

Rapid evolution and strain turnover in the infant gut microbiome

Daisy W. Chen¹ and Nandita R. Garud^{2,3,*}

1. Computation and Systems Biology, University of California, Los Angeles
2. Department of Ecology and Evolutionary Biology, University of California, Los Angeles
3. Department of Human Genetics, University of California, Los Angeles

* correspondence to ngarud@ucla.edu

ABSTRACT

While the ecological dynamics of the infant gut microbiome have been intensely studied, relatively little is known about the evolutionary dynamics in the infant gut microbiome. Here we analyze longitudinal fecal metagenomic data from >700 infants and their mothers over the first year of life and find that the evolutionary dynamics in infant gut microbiomes are distinct from that of adults. We find evidence for almost 100-fold increase in the rate of evolution and strain turnover in the infant gut compared to healthy adults, with the mother-infant transition at delivery being a particularly dynamic period in which gene loss dominates. Within a few months after birth, these dynamics stabilize, and gene gains become increasingly frequent as the microbiome matures. We furthermore find that evolutionary changes in infants show signatures of being seeded by a mixture of *de novo* mutations and transmissions of pre-evolved lineages from the broader family. Several of these evolutionary changes occur in parallel in multiple infants, highlighting candidate genes that may play important roles in the development of the infant gut microbiome. Our results point to a picture of a volatile infant gut microbiome characterized by rapid evolutionary and ecological change in the early days of life.

INTRODUCTION

The infant gut microbiome is a rapidly changing ecosystem that plays a crucial role in infant health, including milk digestion [1], immune system development [2], and prevention of colonization of pathogens [1,3]. Given its importance, there has been substantial research focus on the ecological and functional maturation of the infant gut microbiome [4–14] and the extent that mode of delivery, feeding, and exposure to antibiotics alters its development [11,12,15–20]. Strain tracking methods using single nucleotide variants (SNVs) have revealed that mothers [14,20–22], the family at large [13,23], and hospitals [24,25] play a critical role in seeding the

infant microbiome. By contrast, how these lineages evolve once they colonize the infant remains unknown.

Evolutionary changes are important to characterize because the rise of new genetic variants ultimately drives the emergence of new traits. For example, genetic variants in the human gut microbiome are known to confer traits such as the ability to digest food [26,27], metabolize drugs [28], and evade antibiotics [29]. Recent studies illustrate the pervasiveness of evolutionary changes in adult gut microbiomes, which include changing SNV frequencies and horizontal gene transfers on short time scales of just a few days, weeks, and months [30–37]. Interestingly, rapid evolutionary changes within-species frequently do not detectably alter species relative abundances [35], illustrating that intra-species genetic variation in the microbiome can reveal important changes in the microbiome that species abundances cannot. Given the functional importance of genetic variation in the microbiome, it is necessary that we characterize the typical evolutionary dynamics that occur over the course of infant maturation.

There is reason to believe that the targets of selection as well as the tempo and mode of evolution in infant gut microbiomes are substantially different from those of adults. Compared to adults, microbes in infants confront unique selective pressures. For example, infants have an immature immune system compared to adults [38] and a vastly simpler diet composed primarily of milk in the early months of life before transitioning to solid foods. Moreover, infants harbor simpler ecological communities with low levels of richness [6], which could alter the overall rate of evolution in the community [39,40]. In particular, rates of horizontal gene transfer, a common feature of the adult gut microbiome [30,33,41,42], could be reduced in infants given the simpler community since the broader ecosystem may serve as a reservoir for pre-adapted material. Given these differences, it is possible that infant-specific evolutionary changes are not only present but also necessary for gut microbiota to successfully colonize their new environment.

Here, we track the evolutionary dynamics of the gut microbiome in >700 infants and their mothers over the span of 1 year after birth. We find that rates of evolution and strain replacement are significantly elevated in infants compared to adults, and that these rates decay with age. Immediately after birth, gene losses dominate, though gene gains become increasingly prevalent over the first year of life. Additionally, we find evidence of parallelism of evolution across infants, highlighting candidate genes that may play important roles in the developing infant gut microbiome. In sum, we find that the ecological and evolutionary dynamics in the developing

infant gut microbiome are distinct and more volatile than that of adults, suggesting that both need to be studied in order to fully understand the formation of the gut microbiome in infants.

RESULTS

Microbiome diversity rapidly grows in the first year of life

To quantify the evolutionary dynamics of gut microbiota in infants, we analyzed fecal metagenomes from four cohorts [6,14,15,22], totaling 2399 samples from 762 healthy infants and 337 mothers. Additionally, we analyzed 249 healthy adults from the Human Microbiome Project (HMP) [43,44], and 185 healthy adults from Qin et al. (2012) (**Table 1**) to compare the evolutionary dynamics in adults versus infants. Infants were longitudinally sampled at 2-7 timepoints ranging from birth (meconium) to 1 year post-delivery across cohorts (**Figure S1**), with dense samples within the first week and month of life, as well as every month thereafter.

Dataset	# Hosts	Host Type	# Samples	Max # Timepoint(s) per host
HMP	249	U.S. adults	469	2-3
Qin et al. 2012	185	Chinese adults	185	1
Backhed et al. 2015	98	Swedish infants and mothers	391	4
Yassour et al. 2018	44	Finnish infants and mothers	213	6
Ferretti et al. 2018	25	Italian infants and mothers	119	6
Shao et al. 2019	600	UK infants and mothers	1676	2-7

Table 1. Datasets analyzed in this study. For the infant gut microbiome datasets, “host” refers to all samples from mothers and infants in the same dyad. For the Yassour et al. (2018) dataset, only mothers at time of delivery were included in our analyses.

We used a reference-based mapping approach [21] to call single nucleotide variants (SNVs) and gene copy number variants (CNVs) for sufficiently abundant and prevalent species in our dataset (**Methods**). Summaries of diversity in infant microbiomes at the species and sub-species levels have been reported previously [6,13–15,21,22]. Here, we revisit diversity patterns at the species and nucleotide level to investigate how gut microbiome community complexity changes in the infant over the first year of life. We then leverage these observations to infer the lineage structure within metagenomic samples to be able to make evolutionary inferences.

To understand how diversity changes over life stage, we first assessed Shannon alpha

diversity of gut microbiota from birth to adulthood (**Figure 1A**). Immediately after birth, median alpha diversity in infant meconium is ~ 1.5 and within a day increases to significantly higher values of ~ 2.7 (p value= 1.5×10^{-5} , T-test). However, by day three, median alpha diversity decreases to ~ 1.4 (p value= 4×10^{-6} , T-test). This initial increase followed by a rapid decrease in community diversity has been previously reported [14,46] and is likely the result of an influx of microbes into the infant from maternal and other environmental sources followed immediately by ecological filtering. Thereafter, median diversity steadily increases, and by month 12 it approaches levels observed in adults (HMP) but not surpassing that of mothers at time of delivery. Mothers have significantly higher median alpha diversity (~ 3.3) than HMP adults (~ 2.65) (p value= 10^{-61} , T-test) (**Figures 1A and S2A**), though richness is significantly lower (**Figure S2B**). This indicates that pregnancy significantly changes microbiome diversity in mothers.

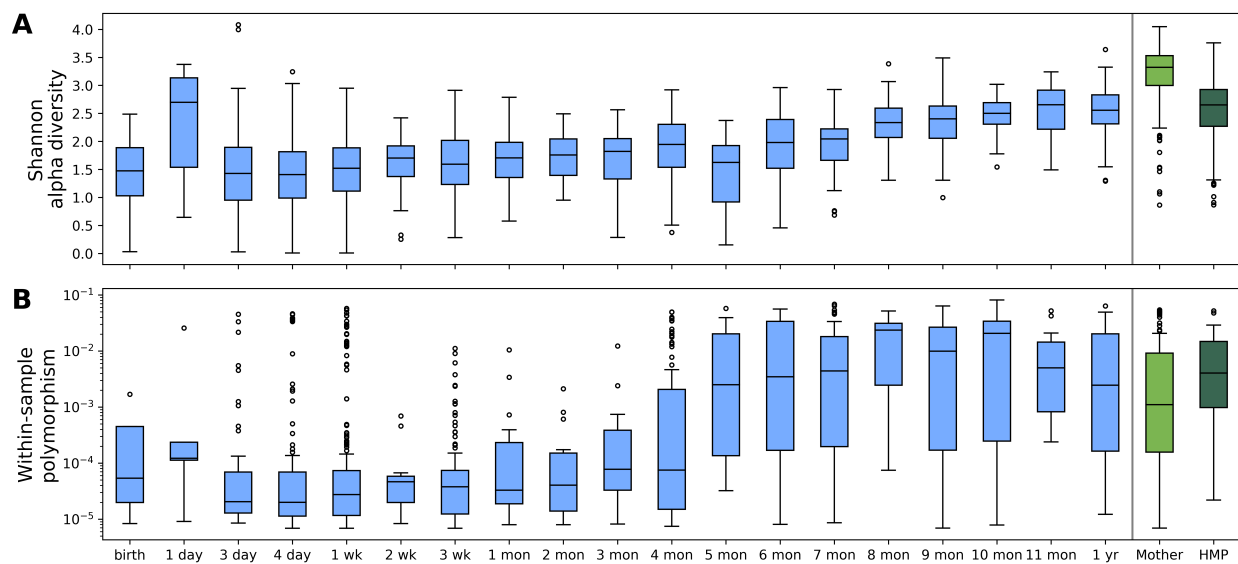


Figure 1. Community and polymorphism diversity increase with life stage. (A) Shannon alpha diversities across life stages over the first year postpartum in infants, as well as in mothers and non-pregnant adults (HMP). Infant samples (blue) are grouped by age; mother samples are from the time of delivery. Sample sizes range from 11-581 for each life stage with a median of 57. In **Figure S2**, alpha diversity in mothers is compared with that of adults in HMP and Qin *et al.* **(B)** Within-sample nucleotide polymorphism levels for *E. coli*. The number of hosts harboring *E. coli* for each age category ranges from 5-231 for each life stage, with a median sample size of 28. In **Figure S3** an additional 12 species within-sample polymorphism levels are plotted.

We next asked whether within-species diversity, or polymorphism rate (measured by the fraction of synonymous sites in core genes with an intermediate frequency polymorphism between 0.2 and 0.8, **Methods**), also increases with life stage. Some species, including *Escherichia coli*, *Bifidobacterium longum* and *Bifidobacterium breve*, experience increases in nucleotide diversity over time (**Figures 1B** and **S3**). Specifically, coincident with the average time of the transition to solid foods at ~5-6 months of age [47], median polymorphism rates increase sharply by an order of magnitude from $<10^{-3}/\text{bp}$ to $\sim 10^{-3}-10^{-2}/\text{bp}$; by contrast, Shannon alpha diversity in **Figure 1A** increases gradually over the first year. This sudden increase in polymorphism rates is likely driven by increasing numbers of strains colonizing the host and not evolution of the resident strains. A mutation rate of $10^{-9}/\text{bp}$ [48] and a generation time of ~1-2 generations/day [49] is too low to produce such rapid increases in levels of polymorphism, whereas ‘oligo-colonization’ of multiple strains of the same species is a far more likely candidate [30].

Rates of polymorphism does not uniformly increase with age for all species. For example, *Bacteroides fragilis* and *Bifidobacterium bifidum* have consistently low median within-host polymorphism of $10^{-4}/\text{bp}$ over most time points in the first year of life. Other species do not display consistent trends and instead show wide variation in levels of polymorphism over time, likely a reflection of the stochasticity of the colonization process (**Figure S3**). Thus, the ecological forces determining colonization success likely vary from species to species.

Since hosts are often oligo-colonized by multiple genetically distinct strains of the same species [30,50], fluctuations in strain frequencies can confound the detection of evolutionary changes from shotgun metagenomic data because both can generate SNV and gene differences over time [30]. To confidently distinguish SNV and gene changes (e.g. horizontal gene transfers) due to evolution from fluctuations in strain frequencies, we leveraged a quasi-phasing (QP) approach that we previously developed [30] to assign genotypes to individual lineages for each species (**Methods**). This approach relies on the fact that samples typically have sufficiently simple lineage structures with a single dominant strain such that alleles can be confidently assigned to that dominant lineage. With this approach, the probability of incorrectly inferring the allelic state of a lineage at any given site is bounded. In practice, quasi phasing excludes samples with high proportions of intermediate frequency polymorphisms because intermediate frequency polymorphisms have a high probability of being misassigned to the correct lineage.

