

# **1 Community coalescence reveals the dynamics of 2 intra-species selection in complex microbial communities**

**3 Sophie Jean Walton<sup>1,2,3</sup>, Qing Xu<sup>2</sup>, Richa Sharma<sup>3</sup>, Hannah R. Gellert<sup>3</sup>, Chih-Fu Yeh<sup>3</sup>, Jonas  
4 Cremer<sup>3</sup>, Katherine S. Xue<sup>4</sup>, Dmitri A. Petrov<sup>3,5,\*</sup>, Benjamin H. Good<sup>2,3,5,\*</sup>**

**5 <sup>1</sup>Biophysics Program, Stanford University, Stanford, CA 94305, USA**

**6 <sup>2</sup>Department of Applied Physics, Stanford University, Stanford, CA 94305, USA**

**7 <sup>3</sup>Department of Biology, Stanford University, Stanford, CA 94305, USA**

**8 <sup>4</sup> Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine CA 92697,**

**9 USA**

**10 <sup>5</sup>Chan Zuckerberg Biohub – San Francisco, San Francisco, CA 94158, USA**

**11**

**12 \*Correspondence: dpetrov@stanford.edu, bhgood@stanford.edu**

## **13 Abstract**

**14 Microbial communities harbor extensive intra-species diversity, but the evolutionary forces that  
15 shape this variation remain poorly understood empirically. Here, we addressed this question by  
16 coalescing *in vitro* gut communities derived from different human donors, revealing the  
17 emergent fitness differences between multiple pairs of conspecific strains as they compete within  
18 a larger, species-rich community. Most pairs of strains were subject strong and  
19 context-dependent selection, even when their parent communities were pre-assembled in the  
20 same abiotic environment. However, these fitness differences often shifted over time due to  
21 biotic interactions within the community, leading to extended coexistence within species. These  
22 results suggest that conspecific strains can fulfill distinct ecological roles when competing within  
23 a diverse community, even when their genomic diversity exhibits the hallmarks of a single  
24 biological species.**

## 25 Introduction

26 Large microbial communities, from the soil to the gut microbiome, harbor diversity over multiple  
27 spatial and taxonomic scales (1–8). Their vast species diversity is accompanied by extensive  
28 genetic variation within each of their component species, with strains from different local  
29 communities typically varying at thousands of genomic loci (6, 9–19). On longer timescales, these  
30 conspecific strains compete with each other as part of a broader global population (9, 20–24).

31 They also compete on short timescales when new strains of an existing species invade from  
32 outside their local community. Examples of these inter-strain competitions are readily observed  
33 in longitudinal sequencing of microbial communities (7, 25–29), and are conjectured to play an  
34 important role in driving their short-term (9, 21, 26, 30) and long-term (9, 14, 21, 23) evolution.  
35 However, the evolutionary forces that govern such local strain competitions remain poorly  
36 understood empirically. A central challenge lies in understanding how natural selection acts on  
37 strains of the same species that compete within larger, species-rich communities.

38 One prominent view is that selection on genetically diverged strains should be very strong, with  
39 one strain nearly instantaneously dominating over the other. This scenario is motivated by  
40 previous work showing that circulating strains often vary in many fitness-relevant traits,  
41 including resource utilization (15, 19), inter-bacterial antagonism (16), and phage susceptibility  
42 (31, 32). The large number of such differences suggests that their cumulative effect on relative  
43 fitness could be very large in any given local environment. However, ecological theory also  
44 predicts that the assembly of highly diverse communities can select for suites of strains that fulfill  
45 very similar functional roles, leading to weak or nearly-neutral competition between strains  
46 within their local community context (33–39). Signatures of this emergent neutrality have  
47 previously been observed at higher levels of taxonomic organization (40), suggesting that they  
48 might also extend to strain-level competition at finer genetic scales.

49 A third possibility is that strains can coexist with each other by competing for different ecological  
50 niches, with their extensive phenotypic differences allowing them to function more like different  
51 species within their local community. This model is motivated by empirical observations  
52 showing that multiple strains of the same species can sometimes persist within a local  
53 community for extended periods of time (26–29, 41–44), even when their genomes are more  
54 consistent with a single biological species (20, 21, 23, 45). However, ecological theory predicts  
55 that competition with other species can severely limit the realized niche of a given species  
56 (39, 46–50), which could make it more difficult for conspecific strains to successfully engage in  
57 niche partitioning when competing within a larger community. Understanding how niche  
58 partitioning emerges in these species-rich settings is a critical open problem, which has  
59 important implications for engineering microbial therapeutics (51–55) and the dynamics of strain  
60 transmission across hosts (25, 30, 44, 56–58).

61 Distinguishing these scenarios in natural microbial communities can be challenging due to the  
62 confounding effects of spatial structure (59, 60), migration (25, 56, 57), and stochastic  
63 environmental variation (7, 25, 26, 59, 61). *In vitro* systems provide a promising alternative, by  
64 enabling quantitative measurements of microbial competition in controlled laboratory  
65 conditions. In particular, several recent studies have shown that anaerobic passaging of human  
66 fecal samples can yield highly diverse, stable, and reproducible microbial communities that can  
67 be serially propagated in liquid culture for extended periods of time (40, 62–65). We reasoned that  
68 these stool-derived communities could provide a natural platform for investigating strain-level

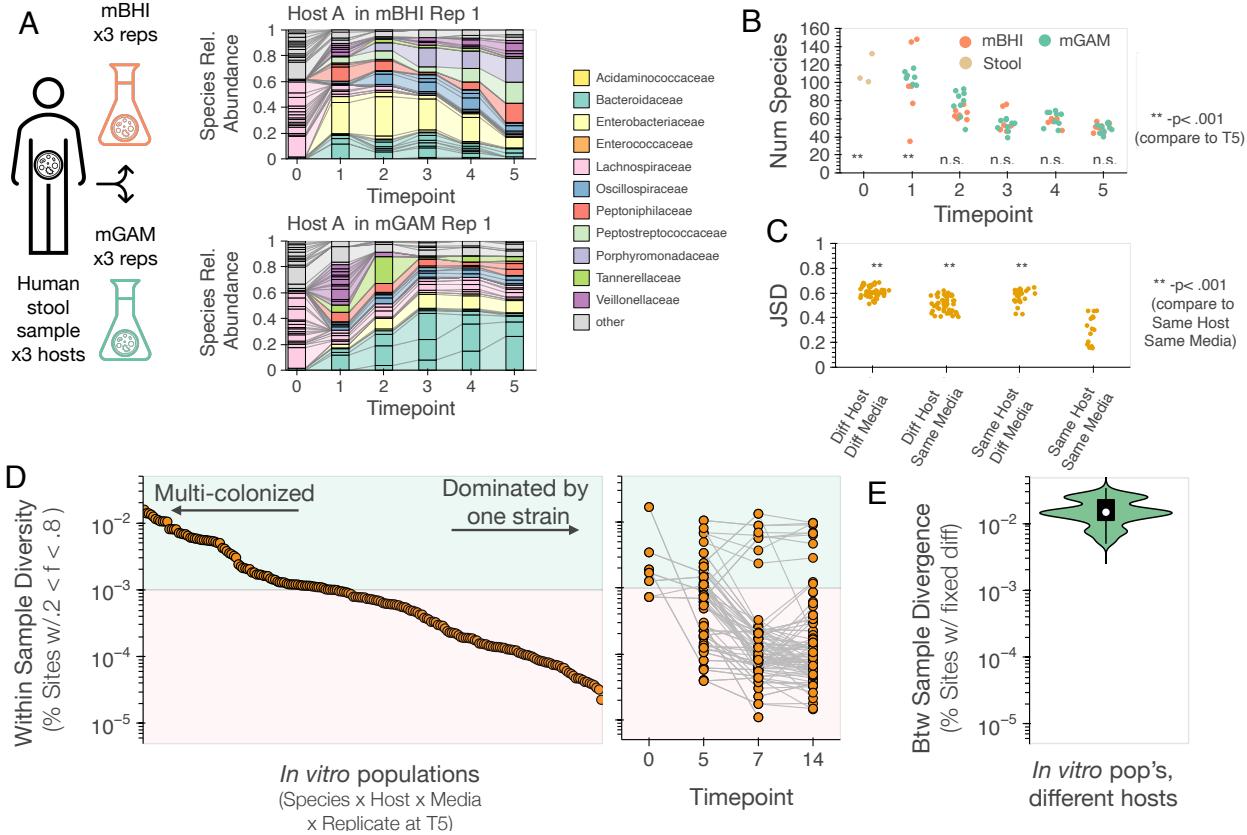
69 competition within a complex community setting, by leveraging the extensive strain-level  
70 variation that already exists within and among different people.  
71 Here we illustrate this approach by pairing strain-resolved metagenomics with serially passaged  
72 stool communities derived from different human donors. By tracking whole-community  
73 mixtures at high temporal and genetic resolution, we measured the emergent fitness differences  
74 between multiple pairs of conspecific strains as they competed within larger, species-rich  
75 communities. These community-scale fitness assays provide a window into the eco-evolutionary  
76 forces acting on circulating strains within complex microbial communities.

## 77 Results

### 78 ***In vitro* communities assembled from human stool samples contain a range of strain-level di-** 79 **versity**

80 To study strain-level competition in a complex community setting, we derived replicate *in vitro*  
81 microbial communities from frozen stool samples of three healthy human donors (Methods). We  
82 inoculated three replicate communities from each host in mBHI media and three in mGAM  
83 media, for a total of 18 initial communities. We propagated each community via 1:200 dilutions  
84 into fresh media every 48 hours for five total passages (~38 generations of growth), and froze the  
85 resulting communities in glycerol for future use. We then performed shotgun metagenomic  
86 sequencing on each timepoint as well as the original stool sample, and used a read mapping  
87 approach (66) to characterize species relative abundances and the genetic variation within each  
88 species (Methods). Consistent with previous results (40, 62–65), we found that communities  
89 rapidly converged to diverse steady states within ~2-3 passages (Fig. 1A-B,  $p < .05$  FDR  
90 correction, Methods) with ~40 species present above 0.1% relative abundance (Fig. 1B). While the  
91 final composition differed from the original stool sample (Fig. 1B;  $p < .001$  permutation test  
92 between replicate variation T5 vs replicate to stool), the overall richness was high  
93 (richness=number of species at relative abundance  $\geq .1\%$ , stool richness IQR=101-132,  
94 community richness at timepoint 5 IQR=46-53), and contained many prominent microbial  
95 families, such as *Bacteroidaceae* (Fig. 1A). The species composition of the assembled communities  
96 was also specific to the original donor sample and assembly media (Fig 1A-C,  $p < .001$   
97 permutation test).

98 To investigate the genetic diversity within species, we used a reference-based approach to  
99 identify single nucleotide variants (SNVs) within the core genomes of each species with at least  
100 5x coverage at a given timepoint (Methods). Consistent with previous observations in human  
101 fecal metagenomes (9, 67), we found that our *in vitro* populations varied widely in the total  
102 number SNVs that were present at intermediate frequencies (Fig. 1D, 20%  $< f < 80\%$ , Methods).  
103 Some populations have just a handful of polymorphisms, while others had to tens of thousands –  
104 comparable to the typical genetic differences between strains in different hosts (Fig. 1E)  
105 (9–11, 67). Previous work has shown that these high-diversity populations correspond to cases  
106 where the host was originally colonized by two or more diverged strains of the same species  
107 (9, 67). By contrast, the lower diversity populations correspond to cases where the species is  
108 dominated by a single resident strain whose genotype can be inferred from the consensus allele  
109 in the population. Using previously established thresholds (9, 67), we classified each population  
110 as single- or multi-colonized if the within-sample diversity was less than or greater than 0.1%.  
111 We found that around half of assessed populations at timepoint 5 showed evidence of



**Figure 1: Assembly of *in vitro* communities from human stool samples yields diverse, stable communities with a range of strain-level variation.**

(A) Stool samples from three healthy hosts were serially passaged in mBHI or mGAM media for a total of 38 generations (Methods). Two example communities are shown. Bars denote species relative abundances estimated from metagenomic sequencing (Methods), colored by bacterial family. (B) Species richness over time, estimated as the total number of species with relative abundance >0.1%. Statistical significance was assessed via a permutation test of median richness at each timepoint (Methods). (C) Jensen-Shannon distance (JSD) between communities at the end of passage 5, demonstrating that community composition is host- and media-specific. Statistical significance was assessed via a permutation test of median JSD of each type of comparison (Methods). (D) Intra-species genetic diversity, estimated from the fraction of polymorphic sites, for all species with sufficient sequencing coverage. Left panel shows genetic diversity at passage 5, while right panel shows the stability over an additional 68 generations (Methods). Many *in vitro* populations exhibit high levels of genetic diversity, comparable to the typical genetic differences between strains in different hosts (E). Horizontal line denotes the threshold used to classify samples as effectively mono-colonized (Methods).

