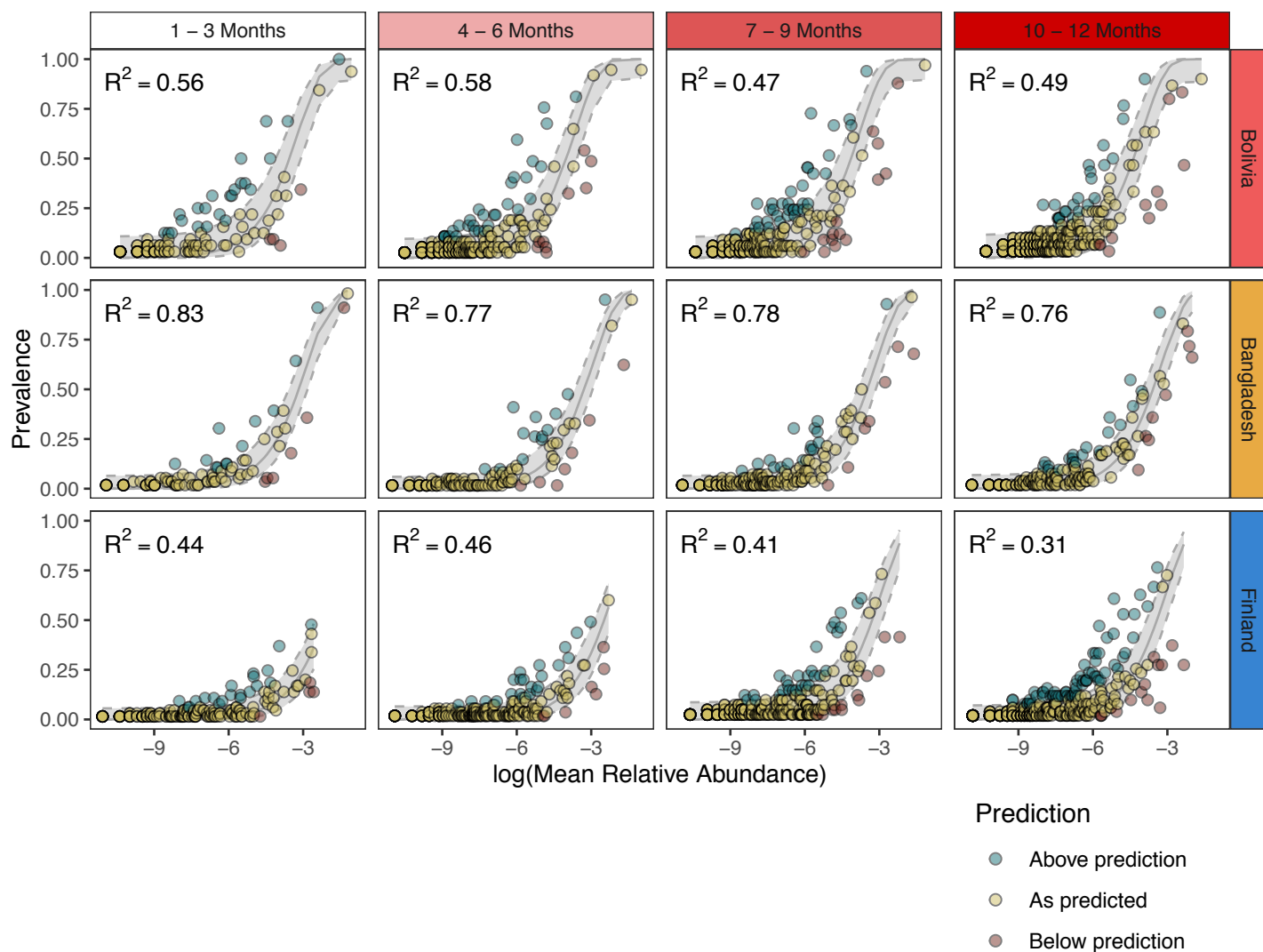
Supplementary Figure 5. **Neutral processes explain microbial community assembly across age groups and geography.** Scatterplot of the prevalence of each ASV in the infant gut versus its mean relative abundance in the regional species pool for each age group and country. The gray line is their predicted distribution (shaded area is 95% confidence interval) based on the neutral community model. Points are colored by the ASV's fit to the model: above prediction– green, below prediction– red, at prediction – yellow.

Supplementary Figure 6

Supplementary Figure 6. **Gut microbiota diversity among Tsimane villages.** (a) Shannon Diversity Index plotted against the PC1 value from Fig. 5b (R^2 = 0.289, p < 0.001). The red line indicates the linear mixed-effects regression while treating subject as a random effect, and shading indicates the 95% confidence interval. The conditional R^2 describes the proportion of the variation explained by both fixed and random factors. (b) Boxplot comparing the average diversity of adult stool microbiota of samples collected from proximal river, distal river, and forest Tsimane villages (p > 0.05 for all comparisons).

Supplementary Figure 7

Supplementary Figure 7. **Bacterial taxa discriminate between village types.** (a) The LDA score of each adult stool sample was calculated from a phylogeny-based form of linear discriminant analysis (treeDA, version 0.0.3). Higher scores indicate samples more likely to have been collected from a proximal river village, while lower scores indicate samples more likely to have been collected from a distal river village. The dashed gray line represents the optimal LDA score cut-off between groups, where it accurately classifies 87.3% of samples. Samples are colored by village, and their shapes represent the village type (proximal river – circles, distal river – triangles, forest – squares). (b) The LDA score of each adult stool sample was calculated from a phylogeny-based form of linear discriminant analysis (treeDA, version 0.0.3). Higher scores indicate samples more likely from a forest village, while lower scores indicate samples more likely from a river village. The dashed gray line represents the optimal LDA score cut-off between groups, where it accurately classifies 91.8% of samples. The colors and shapes are the same as in panel a.

Supplementary Table 1

Village	Age Class	Number of Subjects	Stool Samples	Tongue Swabs	All Samples	Distance from City (Km)	Population*
Proximal River 1	Adult	6	12	8			
	Child	0	0	0	47	14.3	331
	Infant	6	20	7			
Proximal River 2	Adult	4	13	6			
	Child	0	0	0	42	18.5	292
	Infant	4	17	6			
Distal River 1	Adult	13	34	13			
	Child	0	0	0	92	29.8	151
	Infant	11	33	12			
Distal River 2	Adult	9	39	18			
	Child	0	0	0	128	33.5	253
	Infant	10	53	18			
Distal River 3	Adult	5	13	5			
	Child	0	0	0	36	37.5	101
	Infant	5	14	4			
Distal River 4	Adult	14	23	12			
	Child	0	0	0	65	42.5	206
	Infant	12	19	11			
Forest 1	Adult	12	12	0			
	Child	8	8	0	21	31	61
	Infant	1	1	0			
Forest 2	Adult	16	16	0			
	Child	11	11	0	27	61	114
	Infant	0	0	0			
Forest 3	Adult	20	20	0			
	Child	5	5	0	25	68	101
	Infant	0	0	0			

Supplementary Table 1. Demographics of Tsimane villages and study subjects
Villages are listed in order of increasing distance from the closest market town. Age classes are defined as Infant – under 2 years, Child – 2 to 13 years old, Adult – 14 years old and older. The 73 fecal samples collected from forest communities were collected in July 2009, while the remaining fecal and oral samples from proximal and distal river communities were collected between September 2012 and March 2013. *Population sizes for forest villages were collected during a 2009 census, while the remaining were collected in a 2012 census.