Figure S4 shows the distribution of QP samples (host x species quasi-phaseable lineages) across sufficiently prevalent gut bacterial species in our dataset. The distribution of QP samples varies by life stage: among infants, there are 7,063 QP samples; among mothers, there are 1,159 QP samples; and among HMP adults there are 3,544 QP samples from 217 candidate species. Infants harbor a higher proportion of QP samples (0.87 aggregated over all timepoints and species) compared to adults (0.78) (**Figures S4 and S5**), reflecting simpler communities in the infants. To quantify evolutionary changes, we first identified QP sample pairs for consecutive time points from the same infant or adult host as well as mother-infant comparisons from the same dyad. This yielded a total of 2,184 infant-infant, 241 mother-infant (where infant was within the first week of life), and 1,296 adult-adult HMP QP pairs (combinations of host, species, and timepoint pair) across 176 of the 217 most prevalent species. To then track evolutionary changes, we then identified SNVs that change in allele frequency from ≤ 0.2 to ≥ 0.8 in the core and accessory parts of the genome between pairs of QP time points (**Methods**).

Elevated rates of evolutionary change and strain turnover in the infant compared to adult gut microbiome

We quantified SNV differences between sampled time points in four categories of life stages: HMP adults sampled ~6 months apart, infants sampled over <1 month intervals (which fall within the first 3 months after birth), infants sampled over ~3-6 month intervals, and pairs of mother and infant samples within the first week of life. In this last category, we treated mothers as one time point and infants as another for a given dyad.

SNV differences between timepoints in a QP pair can arise from a combination of two processes: strain replacement and evolutionary modification [30]. Reflecting this, SNV differences are distributed bimodally across all four life stage categories (**Figure 2A**). Most within-host QP sample pairs (83%) experience zero SNV changes over timescales of a week or less, but a small percentage (9%) undergo a small number of SNV changes (≤ 20). An even smaller percentage of hosts harbor $\sim 10^4$ SNV differences (7%), which is on the same order of magnitude of the number of SNV differences between unrelated hosts. This between-host comparison serves as a helpful reference for typical nucleotide divergence between resident and invading strains. Thousands of SNV changes accumulating within hosts on 6 month time scales are unlikely to have arisen from evolutionary diversification of a lineage within a host and are

instead consistent with strain replacements [30]. Thus, we classified the samples experiencing ≤ 20 SNV changes as undergoing evolutionary modifications and samples experiencing >500 SNV changes as undergoing strain replacements.

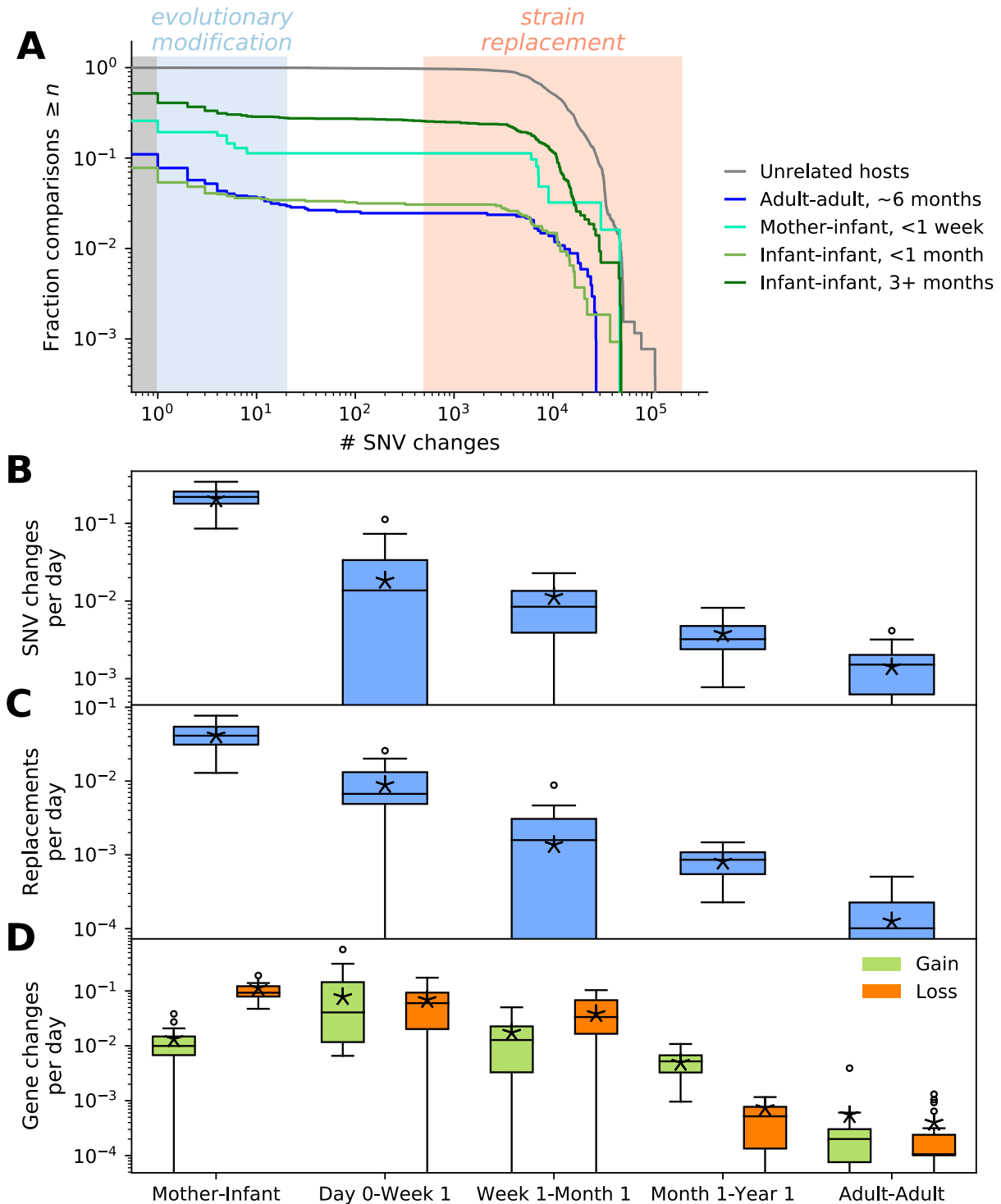


Figure 2. Rates of evolutionary change and replacement decay with life stage. (A) Survival distributions of number of SNV differences (defined as changes in allele frequency from <0.2 to >0.8 between two time points) for each life stage. (B) Rates of number of SNV changes per day for QP pairs experiencing evolutionary modification, (C) rates of number of strain replacements per day for all QP pairs, and (D) rates of gene gains and loss per day for QP pairs experiencing evolutionary modification. Asterisks indicate the mean rate computed over all samples in a life stage. To assess variance, rate estimates for each life stage were bootstrapped 20 times by subsampling with replacement 50 out of all QP pairs available. The mother-infant category only includes comparisons between mother at delivery and her baby sampled within one week of life. To address potential confounding of amount of time between sampling versus life stage, we compared rates for infants and HMP adults sampled 4-8 months apart (**Figure S8**). In **B-D**, all sample sizes are ≥ 54 for each life stage. In **Figure S6**, a scaling relationship between changes and life stage is assessed.

Figure 2 shows that infants have elevated proportions of modification and replacements compared to adults. Among adults, only 8.1% of resident populations in HMP adults sampled over ~ 6 month time scales undergo modification and an even smaller percentage (2.5%) undergo replacement [30]. By contrast, infants sampled ~ 3 -6 months apart have substantially elevated proportions of modification (24.2%) and replacement (25.6%). Mother-infant dyads sampled within the first week also have elevated proportions of modification (14.5%) and replacement (11.3%).

The increased proportions of modifications and replacement events in infants relative to adults suggest that ecological and evolutionary processes occur more frequently and rapidly early in life compared to in adulthood. To confirm this, we next quantified the per-day rates of modification and replacement (**Figure 2B-D**) over various life stages, aggregated over species (**Methods**).

Both modification and replacement rates decay rapidly with life stage and in fact follow a scaling relationship with the age of the host (**Figures 2** and **S6**). During the mother to infant transition, modification rates are ~ 0.2 SNV changes/day. Within infants during the first week of life, the average rate drops 10-fold to ~ 0.02 SNV changes /day, and by adulthood, the rate drops another ~ 10 -fold to ~ 0.001 SNV changes/day, consistent with rates previously estimated in adults [31,51]. Similarly, replacement rates also rapidly decay with the age of the host. During the mother-infant transition the overall strain replacement rate is 4.2×10^{-2} replacements/QP pair/day. Then, within the first week of life, the replacement rate drops to 8.9×10^{-3} replacements/day, and finally in adulthood it drops to 1.3×10^{-4} replacements/day.

We next quantified rates of gene gains and loss per day, representing another mode of

evolutionary change common in adult microbiomes [30–33,41,52–54] (**Methods**). Like the modification and replacement rates, gene change rates also rapidly decay with life stage and follow a scaling relationship (**Figures 2D and S6B**). Interestingly, the rate of gene loss decays more quickly than the rate of gene gain with life stage after the mother-infant transition; furthermore, during the mother-infant transition, gene losses exceed gains by almost 100-fold (0.19 gene losses/day versus 0.013 gene gains/day (p value= $2*10^{-13}$, T-test on bootstrapped rates). This trend reverses later in life: between 1 month and 1 year after birth, gains exceed losses by >10-fold ($4.9*10^{-3}$ gains/day versus $7.4*10^{-4}$ losses/day, p value= $5.9*10^{-11}$, T-test on bootstrapped rates). Eventually, by adulthood, losses and gains occur at similar rates of $\sim 5*10^{-4}$ changes/day. This suggests that during the initial host colonization process at birth, diversity at the gene level is dramatically reduced through gene loss, but later gene-level diversity recovers gradually as the infant gut matures before eventually reaching an equilibrium in adulthood.

Since earlier life stages are generally sampled at shorter consecutive time intervals in our dataset, we assessed whether duration of time between sampling points, rather than life stage, could explain the faster rates of modification, strain replacement, and gene changes in infants. We found that infants sampled over 4-8 month timescales still had significantly elevated rates compared to adults sampled on similar time scales (p values <0.05 , see **Figure S7**), confirming that age, rather than duration of the sampling interval, is a major driver of faster evolution in the infant gut microbiome.

Evolutionary changes in the infant are seeded by *de novo* mutations and pre-evolved lineages from the broader family

To further probe the evolutionary origins of the SNV changes in infants, we asked whether they arise in infants from *de novo* mutation or pre-existing standing variation that may have been seeded by recombination and also could have potentially arisen in the broader family unit before being transmitted to the infant. To assess evidence for the two scenarios, we combined the 428 SNV changes occurring in the 154 modification events in the infant-infant cohort and then assessed the prevalence of sweeping (or “derived”) alleles (**Figure 3**). Here, “prevalence” is defined as proportion of HMP adults that harbor the sweeping allele. We compare the prevalence of sweeping alleles with that of a null distribution of randomly selected sites assuming *de novo* mutation (**Methods**).

The observed distribution of prevalences shows interesting departures from the null expectation. Specifically, under the null, the majority of sweeping alleles are expected to be rare in the broader population. By contrast, the observed distribution of prevalences is bimodal, with infant sweeping alleles being completely absent from adults or present in virtually all HMP adults (**Figure 3A**). This is distinct from the distribution observed in adults in which sweeping SNVs are also enriched for intermediate prevalences (0.1-0.5). The bimodal distribution in infants suggests that evolutionary changes occurring within infants are either *de novo* mutations that are rare in the broader population or reversions to the consensus state in adults, possibly reflecting infant-specific adaptations in the former case or reversions of maladaptations in the latter case.