112 multi-colonization by this criterion ( $n = 82/200$  populations, 41%, Fig. 1D). Since any  
113 non-growing strains would be rapidly diluted out, this indicates that multiple conspecific strains  
114 were able to grow and persist in these *in vitro* communities over time.

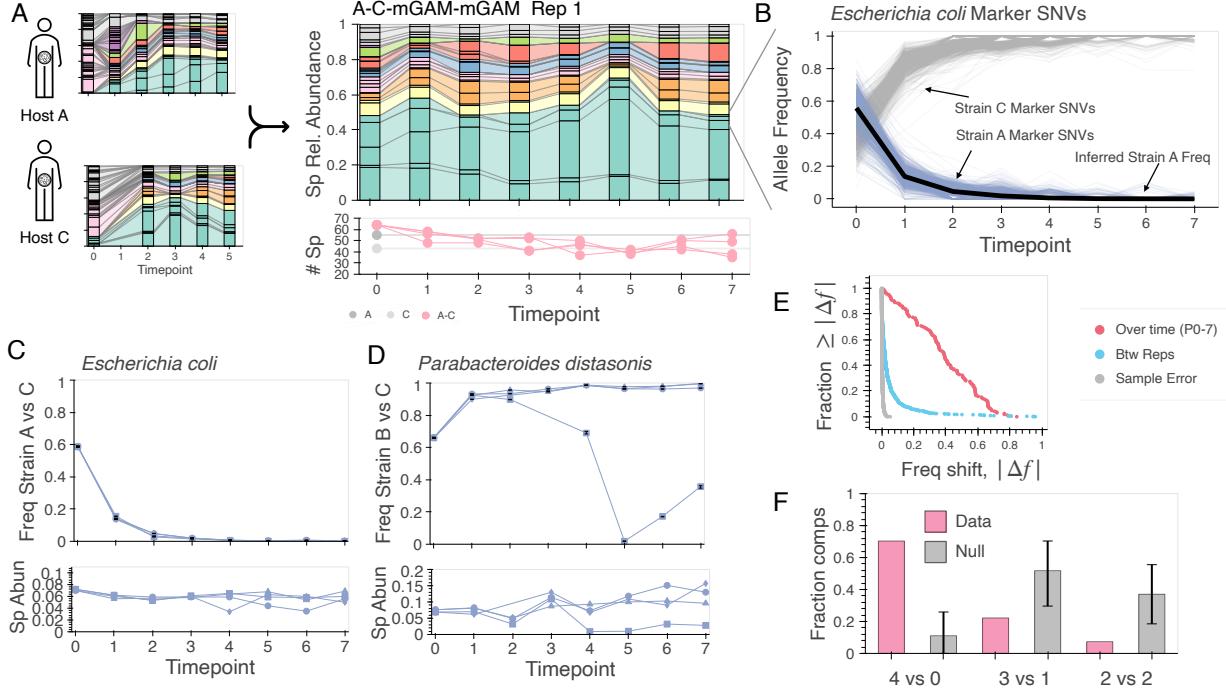
115 We confirmed that these signatures of multi-colonization were not driven by cross-contamination  
116 by testing for SNV sharing across samples from different hosts (Methods, (9, 10, 25)). Moreover,  
117 this intra-species diversity was also consistent across independent replicates from the same  
118 inoculum. Of the 34 populations with at least one multi-colonized replicate, a total of 21 (61.8%)  
119 had another replicate that was also classified as multi-colonized ( $p < .001$ ; permutation test,  
120 Methods). To examine the persistence of multi-colonization, we also revived the frozen  
121 communities at timepoint 5, and passaged them for an additional 9 growth cycles, or 68  
122 generations (Fig. 1E). While intra-species diversity declined in many multi-colonized populations  
123 (22 of 30, 73.3%), a substantial number (8 of 30, 26.7%) continued to exhibit high diversity at the  
124 end of this second round of passaging. These observations show that multiple conspecific strains  
125 can be maintained in spatially homogeneous communities for extended periods of time.

126 **Community coalescence reveals strong selection on conspecific strains from different host com-**  
127 **munities**

128 The multicolonization in Fig. 1D,E could be driven by nearly neutral competition or coexistence  
129 through niche partitioning. To distinguish these scenarios, we sought to perform more controlled  
130 investigations of selection on conspecific strains without a prior history of co-occurrence. We  
131 therefore turned to whole-community competitions [also known as “community collisions” (68)  
132 or “community coalescence” (69, 70)], where two communities are mixed together and  
133 propagated over time. Community coalescence experiments have previously been used to study  
134 species-level behavior (40, 71–74), and are reminiscent of fecal microbiome transplants in  
135 therapeutic contexts (27, 41). Here we sought to use the same idea to study the competition  
136 between strains of the same species that are derived from different host communities.

137 To implement this approach, we took pairs of assembled communities from timepoint 5 of our  
138 initial experiment and mixed them in equal volumes in a round robin design (Fig. 2A). This  
139 yielded a total of 6 initial inocula (3 pairwise mixtures in each of two assembly environments).  
140 We then split each inoculum into eight replicate collision experiments, four of which were  
141 passaged in mBHI and four passaged in mGAM, for 7 additional passages (~50 generations). We  
142 then performed metagenomic sequencing on each community at all timepoints to monitor their  
143 species and strain dynamics over time. Since the assembled communities from different hosts  
144 shared 4–7 species at relative abundance > 5%, we reasoned that this experimental design would  
145 allow us to measure competition between multiple strains in parallel, each of which was  
146 previously selected to grow in at least one *in vitro* community.

147 Consistent with previous work (40), we found that the species composition of each collision  
148 rapidly approached a new steady state within ~3–4 passages (Fig. 2A, Fig. S6 Fig. S7, Methods).  
149 Replicate collisions were highly reproducible and were specific to the initial inoculum and the  
150 media in which they were competed (Fig. S8,  $p < .001$  permutation tests, Methods). Many species  
151 in the initial mixture went extinct in the new steady-state community, such that final richness  
152 was lower than the initial mixture ( $p < .05$  permutation test, initial mixture richness IQR = 71–77,  
153 timepoint 7 richness IQR = 42–58, Fig. S7), and were comparable to autologous control  
154 communities that were passaged for the same amount of time ( $p > .05$  permutation test,



**Figure 2: Community coalescence reveals strong selection on conspecific strains from different host communities.**

**A** Stable communities from different hosts were coalesced through 1:1 community mixtures. (Left, Top Right) Bar plots are colored by family using the same color scheme as in Fig 1. (Bottom Right) Number of species was determined as the number of species with relative abundance  $\geq .1\%$ . Grey lines indicate the number of species in the communities used to inoculate the A-C-mGAM-mGAM collisions. Each colored line represents the dynamics of species extinctions in one replicate A-C-mGAM-mGAM collision. **B** Example SNP dynamics of the *in vitro* *Escherichia coli* population in the A vs C community coalescence in mGAM from communities assembled in mGAM. Each line represents the frequency of the non-reference allele at a site along the *E. coli* genome where strain C and strain A differ. Strain C (grey) and Strain A (lavendar) marker SNPs are sites where the strain C and strain A alleles respectively are the non-reference alleles. The dark black line represents the inferred trajectory of strain A (Methods). **C-D** Species and strain dynamics of replicate A vs C *E. coli* collisions (**C**) and replicate B vs C *P. distasonis* collisions (**D**). In both cases, collisions were performed in mGAM using communities assembled in mGAM. Marker shapes indicate replicates. Black lines in the strain frequency plot correspond to error bars (bootstrapped 95% confidence intervals, Methods) **E** Frequency shifts over time exceed frequency shifts between replicate collisions. Distribution of measurement error (grey), between replicates variation (blue) and frequency shift over passages 0 to 7 in a given replicate (pink). **F** The direction of shifts are consistent across replicates. Null distribution of how many competitions would be expected to have consistent shift directions across all replicates was generated through bootstrapping.

155 timepoint 7 richness IQR = 42-55, Fig. S7).

156 To measure competition within species, we leveraged the fact that two strains from different  
157 hosts typically varied at thousands of distinct marker SNVs (Fig. 1F). By tracking the allele  
158 frequencies of thousands (IQR=3220-12300 SNVs) marker SNVs using the reference-based  
159 pipeline above, one can infer the relative frequencies of the two strains with a resolution of ~ .1  
160 % (Fig. 2B-E, Fig. S9 Methods, IQR = .078-.469 %). For simplicity, we restricted our attention to  
161 the mono-colonized samples in Fig. 1D where the marker SNVs for each strain could be inferred  
162 with a high degree of confidence (Methods). After applying these filters, this resulted in a total of  
163 52 pairwise strain competitions across 12 species.

164 We found that the relative frequencies of the strains often underwent large shifts over the first  
165 few passages of the collision (Fig. 2B-E). The magnitudes of these shifts were significantly larger  
166 than expected from either measurement error or demographic noise (Fig. 2E,  $p < .001$ ,  
167 permutation test), suggesting that they are driven by differential growth rates within the  
168 community. Strain dynamics were also highly similar across replicate competitions, with most  
169 (87.7%) replicates differing by less than 10% from each other (Fig. 2C-E). As a result, the overall  
170 direction of the shift between timepoints 0 and 7 was consistent in ~80% of cases, and was  
171 significantly higher than expected by chance (Fig. 2F,  $p < .001$ , Methods). The magnitudes and  
172 reproducibility of these shifts imply that these intra-species dynamics are driven by strong and  
173 deterministic selection.

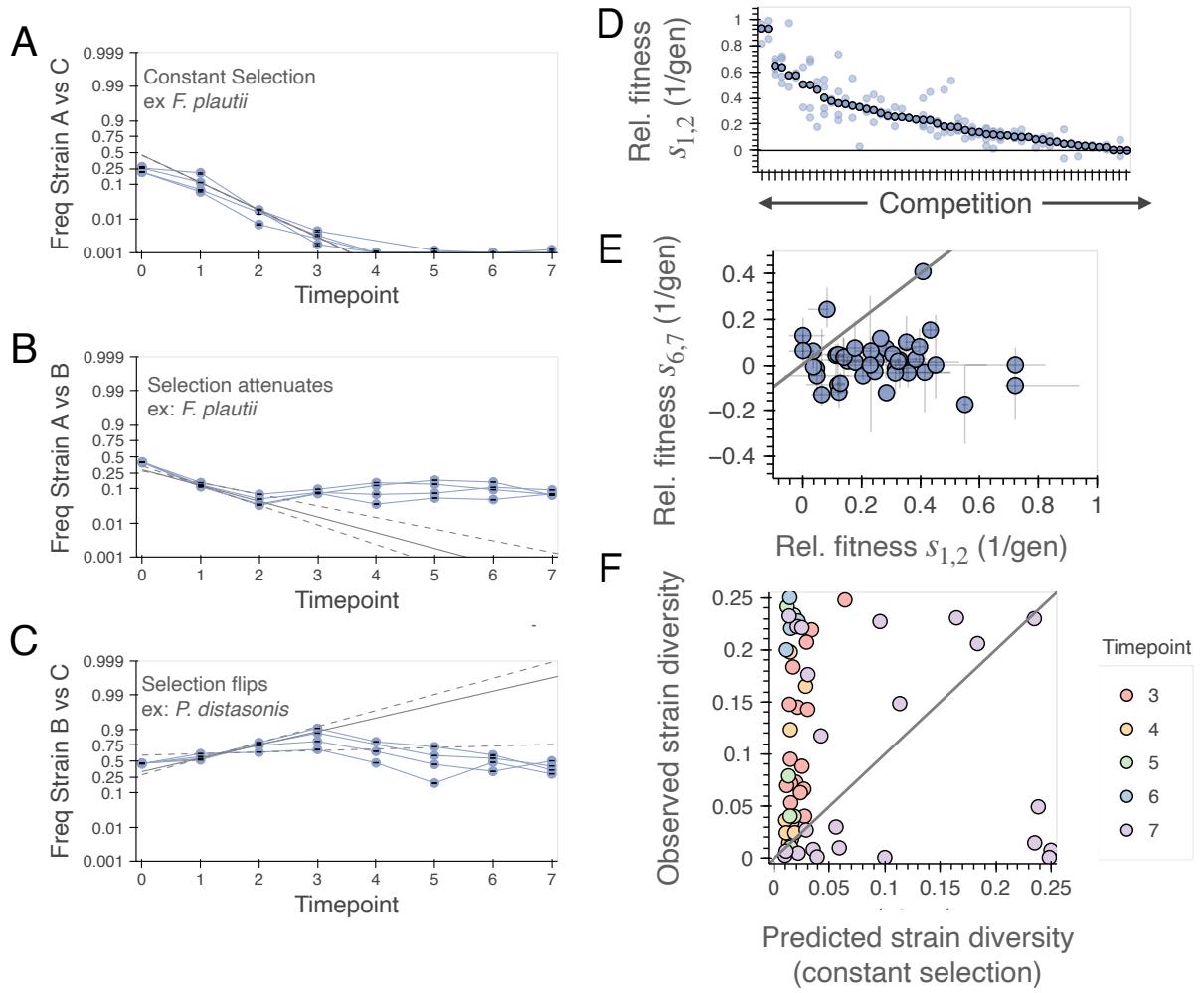
174 While most competitions were consistent across replicates, we also observed some cases where  
175 the trajectories were more variable. For example, in the *Parabacteroides distasonis* competition in  
176 Fig. 2D, one replicate trajectory underwent a large reversal in frequency, while the other three  
177 replicates remained highly similar to each other. As above, the magnitude of this difference was  
178 far larger than expected from measurement error or demographic noise ( $p < .001$ , Methods),  
179 suggesting that it was likely driven by idiosyncratic differences in the selective landscape within  
180 this particular replicate community. Consistent with this hypothesis, we found that the species  
181 abundance of *P. distasonis* also systematically diverged in this replicate (Fig. 2D), suggesting that  
182 it experienced a different growth environment possibly due to subtle shifts in the local biotic  
183 environment. This same pattern also holds more broadly, with a statistically significant  
184 correlation between species- and strain-level deviations between replicates (Fig. S11,  $p < .001$ ,  
185 pearson's correlation). These observations further highlight that conspecific strains can  
186 experience strong selective differences when competing in a larger community, even in  
187 communities that were assembled in the same abiotic environment.

