Trait-based community assembly and succession of the infant gut microbiome

John Guittar[1], Ashley Shade[2,3], Elena Litchman[1,4,5]

Correspondence and requests for materials should be addressed to J.L.G. (email: guittarj@msu.edu).

Keywords: Community assembly, dispersal limitation, ecological succession, functional ecology, gut microbiome, human microbiome, microbial ecology, trait-based ecology.

[1]Kellogg Biological Station, Michigan State University, 3700 E Gull Lake Dr., Hickory Corners, MI 49060, USA.

[2]Department of Microbiology and Molecular Genetics, Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing MI 48840, USA.

[3]Program in Ecology, Evolutionary Biology and Behavior, Michigan State University East Lansing MI 48840, USA.

[4]Program in Ecology, Evolutionary Biology and Behavior, Michigan State University East Lansing MI 48840, USA.

[5]Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA.

Abstract

The human gut microbiome develops over early childhood and aids in food digestion and immunomodulation, but the mechanisms driving its development remain elusive. Here we use data curated from literature and online repositories to examine trait-based patterns of gut microbiome succession in 56 infants over their first three years of life. We also develop a new phylogeny-based approach of inferring trait values that can extend readily to other microbial systems and questions. Our analysis suggests that infant gut succession begins with a functionally variable cohort of taxa, adept at proliferating rapidly within hosts, which gradually matures into a more functionally uniform cohort of taxa adapted to thrive in the anoxic gut and disperse between anoxic patches as oxygen-tolerant spores. Trait-based composition stabilizes after the first year, while taxonomic turnover continues unabated, suggesting functional redundancy. Trait-based approaches powerfully complement taxonomy-based approaches to understand the mechanisms of microbial community assembly and succession.

Introduction

Classical ecological theory posits that successional patterns arise from the combined influence of dispersal, species interactions, and the environment[1,2], and this general framework extends readily to gut communities[3]. Before a microbe can inhabit the colon, the most distal and speciose part of the gastrointestinal tract, it must first be swallowed by the host and survive the acidic conditions of the stomach and small intestine (i.e., it must disperse). A species will persist in the colon only if it can acquire enough resources to reproduce (i.e., it must be competitive) or arrive there in high enough numbers to sustain a population[4]. Microbial colonists may then alter the environment, e.g., by depleting intestinal oxygen[5] or providing opportunities for cross-feeding[6], favoring taxa with different phenotypes as succession proceeds.

Yet successional patterns in the gut may differ from classical successional expectations due to the active influence of the host and the host mother[7,8]. Early colonists are passed directly from the mother during or even before birth[9], and therefore may lack characteristics that would otherwise facilitate early arrival, e.g., via active dispersal, and instead have characteristics selected for in the mother's gut or vaginal environment. Following birth, mothers supply bacterial growth factors in breastmilk and continue to introduce new taxa through physical contact[10]. Meanwhile, the maturing infant is beginning to suppress undesirable taxa through immune response[11], and actively cultivate commensal taxa by providing nitrogen-rich mucus and favorable habitat in the outer mucus layer of the large intestine[12]. Gut community composition is also affected by the introduction of solid food[13], in particular with the introduction of insoluble fiber[14].

Gut community successional patterns will necessarily reflect a combination of dispersal, microbial species interactions, and host physiology and behavior. A present challenge is to determine how the relative influence of these drivers changes over time. One approach to disentangling the mechanisms of community assembly is to examine patterns in trait-based community composition[15]. A trait, in the broadest sense, is defined as a measurable organismal characteristic directly or indirectly linked to fitness or performance[16]. As such, observable shifts in the trait-based composition of a community imply shifts in local environmental conditions favoring different species and/or dispersal limitation (i.e., when a taxon does not colonize a site because it does not arrive). Despite the success and proliferation of trait-based approaches to study community assembly in plant[17,18], animal[19,20], and phytoplankton systems[21], they have only rarely been used for bacterial and archaeal systems[22,23]. This is due partly to the challenges of identifying ecologically relevant traits for a functionally diverse cohort of taxa, and partly to a dearth of curated trait data. But thanks to recent advances in high-throughput molecular techniques, renewed efforts to directly collect phenotypic data[24], and the aggregation of data

from disparate sources[25,26], trait-based approaches to microbial community dynamics are becoming more feasible, especially for well-studied systems like the human gut.

Here we examine trait-based successional patterns in a cohort of 56 infants from Finland and Estonia for which longitudinal microbiome survey data were publicly available[27,28]. We develop a unique approach to inferring microbial trait data, which entails (1) building a phylogeny that contains the taxa from infant gut samples and 13900 other taxa with formally described type specimens and Latin binomials[29], (2) using the Latin binomials to map trait data curated from literature and online repositories onto the tips of the phylogeny, and (3) inferring unknown trait values using hidden state prediction when statistically justified. We then compare taxonomic and trait-based community turnover in time (i.e., over infant development) and space (i.e., across infants) to gain insight into the mechanisms driving successional patterns. We show significant trait-based shifts over the first year of infant development, during which time oxygen-tolerant taxa and flagellated taxa become less abundant, and slower-growing taxa and sporulating taxa become more abundant. Intriguingly, during this time, microbiomes become compositionally more similar across infants. Taxonomic turnover continues after the first year, but is largely redundant with respect to the traits examined. Our results suggest that succession begins with a functionally variable cohort of early arrivers, adept at proliferating rapidly within hosts, which gradually matures into a more functionally uniform cohort of taxa able to both thrive in the anoxic gut environment and disperse between anoxic patches (e.g., guts) as oxygen-tolerant spores.

Results

Trait-based patterns of succession

We observed consistent taxonomic and trait-based shifts in infant gut microbiomes during the first three years of infant life (Fig. 1, Fig. 2). Early succession was dominated by Bacteroidaceae and

Bifidobacteriaceae (Fig. 1a,b), whereas late succession was dominated by Lachnospiraceae, Ruminococcaceae, and (still) Bacteroidaceae (Fig. 1e,f). About three fourths of the operational taxonomic units (OTUs) in this study, defined using a threshold of 97 percent sequence similarity in the 16S rRNA V4 region, exhibited significant positive or negative trends in abundance over succession across all infants. The extensive number of significant trends emphasizes the taxonomically predictable nature of gut microbiome development. Early and late successional specialists differed significantly in their predicted trait values: late successional specialists were less tolerant of oxygen, were more capable of sporulation, and had higher temperature optima than early successional specialists (Supplementary Figure 1).

Community weighted means (CWMs) of several traits trended significantly over the course of succession (Fig. 2), illustrating the functionally predictable nature of gut microbiome development[30]. A CWM is the mean trait value of the OTUs in a community, weighted by their relative abundances. Ecologically speaking, CWMs characterize the dominant traits of a community, and can be thought of both in terms of how they reflect system properties (i.e., as response traits) and how they influence system properties (i.e., as effect traits)[31]. For example, oxygen-tolerant taxa (e.g., facultative anaerobes) present at the onset of succession were rapidly overtaken by obligate anaerobes (Fig. 2i), presumably in response to a drop in gut oxygen concentration due to increased uptake by epithelial cells[32]. Meanwhile, the mean number of B-vitamin pathways present per cell decreased over time (Fig. 2b), contradicting our expectation that human hosts would selectively enrich such taxa over the course of succession to promote the production of these essential nutrients.

Pronounced shifts in two traits potentially related to dispersal ability suggest that dispersal dynamics may play a key role in shaping successional patterns. First, the initial presence and subsequent decline of taxa with flagella (Fig. 2h) could mean that the ability to actively disperse over short distances (i.e., spread within hosts) improves colonization rates during early succession, but that flagella are not as

advantageous in the mature gut. In support of this, unflagellated strains have been shown to be poorer colonizers of chickens' gastrointestinal tracts than flagellated strains[33], and a positive relationship has been drawn between motility and bacterial transmission[34]. Second, the increase in sporulating taxa over time (Fig. 2j, Supplementary Figure 3) may reflect the long-term advantages of being able to disperse among hosts and/or persist within hosts in a dormant state during stressful conditions[24,35]. As succession proceeds and the gut environment becomes increasingly anoxic, obligate anaerobes gain a competitive advantage over facultative anaerobes because they do not need to maintain the machinery for tolerating oxidative stress. However, this advantage comes at the cost of being more vulnerable to oxidative stress while dispersing through oxic environments to colonize new hosts. Sporulating taxa circumvent this potential tradeoff by traversing oxic environments as oxygen-tolerant spores, and then thriving in the gut as obligate anaerobes. The observed increase of sporulating taxa over gut community development, both in total abundance (Fig. 2j) and OTU richness (Supplementary Figure 3), likely reflects the steady arrival and successful colonization of these taxa well-adapted for the anoxic gut environment.

The mean number of 16S rRNA gene copies, a genomic trait associated with the ability to quickly exploit available resources due to higher maximum potential growth rates[36], decreased steadily in gut microbiomes over time (Fig. 2a). A decrease in mean 16S rRNA gene copy number over time is characteristic of primary succession in microbial systems that are initially rich in resources[8], such as a vial of sterile nutrient broth placed in an open-air environment[37]. However, a decrease in mean 16S rRNA gene copy number could also arise if faster-growing taxa thrive on easily-digested milk or formula, the primary carbon source during early succession, and slower-growing taxa only begin to thrive as the primary carbon source shifts towards increasingly complex molecules derived from solid food. In either case, the decrease in mean 16s rRNA gene copy number over time likely reflects a shift from taxa capable of rapid low-efficiency growth to slower high-efficiency growth over succession[23,38].