Rare infant sweeping alleles have a dN/dS value of 1.3 (CI 0.95 to 2.37, **Figure S9, Methods**), indicating that they are indeed potentially adaptive. However, they could also be recent deleterious mutations that reached high frequency by chance during a bottleneck associated with transmission to the infant. In several instances, we observe that private nonsynonymous mutations that arise early in an infant's life revert to an allelic state that is present in all other HMP adults later in life (**Figure 4A**), suggesting that some of these initial mutations may indeed be deleterious. Together, the preponderance of rare alleles with an overall dN/dS >1 as well as frequent reversions of these rare alleles to prevalent allelic states indicate that *de novo* mutations are common in infants.

However, the rest of the infant sweeping alleles, most of which are present in the majority of HMP adults (**Figure 3A**), have a dN/dS value of ~0.6. This lower dN/dS likely reflects that purifying selection has had sufficient time to purge deleterious variants from the population, indicating that *de novo* mutation is not the only mode of evolution in the infant. Instead, the high prevalences and lower dN/dS value together suggest that in many instances, sweeps from standing variation seeded by recombination events are common in the infant as well, as has been observed in adults [30]. Consistent with recombination playing an important role, the gene gains observed in infants cannot have arisen via *de novo* mutation and instead must have arisen via recombination events. In fact, the increase in rate of gene gains over the first year of life indicates that recombination plays a significant role in recovering diversity that is lost during the mother to infant transition.

It is possible that several of these evolutionary events observed in infants occurred in the

mother or broader family unit before the strain was transmitted to the infant. In two mother-infant dyads, multiple alleles sweep to high frequency in the infant after birth before later reverting to allelic states found in the mother (**Figure 4B**). It is unlikely that these are *de novo* mutations that occurred twice in succession since >4 SNVs change twice and include synonymous sites. Interestingly, multiple haplotype configurations are observed over time in these two dyads, suggesting that multiple variants of the same lineage circulate among the members of the family unit. This suggests that infants may be seeded multiple times by their family members and likely experience significant flux of variants of the same strain over the course of their first year.

We conclude that multiple processes contribute to SNV changes within infants. During infancy, new *de novo* mutations arise that are generally absent from the broader population. Additionally, more prevalent allelic changes in the infant are likely seeded by recombination events. Finally, there is evidence for ongoing transmission between mother and infants well after birth, which may also seed evolutionary changes in the infant.

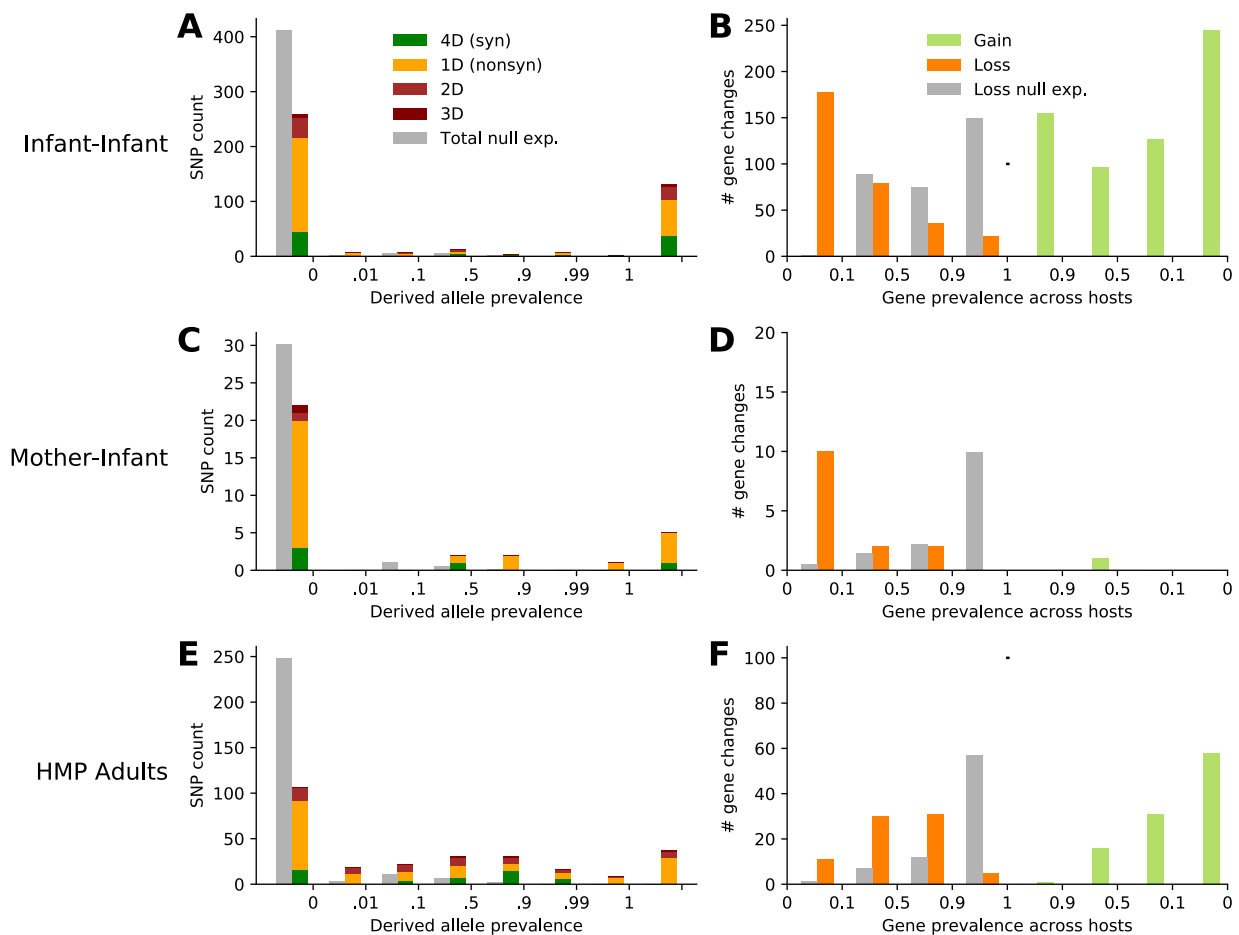
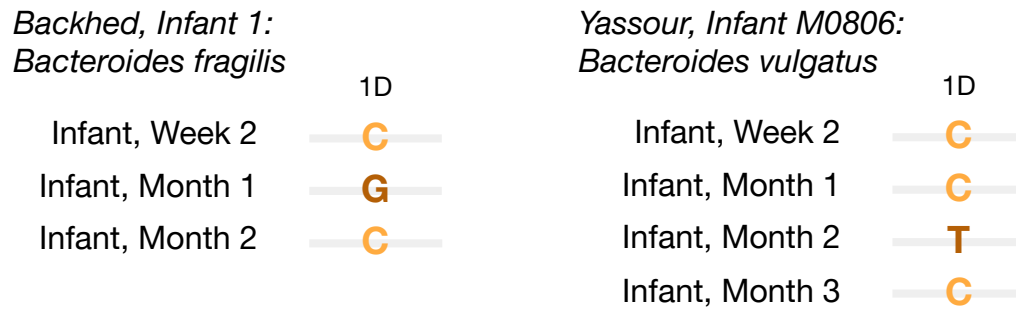
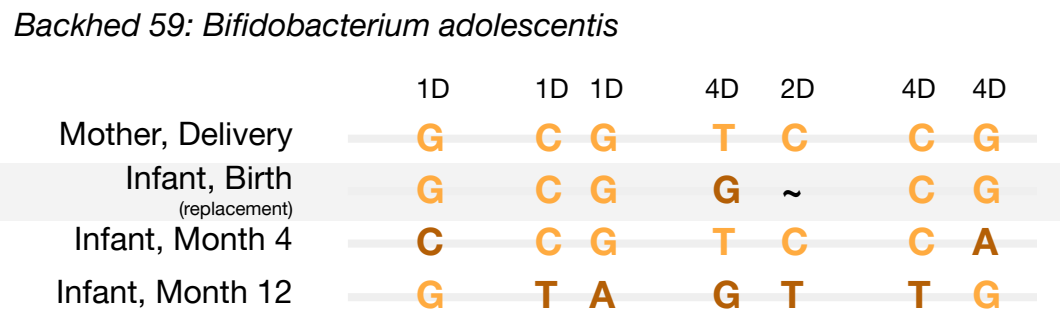


Figure 3. Prevalences of gene and SNV changes. Distribution of prevalences of SNV (A, C, and E) and gene changes (B, D, and F) occurring in infant, mother-infant and HMP adult cohorts. Here prevalence is defined as the fraction of HMP adults that harbor a SNV or gene. Additional cohorts are used to compute prevalence in **Figures S8** and **S9**. The null distributions in grey show the expectation for random *de novo* events (**Methods**). By definition, gene gains cannot occur *de novo*. In A, C and E, the prevalence bin <0 indicates the no HMP adult harbors the sweeping allele other than the person in which the allele arises, and the prevalence bin >1 indicates that the allele is present in all HMP adults.

A Within-infant reversions at single sites



B Reversions from mother-infant transition



Shao C02143: Bifidobacterium bifidum

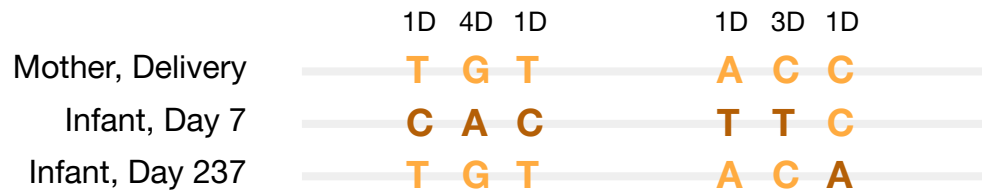


Figure 4: Reversion events in infant gut microbiomes. There were 371 host x species instances with at least 3 QP samples across available infant timepoints and mother at delivery. Of these, 13 experienced modifications at 2 nonoverlapping timepoint pairs. Seven of these 13 experienced reversions; here we illustrate the haplotypes corresponding to these reversions. Sites are ordered by gene's position in the reference genome. **(A)** In 5 of these cases, there is a reversion at a single nonsynonymous (1D or 2D) site to an allelic state that is prevalent in adults; two examples are shown. **(B)** In 2 instances, multiple nucleotides change at both synonymous and nonsynonymous sites. In the host 'Backhed 59', a replacement occurs at birth with respect to mother, and then by month 4 it reverts back to the strain harbored by mother.

Parallelism of SNV changes across hosts

We next scanned for putative loci experiencing positive selection in the infant gut microbiome. To do so, we leveraged a signature of positive selection known as parallel evolution in which multiple independent mutations in the same gene sweep to high frequency in distinct hosts. Signatures of parallel evolution have been successfully used to detect positive selection in human-associated microbes [31,55,56], as well as in laboratory experiments [57–59].

To identify candidates for parallel evolution, we counted the number of mutations that each gene class (defined here as a unique PATRIC gene description, **Methods**) has accumulated across hosts and compared this number with that of a null in which the mutations are randomized across the metagenomes across hosts (**Figure 5A**) (**Methods**). Under the null, fewer than 1 mutation is expected in a given gene class, reflecting the massive size of the metagenome relative to the number of evolutionary changes observed (**Table S1**). By contrast, several gene classes have >1 mutations in the observed data, sometimes occurring in a single host and sometimes in multiple hosts (**Table S1**). To bias our candidate list towards gene classes that are mutated recurrently in multiple hosts, we imposed a threshold requirement that gene classes must undergo an evolutionary modification in at least 4 distinct hosts. With these criteria, we found a total of 14 distinct gene classes (excluding hypothetical proteins) experiencing parallel evolution (**Figure 5B**). Some of these gene classes additionally acquired multiple mutations within hosts; for example, a TonB-dependent transporter part of the SusC family acquired a total of 32 SNV changes across 26 unique hosts (**Figure 5B**).

The mutations in these 14 gene classes have a combined dN/dS of ~1.0. Although this value of dN/dS is consistent with a model without positive or negative selection, it is significantly higher than the dN/dS for other gene classes that are not as recurrently mutated

(~0.6) (p value= 2.3×10^{-35} , T-test), and is even greater than the dN/dS for SNVs that differ in replacement events (~0.1) (**Figure 5C**). This suggests that a higher proportion of mutations in these recurrently evolving gene classes may be potentially adaptive.