### 188 Selection shifts over time, allowing for strain coexistence

189 To more precisely quantify the selective forces acting on the conspecific strains in Fig. 2, we  
190 estimated their standard relative growth rate,

$$s(t) \approx \frac{1}{\Delta t} \log \left( \frac{f(t + \Delta t)}{1 - f(t + \Delta t)} \cdot \frac{1 - f(t)}{f(t)} \right), \quad (1)$$

191 from their frequency trajectories  $f(t)$ ; this provides an operational definition of their relative  
192 fitness in the interval between  $t$  and  $t + \Delta t$ . This difference in per capita growth rates can be  
193 computed for any two lineages within a community, regardless of whether they are competing  
194 within the same ecological niche. In the latter case,  $s(t)$  reduces to the standard competitive



**Figure 3: Selection pressures shift over time, promoting strain coexistence**

(A-C) Examples of constant ((A)) and time varying (B,C) selection. Lavendar lines indicate data, and grey lines represent the median (solid) and highest and lowest (dashed) predicted strain trajectories under a model of constant selection fit using data from timepoints 1 and 2. (D) Selection shifts between the early (timepoitns 1-2) and late (timepoints 6-7) phases of competitions. Each point represents the median selection coefficient across replicates of strains that underwent positive selection between timepoints 1 and 2. Grey lines represent ranges of replicate measurements. (E) Temporal shifts in selection favor the maintenance of strain diversity (product of strain frequencies). Each point represents the the predicted and observed strain frequency for a given competition at the final timepoint in which the predicted minor strain frequency was greater than 1%.

195 fitness from population genetics. In more general contexts,  $s(t)$  will also account for feedbacks  
196 between the mutant strains and their environment (75, 76), including any frequency-dependent  
197 interactions between the focal strains themselves.

198 Using this metric, we found that most pairs of strains exhibited large fitness differences during  
199 the early stages of competition (Fig. 3D). The typical selection coefficients between passages 1  
200 and 2 ranged from 12-35% per generation (IQR), comparable to the largest strain-level fitness  
201 differences that have been inferred from *in vivo* microbiome sequencing (9, 12, 25, 26, 77, 78). If  
202 selection were constant over time, these relative fitness differences are large enough that the fitter  
203 strain would be expected to dominate the population over the next several passages.

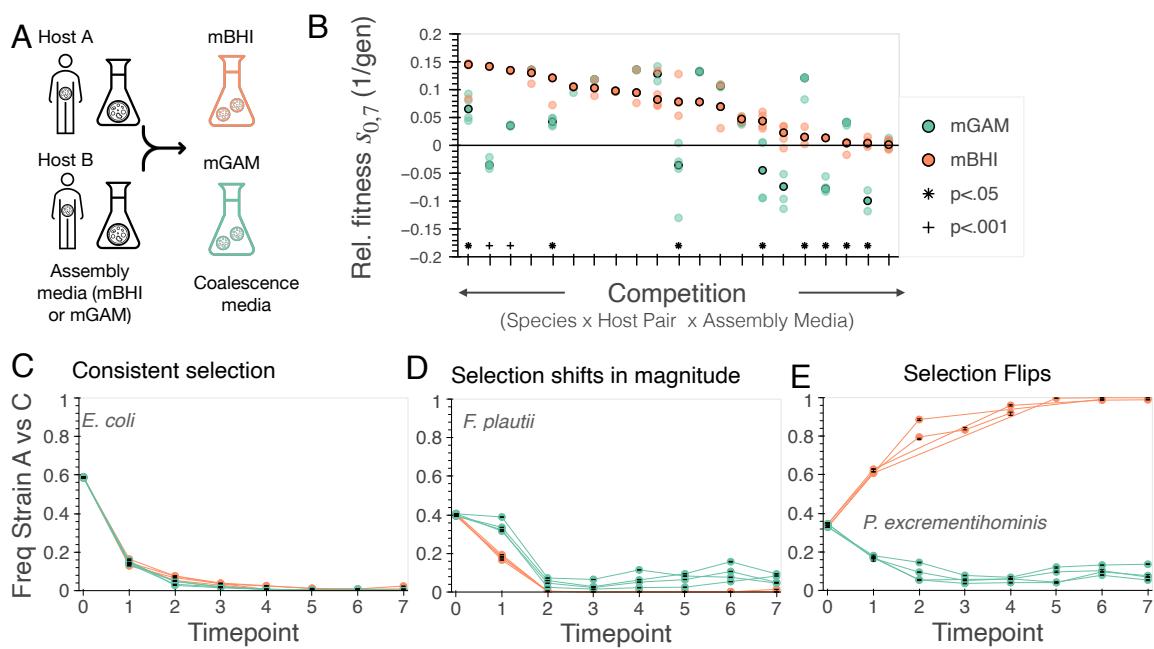
204 Some *in vitro* competitions were consistent with this prediction, with the fitter strain sweeping  
205 within the focal species at a roughly constant rate (e.g. *Flavinobacter plautii* in Fig. 3A). However,  
206 in many cases, we observed that the strain frequencies strongly deviated from a model where  
207 selection is constant over time (Fig. 3B-D, 9/36, Methods,  $q < .05$ , t-test, BH correction and effect  
208 size  $> 1\%$ ). These temporal deviations were often consistent across multiple independent  
209 replicates (Fig. 3B-D, Fig. S13), demonstrating that they were caused by deterministic shifts in the  
210 selective landscape over the course of the competition. Since the abiotic environment was held  
211 fixed throughout these experiments, we conclude that the changing intra-species selection  
212 coefficients must have been caused by shifts in the biotic environment. These feedbacks could  
213 arise through frequency-dependent selection between the focal strains, or more general  
214 interactions with other species in the community.

215 We found that the strength of selection often attenuated over time, such that the fitness  
216 differences at later timepoints were much smaller than the first two passages ( $p < 10^{-3}$   
217 permutation test, median  $|s_{1,2}|$  across replicates IQR =.123-.348 per generation, median  $|s_{6,7}|$   
218 across replicates IQR =.023-.090 per generation; Fig. 3D). In some of these cases (e.g. *Flavinobacter*  
219 *plautii* in Fig. 3B,), the selection coefficients appeared to relax towards zero, suggestive of stable  
220 coexistence (Fig. S13, (76)). In other examples, the direction of selection reversed at later  
221 timepoints (e.g. *Parabacteroides distasonis* in Fig. 3C), such that the strains remained at  
222 intermediate frequencies by the end of passage 7. Interestingly, we found that across all of these  
223 cases, the shifts in  $s(t)$  tended to be biased in the direction of increasing intra-species diversity.  
224 By comparing the number of competitions that fell above versus below their corresponding  
225 constant selection curves, we found that the vast majority of the deviations were biased towards  
226 leaving the strains at intermediate frequencies (Methods, Fig. 3E,  $p < .01$ ). This suggests that the  
227 time-varying selection pressures tended to promote coexistence of conspecific strains within the  
228 context of their larger community.

## 229 Conspecific strains exhibit fitness tradeoffs across abiotic environments

230 Our experimental design also allows us to follow independent replays of the same collision  
231 across two different nutrient environments (Fig. 4A). This allows us to quantify how the  
232 dynamics of intra-species competition vary under shifts in the abiotic environment. Previous  
233 theoretical work has suggested that large communities might self-organize to shield themselves  
234 from external environmental perturbations (39, 79, 80), but the extent to which this occurs in  
235 practice remains unclear.

236 To investigate this question, we focused on the 21 cases where the same species was present at



**Figure 4: Conspecific strains exhibit fitness tradeoffs across abiotic environments.** (A) Community coalescence experiments were performed in two different nutrient conditions, allowing comparisons of strain-level fitness across different abiotic environments. (B) Quantifying shifts in selection across abiotic environments. Solid circles with dark outline indicate the median selection coefficient across replicates in a given median. Lighter colored dots are biological replicates. Significance obtained via an FDR corrected t-test. (C-E) Examples of consistent (C) and inconsistent (D,E) selection across abiotic environments in competitions between strains from hosts A and C.

reasonable abundances in collisions performed in both growth media. We then asked how the underlying strain trajectories in these species differed from each other. In some cases, such as the *E. coli* competition between hosts A vs C (Fig. 4B,C), we observed remarkably consistent trajectories across the two abiotic environments, both in the qualitative shape of the trajectory as well as the quantitative values of  $s(t)$ . In other cases, we observed significant differences in the strain trajectories across the two abiotic environments (Fig. 4B,D,E). Some of these shifts were quantitative in nature. For instance, the *F. plautii* strain from host A declined in frequency in both media when competing against the strain from host C, but the magnitude of the decline was greater in mBHI than in mGAM (Fig. 4D;  $p < .01$  t-test, FDR correction BH). In other cases, the direction of selection flipped between the two abiotic conditions (e.g. *P. excrementihominis* in Fig. 4E,  $p < .001$  t-test, FDR correction BH), or disrupted instances of apparent coexistence (e.g. *P. excrementihominis* in Fig. S17  $p < .001$  t-test, FDR correction BH).