Many traits correlated significantly among taxa (Supplementary Figure 2). The strongest positive correlations were between gene number and genome size, genome size and B-vitamin pathway number, and sporulation and Gram-positive status, while the strongest negative correlations were between optimal growth temperature and oxygen tolerance, Gram-positive status and B-vitamin pathway number, and GC content and 16S rRNA gene copy number. The remaining Pearson correlation coefficients were less than 0.6 or greater than -0.6. On one hand, correlations among traits are noteworthy because they may be independent indicators of a taxon's position on the same ecological tradeoff axis (i.e., they may constitute a trait syndrome). For example, the negative correlation observed between sporulation score and oxygen tolerance represent two approaches for dealing with oxidative stress, either by becoming metabolically dormant until oxidative stress is relaxed, or by carrying the cellular machinery to tolerate it. On the other hand, correlations among traits may simply be artifacts of arbitrary genomic linkage, and not evidence of evolutionary adaptation. As such, the mechanisms we invoke as possible explanations for the trait-based patterns observed in this study are merely hypotheses which hopefully spur further experimental work.

To explore how early exposure to different taxa could affect the trajectory of gut succession, we compared trait-based successional patterns of infants delivered vaginally and by C-section (Fig. 3). We reasoned that any consistent community differences between the two groups of infants would likely arise due to differences in early colonization, i.e., because infants born vaginally were initially colonized by taxa from the mother during delivery, and infants born by C-section were initially colonized by a different cohort of taxa arriving from the ambient environment (e.g., the mother's skin, hospital surfaces). Notable trait-based differences between the microbiomes in C-section infants, relative to those in vaginally delivered infants, were initially elevated numbers of Gram-positive taxa (Fig. 3f), and prolonged persistence of oxygen-tolerant taxa (Fig. 3i). There were also initially elevated mean 16S rRNA gene copy numbers (Fig. 3a) and initially higher prevalence of flagellated taxa (Fig. 3h) in C-section

149 infants, relative to vaginally born infants, but these differences were not statistically significant after
150 accounting for multiple comparisons. At minimum, these results suggest that taxa encountered by
151 infants during vaginal delivery are functionally distinct from those encountered by infants after C-section
152 delivery in the hospital environment. More interestingly, however, they suggest that gut colonization
153 patterns differ depending on the composition of the initial pool of colonizing taxa. Significant trait-based
154 compositional differences by birth mode persisted for up to two years (Fig 3i), corroborating previous
155 research showing that differences in early colonization can have lasting effects on community
156 composition[39,40], a phenomenon also termed priority effects[41,42]. On the other hand, sustained trait-
157 based differences between infants by delivery mode are surprising given recent work which found
158 strong selective forces to quickly discourage the growth of immigrant taxa from the mother's skin or
159 birth canal[43]; hence, our findings suggest that the persistent differences by birth mode may result from a
160 lack of arrival (i.e., dispersal limitation) of gut-adapted taxa from the mother, rather than qualitatively
161 different community filters among infants.
162 Exposure to antibiotics was associated with consistent trait-based shifts in gut microbiome
163 composition (Fig. 3). Specifically, infants exposed to repeated antibiotic treatments had gut taxa that
164 were on average less likely to be Gram-positive (Fig. 3f), smaller (Fig. 3g), and less capable of sporulation
165 (Fig. 3j) than infants exposed to no antibiotics. Decreases in the relative abundances of Gram-positive
166 taxa over time is arguably expected given that Gram-positive taxa lack the protective outer membrane
167 that make Gram-negative bacteria generally more resistant to antibiotics[44]. The drop in mean
168 sporulation score is less expected, given that spores are generally very resistant to antibiotics[24].
169 However, spore formation is far from the only mechanism of antibiotic tolerance in Bacteria, and other
170 strategies may be more effective for survival in the gut environment. For instance, antibiotic treatments
171 usually result in decreases in the relative abundances of spore-forming taxa in the class Clostridia, and
172 increases in the relative abundances of non-spore-forming taxa in the family Enterobacteriaceae[32]. More

generally, consistent with prior work[45], the persistent differences in trait-based community composition between infants that underwent heavy antibiotic treatments and those that did not suggests that these disturbances can exert long-term effects on community structure and function.

Trait variances within infant gut communities decreased over time in seven traits, and increased over time only in three traits (Supplementary Figure 4). The overall decrease in trait-based variance over time indicates that individuals of the gut community became more functionally homogeneous as the infants matured, perhaps due to increasingly strict environmental filtering processes[46] and/or competitive exclusion of poorly adapted taxa[47].

Comparing taxonomic and trait-based successional patterns

To evaluate the degree to which taxonomic changes aligned with trait-based changes, we compared taxonomic and trait-based turnover over time within infants, both in terms of short-term compositional variability (measured as the dissimilarity between subsequent samples) and directional turnover (measured as the dissimilarity between each sample and the final sample collected). Compositional variability was higher in the first year of development, both in terms of OTUs (Fig. 4a) and traits (Fig. 4c), than in the second or third years of development. A decrease in compositional variability over time is a classical feature of many ecological successional systems[48]. To evaluate whether trait-based compositional variability was higher or lower than expected by chance, given the magnitudes of taxonomic variability observed, we compared observed patterns to predictions from null model simulations for which trait values were randomly shuffled among taxa and trait-based compositional variability was re-calculated (see Methods). In other words, we calculated what trait-based compositional variability would look like if the traits in our study were completely decoupled from taxon performance. Differences between observed and null predictions were neither large nor significant (Fig.

196 4c), suggesting that the traits in our study had little influence on compositional variability over
197 succession.

198 An analysis of directional turnover over succession revealed that infant gut communities
199 matured and stabilized faster in their trait-based compositions than in their OTU-based compositions.
200 Specifically, OTU-based directional turnover was relatively steady across all three years of study (Fig.
201 4b), whereas trait-based directional turnover was high only in the first year (Fig. 4d) before dropping to
202 nearly-baseline levels of trait-based compositional variability (Fig 4c). Trait-based directional turnover
203 significantly exceeded null model predictions of trait-agnostic turnover (Fig. 4d), suggesting that infant
204 gut microbiomes stabilize (i.e., cease to exhibit directional turnover) in terms of traits and their
205 associated functions sooner than they stabilize in terms of OTUs, aligning with previous metagenomic
206 work[30]. The fact that OTU-based directional turnover was steady over all three years of infant
207 development despite early convergence in trait-based community composition indicates that late-stage
208 OTU-based turnover was of OTUs that were functionally redundant, at least with respect to the traits
209 examined in this study. Functionally redundant turnover could arise due to variable immigration rates
210 (i.e., if different functionally redundant taxa immigrated into the gut at variable rates over time), or due
211 to ecological drift (i.e., changes in the relative abundances of taxa through stochastic birth/death
212 events). With respect to the latter: even though the gut community has a large number of individuals,
213 which, all else being equal, makes it less susceptible to ecological drift[49], many of its constituent taxa are
214 rare and therefore still vulnerable to stochastic variation in their relative population sizes over time.
215 Future work should quantify immigration rates, and consider other traits as potential drivers of late-
216 stage successional community turnover, such as those relating to metabolism of specific dietary
217 compounds[50], cross-feeding[6], or phage-host interactions[51].
218

Compositional differences across microbiomes

Surprisingly, gut community compositions became more similar (i.e., converged) across infants as they aged (Fig. 5). This ran counter to our expectations that gut community compositions would diverge as infants shifted from subsisting on milk and/or formula (i.e., simple substrates with low resource variability expected among hosts) to solid foods (i.e., complex substrates with higher resource variability expected among hosts), and as interactions between infants and their idiosyncratic home environments accumulated over time. Compositional convergence across infants over development may reflect a process whereby a stochastic cohort of initial taxa colonize infants but are gradually replaced, or supplemented with, taxa better suited for the gut environment. Such initial compositional differences among infants could be generated by stochastic colonization dynamics, differences in the pool of potential immigrants from the infants' mothers, or a combination of the both. Regardless, it is likely that gut community convergence over infant development is partly due to the delayed arrival of taxa well-adapted for the gut environment, i.e., dispersal limitation. Future experimental work should quantify the relative importance of dispersal dynamics and niche availability in driving compositional convergence over time.

Compositional convergence among infant gut communities was more pronounced and abrupt in terms of traits (Fig. 5b) than in OTUs (Fig. 5a), which converged only slightly and gradually over time. Trait-based rates of convergence significantly exceeded null model expectations of trait-agnostic convergence (Fig. 5b), indicating that trait-based convergence was not random with respect to the traits examined in this study. This discrepancy between OTU-based and trait-based patterns of convergence among infants leads to two insights. First, it is another reminder that microbial communities with different OTU-based compositions do not necessarily differ in their functional potentials [30,52]. Second, it means that community succession can be more predictable with respect to traits than OTUs. Together, these results indicate that OTU-based turnover over late succession is largely functionally redundant

with respect to the traits examined. Functional redundancy among gut microbiome taxa may benefit the host by improving community resilience in response to disturbance[53]. Interestingly, mean compositional differences among infants born by C-section were, on average, greater both in terms of OTU-based and trait-based dissimilarity (Supplementary Figure 5). Such differences could arise if the taxa to which C-section infants are initially exposed are more taxonomically and functionally variable than the taxa to which vaginally delivered infants are exposed.