Two of the top ranking gene classes have been previously discovered to undergo parallel evolution, most prominently genes in the starch utilization system (SusC and SusD) family [31]. Of notable interest is the previously mentioned TonB-dependent transporter part of the SusC family, which is implicated in processing of complex glycans [60] and degradation of breast milk-associated human milk oligosaccharides [1], is associated with the colonization of infant gut microbiomes [22], and is posited to undergo adaptation in infant microbiomes [1,61]. Another notable candidate gene class undergoing parallel evolution in our dataset includes beta-galactosidase, which was previously shown to be enriched in infants [1,61–64] and is involved in the digestion of breast milk [65].

Other gene classes undergoing parallel evolution include transcriptional regulators, membrane transporters, and histidine kinases. In particular, the periplasmic ligand-binding sensor domain/histidine kinase is, intriguingly, mutated 12 times in infants but never in adults. While, to our knowledge, these gene classes have not previously been implicated in microbiome adaptation or infant microbiomes, the recurrent evolutionary changes across multiple hosts provide interesting opportunities for follow up analyses.

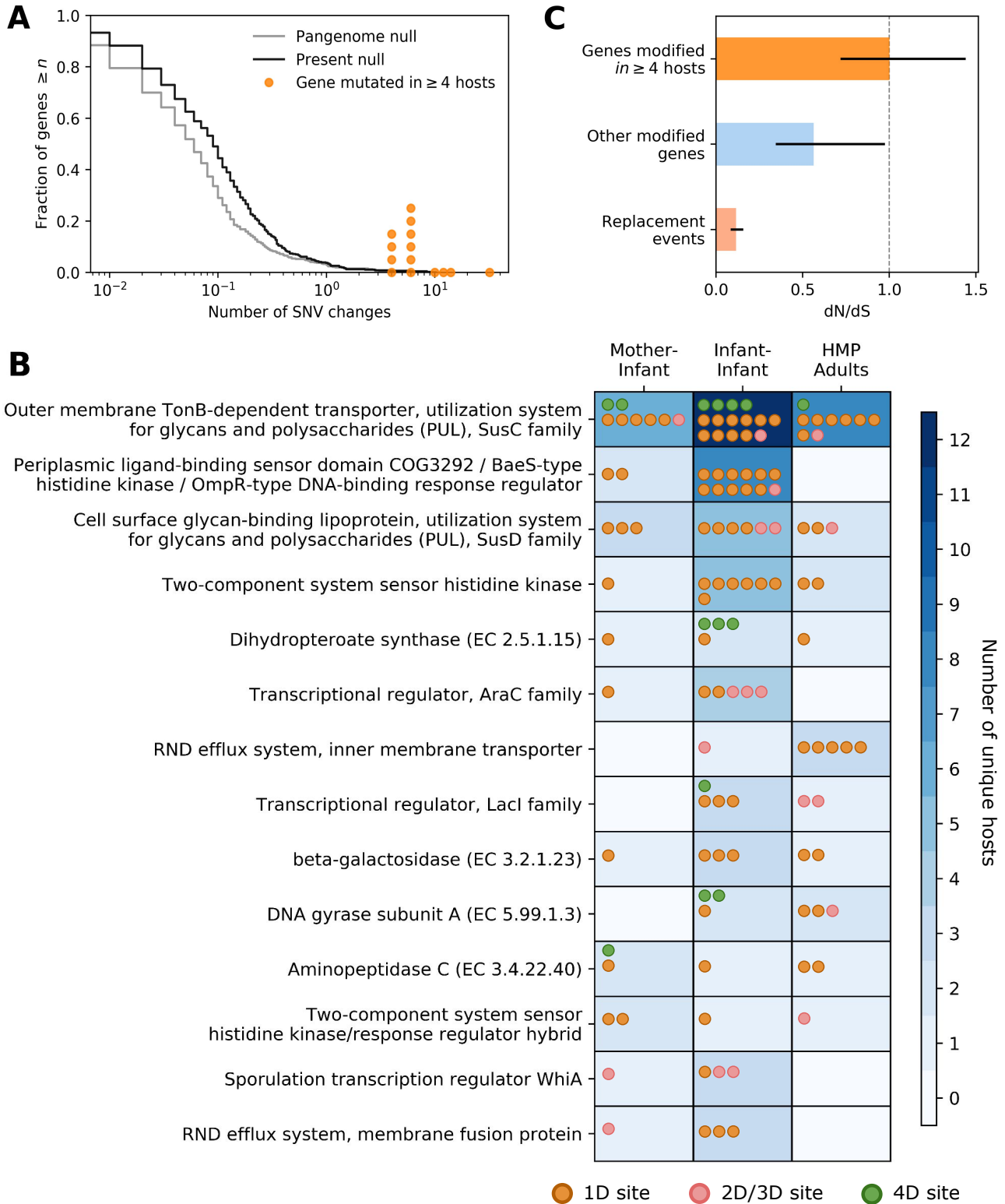


Figure 5: Parallelism of SNV changes across hosts. (A) Mutations are randomized across the metagenomes across hosts to construct two null distributions to assess parallelism (**Methods**).

Observed SNV change counts are shown with orange dots for the 14 gene classes found to mutate in parallel in ≥ 4 hosts. In the cases where there are multiple orange dots, there are multiple gene classes experiencing the same number of SNV changes. **(B)** SNV changes in these 14 gene classes found to mutate in parallel in ≥ 4 hosts are partitioned by life stage and colored by degeneracy (e.g. 1D indicates a nonsynonymous site and 4D indicates a synonymous site). Some functional classes are mutated multiple times within hosts too. For inclusivity, the mother-infant age class here includes all QP sample pairs in which the earliest infant timepoint is taken, irrespective of whether the infant was sampled in the first week of life. **(C)** dN/dS of SNV changes in the 14 gene classes found to mutate in parallel in ≥ 4 hosts is compared with dN/dS of SNV changes in all other gene classes and dN/dS of sites that differ in strain replacements. 95% confidence intervals for bootstrapped dN/dS values are reported as black bars.

DISCUSSION

To date, there has been substantial focus on the ecological and functional development of the infant gut microbiome [4–10,12,14,22], but there is relatively little known about the evolutionary dynamics of microbiota in the infant gut microbiome. Here we examined the evolutionary dynamics of gut microbiota in a cohort of >700 infants [6,14,15,22] and compared them with that of adults. We found that the initial days after birth are marked by almost 100-fold increased rates of evolutionary modification and replacement of resident strains relative to typical rates observed in adults. Over time, these rates settle, but not without first rapid recovery of diversity initially lost during the mother to infant transmission period via elevated rates of gene gains in the first year of life. Many of these evolutionary events show signatures of *de novo* mutation that may potentially be adaptive, but ongoing transmission from the broader family also plays a significant role in seeding evolutionary changes in the infant gut microbiome over time.

Perhaps it is not surprising that infancy is marked by particularly volatile rates of evolutionary change and strain turnover that eventually stabilize with time, given that ecological succession at the species level is also rapid and dynamic in the early days of life. However, with a mechanistic understanding of how individual lineages change at a nucleotide level over time, we may be able to better understand the ecological succession in the infant gut microbiome. For example, some species such as *E. coli* undergo two orders of magnitude increase in nucleotide polymorphism levels over just a span of 1-2 months (**Figure 1**), while other species such as *B. vulgatus* and *E. faecalis* exhibit idiosyncratic patterns at different life stages (**Figure S3**). Interestingly, community-wide level statistics like Shannon alpha diversity miss these sharp transitions in sub-species diversity because Shannon alpha diversity is computed from an

aggregate of abundances across species. These species-specific signatures of ecological and evolutionary change at the nucleotide level may be important for uncovering functional shifts in the microbiome and understanding the roles each species plays in maturation of the human gut microbiome.

We note that several studies (e.g. Niu et al. 2020; Donovan 2020) use the term ‘evolution’ to describe ecological changes in species composition over time. In our paper, we define evolution as a genetic change on the background of a resident lineage on top of which additional changes can accumulate, and we distinguish such evolutionary changes from strain replacement [67]. By distinguishing evolution from strain replacement, we can start to understand at a molecular level the genetic variants needed for microbes to survive in the gut. Additionally, adaptation and co-evolution of infant gut bacteria have also been described in the literature in reference to enrichment of genes that play a role in milk digestion in genus like *Bifidobacterium* [1,63,68,69]. Here, we do not examine the longer-term evolutionary forces that result in functional enrichment in certain strains of gut bacteria and instead focus on short term within-host changes. Despite this, among the genes experiencing parallel evolution across multiple infants are those implicated in milk digestion, suggesting that short-term forces within hosts could be contributing to signals of adaptation that accrue over longer periods.

Although our analysis distinguishes evolution from strain replacement, some evolutionary modifications detected in infants reflect migrations of lineages that may have evolved recently in the mother or another family member. Previous studies have found extensive evidence for transmission of strains from mother and infant, as well as among other family members over time [13–15,20,22,23,68,70–72]. In our study, with the QP approach, we were able to recover evidence of transmission of lineages diverged by only a few SNVs, reflecting evolutionary modifications that occurred in the recent past in the broader family unit. If families indeed circulate strains among each other over several decades, it may then be worth considering the broader family’s microbiome as a larger, inter-connected microbiome. Additional data from multiple co-habiting individuals is needed to fully understand the extent of evolution and transmission of closely related lineages among family members [13,23]. Moreover, data from multiple body sites over multiple time points [14,20] can reveal additional sources of strains colonizing the infant.

It is possible that the genetic changes observed during the mother to infant transmission are confounded by yet another process: bottlenecks. During a bottleneck event, deleterious alleles can rise in frequency due to drift. However, recent work suggests that fluid flow in the adult colon cannot create bottlenecks that can result in large fluctuations in allele frequencies [73] and that, instead, natural selection is a more plausible mechanism of genetic changes occurring over short time scales. Still, the mother-infant transition represents a unique colonization process in which a sterile or nearly sterile environment is colonized by microbes for the first time. Thus, the strength of bottleneck may be stronger in the infant than in established adult guts. In support of bottlenecks potentially playing a role in the evolutionary dynamics of the infant, we do observe, in five host x species instances, private single nucleotide mutations at non synonymous sites reaching high frequency within infants and then reverting back to the prevalent allelic state observed in adults. This suggests either that the initial mutation was temporarily adaptive or rose to high frequency due to drift and then reverted due to deleterious selective effects. There are, however, hallmarks of positive selection driving evolutionary change within infants. dN/dS of rare SNV changes in infants is >1 , and many genes are mutated in parallel across multiple hosts, which also harbor elevated levels of dN/dS ~ 1 compared to other sweeping SNVs. Nonetheless, future work resolving the role of bottlenecks in the infant colonization process will be needed to fully understand the evolutionary processes taking place in infant microbiomes.

Another potential area for future work is to understand the changes in the gut microbiome in pregnancy. In this paper, we computed species diversity in delivering mothers in four cohorts [6,14,15,22] and compared levels with that of healthy non-pregnant adults in the HMP [43,74] and [45] (**Figure S2**). We found that alpha diversity is significantly higher among mothers, but that richness decreases. The increase in alpha diversity is finding is consistent with previous findings by [75]. However, [76] find that Shannon diversity decreases and [77] find that Faith's phylogenetic diversity decreases in pregnancy. Given the importance in being able to predict microbiome changes associated with pre-term birth [78] and the resulting impacts on the developing infant gut microbiome, future studies examining microbiome ecological and evolutionary biomarkers of pregnancy and delivery are needed.

The finding that many species in the infant gut microbiome rapidly evolve is important for understanding how the microbiome is assembled early in life. For example, it will be

important to understand how these frequent evolutionary changes in the infant gut impact the persistence of lineages, ecological interactions, and the overall development of the gut microbiome. Moreover, by incorporating evolution into our understanding of the development of the infant gut microbiome, we may be able to better understand the functional impact of the microbiome on human health.