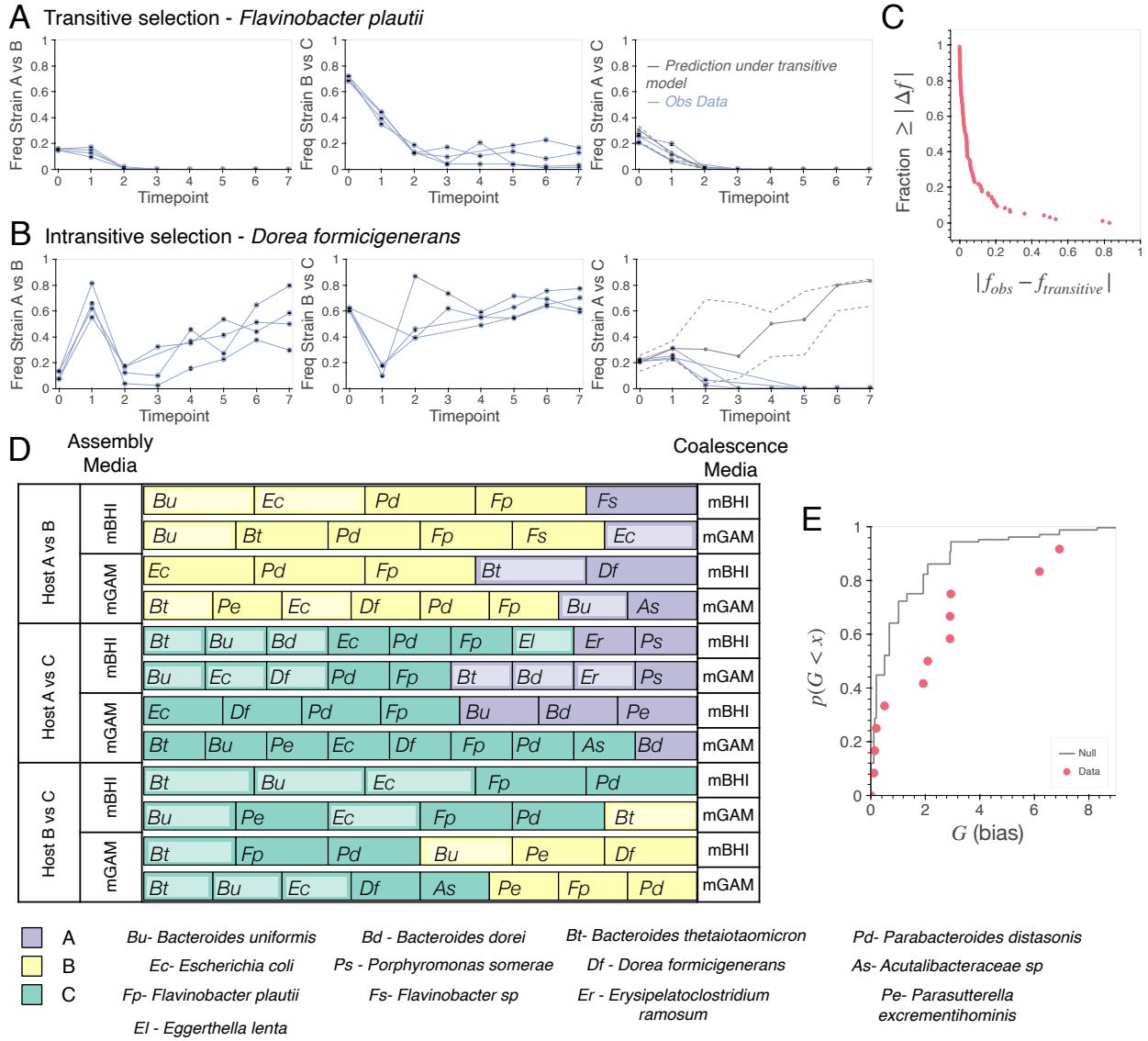
To quantify the prevalence of these effects more broadly, we computed the net selection coefficients between days 0 and 7 and compared them across the two nutrient environments (Methods). We observed significant environmental differences in these net selection coefficients in ~50% of cases ( $n=10/21$ ; FDR corrected t-test; Fig. 4B, Fig. S16, Fig. S17), half of which ( $n=5/10$ ) involved a reversal in the overall sign of selection (Fig. S16). These data show that intra-species selection can be highly contingent on shifts in the abiotic nutrient environment. In a community context, these genotype-by-environment interactions could be mediated both by direct effects on the focal strains, or indirect effects from shifts in the other community members (Fig. S8).

## Deviations from transitivity reveal community-specific selection pressures

We also asked how intra-species competition depends on the biotic environment. While the results in Fig. 3 show that selection is often sensitive to the surrounding community, these shifting selection coefficients could be caused by biotic interactions that are common to all communities assembled in the same environment, as well as idiosyncratic interactions (e.g. the identity of the competing strain) that vary across communities from different hosts. We can test for these latter effects indirectly by examining the transitivity of strain-level selection across collisions from different pairs of hosts. If the underlying fitness landscape was consistent across communities, then we would expect that the selection coefficients between different pairs of strains should be transitive at each point in time. By contrast, any signatures of non-transitivity would indicate that their relative growth rates are varying in a community-specific manner.

A striking example of non-transitivity is illustrated by the *Dorea formicigenerans* populations in Fig. 5B. In this species, the competitions between hosts A-vs-B and B-vs-C suggest that the strain from host A should dominate over host C strain if the underlying fitness landscape was transitive. However, the actual data showed the opposite trend, with the strain from host C dominating over the host A strain in a rock-paper-scissors-like manner (Fig. 5B).

These rock-paper-scissors dynamics were uncommon among the broader collection of species in our dataset. Of the 13 cases where all three pairwise competitions could be observed, Fig. 5B was the only example where the sign of selection clearly flipped. We therefore tested for more quantitative deviations from transitivity by leveraging the relative fitness model in Eq. (1). For each of the 13 trios, we used two pairwise competitions to infer the relative fitness trajectories for two of the three pairs of strains at each point in time (Methods). Under the assumption of transitivity, this yields a time-varying prediction for the third pairwise competition (dashed lines



**Figure 5: Evaluating the role of community-specific interactions on intra-species competition.** Examples of competition trios exhibiting transitive (A) and intransitive (B) selection. Each column shows the dynamics of conspecific strains in one competition for a trio of competitions. The third column also includes the predicted dynamics under the transitive model (grey). The higher and lower lines dashed represents upper and lower limits of predictions generated through bootstrapping. C Distribution of the absolute differences in observed frequency ( $f_{obs}(t)$ ) and predicted frequency under the transitive model ( $f_{transitive}(t)$ ) for each competition trio at each timepoint. D Evaluating correlations with subject of origin in competition outcomes. Each row represents a pairwise coalescence. Entries are colored by the host in which the winning strain was derived from. Entries with a light box indicate that one or both of the initial populations used to inoculate collisions was multi-colonized. E Distribution of bias statistics ( $G$ ) for each competition (points) compared a null model where strain winners are selected randomly from each host with equal probability.

280 in Fig. 5A-C) that has only one additional parameter (corresponding to the initial relative  
281 frequency). By comparing these predictions against the data, we identified some cases ( $n = 6/13$ )  
282 where the observed trajectories closely matched the predictions of the transitive model (Fig. 5A,  
283 C, Fig. S18 difference of at least 5% between inter-replicate variation and transitive model  
284 predictions). However, we also observed several additional cases ( $n = 7$ ) in which there were  
285 quantitative deviations from transitivity even when the overall sign of selection was preserved  
286 (Fig. 5C, Fig. S19). These observations, together with our earlier observations in Figs. 2D and 3,  
287 demonstrate that strain-level competition can be contingent on the local biotic context, even in  
288 communities that were originally assembled in the same abiotic environment.

### 289 Limited strain-level cohesion during community coalescence

290 Finally, we sought to examine whether strain-level competition was correlated across multiple  
291 species in the same community. Such correlated outcomes could arise from direct interactions  
292 (e.g. if one strain consumes a metabolite produced by another) (69, 71). They could also arise  
293 indirectly if one community is better adapted to the environment where coalescence occurs (73).

294 To test for these effects, we calculated the net selection coefficient between timepoints 0 and 7 for  
295 all strains in a given collision, and asked whether the sign of selection was biased towards one  
296 subject of origin (Fig. 5D,E, Methods.). Although, the typical host bias across all competitions  
297 was elevated compared to a null model where the winners were randomly chosen from each host  
298 (Methods, Fig. 5E, median obs host bias = 2.01, null host bias = 0.51,  $p = 0.002$ ), no individual  
299 competition showed evidence for significant host bias ( $q > .05$ ,  $\chi^2$  test, BH FDR correction, IQR  
300 of G-statistic = 0.187 - 2.91, Fig 5D). In particular, we found that the winning strains in each  
301 competition often originated from different hosts. This suggests that strain-level cohesion played  
302 a limited role our experiments, despite the past evolutionary history of selection within each host.

## 303 Discussion

304 Our results show that community coalescence can provide a scalable approach for measuring  
305 intra-species competition within larger microbial communities. By colliding *in vitro* gut  
306 communities derived from healthy human donors, we quantified the emergent selection  
307 pressures on conspecific strains across a range of biotic and abiotic conditions. Our statistical  
308 characterization of this landscape revealed that strains from different hosts are generally subject  
309 to strong and context-dependent selection pressures, even when their original communities were  
310 assembled in the same abiotic conditions. We also found that the relative fitnesses shifted over  
311 time due to ecological interactions within the community, enabling long-term coexistence of  
312 multiple conspecific strains in the same spatial location. While our data do not directly  
313 distinguish between stable and unstable coexistence, they nevertheless support the view that  
314 some conspecific strains can coexist with each other by partitioning the available resources.

315 Our results are consistent with recent theoretical predictions showing that mutant strains can  
316 often coexist within larger communities, even when they lack an exclusive spatial or metabolic  
317 niche (39, 81). However, on longer evolutionary timescales, the conspecific strains in our study  
318 exhibit few genome-wide signatures of stable ecological differentiation (or “ecotypes”) (82), with  
319 high levels of gene flow that are more consistent with a single biological species (20, 21, 23, 45).  
320 Understanding how this long-term genetic cohesion emerges in a world of widespread local

321 niche differentiation remains an interesting evolutionary puzzle. Our results suggest that local  
322 coexistence could help facilitate this genetic mixing in ecologically diverse communities like the  
323 human gut, by providing more opportunities for conspecific strains to engage in horizontal gene  
324 transfer and recombination (21, 53, 83).

325 Our competition experiments were performed in a simplified environment that omits important  
326 factors, such as spatial structure or host immunity, that are thought to influence strain-level  
327 selection in natural gut microbiomes. While these simplifications allowed us to show that  
328 conspecific strains can coexist even in the absence of these additional factors, further work is  
329 needed to understand how the fitnesses differences measured in our community experiments  
330 relate to the specific selection pressures that these strains experience *in vivo*. Future experiments  
331 could incorporate these factors in a controlled manner, by performing similar community  
332 collisions in spatially structured environments (84) or in gnotobiotic mice (85).

333 Our experiments also focused on the standing genetic variation present among gut bacterial  
334 strains from different hosts. On longer timescales, these initial strains will start to acquire further  
335 genetic changes as they begin to adapt to their local community environment (12, 78, 86). These  
336 processes of local evolution and inter-strain competition, when aggregated over the larger  
337 collection of human microbiomes, ultimately generate the phenotypic and genotypic diversity  
338 within the global strain pool. Understanding the evolutionary rules that shape this phenotypic  
339 landscape — and how they impact competition on local scales — will be critical for  
340 understanding how large microbial communities will evolve.

341 **Methods**

342 **Collection of stool samples**

343 This study was approved by the Stanford University Institutional Review Board as Protocol  
344 IRB-64602, and written, informed consent was obtained from all participants. Healthy adult  
345 individuals were recruited from Stanford University and the surrounding community. Exclusion  
346 criteria included age under 18, pregnancy or nursing, chronic illness and routine use of  
347 prescription medications known to influence the gastrointestinal tract, as well as antibiotic usage  
348 in the past three months. Stool samples were collected using the protocol described in Ref. (25).  
349 Each participant collected one stool sample by depositing ~20 mL of stool into a vial. Stool  
350 samples were frozen immediately after collection in donors' home freezers and were stored there  
351 for up to one week until they were transferred to storage at -80 degrees C in the laboratory. We  
352 initially collected stool samples from 8 study participants. After an initial round of sequencing  
353 and species profiling (see below), we selected 4 of these samples for further *in vitro* passaging  
354 based on the abundances of their *Bacteroidaceae* species.

355 **Assembly of *in vitro* communities from stool samples**

356 *In vitro* communities were assembled from frozen stool samples using the methods described in  
357 Refs. (40, 62, 63). All culturing experiments were performed in an anaerobic chamber (Coy  
358 instruments, H<sub>2</sub> 2.4%, CO<sub>2</sub> 6.5%) with pre-reduced media and solutions. Small fragments (~50  
359 mg) of frozen stool were first resuspended in 1 mL of pre-reduced, filter-sterilized PBS. This  
360 initial stool suspension was then used to create six initial cultures, by adding 20 µL aliquots of  
361 the stool suspension to 180 mL of pre-reduced mGAM (HyServe No 1005433) or Brain-Heart  
362 Infusion (BHI, BD Biosciences 237200) supplemented with 0.2 mg/mL L-tryptophan, 1 mg/mL  
363 L-arginine, 0.5 mg/mL L-cysteine, 5 µg/mL vitamin K, .5 µg/mL hemin (mBHI) in a 96-well  
364 deep-well plate, for a total of three replicates per subject and media condition. After inoculation,  
365 communities were grown at 37°C and were passaged via a 1:200 dilution into fresh pre-reduced  
366 media every 48 hr for a total of 5 passages (~38 generations). Cultures were not shaken during  
367 growth, but were mixed via pipetting immediately prior to each dilution. At the end of the  
368 experiment, 50 µL of each saturated culture was frozen in 25% glycerol and stored in mylar bags  
369 at -80 °C for future use. This yielded a total of 24 initial communities (Table S2).