Discussion

As in the ecological studies of macroorganisms, trait-based analysis of gut microbiome succession offers insights into the mechanisms of community assembly, such as dispersal limitation and ecological filtering, and the balance between stochastic and deterministic forces. The stabilization of trait-based community composition after the first year of development (Fig. 4), and the drop in variance of community trait values for most traits over time (Supplementary Figure 4), both suggest that succession is at least partially functionally deterministic, with early dynamics potentially reflecting stochastic colonization during the birthing process, followed by the gradual colonization and enrichment of a more functionally uniform cohort of taxa better adapted for the mature gut environment. Rates of OTU-based directional turnover remained steady over the first three years of succession (Fig. 4b), even though trait-based directional turnover essentially stabilized after only one year (Fig. 4d), underscoring the fact that OTU-based compositional changes need not imply changes in trait-based composition[54]. However, there are surely aspects of community assembly that cannot be understood using only the traits used in this study, and future work should expand the number of traits considered. Moreover, because our study is observational, we cannot distinguish between an OTU that fails to disperse to a potential host and an OTU that arrives but fails to establish, so future research should also explore the relationship between

OTU arrival and detection in fecal samples to better disentangle dispersal limitation and niched-based differences among taxa.

Comparisons of trait-based patterns between cohorts of infants are an opportunity to understand the effects of specific events (e.g., delivery mode, antibiotic exposure), and serve as natural experiments that can reveal how gut communities respond to, and recover from, systematic disturbances. In our analysis, for example, delivery mode resulted in sustained differences in community composition, indicating that priority effects can play an important role in gut community assembly[41,42], a result that likely extends to other types of disturbance during early life, such as gastrointestinal illness or malnutrition. Similarly, repeated antibiotics treatments led to significant differences in trait-based community composition (Fig. 3), suggesting that gut communities are not infinitely functionally resistant and/or that tradeoffs exist between antibiotic resistance and other traits[52]. Understanding trait-based differences between other cohorts, such as healthy vs. diseased[55], or on and off specific diets[56], could provide insight into additional factors shaping gut microbiome community assembly. For example, the unhealthy, dysbiotic gut may have a higher prevalence of microaerobic and biofilm-forming species[57], a difference that could be detected using trait-based analyses. Trait-based approaches, which link organismal structures to ecological functions, are poised to advance our mechanistic understanding of the gut microbiome, and their usefulness will only increase as we improve our knowledge of how traits mediate microbial interactions and as we increase the depth and breadth of microbial trait databases.

Methods

Infant microbiome sampling and sequence processing

Our foremost aim in this study was to characterize general patterns of gut primary succession that hold true regardless of host-related differences. As such, unless otherwise noted, we include all infants in our analyses, regardless of delivery mode or other host differences specific to each included study.

Longitudinal infant gut microbiome data were compiled from two studies from the DIABIMMUNE study group (https://pubs.broadinstitute.org/diabimmune), one focused on the effects of antibiotics on gut community development[28], and the other focused on the effects of type-1 diabetes on gut community development[27]. In the antibiotics study, infants either had nine or more antibiotic on gut community development courses, or no antibiotic courses[28]. In the type-1 diabetes study, infants tested positive for HLA DR-DQ alleles conferring risk of type-1 diabetes; of the infants which met our sampling criteria (see below), three developed type-1 diabetes during the sampling period[27].

Stool samples of infants were collected by participants' parents and stored in their house freezers until the next scheduled visit to the local study center. Samples were then shipped on dry ice to the DIABIMMUNE Core Laboratory, where they were stored at -80°C until being sent to the Broad Institute for DNA extraction and 16S rRNA amplicon sequencing. Sequencing was performed on the Illumina HiSeq 2500 platform using the 515F and 806R primers. Of 74 infants across the two studies, only those with at least 12 samples and those which extended more than 30 months were used in this study, yielding 56 infants with 12 - 36 sampling points (mean = 26.45; median = 27) taken at semi-regular intervals over the first 3 years of infant life (Supplementary Figure 6). All subjects were from Finland, except one from Estonia.

Infants varied in their modes of delivery and antibiotic histories, providing an opportunity to explore the potential effects of these natural experiments on trait-based gut community composition. To this end, infants were divided into three groups: 1) High antibiotic exposure (N = 18), if they underwent at least 50 days of antibiotic treatment and were delivered vaginally, 2) C-section delivery (N = 6), if they were delivered by C-section and underwent two or fewer rounds of antibiotics, and 3) a control group that was delivered vaginally and received no antibiotic treatments (N = 18). In some instances, antibiotic treatment durations were not reported, in which case we assumed seven days per treatment. Twelve types of antibiotics were administered for a variety of ailments, with the most

common being amoxicillin, trimethoprim and sulfadiazine aimed at treating acute ear infections. Infant metadata, drawn from the two studies from which sequence data for this study are drawn[27,28], is available in Supplementary Data 1.

Sequence processing was done using USEARCH version 10.0.240[58]. Raw sequencing data were downloaded from the DIABIMMUNE website https://pubs.broadinstitute.org/diabimmune/. Chimeras and reads flagged with more than one error were excluded, and the remaining reads were truncated to 250 bp, the expected overlap when using 515F and 806R primers. Reads were clustered into operational taxonomic units (OTUs) at 97 percent sequence identity using the UPARSE-OTU algorithm (Supplementary Data 2). Representative sequences from each OTU were mapped to the SILVA v123 database[59] to determine potential taxonomic identities (Supplementary Data 3). To avoid bias in sampling effort, samples were rarefied to 5000 sequences, and seven samples with fewer than 5000 sequences were removed.

Assembling trait data

We compiled data on 16 genomic, physiological, and life history traits of bacteria from public databases and individual studies (Table 1, Supplementary Data 4). All trait data were associated with taxa with full Latin binomials (i.e., Genus and Species labels) that appeared either in the SILVA-derived taxonomy file for the combined gut community samples or in the curated taxonomy file from the 132 release of the Living Tree Project[29]. Altogether, these amounted to 57,543 collected trait data spread across 10,906 taxa. When a taxon had more than one trait value, the mean or mode was used, depending on whether the trait was quantified continuously or discretely.

Descriptions and data sources for each trait are listed briefly in Table 1, but here we elaborate with a few additional details: 1) The numbers of B-vitamin synthesis pathways in the genome were drawn from ref. 60 and are based on genome annotations from the pubSEED platform[61]. 2) In some

cases, optimal temperature was calculated as the mean of lower and upper temperature ranges, consistent with ref. 26. 3) IgA binding affinity refers to the degree that immunoglobulin A bound to specific bacterial taxa, and was quantified using an IgA coating index calculated in ref. 62 using flow-cytometry-based bacterial cell sorting and 16S rRNA sequencing to characterize the coating load of IgA on specific taxa from fecal samples in a murine model. 4) Sporulation score indicates the tendency of taxa to sporulate, and was calculated in ref. 24 as a continuous score ranging from zero to one that depended on a combination of targeted phenotypic culturing and whole-genome sequencing from stool samples. When possible, we used sporulation scores from ref. 24. When sporulation scores from ref. 24 were unavailable for a given Latin binomial, we drew on sporulation data from other repositories (Table 1), which were generally binary, either noting the presence or absence of spores; when spores were present, taxa were given sporulation scores of 0.549, equal to the median sporulation score of taxa with sporulation scores greater than zero in ref. 24; when spores were not observed, taxa were given sporulation scores of zero.

Predicting unknown trait data

We estimated unknown phenotypes and genotypes using hidden state prediction methods based on phylogenetic inference (Supplementary Data 5). Specifically, we generated a phylogenetic tree with the 3,311 OTUs from our USEARCH pipeline (before any taxa were lost due to rarefying) and the 13,900 OTUs from the 132 release of the Living Tree Project (LTP)[29] (Supplementary Figure 7; Supplementary Data 6). The topology of the tree reflects percent sequence similarity among taxa in the 16S rRNA V4 region, and was generated using agglomerative clustering of a distance matrix based on the U-sort heuristic[58]. Because LTP representative sequences were of the entire 16S rRNA gene (i.e., the ribosomal small subunit), they were truncated to the 250 bp of the V4 region using 515F and 806R primers before generating the distance matrix. Trait data were then mapped onto the tips of the phylogenetic tree with

Latin binomials. The LTP database was uniquely well-suited to interface with literature-derived trait data because each sequence represents a type strain with Genus and Species annotations drawn from the literature, not inferred phylogenetically.

Missing trait values were estimated using three hidden state prediction algorithms: independent contrasts, subtree averaging, and weighted squared-change parsimony, each calculated using the R package Castor version 1.3.4[63]. The three methods have different strengths and weaknesses[63,64], but their predictions correlated strongly (Supplementary Table 1), lending confidence to our results. We ultimately used weighted square-change parsimony for our analysis, which recursively calculates locally parsimonious states for each node based on its descending subtree, until reaching a parsimonious state estimate for the tree root[65]. Because all trait values were either numeric or converted to numeric (e.g., Gram-negative = 0 and Gram-positive = 1), state predictions for discrete traits could be fractional (e.g., a Gram-positive score of 0.5), reflecting their probabilistic uncertainty.