METHODS

Data

The raw sequencing reads for the metagenomic samples used in this study were downloaded from Bäckhed et al. (2015) (accession number PRJEB6456); Ferretti et al. (2018) (accession number PRJNA352475); Yassour et al. (2018) (accession number PRJNA475246); Shao et al. (2019) (accession number PRJEB32631); Human Microbiome Project Consortium 2012 and Lloyd-Price et al. (2017) (URL: <https://aws.amazon.com/datasets/human-microbiome-project/>); and Qin et al. (2012) (accession number PRJNA422434).

Estimation of nucleotide and gene content of metagenomic samples

We used MIDAS (Metagenomic Intra-Species Diversity Analysis System) [21] to estimate within-species nucleotide and gene content of each metagenomic shotgun sequencing sample. MIDAS utilizes a reference database of 31,007 bacterial genomes, clustered into 5,952 species which covers roughly 50% of species found in “urban” human stool metagenomes.

We followed the parameters and filters described in Garud et al. (2019) to estimate species abundances, SNVs, and gene copy numbers and to filter samples, genes, and specific positions with insufficient or anomalous coverage.

For each species, we computed “core genes”, defined as genes in the MIDAS reference database that are present in at least 90% of samples within a given cohort (infants or adults). We separated the computation of core genes by cohort because several species that are prevalent in infants are absent in adults, and vice versa. In our downstream analyses, we considered the union of core genes with respect to adults (mother + HMP) and infants.

Quasi-phasing and inference of rates of SNV changes, gene changes, and replacements over time

We followed the approach in Garud et al. (2019) to identify “quasi-phaseable” (QP) samples. We then compared pairs of QP samples collected from consecutive time points with available data from the same host to identify SNV and gene changes between QP pairs. A consecutive time point consists of two time points in the mother-infant dyad that are both QP for a given species such that there is no intervening timepoint that is also QP. Note that “mother-infant” timepoint pairs were restricted to comparisons between mother timepoints at delivery and infant timepoints within the first week in order to best approximate gut microbiome dynamics of the mother-infant transition at birth and exclude changes that occur later on within the infant. The only exception to this “mother-infant” definition is in Figure 5B, in which the mother-infant age class includes all QP sample pairs comprising mothers and the earliest infant timepoint, irrespective of whether the infant was sampled in the first week of life. This enables all data to be considered for the parallelism analysis.

SNV changes were computed by identifying SNV allele frequency changes from ≤ 0.2 to ≥ 0.8 . With these strict thresholds, in addition to other thresholds and filters as described in Garud et al. (2019), less than 1 SNV change per genome is expected by chance due to sampling error. Since SNV changes can be generated by either modification or replacement events, we classified QP pairs as undergoing a modification or no change if they had ≤ 20 SNV differences, or a replacement if they had ≥ 500 SNV differences.

We also computed gene gain and losses between QP pairs. These were computed by identifying genes with copy number $c \leq 0.05$ (indicating gene absence) in one sample and $0.6 \leq c \leq 1.2$ in another (indicating single copy gene presence). All additional thresholds and filters as described in Garud et al. (2019) were applied.

Per-day rates of SNV changes, replacements, and gene changes

To quantify per-day modification rates, we divided the number of SNV changes and gene changes, respectively, observed among non-replacement QP pairs by the number of days elapsed between sampling time points. To quantify per-day replacement rates, we divided the number of replacement events by the number of available QP pairs and the number of days elapsed between sampling timepoints. In the case of mother-meconium infant time point pairs, a duration of 1 day

was assigned. To assess variance, we performed bootstrapping of rates by repeatedly taking random subsamples of size 50 (with replacement) of all QP pairs in a category.

To quantify the scaling relationship between rate of change versus life stage, we performed a linear regression on logged mean SNV change, replacement and gene gain/loss rates computed per timepoint pair category as a function of days since birth. To compute days since birth, the median timepoint was computed for each life stage (mother-infant within the first week, day 0-week 1, week 1-month 1, month 1-year 1). In the case of HMP adults, a value of 40 years was assigned, reflecting the age range of HMP participants [43,74].

Prevalence of sweeping alleles

Derived allele prevalences were computed with respect to three prevalence cohorts: HMP adults, infants, and mothers (**Figures 3, S8, S9**). Replicating the analysis in Garud et al. (2019), we define population prevalence of an allele as the fraction QP samples with the allele. If a host has QP samples for multiple time points, that host's contribution to total prevalence is the fraction of timepoints possessing the allele. Private SNVs were assigned a prevalence of 0. We computed a null distribution for SNV change prevalences by randomly drawing the number of SNV modifications observed in data from all SNV opportunities in the genome, bootstrapped 10 times.

Prevalence of gene gains and losses

Gene prevalences were computed with respect to three prevalence cohorts of HMP adults, infants and mothers. As in Garud et al. (2019), we define population prevalence of a non-shared gene as the proportion of all samples harboring the gene with copy number ≤ 3 and ≥ 0.3 ; that is, we exclude samples with unusually high gene copy number from consideration and use a threshold of 0.3 to determine gene presence. We computed a null distribution for prevalences of gene losses by randomly drawing the number of genes lost from all genes present in the pangenome, bootstrapped 10 times. The null expectation for gene gains was zero across prevalence bins as there are no *de novo* gene gains by definition.

dN/dS computation for changing SNVs

To investigate whether or not SNV changes in modification events are adaptive, we

estimated dN/dS. dN/dS was computed as the ratio between number of observed nonsynonymous (1D) SNV changes divided by nonsynonymous opportunities and number of observed synonymous (4D) SNV changes divided by nonsynonymous opportunities per QP pair under consideration. We also bootstrapped dN/dS estimates by sampling a binomial distribution 10,000 times with number of trials n equal to total number of nonsynonymous or synonymous SNV changes and success probability p equal to the proportion that are either nonsynonymous or synonymous; this resulted in bootstrapped nonsynonymous and synonymous SNV change counts which were divided by the same nonsynonymous and synonymous SNV opportunities. We reported 95% confidence interval for dN/dS estimates by the 2.5% and 97.5% quantiles of the bootstrapped dN/dS values.

Parallelism of evolutionary SNV changes across hosts

To assess parallelism of evolutionary changes across hosts, we enumerated the number of observed SNV changes per PATRIC gene description. A PATRIC gene description is a unique string that describes a gene. This same unique string can potentially be present across multiple species (**Table S1**), though like in Garud et al. (2019), genes shared across species boundaries (i.e. possess $\geq 95\%$ nucleotide identity with a gene in a different species in the MIDAS database) have been filtered from our dataset.

We evaluated whether the observed number of SNV changes per gene class is greater than expected under a null distribution in which the number of observed SNV changes are randomly shuffled. Specifically, we computed two null distributions to assess whether SNV changes in a specific gene occur more often than expected. The two nulls were constructed as follows: the same number of observed changes are randomly distributed across (1) all genes that are present in either timepoint in a given QP pair and (2) all genes that are present in the MIDAS pangenome of the species. Genes are considered “present” if they have copy number 0.5-2, and a candidate gene is counted twice if present in both samples of the QP pair. We bootstrapped null distribution estimates 100 times.

ACKNOWLEDGEMENTS

We sincerely thank William R. Shoemaker, Richard Wolff, Alison Feder, Sandeep Venkataram, and Leah Briscoe for their critical comments on the manuscript and Benjamin Good and Katherine Pollard for early discussions on this project. We thank members of the Garud lab for their feedback during the development of this paper. NRG is funded by the Paul Allen Institute, a UCLA Hellman fellowship, a UCLA Faculty Career Development award, and a Scialog MND research grant. DWC received funding support from NIH R25 MH 109172.

REFERENCES

- 1 Sela, D.A. *et al.* (2008) The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18964–18969
- 2 Smith, P.M. *et al.* (2013) The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science* (80-.). 341,
- 3 Nicholson, J.K. *et al.* (2012) Host-gut microbiota metabolic interactions. *Science* (80-.). 336, 1262–1267
- 4 Stewart, C.J. *et al.* (2018) Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562,
- 5 Niu, J. *et al.* (2020) Evolution of the Gut Microbiome in Early Childhood: A Cross-Sectional Study of Chinese Children. *Front. Microbiol.* 11,
- 6 Bäckhed, F. *et al.* (2015) Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* DOI: 10.1016/j.chom.2015.04.004
- 7 Sprockett, D.D. *et al.* (2020) Microbiota assembly, structure, and dynamics among Tsimane horticulturalists of the Bolivian Amazon. *Nat. Commun.* 11,
- 8 Chu, D.M. *et al.* (2017) Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat. Med.* 23,
- 9 Palmer, C. *et al.* (2007) Development of the human infant intestinal microbiota. *PLoS Biol.* 5,
- 10 Vallès, Y. *et al.* (2014) Microbial Succession in the Gut: Directional Trends of Taxonomic and Functional Change in a Birth Cohort of Spanish Infants. *PLoS Genet.* 10,
- 11 Yassour, M. *et al.* (2016) Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci. Transl. Med.* 8,
- 12 Koenig, J.E. *et al.* (2011) Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci. U. S. A.* 108,
- 13 Korpela, K. *et al.* (2018) Selective maternal seeding and environment shape the human gut microbiome. *Genome Res.* 28, 561–568
- 14 Ferretti, P. *et al.* (2018) Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe* 24, 133-145.e5
- 15 Shao, Y. *et al.* (2019) Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* 574, 117–121
- 16 Karlsson, F.H. *et al.* (2013) Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 498, 99–103
- 17 Bokulich, N.A. *et al.* (2016) Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci. Transl. Med.* 8,
- 18 Dominguez-Bello, M.G. *et al.* (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. U. S. A.* 107,
- 19 Fehr, K. *et al.* (2020) Breastmilk Feeding Practices Are Associated with the Co-Occurrence of Bacteria in Mothers' Milk and the Infant Gut: the CHILD Cohort Study. *Cell Host Microbe* 28,
- 20 Mitchell, C. *et al.* (2020) Delivery mode impacts newborn gut colonization efficiency. *bioRxiv* DOI: 10.1101/2020.01.29.919993
- 21 Nayfach, S. *et al.* (2016) An integrated metagenomics pipeline for strain profiling reveals

- novel patterns of bacterial transmission and biogeography. *Genome Res.* 26, 1612–1625
- 22 Yassour, M. *et al.* (2018) Strain-Level Analysis of Mother-to-Child Bacterial
Transmission during the First Few Months of Life. *Cell Host Microbe* 24, 146-154.e4
- 23 Hildebrand, F. *et al.* (2021) Dispersal strategies shape persistence and evolution of human
gut bacteria. *Cell Host Microbe* 29,
- 24 Raveh-Sadka, T. *et al.* (2016) Evidence for persistent and shared bacterial strains against a
background of largely unique gut colonization in hospitalized premature infants. *ISME J.*
10,
- 25 Brooks, B. *et al.* (2017) Strain-resolved analysis of hospital rooms and infants reveals
overlap between the human and room microbiome. *Nat. Commun.* 8,
- 26 Kenny, D.J. *et al.* (2020) Cholesterol Metabolism by Uncultured Human Gut Bacteria
Influences Host Cholesterol Level. *Cell Host Microbe* 28, 245-257.e6
- 27 Hehemann, J.H. *et al.* (2010) Transfer of carbohydrate-active enzymes from marine
bacteria to Japanese gut microbiota. *Nature* 464, 908–912
- 28 Spanogiannopoulos, P. *et al.* The microbial pharmacists within us: A metagenomic view
of xenobiotic metabolism. *Nature Reviews Microbiology*, 14. (2016) , 273–287
- 29 Gumpert, H. *et al.* (2017) Transfer and persistence of a multi-drug resistance plasmid in
situ of the infant gut microbiota in the absence of antibiotic treatment. *Front. Microbiol.* 8,
- 30 Garud, N.R. *et al.* (2019) Evolutionary dynamics of bacteria in the gut microbiome within
and across hosts. *PLoS Biol.* 17, e3000102
- 31 Zhao, S. *et al.* (2019) Adaptive Evolution within Gut Microbiomes of Healthy People.
Cell Host Microbe 25, 656-667.e8
- 32 Yaffe, E. and Relman, D.A. (2020) Tracking microbial evolution in the human gut using
Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nat.*
Microbiol. 5,
- 33 Groussin, M. *et al.* (2021) Elevated rates of horizontal gene transfer in the industrialized
human microbiome. *Cell* 184,
- 34 Poyet, M. *et al.* (2019) A library of human gut bacterial isolates paired with longitudinal
multiomics data enables mechanistic microbiome research. *Nat. Med.* 25, 1442–1452
- 35 Roodgar, M. *et al.* (2021) Longitudinal linked-read sequencing reveals ecological and
evolutionary responses of a human gut microbiome during antibiotic treatment. *Genome*
Res. 31,
- 36 Jiang, X. *et al.* (2019) Invertible promoters mediate bacterial phase variation, antibiotic
resistance, and host adaptation in the gut. *Science* (80-.). 363,
- 37 Ghalayini, M. *et al.* (2018) Evolution of a dominant natural isolate of *Escherichia coli* in
the human gut over the course of a year suggests a neutral evolution with reduced
effective population size. *Appl. Environ. Microbiol.* 84,
- 38 Sjögren, Y.M. *et al.* (2009) Influence of early gut microbiota on the maturation of
childhood mucosal and systemic immune responses: Gut microbiota and immune
responses. *Clin. Exp. Allergy* 39, 1842–1851
- 39 Venkataram, S. *et al.* (2021) Early adaptation in a microbial community is dominated by
mutualism-enhancing mutations. *bioRxiv*
- 40 Post, D.M. and Palkovacs, E.P. (2009) Eco-evolutionary feedbacks in community and
ecosystem ecology: Interactions between the ecological theatre and the evolutionary play.
Philos. Trans. R. Soc. B Biol. Sci. 364,
- 41 Lin, M. and Kussell, E. (2019) Inferring bacterial recombination rates from large-scale