370 **Community coalescence experiments**

371 Coalescence experiments were performed using a modified version of the protocol described in  
372 Ref. (40). For each nutrient condition, thawed glycerol stocks from one replicate of each  
373 community at passage 5 were inoculated into 200 µL of fresh, pre-reduced media and were  
374 passaged as before for two additional growth cycles. In the second growth cycle, communities  
375 were grown in a larger volume (475 µL) to produce sufficient cell counts for the coalescence  
376 experiment. At the end of the second growth cycle, pairwise mixtures were formed by mixing  
377 100 µL of each saturated culture to create an initial inoculum, with two replicate inocula created  
378 for each pair of "parent" communities. From each inoculum, four replicate coalescence  
379 experiments were founded by adding 1 µL of the initial inoculum to 199 µL of fresh mGAM  
380 (two replicates) or mBHI (two replicates); the rest of the inoculum was frozen at -80 °C for  
381 sequencing. We also founded single-community control lines (two replicates per growth  
382 medium) using 1 µL of the saturated culture at the end of the second growth cycle. This yielded a

383 total of 96 collision experiments and 32 controls (see Table S3). Communities were then  
384 propagated for seven additional passages using the protocol described above, with frozen  
385 samples preserved at the end of each passage for sequencing.

386 **DNA extraction and metagenomic sequencing**

387 For the initial sequencing of stool samples, DNA was extracted from a small (~50mg) fragment  
388 of stool using the Qiagen DNAeasy PowerSoil Kit. This same extraction method was also used to  
389 obtain DNA from the resuspended stool inocula that were used to found our initial *in vitro*  
390 communities. DNA from subsequent passages was extracted from 50  $\mu$ L of frozen culture using  
391 the Qiagen DNAeasy Ultraclean 96 Microbial Kit. Metagenomic sequencing libraries were  
392 prepared using the Nextera DNA Flex Library Prep kit. Sequencing was performed on an  
393 Illumina NovaSeq with read lengths of 2x150bp and a target depth of ~2.5 Gbp per sample  
394 (Table S1). This target depth was chosen so that species with a relative abundance of >1% would  
395 have an expected coverage of >5x.

396 **Species and SNP profiling of metagenomic samples**

397 Raw sequencing reads were trimmed using Skewer (87) and reads mapping to the human  
398 genome (GRCh38) were filtered using Bowtie 2 (88). Remaining reads will be made available in  
399 the SRA upon publication.

400 Species relative abundances were estimated using the species module of MIDAS2 (66), which  
401 maps the trimmed sequencing reads against a set of universal, single copy marker genes  
402 representing different bacterial species from the UHGG database (5). Relative abundances were  
403 reported for all species that had at least two marker genes with at least two mapped reads in a  
404 given sample. We used these estimated species abundances for all species-level analyses, such as  
405 calculating species richness or Jensen-Shannon distance (see below).

406 Single nucleotide variants (SNVs) were identified using a variant of the MIDAS2 snps module.  
407 We first generated a list of the 150 most abundant species, based on their maximum relative  
408 abundance across all samples. We then used the MIDAS2 snps module to map the raw  
409 sequencing reads from each sample against a set of dereplicated reference genomes representing  
410 these different microbial species. For each site in the associated reference genomes, MIDAS2  
411 reports the number of reads mapping to each allele, as inferred from the read pileups in that  
412 sample. To enable accurate tracking of mutations over time, we modified the source code of  
413 MIDAS2 to report read counts for all alleles, independent of their total coverage within a single  
414 sample. We then used the MIDAS2 merge module to combine allele counts across all samples in  
415 our study. We only included populations with a median vertical depth of 1 and a horizontal  
416 depth of .5, and excluded all sites whose total coverage differed from the median vertical depth  
417 by more than 2.5x. For our downstream analyses, we restricted our attention to populations  
418 where the median non-zero coverage along the genome ( $\bar{D}$ ) was greater than 5x (10, 25).

419 **Quantifying species composition and community stability**

420 For each sample, we estimated the total species richness using the number of species that were  
421 present with an estimated relative abundance greater than 0.1% (Figs. 1B, 2B, S1, S2, S3, S4, S5,S6,

<sup>422</sup> S7, S8). We also computed the compositional dissimilarity between pairs of communities using  
<sup>423</sup> the Jensen-Shannon distance (JSD),

$$JSD_{ij} = \sqrt{\sum_{\mu} \left[ f_{\mu,i} \log_2 \left( \frac{2f_{\mu,i}}{f_{\mu,i} + f_{\mu,j}} \right) + f_{\mu,j} \log_2 \left( \frac{2f_{\mu,j}}{f_{\mu,i} + f_{\mu,j}} \right) \right]}, \quad (2)$$

<sup>424</sup> where  $f_{\mu,i}$  is the relative abundance of species  $\mu$  in host  $i$  (Figs. 1C, S6, S8).

<sup>425</sup> We used these metrics to quantify the time to stabilization during community assembly and  
<sup>426</sup> coalescence (Figs 1B, S6, S7). To assess convergence using species richness, we compared the  
<sup>427</sup> distribution of species richness at a given timepoint to the distribution at the final timepoint  
<sup>428</sup> (passage 5 for community assembly and passage 7 for community coalescence). Statistical  
<sup>429</sup> significance was assessed via permutation tests of the median values of each distribution.

<sup>430</sup> To assess convergence using JSD, we used a similar procedure to compare the distribution of  
<sup>431</sup> JSDs between the two final timepoints (4-5 for assembly, and 6-7 for coalescence) to the  
<sup>432</sup> distribution between pairs of earlier consecutive timepoints. We also quantified community  
<sup>433</sup> specificity by comparing the distribution of JSDs between pairs of replicate communities to the  
<sup>434</sup> JSDs of communities assembled or coalesced in different conditions (Figs 1B, S8).

### <sup>435</sup> Quantifying intra-species diversity within and between samples

<sup>436</sup> For each species population in each sample, we estimated the intra-sample genetic diversity by  
<sup>437</sup> calculating the fraction of non-filtered sites in which the major allele was present at an  
<sup>438</sup> intermediate frequency ( $f < 0.8$ ). We also calculated a measure of the genetic divergence between  
<sup>439</sup> pairs of samples, defined as the fraction of non-filtered sites in which an allele was present at  
<sup>440</sup> high frequency ( $f > 0.8$ ) in one sample and low frequency ( $f < 0.2$ ) in the other (Fig. S9). Many  
<sup>441</sup> *in vitro* populations had an intra-sample diversity much lower than their average divergence to  
<sup>442</sup> samples from other hosts (Fig. 1D), suggesting that these samples were likely dominated by a  
<sup>443</sup> single colonizing strain (67, 89). Following Ref. (9), we used these data to classify a given *in vitro*  
<sup>444</sup> population as effectively mono-colonized (“quasi-phaseable”) if its intra-sample diversity was  
<sup>445</sup> less than  $10^{-3}$  (approximately 10-fold lower than the typical between-subject divergence).

### <sup>446</sup> Testing for cross-contamination between samples

<sup>447</sup> We checked for signatures of cross-contamination during passaging by maintaining a set of blank  
<sup>448</sup> wells on each 96-well plate. We also checked for signatures of contamination in our genetic data  
<sup>449</sup> using generalization of the approach in Ref. (25), which searches for unexpected instances of  
<sup>450</sup> strain sharing between communities that were derived from different hosts. Strain sharing was  
<sup>451</sup> assessed by computing the genetic divergence metric defined above, and searching for cases in  
<sup>452</sup> which we observed fewer than 10 fixed differences (sites where an allele is present at >80%  
<sup>453</sup> frequency in one sample and <20% frequency in another). Although it is possible that two  
<sup>454</sup> unrelated subjects contain genetically similar strains by chance, previous work has shown that  
<sup>455</sup> two random subjects are unlikely to share multiple highly similar strains (9, 25). We considered  
<sup>456</sup> there to be sufficient evidence for contamination if multiple species showed evidence of strain  
<sup>457</sup> sharing with a sample from another subject (excluding the collided subject in samples from the  
<sup>458</sup> community coalescence arm). This procedure revealed that one of the four initial fecal  
<sup>459</sup> suspensions was likely contaminated with bacteria from another subject. To be conservative, we

460 removed all samples involving this subject from all of our downstream analyses. This left a total  
 461 of three initial hosts, with 48 community coalescence experiments and 24 autologous controls.

## 462 Tracking the relative frequencies of conspecific strains during community coalescence

463 For each community coalescence experiment, we identified the subset of species that were  
 464 present at sufficient coverage ( $\bar{D} > 5$ ) in both parent communities at the end of the community  
 465 assembly phase. We then quantified the frequencies of the strains within these shared species  
 466 during the resulting coalescence experiment. Strain frequencies were estimated using a  
 467 generalization of the “quasi-phasing” approach in Refs. (9) and (25), leveraging the fact that both  
 468 parent communities were measured separately prior to mixing.

469 For most of our analyses, we restricted our attention to species where both parent populations  
 470 were classified as quasi-phaseable according to the criterion described in the previous section (i.e.  
 471 within-sample diversity  $< 10^{-3}$ ). For these populations, previous work has shown that the  
 472 genotype of the dominant strain can be inferred with a high degree of confidence, simply by  
 473 searching for alleles that are present at a sufficiently high frequency (e.g.  $f > 0.8$ ; 9). We used this  
 474 idea to identify a collection of marker SNVs that distinguish the dominant strains in each of the  
 475 pre-mixture communities. A SNV was counted as a marker SNV for a given strain if the  
 476 non-reference allele was present at >80% frequency in its pre-mixture population and <20%  
 477 frequency in the other population that it was collided with. We further refined these sets of  
 478 marker SNVs by removing any SNVs that deviated from the median marker SNV frequency by  
 479 >50% frequency in >25% of the samples from the coalescence experiment; Fig. S9B shows that  
 480 this typically removes ~1000 putative marker SNVs from each pairwise collision. The resulting  
 481 sets of marker SNVs were then used to track the relative frequencies of the conspecific strains  
 482 during coalescence.

483 The frequency of each strain was estimated in each subsequent timepoint using a piecewise  
 484 estimator designed to reduce the impact of mapping errors and other sequencing artifacts. We  
 485 first computed the median frequency across all the the marker SNVs for a given strain,

$$f_{i,t}^{\text{median}} \equiv \underset{\ell \in \mathcal{M}_i}{\text{median}} f_{\ell,t}, \quad (3)$$

486 where  $f_{\ell,t}$  is the frequency of the non-reference allele at site  $\ell$  in timepoint  $t$  and  $\mathcal{M}_i$  is the set of  
 487 marker SNVs corresponding to strain  $i$ . We then estimated the frequency of strain  $i$  using the  
 488 piecewise formula

$$f_{i,t} = \begin{cases} f_{i,t}^{\text{median}} & \text{if } 0 < f_{i,t}^{\text{median}} < 1, \\ \frac{\sum_{\ell \in \mathcal{M}_i: A_{\ell,t}=1} 1}{\sum_{\ell \in \mathcal{M}_i: A_{\ell,t}=0} D_{\ell,t}} & \text{if } f_{i,t}^{\text{median}} = 0, \\ 1 - \frac{\sum_{\ell \in \mathcal{M}_i: A_{\ell,t}=D_{\ell,t}-1} 1}{\sum_{\ell \in \mathcal{M}_i: A_{\ell,t}=D_{\ell,t}} D_{\ell,t}} & \text{if } f_{i,t}^{\text{median}} = 1, \end{cases} \quad (4)$$

489 where  $A_{\ell,t}$  is the number of reads supporting the alternate allele at site  $\ell$  at timepoint  $t$  and  $D_{\ell,t}$  is  
 490 the total coverage. In other words, we estimate the frequency of a strain as the median frequency  
 491 of its set of distinguishing snps, unless the median frequency was 0 or 1. In those cases we  
 492 estimated the frequency of a strain using the ratio of counts of sites with 0 reads vs 1 read of the  
 493 appropriate allele. We then re-normalized the relative frequencies of strains such that

494  $f_{i,t} + f_{j,t} = 1$ . Deviations from this normalization can be caused by sequencing noise as well as  
495 the presence of other low frequency strains. To be conservative, we eliminated all timepoints  
496 where  $f_{i,t} + f_{j,t}$  deviated from 1 by more than 10% (Fig. S9C), which is larger than what would be  
497 expected from sampling noise alone.