Methods of hidden state prediction offer estimates for all taxa with hidden states, even when there is not sufficient confidence to warrant estimation. To mitigate this, we examined how trait dissimilarity varied with increasing phylogenetic distance, and only used predictions when there were closely-related taxa with known trait values (refer to Supplementary Figure 8 for a graphical depiction of the approach. Specifically, for each trait, the phylogenetic tree was pruned such that only OTUs (i.e., tree tips) that could be linked to direct trait observations remained. Next, differences in trait values and phylogenetic distance (i.e., percent 16S rRNA V4 sequence similarity) were calculated for all OTU pairs. In some cases, the number of OTU pairs was prohibitively large, in which cases only 10,000 pairs were randomly selected at each 0.005 increment of phylogenetic distance. Five generic models were then used to predict trait differences, $|y|$, as a function of phylogenetic distance, x, and the best fitting model of trait evolution was selected by AIC. The models included: (1) Null: $|y| \sim 1$; (2) Linear regression: $|y| \sim x$; (3) Logarithmic regression: $|y| \sim \log(x)$; (4) Asymptotic regression: $|y| \sim a(1 - \mathrm{e}^{(-\mathrm{e}^{b}x)})$, where a

and b were determined using a self-starting nonlinear least squares approach, and the model fit was constrained to pass through the origin; and (5) Logistic regression: $|y| \sim \frac{a}{1+e^{(\frac{b-x}{c})}}$, where a, b, and c were determined using a self-starting nonlinear least squares approach. Null models provided the best fit for aggregation score, IgA binding affinity, pH optimum, and salt optimum, indicating that for these traits trait data should not be estimated at any phylogenetic distance. For the remaining 12 traits, trait predictions were used only when taxa associated with direct trait observations occurred within trait-specific thresholds of phylogenetic distance; we defined these thresholds as the points at which model predictions rose to 90 percent of null expectations (Table 2, Supplementary Figure 8). Null expectations equaled the mean trait-based differences of all OTU pairs with more than 0.1 phylogenetic distance between them, for each trait. Overall, for the traits that were amenable to hidden state prediction, this approach yielded trait predictions for 78.7 percent (16S rRNA gene copy number) to 99.9 percent (Temperature optimum) of sequences used in this study (Supplementary Figure 9). We assessed statistical independence among traits predictions using Pearson correlation coefficients; p-values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure.

Trait-based successional patterns within and across infants

Trait-based successional patterns were evaluated at both the OTU-level and the community-level (i.e., on the level of individual samples). For the OTU-level analysis, OTUs were assigned one of three successional stages based on results of linear models of OTU abundances over time across all infants: early successional OTUs were defined as those with statistically significant negative trends in abundance over time ($p < 0.05$, $\beta < 0$); late successional OTUs were defined as those with positive trends in abundance over time ($p < 0.05$, $\beta > 0$); otherwise, taxa were combined into a single category which included OTUs with sporadic, unvarying, or hump-shaped patterns of abundance over time. Statistical

409 differences in the trait values of OTUs in the three groups were evaluated with Welch t-tests; p-values
410 were adjusted for multiple comparisons using the Benjamini-Hochberg procedure.
411 Trait-based differences at the community level were quantified using CWMs. A CWM is the
412 mean trait value of the species or OTUs in a community, weighted by their abundances. Here, a CWM is
413 formally equal to $\sum_{i=1}^{S} p_i x_i$, where p_i is the abundance of OTU i ($i = 1,2,\dots S$), and x_i is the trait value
414 for OTU i. We used Welch t-tests to test for differences in CWMs between infants treated with and
415 without antibiotics, and infants delivered by C-section and vaginally, for each six-month period of infant
416 development; p-values were adjusted for multiple comparisons using the Benjamini-Hochberg
417 procedure.
418
419 *Comparison of taxonomic and trait-based turnover*
420 We quantified differences in microbiome community compositions in two ways. First, we used Bray-
421 Curtis dissimilarity to quantify differences in the OTU-based compositions of samples[66]. Second, we
422 quantified trait-based differences among communities with multidimensional Euclidean distance[67].
423 Specifically, Euclidean distance between two communities was calculated by (1) scaling trait values by
424 their standard deviations to give each trait equal weight, (2) calculating the CWMs of each trait for both
425 communities, and then (3) using the Pythagorean theorem to determine the distance between the two
426 communities in n-dimensional trait space.
427 We examined OTU-based and trait-based community changes over time in two ways. First, to
428 quantify changes in short-term compositional variability over infant development, we examined
429 compositional differences of subsequent samples from the same infant, at intervals approximately
430 between one to three months. Second, to quantify rates of directional turnover over infant
431 development, we examined compositional differences between samples and the final sample from each
432 infant. To determine whether trait-based rates of compositional variability and directional turnover

exceeded those expected by chance, we compared observed rates of trait-based turnover to null models of trait-agnostic community change. Specifically, we generated 1000 mock versions of our data with trait values randomly shuffled among OTUs, and recalculating pairwise sample dissimilarities. In other words, null models reflect what trait-based turnover would have been if organismal traits were unrelated to performance. We tested for statistical differences between observed and null turnover rates within six-month periods using Welch t-tests.

To determine if community composition converged or diverged across infants as development progressed, we divided samples into one-month slices and calculated mean OTU-based and trait-based distances for all pairwise combinations of samples, excluding pairs of samples from the same infant. To determine whether observed rates of trait-based compositional convergence/divergence across infants differed from those expected by chance, we compared our observations to null models of trait-agnostic community changes over time. Similar to our analysis of trait-based turnover within infants, null models were performed by randomly shuffling trait values among OTUs and recalculating pairwise sample dissimilarities. We tested for statistical differences between observed and null model rates of convergence within six-month periods using Welch t-tests.

Data availability

Raw sequencing data are available online at the NCBI project accession numbers PRJNA231909 [https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA231909] and PRJNA290381 [https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA290381]. Custom scripts used in the bioinformatic pipeline and statistical analyses are available at https:/github.com/ShadeLab/microbiome_trait_succession. All relevant data used in this study are included as supplementary data files, available at https://figshare.com/projects/Trait-based_succession_of_the_infant_gut_microbiome/58202.

Author contributions

JG, AS and EL conceived the study, JG developed methods, analyzed data, and wrote the manuscript, and all authors discussed analysis and revised the manuscript.

Competing interests

The authors declare no competing interests.

Acknowledgments

This work was supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research. AS acknowledges support from the National Science Foundation under Grant No DEB#1749544. JG was supported by the Michigan State University Foundation funding to EL.

References

1. Pickett, S. T. A., Collins, S. L. & Armesto, J. J. Models, mechanisms and pathways of succession. *Bot. Rev.* **53**, 335–371 (1987).
2. Tilman, D. Constraints and Tradeoffs: Toward a Predictive Theory of Competition and Succession. *Oikos* **58**, 3 (1990).
3. Lozupone, C. *et al.* Identifying genomic and metabolic features that can underlie early successional and opportunistic lifestyles of human gut symbionts. *Genome Res.* **22**, 1974–1984 (2012).
4. Leibold, M. A. *et al.* The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* **7**, 601–613 (2004).
5. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
6. Belenguer, A. *et al.* Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. *Appl. Environ. Microbiol.* **72**, 3593–9 (2006).
7. Adair, K. L. & Douglas, A. E. Making a microbiome: the many determinants of host-associated microbial community composition. *Curr. Opin. Microbiol.* **35**, 23–29 (2017).
8. Fierer, N., Nemergut, D., Knight, R. & Craine, J. M. Changes through time: Integrating microorganisms into the study of succession. *Res. Microbiol.* **161**, 635–642 (2010).
9. Ardissone, A. N. *et al.* Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS One* **9**, 1–8 (2014).
10. Fernández, L. *et al.* The human milk microbiota: Origin and potential roles in health and disease. *Pharmacol. Res.* **69**, 1–10 (2013).
11. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during

health and disease. *Nature Reviews Immunology* **9**, 313–323 (2009).

12. Li, H. *et al.* The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.* **6**, (2015).

13. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2013).

14. Vallès, Y. *et al.* Microbial Succession in the Gut: Directional Trends of Taxonomic and Functional Change in a Birth Cohort of Spanish Infants. *PLoS Genet.* **10**, (2014).

15. McGill, B. J., Enquist, B. J., Weiher, E. & Westoby, M. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* **21**, 178–185 (2006).

16. Violle, C. *et al.* Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).

17. Cornwell, W. & Ackerly, D. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecol. Monogr.* **79**, 109–126 (2009).

18. Guittar, J. L., Dear, F. & Vaughan, C. Scarlet Macaw (Ara macao, Psittaciformes: Psittacidae) nest characteristics in the Osa Peninsula Conservation Area (ACOSA), Costa Rica. *Rev. Biol. Trop.* **57**, 387–93 (2009).

19. Frimpong, E. A. & Angermeier, P. L. Traits-based approaches in the analysis of stream fish communities. *Am. Fish. Soc. Symp.* **73**, 109–136 (2010).

20. Luck, G. W. *et al.* Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. *J. Anim. Ecol.* **81**, 1065–1076 (2012).

21. Litchman, E. & Klausmeier, C. A. Trait-Based Community Ecology of Phytoplankton. *Annu. Rev. Ecol. Evol. Syst.* **39**, 615–639 (2008).

22. Allison, S. D. A trait-based approach for modelling microbial litter decomposition. *Ecol. Lett.* **15**, 1058–1070 (2012).