- sequencing datasets. *Nat. Methods* 16, 199–204
- 42 Sakoparnig, T. *et al.* (2019) Whole genome phylogenies reflect long-tailed distributions of recombination rates in many bacterial species. *bioRxiv* DOI: 10.1101/601914
- 43 Lloyd-Price, J. *et al.* (2017) Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 550, 61–66
- 44 Human Microbiome Project Consortium, T. (2012) A framework for human microbiome research The Human Microbiome Project Consortium*. *Nature* 486,
- 45 Qin, J. *et al.* (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60
- 46 Wampach, L. *et al.* (2017) Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life. *Front. Microbiol.* 8, 738
- 47 Van Dijk, M. *et al.* (2012) The dynamics of feeding during the introduction to solid food. *Infant Behav. Dev.* 35,
- 48 Barrick, J.E. and Lenski, R.E. Genome dynamics during experimental evolution. , *Nature Reviews Genetics*, 14. (2013) , 827–839
- 49 Sender, R. *et al.* (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* 14,
- 50 Truong, D.T. *et al.* (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903
- 51 Didelot, X. *et al.* (2012) Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol.* 13,
- 52 Smillie, C.S. *et al.* (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244
- 53 Zlitni, S. *et al.* (2020) Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. *Genome Med.* 12,
- 54 Coyne, M.J. *et al.* (2014) Evidence of extensive DNA transfer between bacteroidales species within the human gut. *MBio* 5, e01305-14
- 55 Lieberman, T.D. *et al.* (2011) Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat. Genet.* 43, 1275–1280
- 56 Feder, A.F. *et al.* The clarifying role of time series data in the population genetics of HIV. , *PLoS Genetics*, 17. (2021)
- 57 Wichman, H.A. *et al.* (1999) Different trajectories of parallel evolution during viral adaptation. *Science (80-.)*. 285,
- 58 Woods, R. *et al.* (2006) Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 103,
- 59 Barroso-Batista, J. *et al.* (2014) The First Steps of Adaptation of *Escherichia coli* to the Gut Are Dominated by Soft Sweeps. *PLoS Genet.* 10, e1004182
- 60 Martens, E.C. *et al.* Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. , *Journal of Biological Chemistry*, 284. (2009)
- 61 Kujawska, M. *et al.* (2020) Succession of *Bifidobacterium longum* Strains in Response to a Changing Early Life Nutritional Environment Reveals Dietary Substrate Adaptations. *iScience* 23,
- 62 Ambrogi, V. *et al.* (2019) Characterization of GH2 and GH42 β -galactosidases derived from bifidobacterial infant isolates. *AMB Express* 9,
- 63 Duranti, S. *et al.* (2019) *Bifidobacterium bifidum* and the infant gut microbiota: an

- intriguing case of microbe-host co-evolution. *Environ. Microbiol.* 21,
- 64 Lawson, M.A.E. *et al.* (2020) Breast milk-derived human milk oligosaccharides promote
Bifidobacterium interactions within a single ecosystem. *ISME J.* 14, 635–648
- 65 Kitaoka, M. (2012) Bifidobacterial enzymes involved in the metabolism of human milk
oligosaccharides. *Adv. Nutr.* 3,
- 66 Donovan, S.M. Evolution of the gut microbiome in infancy within an ecological context. ,
Current Opinion in Clinical Nutrition and Metabolic Care, 23. (2020)
- 67 Garud, N.R. and Pollard, K.S. (2020) Population Genetics in the Human Microbiome.
Trends Genet. 36, 53–67
- 68 Milani, C. *et al.* (2015) Exploring vertical transmission of bifidobacteria from mother to
child. *Appl. Environ. Microbiol.* DOI: 10.1128/AEM.02037-15
- 69 Duranti, S. *et al.* (2017) Maternal inheritance of bifidobacterial communities and
bifidophages in infants through vertical transmission. *Microbiome* 5,
- 70 Koo, H. *et al.* (2020) An individualized mosaic of maternal microbial strains is transmitted
to the infant gut microbial community. *R. Soc. Open Sci.* 7,
- 71 Asnicar, F. *et al.* (2017) Studying Vertical Microbiome Transmission from Mothers to
Infants by Strain-Level Metagenomic Profiling. *mSystems* 2,
- 72 Siranosian, B.A. *et al.* (2020) Acquisition, transmission and strain diversity of human gut-
colonizing crAss-like phages. *Nat. Commun.* 11,
- 73 Ghosh, O.M. and Good, B.H. (2021) Emergent evolutionary forces in spatial models of
luminal growth in the human gut microbiota. *bioRxiv*
- 74 Methé, B.A. *et al.* (2012) A framework for human microbiome research. *Nature* 486, 215–
221
- 75 Jašarević, E. *et al.* (2017) Stress during pregnancy alters temporal and spatial dynamics of
the maternal and offspring microbiome in a sex-specific manner. *Sci. Rep.* 7,
- 76 Goltsman, D.S.A. *et al.* (2018) Metagenomic analysis with strain-level resolution reveals
fine-scale variation in the human pregnancy microbiome. *Genome Res.* 28,
- 77 Koren, O. *et al.* (2012) Host remodeling of the gut microbiome and metabolic changes
during pregnancy. *Cell* 150,
- 78 Vinturache, A.E. *et al.* Maternal microbiome - A pathway to preterm birth. , *Seminars in
Fetal and Neonatal Medicine*, 21. (2016)

SUPPLEMENT

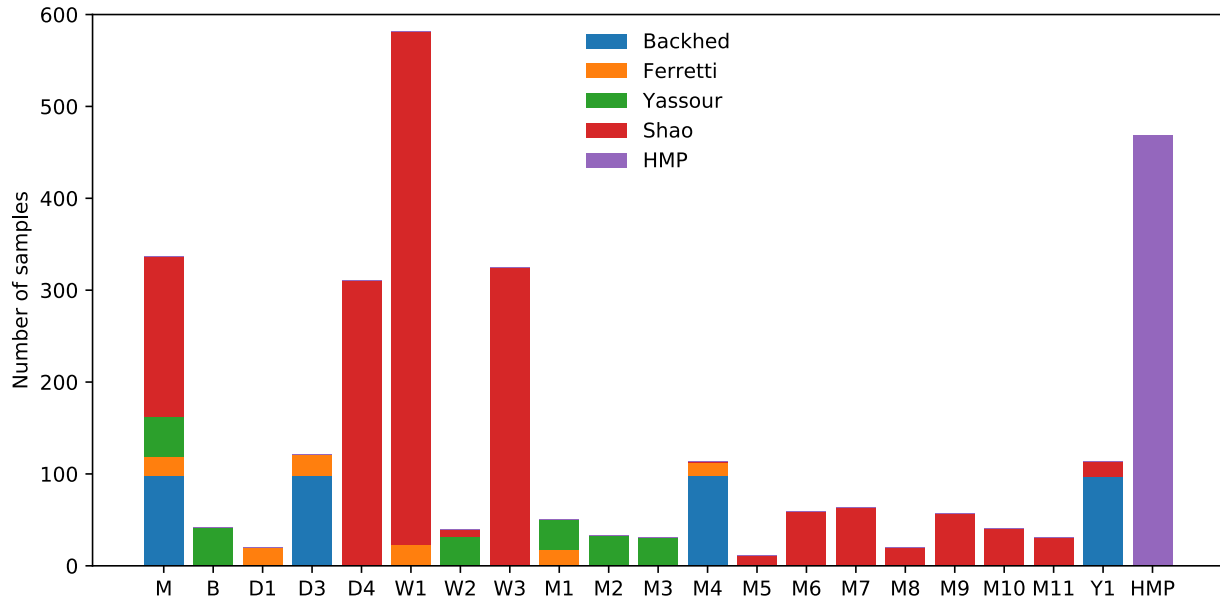


Figure S1: Sampling timepoints and sample sizes for infants and mothers analyzed in this paper. Mother sampling times, which can range from 3 months prior delivery, time of delivery, and 3 months post delivery are grouped into a single category. Some infants from [15] were sampled multiple times between day 7 and 14, and thus were grouped in the Week 1 category.

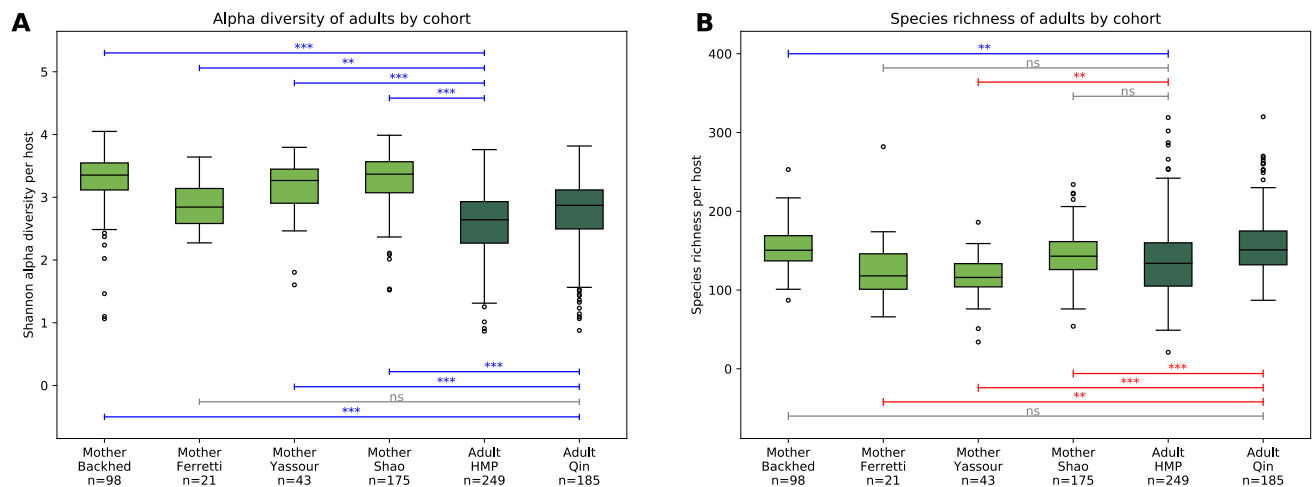


Figure S2: (A) Species alpha diversity and **(B)** richness in mothers' gut microbiomes at time of delivery versus healthy non-pregnant adults. A total of 771 adults were compared across six studies to assess generalizability of results. We tested for significant differences in alpha diversity and richness between mothers at delivery and non-pregnant adults. T-tests for significance are represented as colored intervals; blue corresponds to mothers having higher levels of diversity compared to non-pregnant adults, red corresponds to lower levels of diversity. T-tests with a p value < 0.05 , ≤ 0.01 , and ≤ 0.001 are annotated with one, two and three

asterisks respectively. To assess overall significance, we fit a linear model with pregnancy status as the predictor and study and host as random effects (α diversity \sim pregnancy status + (1|study) + (1|host)) and obtained a coefficient of 0.5 for pregnancy status with a p value of 0.0188.