498 We assessed the uncertainty in our frequency estimates by bootstrapping sets of marker SNVs.  
499 We generated 1000 bootstrapped replicates for each strain frequency and estimated the 95%  
500 confidence intervals using the 2.5th and 97.5th percentiles of the bootstrapped distribution. We  
501 used these estimates to plot the confidence intervals in Fig. 2E.

## 502 Validation on simulated metagenomes

503 To evaluate the limits of our strain tracking approach, we applied our frequency estimation  
504 pipeline to simulated metagenomic datasets designed to mimic the conditions of our experiment.  
505 For each simulated metagenome, we used the read simulator InSilicoSeq v2.0.1 (90) to generate  
506 ~7 million paired end reads under the NovaSeq error model. Simulated metagenomes were  
507 constructed using genomes from the UHGG database (5) using the observed species abundances  
508 from two different coalescence experiments, which both contained a population of *Bacteroides*  
509 *thetaiotaomicron* (*Bt*). For all species except *Bt*, we generated synthetic reads from the MIDAS2  
510 reference genome (66) in proportion to their observed frequency in the sample. For the *Bt*  
511 populations, we simulated multi-strain population by generating a mixture of reads from the *Bt*  
512 reference genome and another *Bt* genome from the UHGG database (GUT\_GENOME000472 or  
513 GUT\_GENOME001637), which were chosen to span a range of genetic distances from the  
514 reference. We simulated a range of mixture ratios from 1:10<sup>5</sup> to 10<sup>5</sup>:1 while varying the coverage  
515 of the *Bt* population from 5x-30x (by altering the relative abundance of *Bt* and re-normalizing the  
516 abundances of the other species). We then estimated the frequencies of the non-reference *Bt*  
517 strain using the procedure described above (Fig. S10).

518 As expected, the accuracy of the frequency estimates strongly depended on the corresponding  
519 number of marker SNVs. For the more diverged *Bt* strain, we typically identified around 20,000  
520 marker SNVs, and could accurately detect the strain at frequencies as low as 10<sup>-5</sup> when the *Bt*  
521 coverage was at least 5x. By contrast, the less-diverged *Bt* strain typically yielded around 200  
522 marker SNVs, and the resolution of our frequency estimates plateaued around ~10<sup>-3</sup> for the  
523 lowest depth samples. We used these simulation results to establish the thresholds for our  
524 frequency estimation pipeline. In particular, we required each species to have a total coverage of  
525 at least 5x to perform strain frequency inference, and set our strain detection floor to  $f_{\min} \equiv 10^{-3}$ .

## 526 Quantifying strain-level selection coefficients

527 We used the frequency estimates above to compute the relative fitness between pairs of strains  
528 using the formula in Eq. (1). For a given pair of timepoints  $t_1$  and  $t_2$ , we estimated the average  
529 selection coefficient within that window using the formula

$$s_{t_1, t_2} = \frac{1}{t_2 - t_1} \log \left( \frac{f_{t_2}}{1 - f_{t_2}} \cdot \frac{1 - f_{t_1}}{f_{t_1}} \right) \quad (5)$$

530 Since this equation is undefined for  $f_t = 0$  or  $f_t = 1$ , we capped the strain frequencies at our  
531 estimated detection limit of 10<sup>-3</sup> to avoid arithmetic overflow.

532 We used these estimates to test for time-varying selection by examining whether selection at an earlier epoch ( $t_{early,1}, t_{early,2}$ ) was predictive of dynamics at later timepoints ( $t_{late,1}, t_{late,2}$ ). To minimize statistical artifacts such as regression to the mean (91), we used a bootstrapping approach to assess statistical significance by dividing the corresponding marker SNVs into several independent groups.

537 For each bootstrap replicate  $b$ , we randomly assigned 1/4 of the marker SNVs to each of the four timepoints ( $t_{early,1}, t_{early,2}, t_{late,1}, t_{late,2}$ ) by sampling without replacement, and computed a set of four independent frequency estimates,  $f_{t_{early,1}}^b, f_{t_{early,2}}^b, f_{t_{late,1}}^b$ , and  $f_{t_{late,2}}^b$ . We then used these independent frequency estimates to compute an estimated selection coefficient for both epochs, which were used to plot the data points in Fig. 3E.

542 We also used the estimated selection coefficient from the first epoch to estimate an extrapolated frequency at later timepoints under a null model of constant selection:

$$f_{t,pred}^b = \frac{f_{t_{early,2}}^b e^{s_{t_{early,1}}^b - t_{early,2}(t - t_{early,2})}}{1 - f_{t_{early,2}}^b + f_{t_{early,2}}^i e^{s_{t_{early,1}}^b - t_{early,2}(t - t_{early,2})}} \quad (6)$$

544 We used this constant-selection null model to test whether the shifts in  $s(t)$  tended to be biased in the direction of coexistence. To do so, we used the extrapolated trajectory in Eq. (6) to find the latest timepoint ( $t^*$ ) where the expected strain frequency was above a minimum threshold where it could still be reliably observed. We set this threshold to  $f^* = 10^{-2}$ , so that it was about 10-fold higher than our estimated limit of detection of from our *in silico* simulations ( $f_{min} \sim 10^{-3}$ ). We then compared how the observed strain diversity at  $t^*$ , measured by the heterozygosity  $f_{t^*}^i(1 - f_{t^*}^i)$ , compared to the expected value for the constant selection prediction  $f_{t^*,pred}^i(1 - f_{t^*,pred}^i)$ , as illustrated in Fig. 3F. Points above this line indicate that the deviations from constant selection led to a higher value of intra-species diversity, while points below the line imply the opposite. Across all species, we observed an excess of points above the constant selection line ( $n = 46/61$ ; Fig. 3F). We assessed the significance of this trend by comparing our data to a null model in which diversity is equally likely to be above or below what is predicted under constant selection. In other words we compared the fraction of competitions with elevated diversity to a model where the likelihood of having elevated diversity was 1/2. We found that the number of competitions in which we observed an excess of points above the line significantly exceeded the null expectation ( $p < .001$ )

## 560 Quantifying signatures of non-transitive selection

561 We used a similar approach to evaluate the transitivity of strain-level selection coefficients between different pairs of strains. We let  $s_{A|B}(t)$  denote the (possibly time-varying) selection coefficient of the strain from host  $A$  when competing against strain  $B$  (Eq. 1). This time-varying selection coefficient could be driven by biotic factors that are common to all community collisions, as well as idiosyncratic features that are specific to the community context of one collision (including specific interactions between strains  $A$  and  $B$ ). To distinguish between these scenarios, we can consider competitions with a third strain  $C$ , which yield an additional pair of selection coefficients,  $s_{A|C}(t)$  and  $s_{B|C}(t)$ . If the relative fitness only depends on properties of the biotic environment that are common to all three collisions, then the selection coefficients must be

570 transitive:

$$s_{A|C}(t) = s_{A|B}(t) + s_{B|C}(t). \quad (7)$$

571 Combining this expression with Eq. (1) yields a corresponding relationship between the three  
572 relative frequencies,

$$\log\left(\frac{f_{A|C}(t)}{1 - f_{A|C}(t)}\right) = \log\left(\frac{f_{A|B}(t)}{1 - f_{A|B}(t)}\right) + \log\left(\frac{f_{B|C}(t)}{1 - f_{B|C}(t)}\right) + c. \quad (8)$$

573 where  $c$  is a constant of integration. We used this expression along with observed trajectories  
574  $f_{A|B}(t)$ ,  $f_{B|C}(t)$ , and  $f_{A|C}(t)$  to test for evidence of non-transitivity.

575 To do so, we chose the integration constant  $c$  to best fit the observed trajectories at timepoints 0  
576 and 1. This yields a parameter free prediction for the remaining timepoints, which are shown in  
577 Fig. 5A-C. Confidence intervals for these predictions were estimated by calculating upper and  
578 lower bounds for  $c$  by bootstrapping replicate competitions.

### 579 Testing for strain-level cohesion during community coalescence

580 We tested for evidence of strain-level community cohesion by quantifying the correlations  
581 between strain-level competition outcomes across multiple species in a given community  
582 collision. Strong correlations would correspond to scenarios where strains from a given host tend  
583 to win or lose as a group. To quantify these correlations, we calculated the net selection  
584 coefficient between timepoints 0 and 7 for all of the focal species in a given collision (Fig. 5D) and  
585 defined the winning strain as the one whose median selection coefficient across replicates was  
586 positive. We quantified the bias in these outcomes using a standard “G-statistic” metric,

$$G_{ij} = n_i \log(2n_i/n) + n_j \log(2n_j/n), \quad (9)$$

587 where  $n_i$  and  $n_j$  denote the numbers of winning strains from parent communities  $i$  and  $j$ , We  
588 respectively, and  $n = n_a + n_b$  is the total number of assessed species (Fig. 5E, Fig. S21). We  
589 assessed the significance of the observed bias scores by generating bootstrapped distributions of  
590  $G$  statistics by drawing  $n_i$  from a binomial distribution with success probability  $p = 1/2$ , using  
591 the observed values of  $n$ . We found that the median G value across communities was not  
592 significantly elevated over the null expectation ( $p = .490$ , Fig. S21).

593 To increase statistical power, we expanded this analysis to incorporate data from initially  
594 multi-colonized populations. We did so using a similar approach that we used to track strains  
595 from mono-colonized samples above. However, in this case the sets of distinguishing SNPs we  
596 infer are markers for the entire population, rather than for an individual haplotype. We then  
597 estimated the frequencies of each initial population from each parent community using the same  
598 methods as described above. For this subset of populations, we manually verified the resulting  
599 SNV trajectories to ensure that SNV trajectories from the same remained correlated over  
600 subsequent timepoints in the collision, as expected for a true strain.

601 Repeating our cohesion analysis yielded

602 for this expanded set of strains yielded similar results (Fig 5D,E), suggesting that there is limited  
603 evidence for strain-level cohesion at community-wide scales.

604

605 **Data and Code Availability:** Raw sequencing data will be made available on the Sequencing  
606 Read Archive (SRA) upon publication. All scripts for data analysis and figure generation are  
607 available on Github (<https://github.com/sophiejwalton/coalescence-pilot-mgx>).

608 **Acknowledgments**

609 We thank Jean Vila, Elisa Visher, Mark Bitter, Leron Perez, Erin Mordecai, Bernard Kim, Tadashi  
610 Fukami, Ami Bhatt, Carlos Voogdt, Jamie A Lopez, and members of the Good, Petrov, Fukami,  
611 and Cremer labs for useful discussions and feedback. We also thank the anonymous donors who  
612 generously contributed their fecal samples to this project. Computational analyses were  
613 performed on the Sherlock High-Performance Computing Cluster, managed by the Stanford  
614 Research Computing Center. Sequencing was performed by the Chan Zuckerberg Biohub  
615 Genomics Core. S.J.W. acknowledges support from a Hertz Foundation Fellowship and a  
616 Stanford Graduate Fellowship. H.R.G acknowledges support from a Stanford Graduate  
617 Fellowship and NIH NIGMS award T32GM154663. K.S.X. acknowledges support from a James  
618 McDonnell Foundation Postdoctoral Fellowship in Understanding Dynamic and Multi-Scale  
619 Systems, a Jane Coffin Childs Memorial Fund Postdoctoral Fellowship, and NIH Grant No. R21  
620 AI168860. D.A.P. acknowledges support from NIH NIGMS Grant No. 5R35GM118165-07. B.H.G  
621 acknowledges support from NIH NIGMS Grant No. R35GM146949. D.A.P. and B.H.G. are Chan  
622 Zuckerberg Biohub Investigators.

623

624 **Author contributions:** S.J.W., J.C., K.S.X., D.A.P. and B.H.G. conceived the project and designed  
625 experiments; S.J.W., R.S., H.R.G, and C.Y. performed experiments; S.J.W., Q.X., D.A.P. and B.H.G  
626 analyzed data; S.J.W., D.A.P., and B.H.G. wrote the initial manuscript draft, which all authors  
627 reviewed.

628

629 **Competing interests:** The authors declare no competing interests.