23. Ortiz-Álvarez, R., Fierer, N., de Los Ríos, A., Casamayor, E. O. & Barberán, A. Consistent changes in

the taxonomic structure and functional attributes of bacterial communities during primary succession. *ISME J.* **12**, 1658–1667 (2018).

24. Browne, H. P. *et al.* Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543–546 (2016).

25. Söhngen, C. *et al.* Bac Dive – The Bacterial Diversity Metadatabase in 2016. *Nucleic Acids Res.* **44**, D581–D585 (2016).

26. Barberán, A., Caceres Velazquez, H., Jones, S. & Fierer, N. Hiding in Plain Sight: Mining Bacterial Species Records for Phenotypic Trait Information. *mSphere* **2**, e00237-17 (2017).

27. Kostic, A. D. *et al.* The Dynamiccs of the human gut microbiome in development and in progression towards type 1 diabetes. **17**, 260–273 (2016).

28. Yassour, M. *et al.* Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci. Transl. Med.* **8**, 343ra81 (2016).

29. Muñoz, R. *et al.* Release LTPs104 of the All-Species Living Tree. *Syst. Appl. Microbiol.* **34**, 169–170 (2011).

30. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl**, 4578–4585 (2011).

31. Lavorel, S. & Garnier, E. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Funct. Ecol.* **16**, 545–556 (2002).

32. Rivera-Chávez, F. *et al.* Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. *Cell Host Microbe* **19**, 443–454 (2016).

33. Nachamkin, I., Yang, X. H. & Stern, N. J. Role of Campylobacter jejuni flagella as colonization factors for three- day-old chicks: Analysis with flagellar mutants. *Appl. Environ. Microbiol.* **59**, 1269–1273 (1993).

34. Josenhans, C. & Suerbaum, S. The role of motility as a virulence factor in bacteria. *Int. J. Med.*

Microbiol. **291**, 605–614 (2002).

35. Lennon, J. & Jones, S. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.* (2011).

36. Klappenbach, J. A., Dunbar, J. M., Thomas, M. & Schmidt, T. M. rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria. *Appl. Envir. Microbiol.* **66**, 1328–1333 (2000).

37. Nemergut, D. R. *et al.* Decreases in average bacterial community rRNA operon copy number during succession. *ISME J.* **10**, 1147–1156 (2016).

38. Litchman, E., Edwards, K. F. & Klausmeier, C. A. Microbial resource utilization traits and trade-offs: Implications for community structure, functioning, and biogeochemical impacts at present and in the future. *Front. Microbiol.* **6**, 254 (2015).

39. Salminen, S., Gibson, G. R., McCartney, A. L. & Isolauri, E. Influence of mode of delivery on gut microbiota composition in seven year old children. *Gut* **53**, 1388–9 (2004).

40. Azad, M. B. *et al.* Gut microbiota of healthy Canadian infants: Profiles by mode of delivery and infant diet at 4 months. *CMAJ* **185**, 385–394 (2013).

41. Fukami, T. & Nakajima, M. Community assembly: Alternative stable states or alternative transient states? *Ecol. Lett.* **14**, 973–984 (2011).

42. Sprockett, D., Fukami, T. & Relman, D. A. Role of priority effects in the early-life assembly of the gut microbiota. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 197–205 (2018).

43. Ferretti, P. *et al.* Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe* **24**, 133–145.e5 (2018).

44. Mcdonnell, G. & Russell, A. D. Antiseptics and disinfectants: Activity, action, and resistance. *Clin. Microbiol. Rev.* **12**, 147–179 (1999).

45. Jernberg, C., Löfmark, S., Edlund, C. & Jansson, J. K. Long-term impacts of antibiotic exposure on

the human intestinal microbiota. *Microbiology* **156**, 3216–3223 (2010).

46. Cavender-Bares, J., Ackerly, D., Baum, D. & Bazzaz, F. Phylogenetic overdispersion in Floridian oak communities. *Am. Nat.* **163**, 823–843 (2004).

47. Mayfield, M. M. & Levine, J. M. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecol. Lett.* **13**, 1085–93 (2010).

48. Anderson, K. J. Temporal Patterns in Rates of Community Change during Succession. *Am. Nat.* **169**, 780–793 (2007).

49. Louca, S. *et al.* Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* **2**, 936–943 (2018).

50. Corfe, B. M., Harden, C. J., Bull, M. & Garaiova, I. The multifactorial interplay of diet, the microbiome and appetite control: current knowledge and future challenges. *Proc. Nutr. Soc.* **74**, 235–244 (2015).

51. Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci.* **110**, 20236–20241 (2013).

52. Allison, S. & Martiny, J. Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci.* **105**, 11512–11519 (2008).

53. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837–848 (2006).

54. Strickland, M. S., Lauber, C., Fierer, N. & Bradford, M. A. Testing the functional significance of microbial community composition. *Ecology* **90**, 441–451 (2009).

55. Sekirov, I., Russell, S. & Antunes, L. Gut microbiota in health and disease. *Physiol. Rev.* **90**, 859–904 (2010).

56. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15**, R89 (2014).

57. Srivastava, A., Gupta, J., Kumar, S. & Kumar, A. Gut biofilm forming bacteria in inflammatory bowel disease. *Microbial Pathogenesis* **112**, 5–14 (2017).

58. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **27**, 2194–2200 (2010).

59. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: imporved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).

60. Magnúsdóttir, S., Ravcheev, D., De Crécy-Lagard, V. & Thiele, I. Systematic genome assessment of B-vitamin biosynthesis suggests cooperation among gut microbes. *Front. Genet.* **6**, (2015).

61. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, (2014).

62. Palm, N. W. *et al.* Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000–1010 (2014).

63. Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2017).

64. Zaneveld, J. R. R. & Thurber, R. L. V. Hidden state prediction: A modification of classic ancestral state reconstruction algorithms helps unravel complex symbioses. *Front. Microbiol.* **5**, 1–8 (2014).

65. Maddison, W. P. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Syst. Biol.* **40**, 304–314 (1991).

66. Beals, E. W. Bray-Curtis Ordination: an effective strategy for analysis of multivariate ecological data. *Adv. Ecol. Res.* **14**, 1–55 (1984).

67. Petchey, O. L. & Gaston, K. J. Functional diversity (FD), species richness and community composition. *Ecol. Lett.* **5**, 402–411 (2002).

68. Stoddard, S. F., Smith B.J., R., H., Roller, B. R. K. & T.M., S. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development.

Nucleic Acids Res. 2014 (2015).

69. Genome [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. Available from: https://www.ncbi.nlm.nih.gov/genom

70. Mukherjee, S. *et al.* Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky977

Figure 1 | OTU and family-level abundance patterns over succession. Colors reflect successional status; taxa were categorized as early/late successional if their abundances across all infants trended significantly negative/positive ($p < 0.05$) over time based on linear regressions; OTUs that did not trend significantly over time were grouped into the mid-successional or no-trend category. **(a) (c) (e)** Percent total abundances of OTUs of each successional category over time. **(b) (d) (f)** Proportions of total abundances of the top five most-abundant families in each successional group.

Figure 2 | Abundance-weighted trait means over succession. Filled red circles show average abundance-weighted trait means of samples in that month. N is equal to the number of samples in each month, and ranges from 27 to 59. Vertical red lines show 95 percent confidence intervals. Black trendlines were fit using generalized additive models.

Figure 3 | Trait-based successional patterns differ by delivery mode and antibiotic history. Abundance-weighted trait means over infant microbiome succession, grouped by infant delivery mode and antibiotic history. Filled circles show average abundance-weighted trait means of samples within six-month periods in each cohort of infants. N is equal to the number of samples in each six-month period; there were between 25 to 31 total samples from six infants delivered by C-section who received little to no antibiotics, 66 to 91 total samples from 18 infants who were treated with antibiotics for at least 50 days, and 72 to 93 total samples from 18 control infants that were delivered vaginally and received no antibiotics. Vertical lines show 95 percent confidence intervals. Asterisks denote significance based on Welch t-tests performed between each treatment group and the control group (*: adjusted $p < 0.05$; **: adjusted $p < 0.01$; ***: adjusted $p < 0.001$).