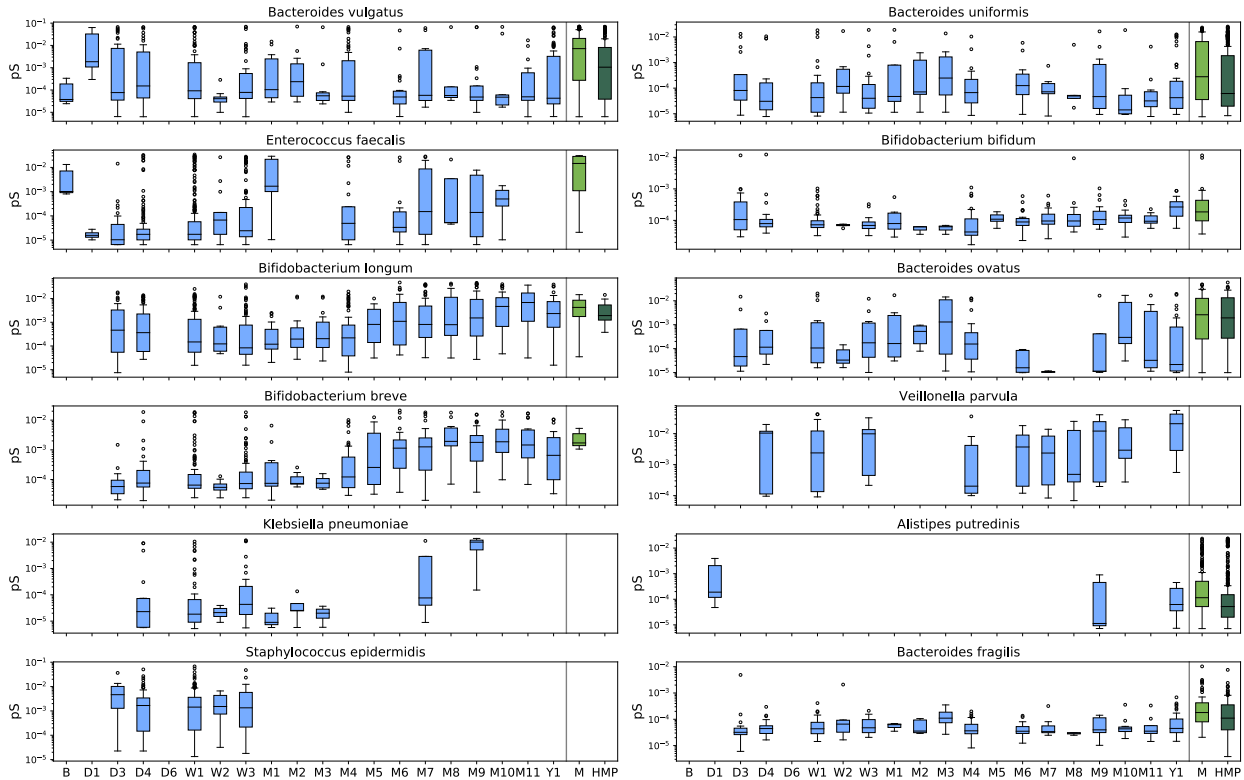


Figure S3: Within-sample polymorphism over different life stages for the top 12 most prevalent species in the gut microbiome across our datasets, excluding *E. coli*, which is shown in **Figure 1B**. The same life stages as in Figure 1B are shown, where “B” indicates birth, “D1” indicates day 1, “W1” indicates week 1, “M1” indicates month 1, “Y1” indicates year 1, and so on. “M” indicates mother samples at delivery and “HMP” includes all HMP1-2 adult samples. Each boxplot is shown only if sample size $n \geq 3$.

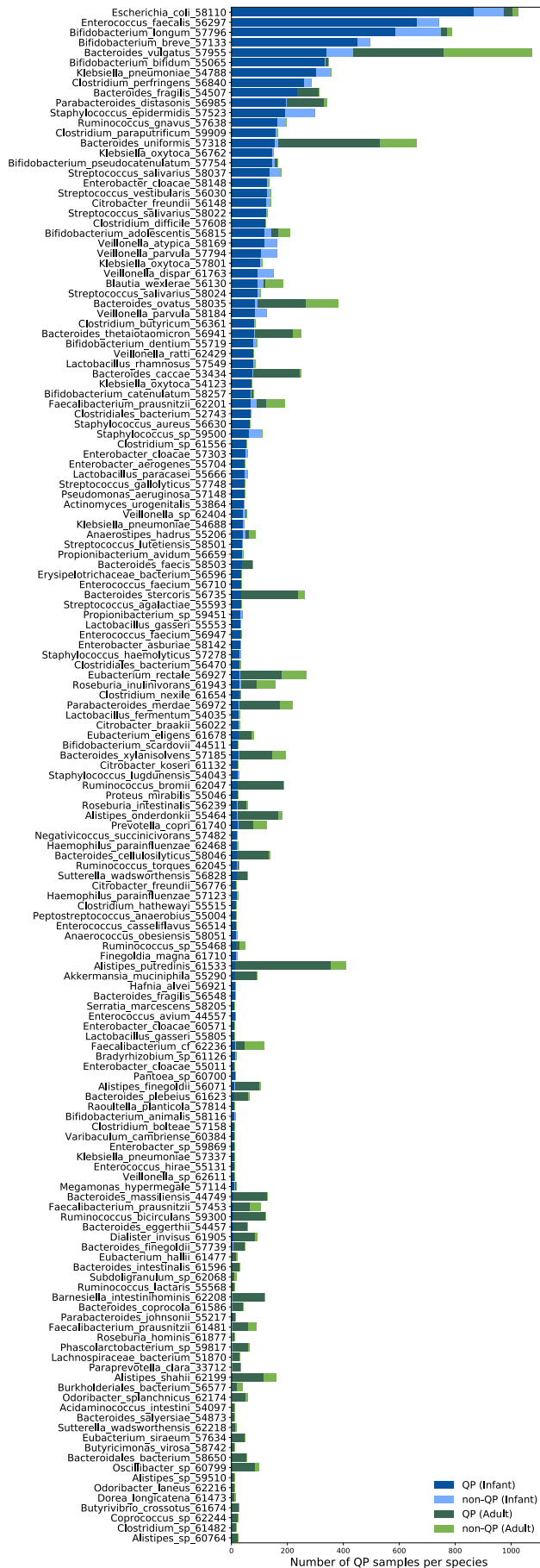


Figure S4: Number of QP vs. non-QP infant and adult (including mother) samples. The 153 species that have greater than 10 QP samples across all hosts are shown; species are ordered by the number of QP samples in infants

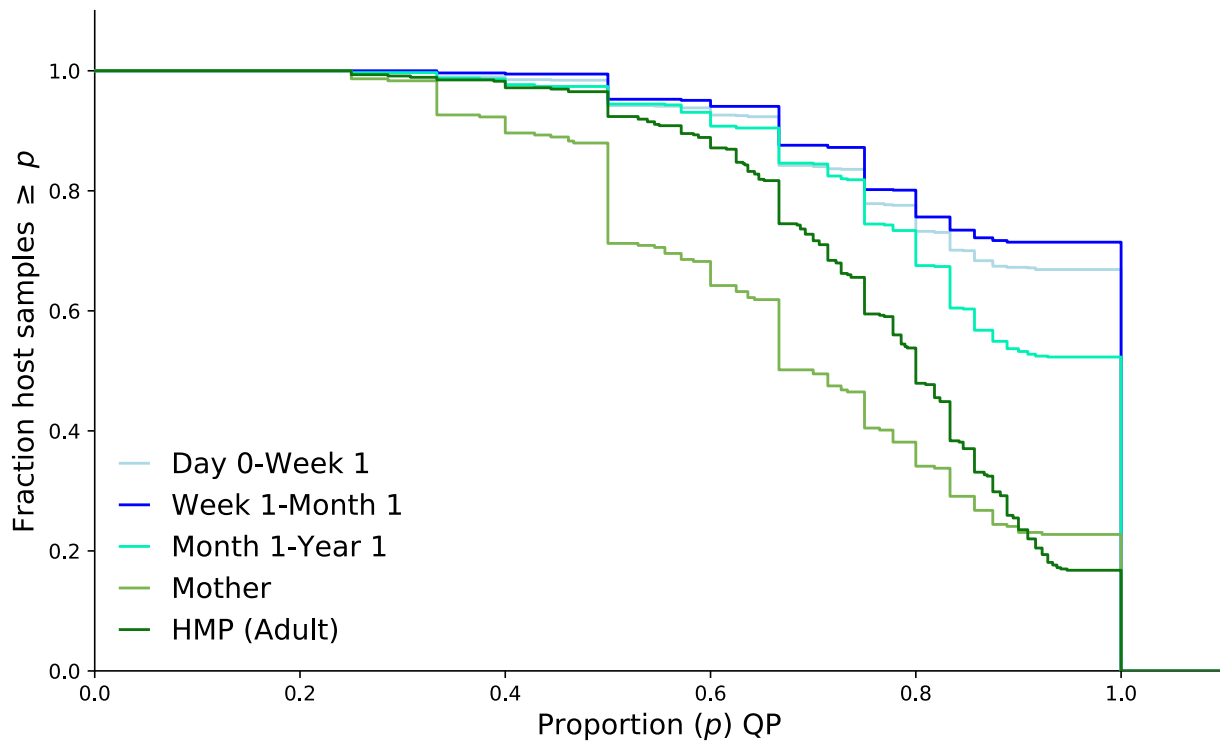


Figure S5: Proportions of high coverage species that are quasi-phaseable (QP) per host sample, categorized by life stage. Besides 'Mother' and 'HMP', category label refers to life stage within infants. Here, 'host sample' is defined as a single metagenomic sample.