630 **References**

- 631 [1] Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al.  
632 Structure, function and diversity of the healthy human microbiome. *Nature*.  
633 2012;486(7402):207–214. doi:10.1038/nature11234.
- 634 [2] Jami E, Israel A, Kotser A, Mizrahi I. Exploring the bovine rumen bacterial community from  
635 birth to adulthood. *The ISME journal*. 2013;7(6):1069–1079.
- 636 [3] Bittleston LS, Wolock CJ, Yahya BE, Chan XY, Chan KG, Pierce NE, et al. Convergence  
637 between the microcosms of Southeast Asian and North American pitcher plants. *Elife*.  
638 2018;7:e36741.
- 639 [4] Vorholt JA. Microbial life in the phyllosphere. *Nature reviews microbiology*.  
640 2012;10(12):828–840.
- 641 [5] Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of  
642 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology*;39(1):105–114. doi:10.1038/s41587-020-0603-3.
- 644 [6] Crits-Christoph A, Olm MR, Diamond S, Bouma-Gregson K, Banfield JF. Soil bacterial  
645 populations are shaped by recombination and gene-specific selection across a grassland  
646 meadow. *The ISME journal*. 2020;14(7):1834–1846.
- 647 [7] Rohwer RR, Kirkpatrick M, Garcia SL, Kellom M, McMahon KD, Baker BJ. Two decades of  
648 bacterial ecology and evolution in a freshwater lake. *Nature Microbiology*.  
649 2025;10(1):246–257.
- 650 [8] Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, et al. Functional overlap of  
651 the *Arabidopsis* leaf and root microbiota. *Nature*. 2015;528(7582):364–369.  
652 doi:10.1038/nature16192.
- 653 [9] Garud NR, Good BH, Hallatschek O, Pollard KS. Evolutionary dynamics of bacteria in the  
654 gut microbiome within and across hosts. *PLOS Biology*. 2019;17(1):1–29.  
655 doi:10.1371/journal.pbio.3000102.
- 656 [10] Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF.  
657 inStrain profiles population microdiversity from metagenomic data and sensitively detects  
658 shared microbial strains. *Nature Biotechnology*;39(6):727–736.  
659 doi:10.1038/s41587-020-00797-0.
- 660 [11] Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, et al. The Prevotella  
661 copri complex comprises four distinct clades underrepresented in westernized populations.  
662 *Cell host & microbe*. 2019;26(5):666–679.
- 663 [12] Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, et al. Adaptive  
664 Evolution within Gut Microbiomes of Healthy People. *Cell Host & Microbe*.  
665 2019;25(5):656–667.e8. doi:<https://doi.org/10.1016/j.chom.2019.03.007>.
- 666 [13] Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across  
667 human gut microbiome species. *Cell*. 2015;160(4):583–594.

- 668 [14] Suzuki TA, Fitzstevens JL, Schmidt VT, Enav H, Huus KE, Ngwese MM, et al.  
669 Codiversification of gut microbiota with humans. *Science*. 2022;377(6612):1328–1332.  
670 doi:10.1126/science.abm7759.
- 671 [15] Park SY, Rao C, Coyte KZ, Kuziel GA, Zhang Y, Huang W, et al. Strain-level fitness in the  
672 gut microbiome is an emergent property of glycans and a single metabolite. *Cell*.  
673 2022;185(3):513–529.
- 674 [16] Chatzidaki-Livanis M, Geva-Zatorsky N, Comstock LE. *< i>Bacteroides fragilis</i>* type VI  
675 secretion systems use novel effector and immunity proteins to antagonize human gut  
676 Bacteroidales species. *Proceedings of the National Academy of Sciences*.  
677 2016;113(13):3627–3632. doi:10.1073/pnas.1522510113.
- 678 [17] Jiang X, Hall AB, Arthur TD, Plichta DR, Covington CT, Poyet M, et al. Invertible promoters  
679 mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut.  
680 *Science*. 2019;363(6423):181–187.
- 681 [18] Sorbara MT, Littmann ER, Fontana E, Moody TU, Kohout CE, Gjonbalaj M, et al. Functional  
682 and genomic variation between human-derived isolates of Lachnospiraceae reveals  
683 inter-and intra-species diversity. *Cell host & microbe*. 2020;28(1):134–146.
- 684 [19] Pudlo NA, Urs K, Crawford R, Pirani A, Atherly T, Jimenez R, et al. Phenotypic and  
685 genomic diversification in complex carbohydrate-degrading human gut bacteria.  
686 *mSystems*. 2022;7(1):e00947–21.
- 687 [20] Olm MR, Crits-Christoph A, Diamond S, Lavy A, Carnevali PBM, Banfield JF. Consistent  
688 Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. *mSystems*.  
689 2020;5(1):10.1128/msystems.00731–19. doi:10.1128/msystems.00731-19.
- 690 [21] Liu Z, Good BH. Dynamics of bacterial recombination in the human gut microbiome. *PLOS  
691 Biology*. 2024;22(2):1–22. doi:10.1371/journal.pbio.3002472.
- 692 [22] Ochman H, Lerat E, Daubin V. Examining bacterial species under the specter of gene  
693 transfer and exchange. *Proceedings of the National Academy of Sciences*.  
694 2005;102(suppl\_1):6595–6599. doi:10.1073/pnas.0502035102.
- 695 [23] Patel AM, Ochman H. Recombination and the Species Structure of the Genus *Bacteroides*.  
696 *bioRxiv*. 2025;doi:10.1101/2025.05.07.652691.
- 697 [24] Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global  
698 network of gene exchange connecting the human microbiome. *Nature*;480(7376):241–244.  
699 doi:10.1038/nature10571.
- 700 [25] Xue KS, Walton SJ, Goldman DA, Morrison ML, Verster AJ, Parrott AB, et al. Prolonged  
701 delays in human microbiota transmission after a controlled antibiotic perturbation. *bioRxiv*.  
702 2023;doi:10.1101/2023.09.26.559480.
- 703 [26] Roodgar M, Good BH, Garud NR, Martis S, Avula M, Zhou W, et al. Longitudinal  
704 linked-read sequencing reveals ecological and evolutionary responses of a human gut  
705 microbiome during antibiotic treatment. *Genome Research*;31(8):1433–1446.  
706 doi:10.1101/gr.265058.120.

- 707 [27] Smillie CS, Sauk J, Gevers D, Friedman J, Sung J, Youngster I, et al. Strain Tracking Reveals  
708 the Determinants of Bacterial Engraftment in the Human Gut Following Fecal Microbiota  
709 Transplantation. *Cell Host Microbe*. 2018;23(2):229–240.e5.  
710 doi:<https://doi.org/10.1016/j.chom.2018.01.003>.
- 711 [28] Wolff R, Shoemaker W, Garud N. Ecological Stability Emerges at the Level of Strains in the  
712 Human Gut Microbiome;14(2):e02502–22. doi:10.1128/mbio.02502-22.
- 713 [29] Xiao X, Singh A, Giometto A, Brito IL. Segatella clades adopt distinct roles within a single  
714 individual's gut. *npj Biofilms and Microbiomes*. 2024;10(1):114.
- 715 [30] Chen DW, Garud NR. Rapid evolution and strain turnover in the infant gut microbiome.  
716 *Genome Research*;32(6):1124–1136. doi:10.1101/gr.276306.121.
- 717 [31] Reyes A, Wu M, McNulty NP, Rohwer FL, Gordon JI. Gnotobiotic mouse model of  
718 phage–bacterial host dynamics in the human gut. *Proceedings of the National Academy of  
719 Sciences*. 2013;110(50):20236–20241.
- 720 [32] Shen J, Zhang J, Mo L, Li Y, Li Y, Li C, et al. Large-scale phage cultivation for commensal  
721 human gut bacteria. *Cell Host & Microbe*. 2023;31(4):665–677.
- 722 [33] Gravel D, Canham CD, Beaudet M, Messier C. Reconciling niche and neutrality: the  
723 continuum hypothesis. *Ecology Letters*. 2006;9(4):399–409.  
724 doi:<https://doi.org/10.1111/j.1461-0248.2006.00884.x>.
- 725 [34] Scheffer M, van Nes EH. Self-organized similarity, the evolutionary emergence of groups of  
726 similar species. *Proceedings of the National Academy of Sciences*. 2006;103(16):6230–6235.  
727 doi:10.1073/pnas.0508024103.
- 728 [35] Holt RD. Emergent neutrality. *Trends in ecology & evolution*. 2006;21(10):531–533.
- 729 [36] D'Andrea R, Gibbs T, O'Dwyer JP. Emergent neutrality in consumer-resource dynamics.  
730 *PLOS Computational Biology*. 2020;16(7):1–17. doi:10.1371/journal.pcbi.1008102.
- 731 [37] Hubbell SP. Neutral theory and the evolution of ecological equivalence. *Ecology*.  
732 2006;87(6):1387–1398.
- 733 [38] Posfai A, Taillefumier T, Wingreen NS. Metabolic Trade-Offs Promote Diversity in a Model  
734 Ecosystem. *Phys Rev Lett*. 2017;118:028103. doi:10.1103/PhysRevLett.118.028103.
- 735 [39] McEnany J, Good BH. Predicting the first steps of evolution in randomly assembled  
736 communities. *Nature Communications*;15(1):8495. doi:10.1038/s41467-024-52467-3.
- 737 [40] Goldman DA, Xue KS, Parrott AB, Lopez JA, Vila JCC, Jeeda RR, et al. Competition for  
738 shared resources increases dependence on initial population size during coalescence of gut  
739 microbial communities. *Proceedings of the National Academy of Sciences*.  
740 2025;122(11):e2322440122. doi:10.1073/pnas.2322440122.
- 741 [41] Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, et al. Durable coexistence of  
742 donor and recipient strains after fecal microbiota transplantation. *Science*.  
743 2016;352(6285):586–589. doi:10.1126/science.aad8852.

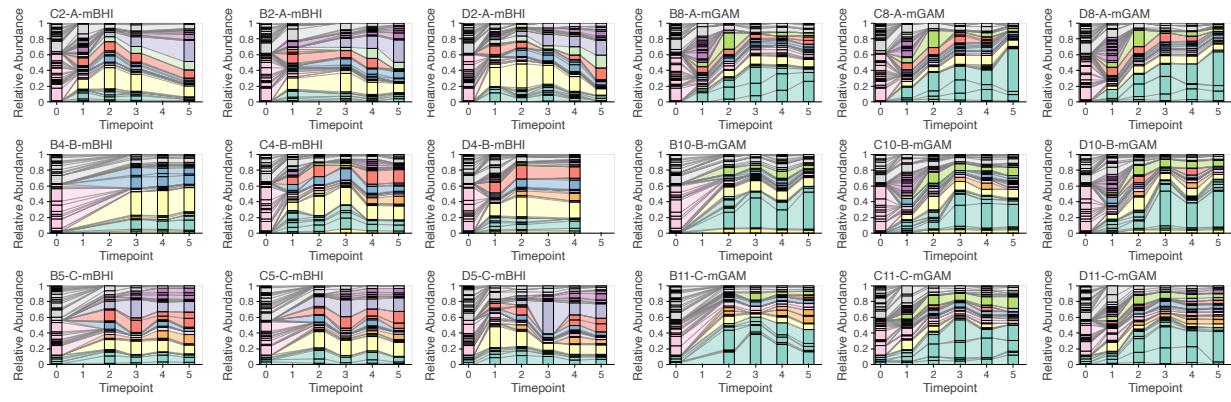
- 744 [42] Morel-Journel T, Lehtinen S, Cotto O, Amia R, Dion S, Figueroa C, et al.  
745      Residence-colonization trade-off and niche differentiation enable coexistence of Escherichia  
746      coli phylogroups in healthy humans. *The ISME journal*. 2025;19(1):wraf089.
- 747 [43] Goyal A, Bittleston LS, Leventhal GE, Lu L, Cordero OX. Interactions between strains  
748      govern the eco-evolutionary dynamics of microbial communities. *eLife*. 2022;11:e74987.  
749      doi:10.7554/eLife.74987.
- 750 [44] Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, et al. *Arabidopsis*  
751      thaliana and *Pseudomonas* pathogens exhibit stable associations over evolutionary  
752      timescales. *Cell host & microbe*. 2018;24(1):168–179.
- 753 [45] Bobay LM, Ochman H. Biological species are universal across life's domains. *Genome*  
754      biology and evolution. 2017;9(3):491–501.
- 755 [46] MacArthur R, Levins R. The Limiting Similarity, Convergence, and Divergence of Coexisting  
756      Species. *The American Naturalist*. 1967;101(921):377–385. doi:10.1086/282505.
- 757 [47] MacArthur R. Species Packing, and What Competition Minimizes. *Proceedings of the*  
758      *National Academy of Sciences*. 1969;64(4):1369–1371. doi:10.1073/pnas.64.4.1369.
- 759 [48] May RM, Arthur RHM. Niche Overlap as a Function of Environmental Variability.  
760      *Proceedings of the National Academy of Sciences*. 1972;69(5):1109–1113.  
761      doi:10.1073/pnas.69.5.1109.
- 762 [49] May RM. On the theory of niche overlap. *Theoretical Population Biology*. 1974;5(3):297–332.  
763      doi:[https://doi.org/10.1016/0040-5809\(74\)90055-0](https://doi.org/10.1016/0040-5809(74)90055-0).
- 764 [50] Abrams P. The theory of limiting similarity. *Annual review of ecology and systematics*.  
765      1983;14:359–376.
- 766 [51] Kearney SM, Gibbons SM, Erdman SE, Alm EJ. Orthogonal dietary niche enables reversible  
767      engraftment of a gut bacterial commensal. *Cell reports*. 2018;24(7):1842–1851.
- 768 [52] Maldonado-Gómez M, Martínez I, Bottacini F, O'Callaghan A, Ventura M, van't Sinderen D,  
769      et al. Stable Engraftment of *Bifidobacterium longum* AH1206 in the Human Gut Depends  
770      on Individualized Features of the Resident Microbiome. *Cell Host & Microbe*;20(4):515–526.  
771      doi:10.1016/j.chom.2016.09.001.
- 772 [53] Whitaker WR, Russ ZN, Stanley Shepherd E, Popov LM, Louie A, Lam K, et al. Controlled  
773      colonization of the human gut with a genetically engineered microbial therapeutic. *Science*.  
774      2025;389(6757):303–308.
- 775 [54] Northen TR, Kleiner M, Torres M, Kovács ÁT, Nicolaisen MH, Krzyżanowska DM, et al.  
776      Community standards and future opportunities for synthetic communities in  
777      plant-microbiota research. *Nature microbiology*. 2024;9(11):2774–2784.
- 778 [55] Mehlferber EC, Debray R, Conover AE, Sherman JK, Kaulbach G, Reed R, et al.  
779      Phyllosphere microbial associations improve plant reproductive success. *Frontiers in Plant*  
780      *Science*. 2023;14:1273330.
- 781 [56] Valles-Colomer M, Blanco-Míguez A, Manghi P, Asnicar F, Dubois L, Golzato D, et al. The  
782      person-to-person transmission landscape of the gut and oral microbiomes. *Nature*.  
783      2023;614(7946):125–135.