643 **Figure 4 | Trait-based composition stabilizes earlier than taxonomic composition.** Filled circles show
644 mean pairwise compositional dissimilarities of gut microbiome samples collected from individual infants,
645 averaged within six-month periods for each infant, and then across infants. OTU-based dissimilarity was
646 calculated using Bray-Curtis dissimilarity. Trait-based dissimilarity was calculated using multidimensional
647 Euclidean distance after scaling the distributions of values for each trait to ensure equal contribution. **(a)**
648 Mean OTU-based dissimilarity between subsequent samples declines slightly over time. **(b)** Mean OTU-
649 based dissimilarity between samples and the final samples taken from each infant decreases steadily
650 throughout the sampling period until finally reaching baseline levels of between-sample dissimilarity in
651 the last six-month period, seen in **a**. **(c)** Mean trait-based dissimilarity between subsequent samples
652 appears elevated in the first year, but does not differ significantly from null model predictions that
653 assumed trait-agnostic turnover. **(d)** Mean trait-based dissimilarity between samples and the final
654 samples taken from each infant decreases rapidly and approaches baseline-levels of between-sample
655 dissimilarity within the first year, seen in **c**. Moreover, trait-based community composition converges
656 towards that of the final sample significantly faster than null model expectations, illustrating the non-
657 random nature of community turnover over succession. In all panels, N equals 56, the number of
658 infants. Vertical lines show 95 percent confidence intervals. Asterisks denote significance between
659 observed and null model predictions based on Welch t-tests (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).
660
661 **Figure 5 | Infants' microbiomes converge compositionally over time.** Filled circles show mean
662 compositional dissimilarities of gut microbiomes across infants within each six-month periods. Mean
663 dissimilarities were calculated by first taking the mean dissimilarity of all sample pairs, except those
664 from the same infant, in each of the first 36 months of development (for these means, N ranges from 88
665 to 3410), and then taking their means within each six month period; hence, for each circle, N equals 6.
666 OTU-based dissimilarity was calculated using Bray-Curtis dissimilarity. Trait-based dissimilarity was

667 calculated using multidimensional Euclidean distance after scaling the distributions of values for each
668 trait to ensure equal contribution. **(a)** OTU-based dissimilarity among infants decreased slightly over
669 time, indicating a modest convergence in taxonomic composition. **(b)** Trait-based dissimilarity among
670 infants fell quickly over the first 18 months and then remained relatively static thereafter, indicating
671 rapid convergence in trait-based composition during early succession. The magnitude of trait-based
672 compositional convergence across infants was significantly greater than predicted by a null model
673 assuming trait-agnostic turnover. Vertical lines show 95 percent confidence intervals. Asterisks denote
674 significance between observed and null model predictions based on Welch t-tests (*: $p < 0.05$; **: $p <$
675 0.01; ***: $p < 0.001$).
676

Trait	Description / Units	Sources
Aggregation score	0 (never) to 1 (observed aggregation)	BacDive[25]; IJSEM[26]
B vitamins	No. B-vitamin pathways in genome	Ref. 60
16S gene copies	No. in 16S rRNA gene copies in genome	rrnDB[68]
GC content	Percent guanine and cytosine in genome	IJSEM[26]; NCBI[69]
Gene number	No. genes in genome	NCBI[69]
Genome size	Genome size in megabases	NCBI[69]
Gram-positive	0 (Gram-negative) to 1 (Gram-positive)	BacDive[25]; GOLD[70]; IJSEM[26]
IgA binding affinity	log ([IgA+]/[IgA-] + 1)	Ref. 62
Length	log (μm)	BacDive[25]; GOLD[70]; IJSEM[26]
Motility	0 (never motile) to 1 (always motile)	BacDive[25]; GOLD[70]; IJSEM[26]
Oxygen tolerance	0 (obligate anaerobe) to 5 (obligate aerobe)	BacDive[25]; GOLD[70]; IJSEM[26]
pH optimum	pH	GOLD[70]; IJSEM[26]
Salt optimum	g per l	IJSEM[26]
Sporulation score	0 (never sporulates) to 1 (sporulates easily)	BacDive[25]; GOLD[70]; IJSEM[26]; ref. 24
Temperature optimum	°C	IJSEM[26]
Width	log (μm)	BacDive; GOLD[70]; IJSEM[26]

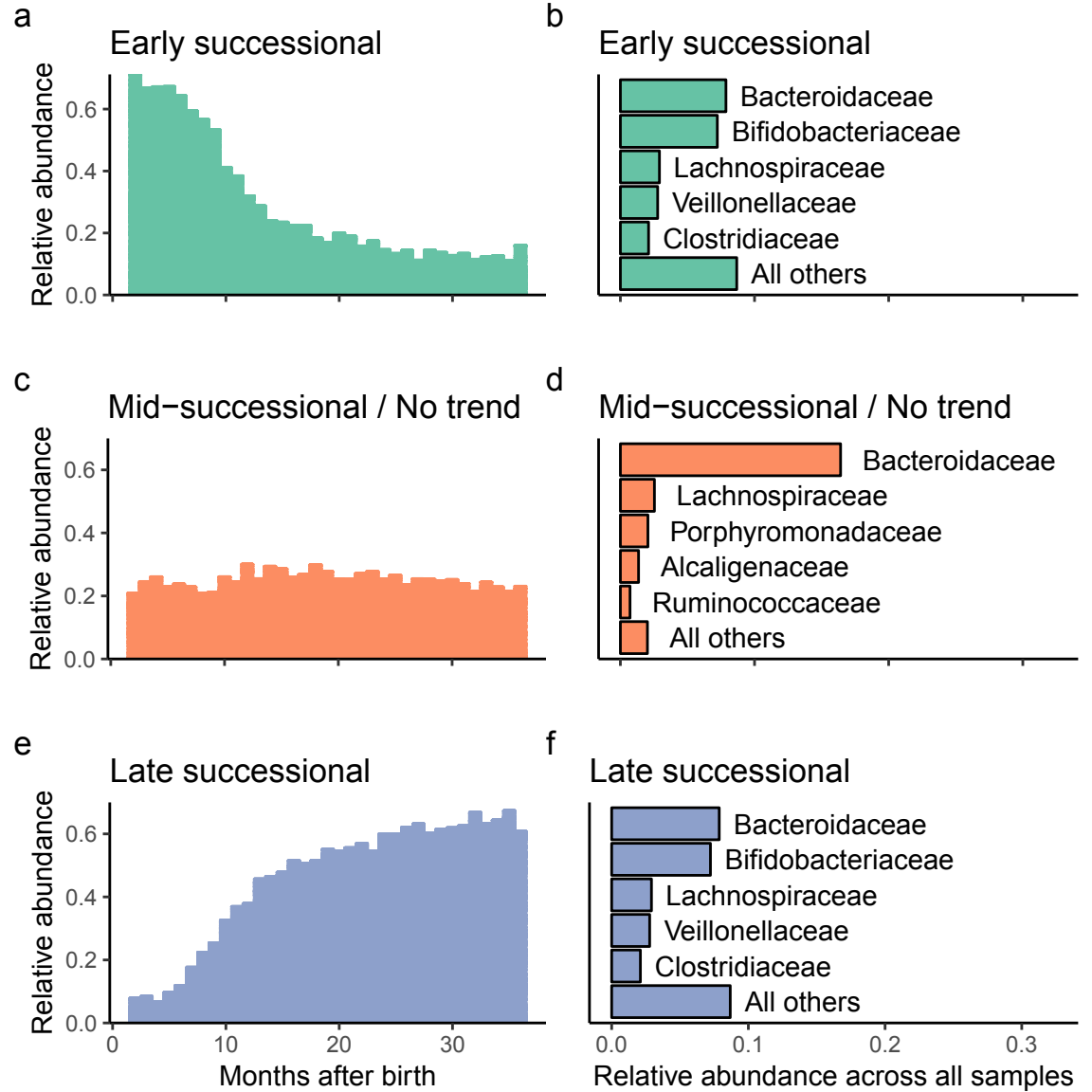
Table 1 | Sources of trait data gathered in this study. IgA: Immunoglobulin A; BacDive: Bacterial Diversity Metadatabase. IJSEM: International Journal of Systematic and Evolutionary Microbiology. GOLD: Genomes OnLine Database: Joint Genome Institute; NCBI: National Center for Biotechnology Information; rrnDB: the ribosomal RNA operon copy number database.