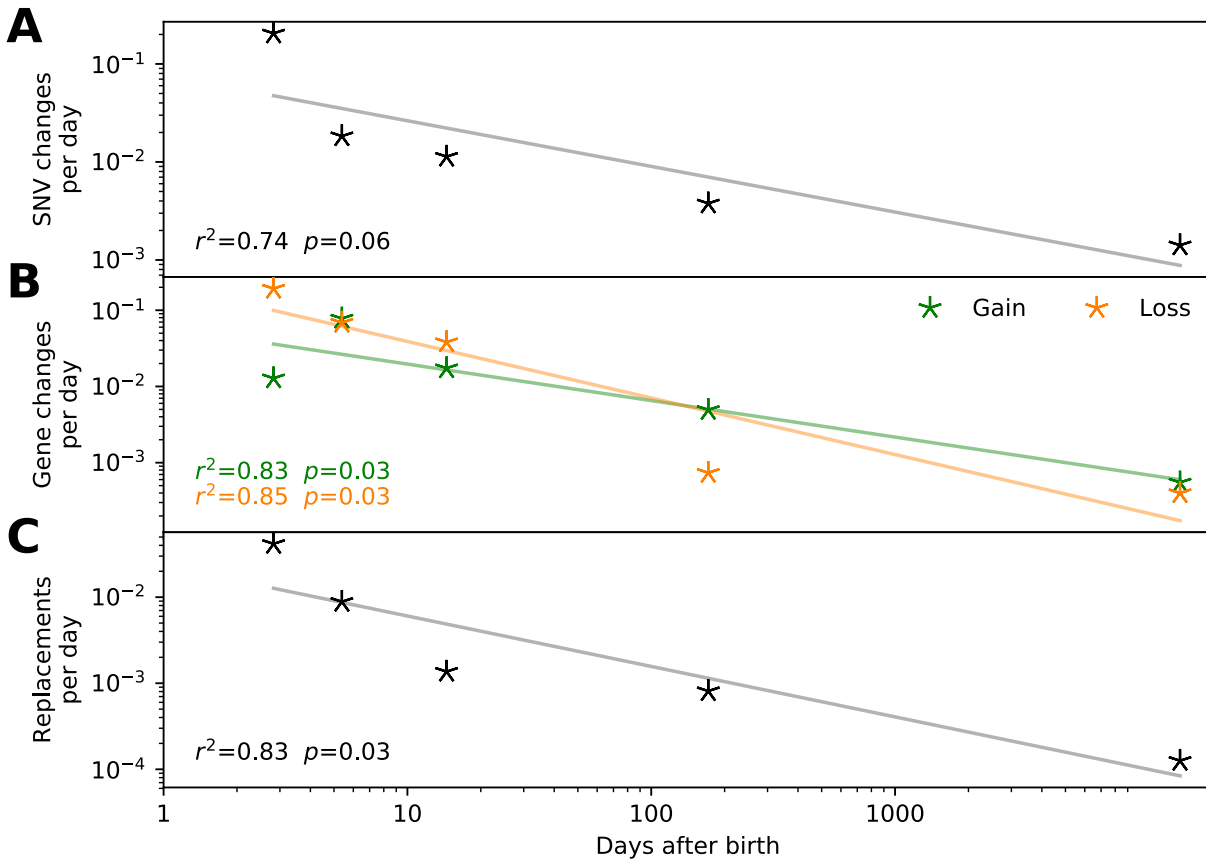


Figure S6: Decay in rates of SNV change, gene gain, gene loss and replacement rates over life stage. The same timepoint pair categories as in **Figure 2** were considered: mother-infant within the first week, infant-infant within the first week, infant-infant week 1-month 1, infant-infant month 1-year 1, and adult-adult. Days after birth were assigned to be the median day in a given life stage interval. Mothers were assigned day 0, infant meconium samples were assigned day 1, and adult timepoints were arbitrarily assigned to be approximately 40 years after birth.

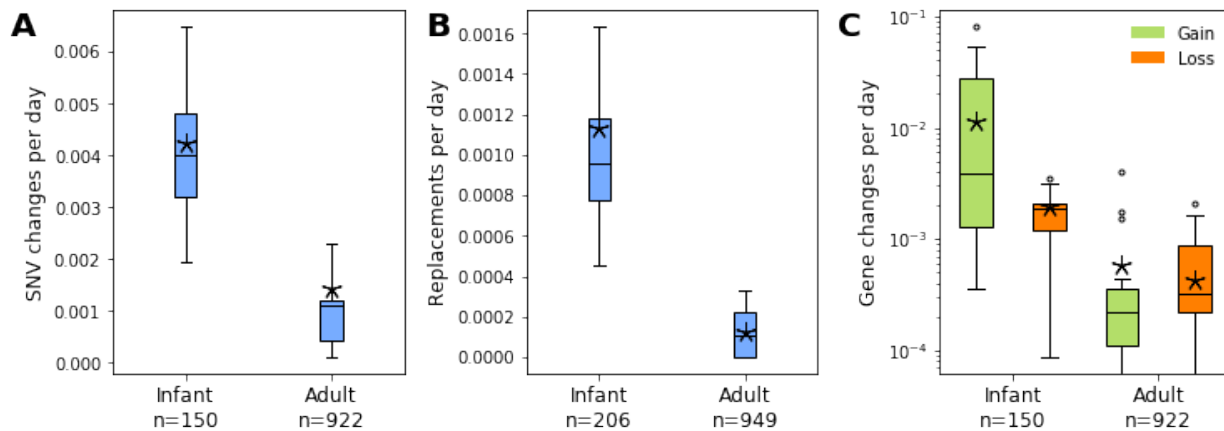


Figure S7: Bootstrapped (A) SNV change, (B) replacement, and (C) gene change rates for all infant-infant vs. HMP adult-adult QP pairs matched for duration between timepoints of 4 to 8 months. T-tests show significant difference between infant and adult distributions for SNV

change rates (p value= 2.1×10^{-13}), replacement rates (p value= 7.7×10^{-15}), gene gain rates (p value= 5.6×10^{-3}), and gene loss rates (p value= 3.8×10^{-5}).

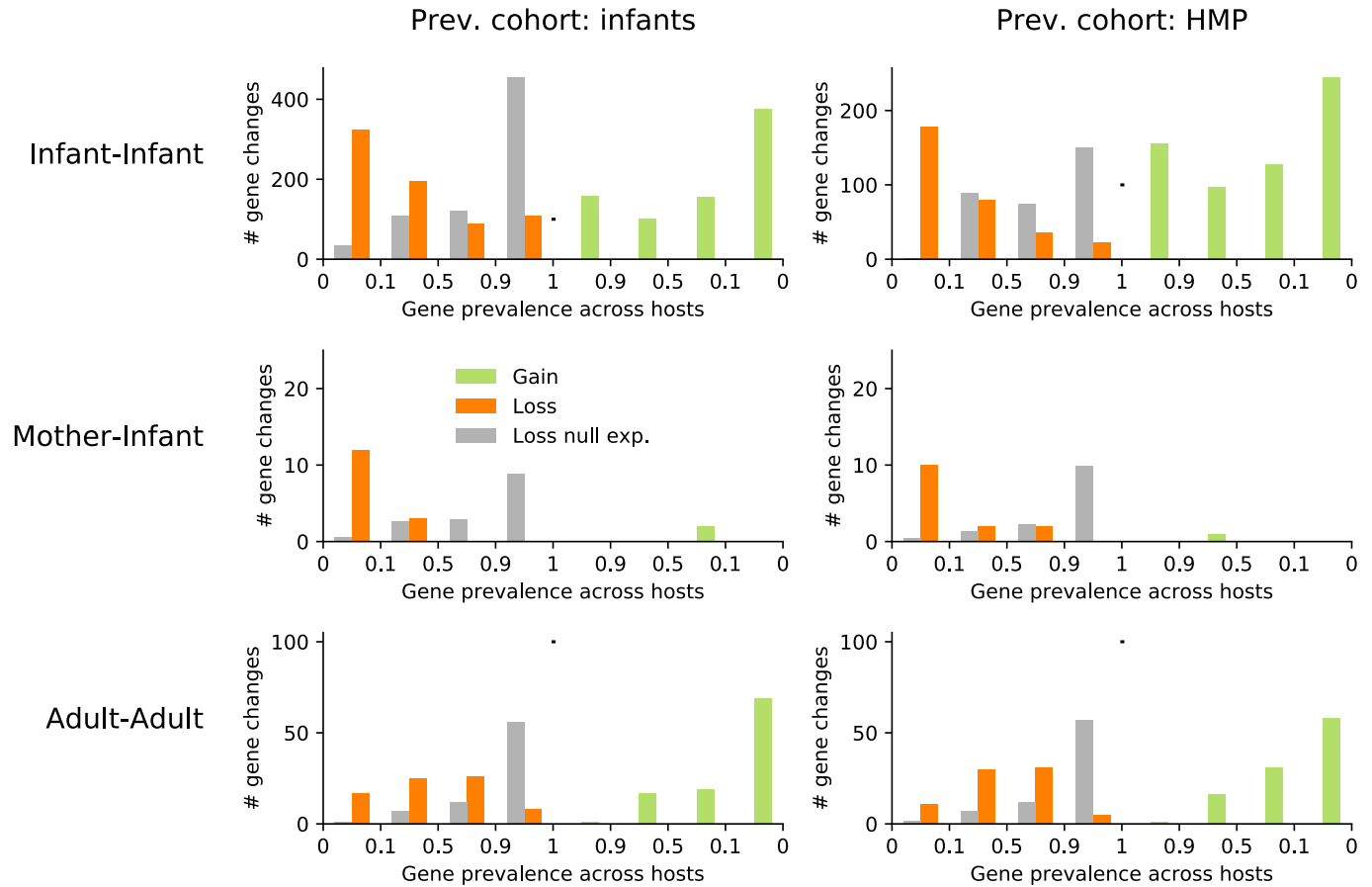


Figure S8: Prevalence of genes that are gained or lost in putative SNV modification events with respect to infant and HMP adult cohorts.

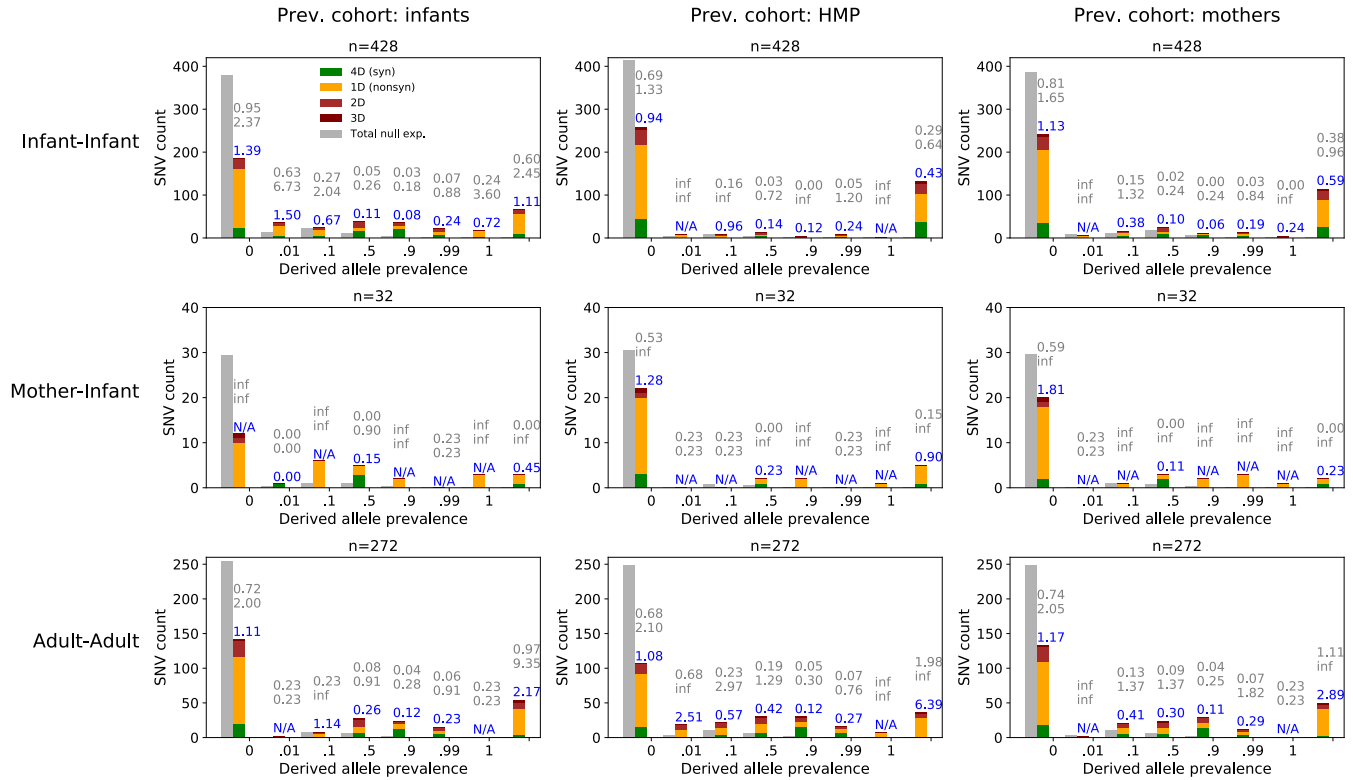


Figure S9: Prevalence of sweeping SNVs in the infant, HMP adult, and mother cohorts. The top left plot shows the prevalence of SNVs sweeping in infants with respect to a prevalence cohort defined by infants. By contrast, the top right plot shows the prevalence of same SNVs sweeping in infants, but with respect to a prevalence cohort defined by mothers. dN/dS of each prevalence bin are reported in blue with 95% confidence intervals reported in gray.

Table S1: Numbers of SNV changes in evolutionary modification events, numbers of unique hosts experiencing such SNV changes, and null expectations grouped by PATRIC gene ID.