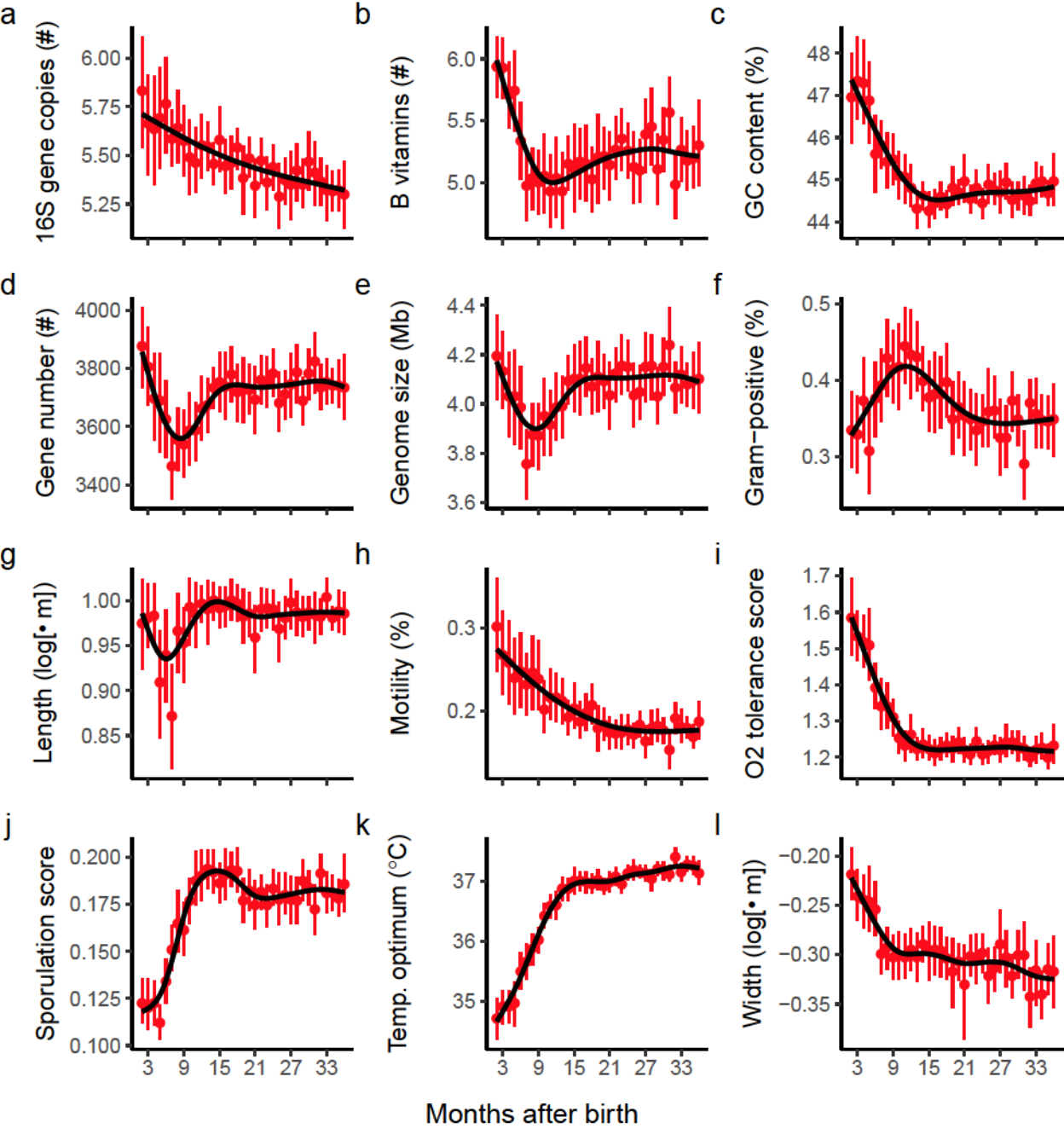
677

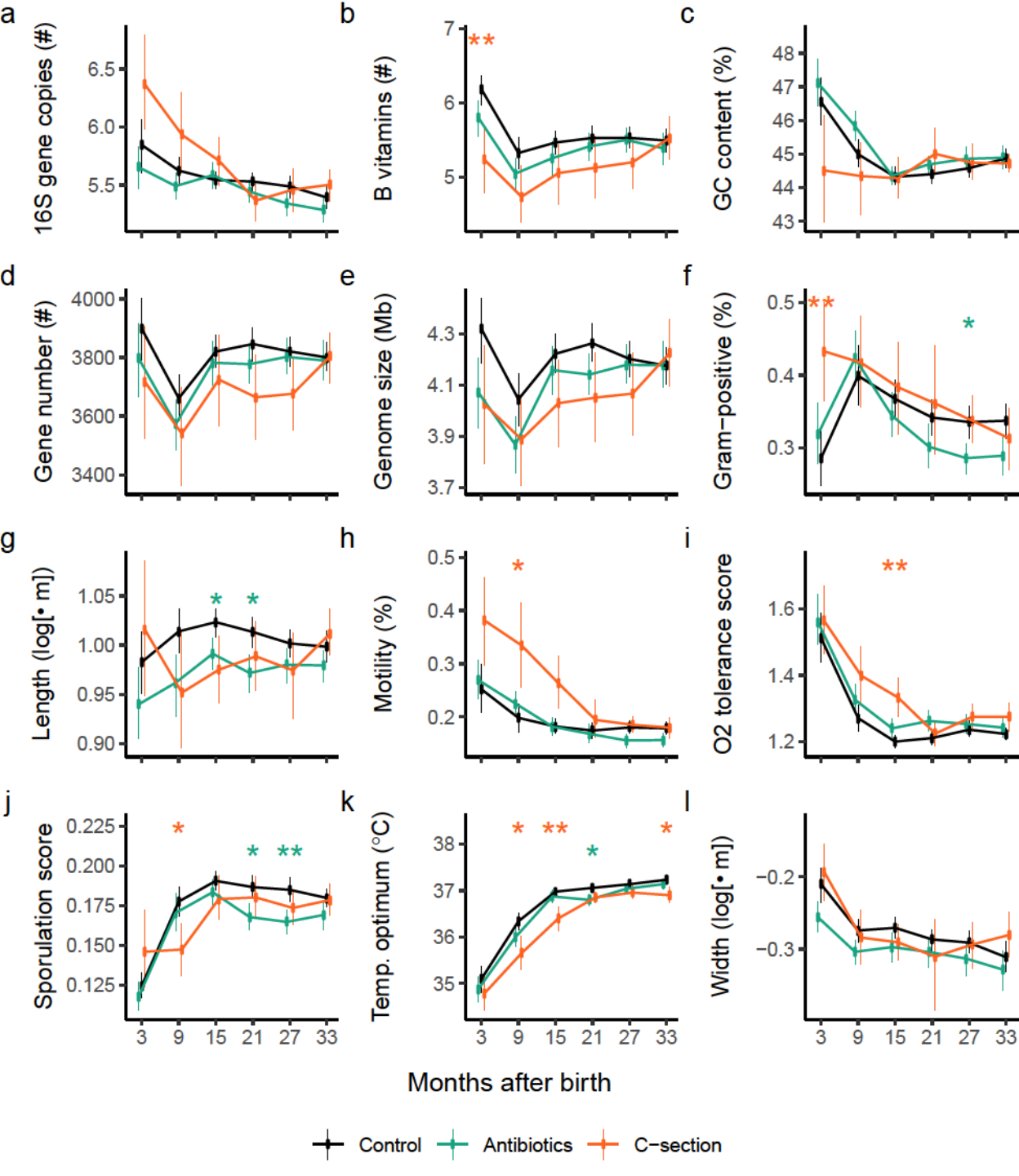
Trait	Max. distance
Aggregation score	0.00
B vitamins	0.06
16S gene copies	0.05
GC content	0.12
Gene number	0.08
Genome size	0.09
Gram-positive	0.10
IgA binding affinity	0.00
Length	0.12
Motility	0.08
Oxygen tolerance	0.11
pH optimum	0.00
Salt optimum	0.00
Sporulation score	0.06
Temperature optimum	0.14
Width	0.12

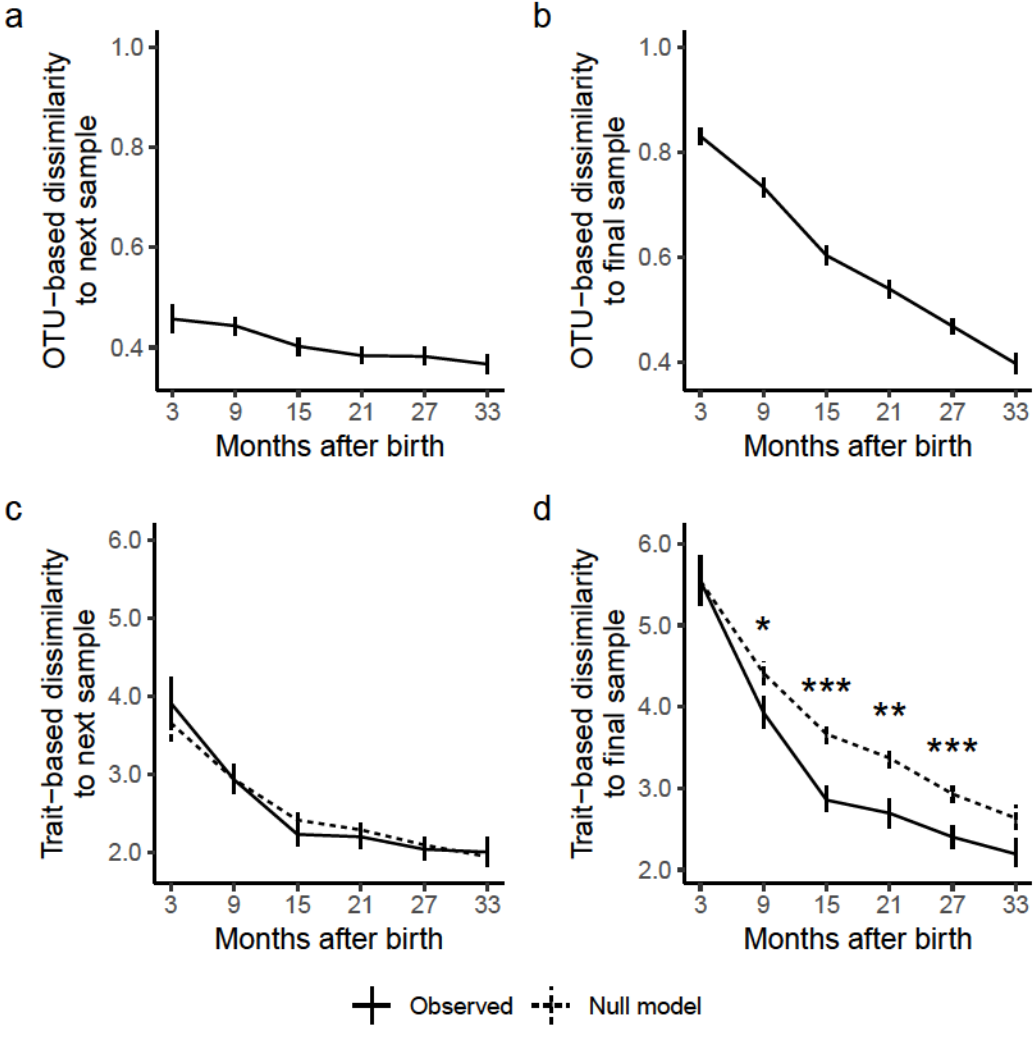
Table 2 | Maximum phylogenetic distances used to infer trait values. Percent sequence dissimilarities (i.e., phylogenetic distances) in the 16S rRNA V4 region at which statistical support for trait conservatism disappears for each trait (see Methods and Supplementary Figure 9).

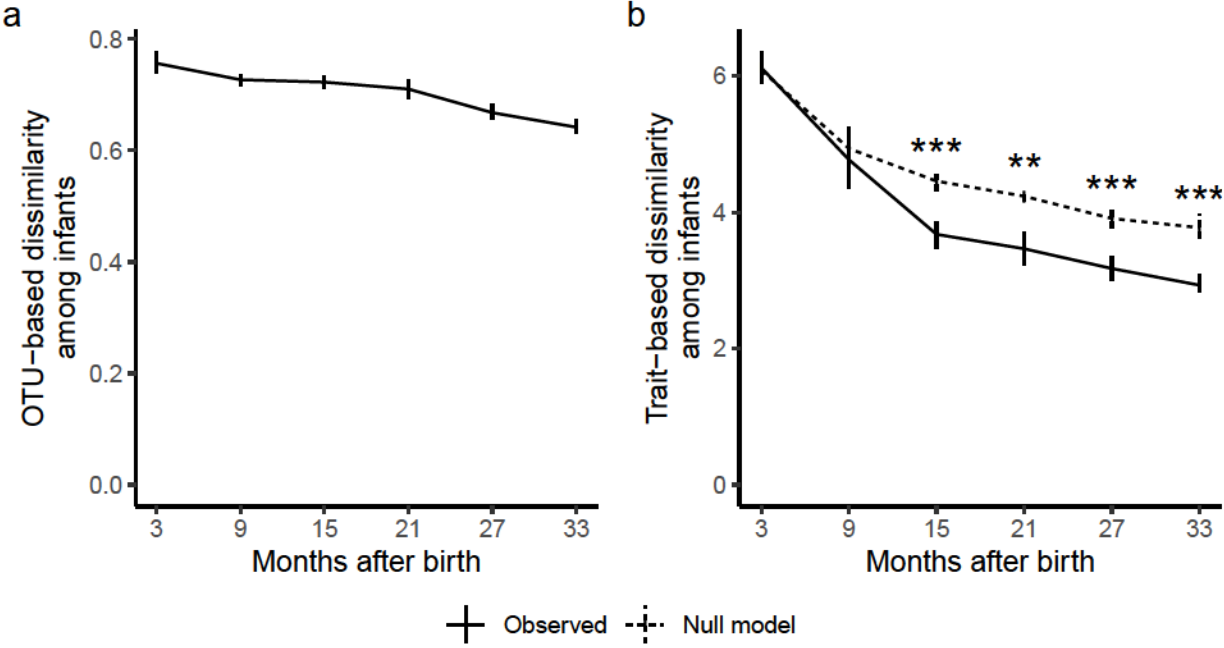
678
679


a
Early successional
Relative abundance
0.0
0.2
0.4
0.6
b
Early successional
Bacteroidaceae
Bifidobacteriaceae
Lachnospiraceae
Veillonellaceae
Clostridiaceae
All others
c
Mid−successional / No trend
Relative abundance
0.0
0.2
0.4
0.6
d
Mid−successional / No trend
Bacteroidaceae
Lachnospiraceae
Porphyromonadaceae
Alcaligenaceae
Ruminococcaceae
All others
e
Late successional
Relative abundance
0.0
0.2
0.4
0.6
0
10
20
30
Months after birth
f
Late successional
Bacteroidaceae
Bifidobacteriaceae
Lachnospiraceae
Veillonellaceae
Clostridiaceae
All others
0.0
0.1
0.2
0.3
Relative abundance across all samples


a
16S gene copies (#)
6.00
5.75
5.50
5.25
b
B vitamins (#)
6.0
5.5
5.0
c
GC content (%)
48
47
46
45
44
d
Gene number (#)
4000
3800
3600
3400
e
Genome size (Mb)
4.4
4.2
4.0
3.8
3.6
f
Gram−positive (%)
0.5
0.4
0.3
g
Length (log[• m])
1.00
0.95
0.90
0.85
h
Motility (%)
0.3
0.2
i
O2 tolerance score
1.7
1.6
1.5
1.4
1.3
1.2
j
Sporulation score
0.200
0.175
0.150
0.125
0.100
k
Temp. optimum (°C)
37
36
35
l
Width (log[• m])
−0.20
−0.25
−0.30
−0.35
3
9
15
21
27
33
Months after birth


a
16S gene copies (#)
6.5
6.0
5.5
b
B vitamins (#)
**
7
6
5
c
GC content (%)
48
47
46
45
44
43
d
Gene number (#)
4000
3800
3600
3400
e
Genome size (Mb)
4.3
4.1
3.9
3.7
f
Gram−positive (%)
**
*
0.5
0.4
0.3
g
Length (log[• m])
* *
1.05
1.00
0.95
0.90
h
Motility (%)
*
0.5
0.4
0.3
0.2
i
O2 tolerance score
**
1.6
1.4
1.2
j
Sporulation score
*
* **
0.225
0.200
0.175
0.150
0.125
k
Temp. optimum (°C)
* ** * *
38
37
36
35
l
Width (log[• m])
−0.2
−0.3
3 9 15 21 27 33
Months after birth
Control
Antibiotics
C−section


a
OTU−based dissimilarity to next sample
1.0
0.8
0.6
0.4
3
9
15
21
27
33
Months after birth
b
OTU−based dissimilarity to final sample
1.0
0.8
0.6
0.4
3
9
15
21
27
33
Months after birth
c
Trait−based dissimilarity to next sample
6.0
5.0
4.0
3.0
2.0
3
9
15
21
27
33
Months after birth
d
Trait−based dissimilarity to final sample
6.0
5.0
4.0
3.0
2.0
*
***
**
***
3
9
15
21
27
33
Months after birth
Observed
Null model


a
b
OTU−based dissimilarity among infants
Trait−based dissimilarity among infants
0.0
0.2
0.4
0.6
0.8
0
2
4
6
3
9
15
21
27
33
Months after birth
***
**
***
***
Observed
Null model

Description of Additional Supplementary Files

File name: Supplementary Data 1
Description: **Relative abundances of OTUs in infant microbiome samples.** Raw 16S rRNA V4 amplicon sequencing data from infant gut microbiome samples were processed using a USEARCH pipeline and clustered into operational taxonomic units (OTUs) at 97 percent similarity. Samples with fewer than 5000 initial sequences were excluded, and the remaining samples were rarefied to 5000 sequences, resulting in a drop from 3311 to 2416 OTUs. The 'subject' column can be used with Supplementary Data 2 to find associated subject metadata (e.g., delivery method, antibiotic treatment history), and the 't' column refers to the month of sampling. Available at https://doi.org/10.6084/m9.figshare.7499498.

File name: Supplementary Data 2
Description: **Infant metadata.** Information for each infant subject includes (1) the approximate total number of days of antibiotic treatment, assuming the duration of any given antibiotic treatment was seven days unless otherwise noted, (2) the country of origin, (3) the mode of delivery, (4) the treatment group assignment used in this study, and (5) references to the prior studies from which the metadata were drawn. Available at https://doi.org/10.6084/m9.figshare.7499525. Additional subject metadata is available in Kostic et al. 2016 (ref. 27 in the main text) and Yassour et al. 2016 (ref. 28 in the main text).

File name: Supplementary Data 3
Description: **SILVA-based taxonomic identities of infant microbiome OTUs.** Representative sequences of operational taxonomic units (OTUs) from infant microbiome samples were mapped to the SILVA v123 database using USEARCH version 10.0.240 to assign potential taxonomic identities. Data include all 3311 OTUs identified in our pipeline, before any were dropped due to rarefying. Available at https://doi.org/10.6084/m9.figshare.7500422.

File name: Supplementary Data 4
Description: **Phylogenetic tree of OTUs from this study and the Living Tree Project.** A phylogenetic tree in Newick format with tips for the 3,311 operational taxonomic units (OTUs) identified in the infant gut microbiome samples from this study identified using a USEARCH pipeline, and the 13,900 OTUs from the 132 release of the Living Tree Project. The topology of the tree reflects percent sequence similarity among taxa in the 16S rRNA V4 region (refer to Methods). Available at https://doi.org/10.6084/m9.figshare.7500563.

File name: Supplementary Data 5
Description: **Trait data mined from literature and online data repositories.** Trait data derived from the sources listed in Table 1 with Latin binomials that matched either those from the SILVA-based taxonomic identifications of OTUs found in infant gut microbiome samples, or from type specimens in the 132 release of the Living Tree Project. When observations existed for the same taxon across more than one data source, means were used. Available at https://doi.org/10.6084/m9.figshare.7501001.

File name: Supplementary Data 6
Description: **Trait data used in our analyses.** Data included trait values that could be directly associated to taxa in our study based on matching Latin binomials, and trait value estimates based on hidden state prediction using weighted square-change parsimony (refer to Methods and Supplementary Figure 8). Because we converted all trait values to numeric (e.g., Gram-negative = 0 and Gram-positive = 1), state predictions for initially discrete traits were allowed to be fractional (e.g., a Gram-positive score of 0.5), reflecting their probabilistic uncertainty. Available at https://doi.org/10.6084/m9.figshare.7501121.