- 784 [57] Siranosian BA, Brooks EF, Andermann T, Rezvani AR, Banaei N, Tang H, et al. Rare  
785 transmission of commensal and pathogenic bacteria in the gut microbiome of hospitalized  
786 adults. *Nature Communications*;13(1):586. doi:10.1038/s41467-022-28048-7.
- 787 [58] Lou YC, Olm MR, Diamond S, Crits-Christoph A, Firek BA, Baker R, et al. Infant gut strain  
788 persistence is associated with maternal origin, phylogeny, and traits including surface  
789 adhesion and iron acquisition. *Cell Reports Medicine*. 2021;2(9).
- 790 [59] Tropini C, Earle KA, Huang KC, Sonnenburg JL. The gut microbiome: connecting spatial  
791 organization to function. *Cell host & microbe*. 2017;21(4):433–442.
- 792 [60] Ghosh OM, Good BH. Emergent evolutionary forces in spatial models of luminal growth  
793 and their application to the human gut microbiota. *Proceedings of the National Academy of  
794 Sciences*. 2022;119(28). doi:10.1073/pnas.2114931119.
- 795 [61] Karasov TL, Neumann M, Leventhal L, Symeonidi E, Shirsekar G, Hawks A, et al.  
796 Continental-scale associations of *Arabidopsis thaliana* phyllosphere members with host  
797 genotype and drought. *Nature Microbiology*. 2024;9(10):2748–2758.
- 798 [62] Aranda-Díaz A, Willis L, Nguyen TH, Ho PY, Vila J, Thomsen T, et al. Assembly of  
799 stool-derived bacterial communities follows “early-bird” resource utilization dynamics. *Cell  
800 Systems*. 2025;16(4).
- 801 [63] Aranda-Díaz A, Ng KM, Thomsen T, Real-Ramírez I, Dahan D, Dittmar S, et al.  
802 Establishment and characterization of stable, diverse, fecal-derived *in vitro* microbial  
803 communities that model the intestinal microbiota. *Cell Host Microbe*. 2022;30(2):260–272.e5.  
804 doi:<https://doi.org/10.1016/j.chom.2021.12.008>.
- 805 [64] Celis AI, Relman DA, Huang KC. The impact of iron and heme availability on the healthy  
806 human gut microbiome *in vivo* and *in vitro*. *Cell Chemical Biology*. 2023;30(1):110–126.e3.  
807 doi:<https://doi.org/10.1016/j.chembiol.2022.12.001>.
- 808 [65] Javdan B, Lopez JG, Chankhamjon P, Lee YCJ, Hull R, Wu Q, et al. Personalized Mapping of  
809 Drug Metabolism by the Human Gut Microbiome. *Cell*. 2020;181(7):1661–1679.e22.  
810 doi:<https://doi.org/10.1016/j.cell.2020.05.001>.
- 811 [66] Zhao C, Dimitrov B, Goldman M, Nayfach S, Pollard KS. MIDAS2: Metagenomic  
812 Intra-species Diversity Analysis System. *Bioinformatics*;39(1):btac713.  
813 doi:10.1093/bioinformatics/btac713.
- 814 [67] Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population  
815 structure and genetic diversity from metagenomes. *Genome research*. 2017;27(4):626–638.
- 816 [68] Merritt J, Kuehn S. When communities collide. *eLife*. 2016;5:e18753.
- 817 [69] Gilpin M. Community-level competition: asymmetrical dominance. *Proceedings of the  
818 National Academy of Sciences*. 1994;91(8):3252–3254. doi:10.1073/pnas.91.8.3252.
- 819 [70] Tikhonov M. Community-level cohesion without cooperation. *eLife*. 2016;5:e15747.  
820 doi:10.7554/eLife.15747.
- 821 [71] Diaz-Colunga J, Lu N, Sanchez-Gorostiaga A, Chang CY, Cai HS, Goldford JE, et al.  
822 Top-down and bottom-up cohesiveness in microbial community coalescence. *Proceedings  
823 of the National Academy of Sciences*;119(6):e2111261119. doi:10.1073/pnas.2111261119.

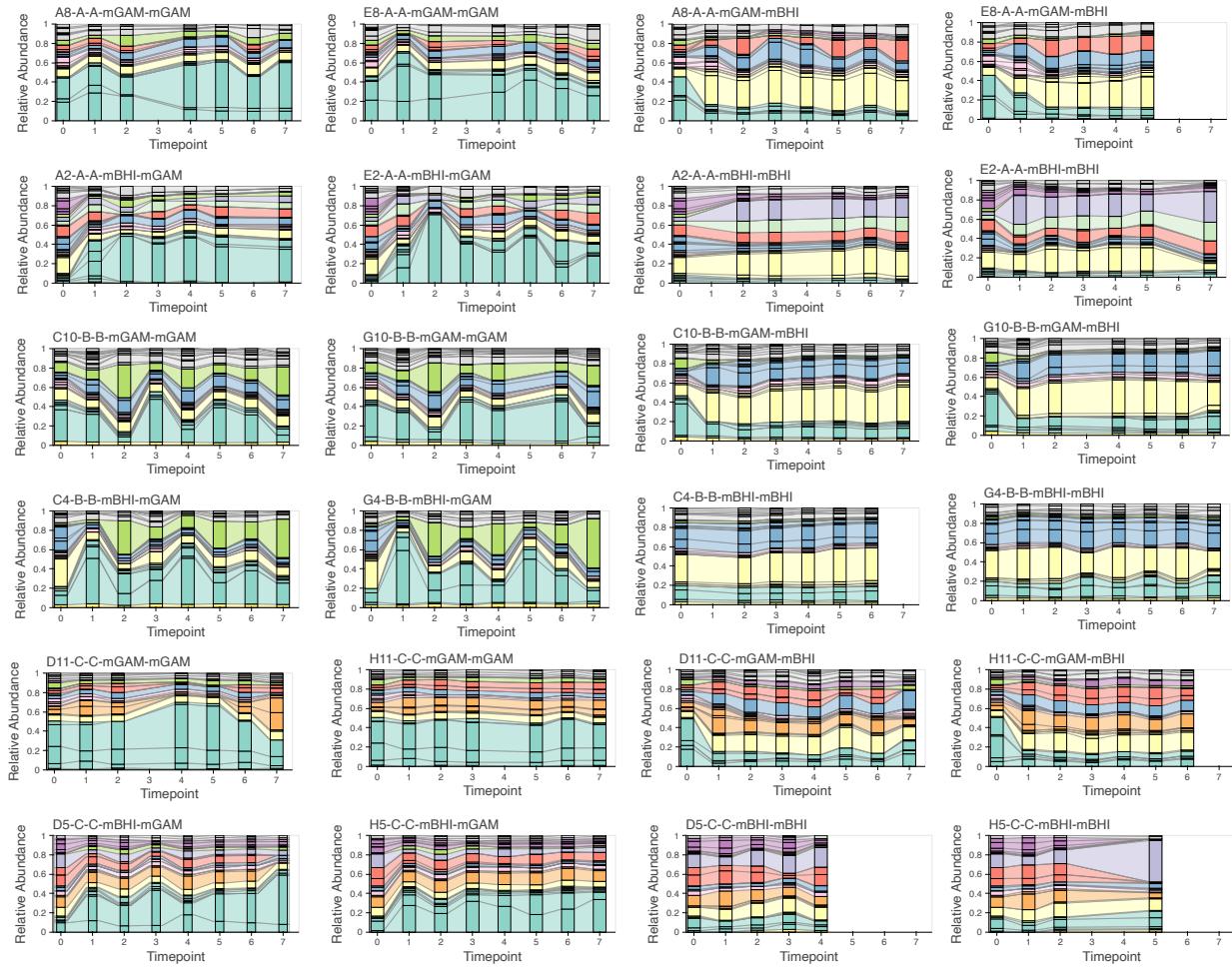
- 824 [72] Morella NM, Weng FCH, Joubert PM, Metcalf CJE, Lindow S, Koskella B. Successive  
825 passaging of a plant-associated microbiome reveals robust habitat and host  
826 genotype-dependent selection. *Proc Natl Acad Sci U S A*. 2020;117(2):1148–1159.
- 827 [73] Meyer KM, Muscettola IE, Vasconcelos ALS, Sherman JK, Metcalf CJE, Lindow SE, et al.  
828 Conspecific versus heterospecific transmission shapes host specialization of the  
829 phyllosphere microbiome. *Cell Host Microbe*. 2023;31(12):2067–2079.e5.  
830 doi:<https://doi.org/10.1016/j.chom.2023.11.002>.
- 831 [74] Sierociński P, Milferstedt K, Bayer F, Grofkopf T, Alston M, Bastkowski S, et al. A Single  
832 Community Dominates Structure and Function of a Mixture of Multiple Methanogenic  
833 Communities. *Current Biology*;27(21):3390–3395.e4. doi:10.1016/j.cub.2017.09.056.
- 834 [75] Chesson P. Quantifying and testing coexistence mechanisms arising from recruitment  
835 fluctuations. *Theoretical Population Biology*. 2003;64(3):345–357.  
836 doi:[https://doi.org/10.1016/S0040-5809\(03\)00095-9](https://doi.org/10.1016/S0040-5809(03)00095-9).
- 837 [76] Chesson P. Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology,  
838 Evolution, and Systematics*. 2000;31(Volume 31, 2000):343–366.  
839 doi:<https://doi.org/10.1146/annurev.ecolsys.31.1.343>.
- 840 [77] Yaffe E, Dethlefsen L, Patankar AV, Gui C, Holmes S, Relman DA. Brief antibiotic use drives  
841 human gut bacteria towards low-cost resistance. *Nature*;641(8061):182–191.  
842 doi:10.1038/s41586-025-08781-x.
- 843 [78] Vasquez KS, Wong DPGH, Pedro MF, Yu FB, Jain S, Meng X, et al. High-resolution lineage  
844 tracking of within-host evolution and strain transmission in a human gut symbiont across  
845 ecological scales. *bioRxiv*. 2024;doi:10.1101/2024.02.17.580834.
- 846 [79] Good BH, Rosenfeld LB. Eco-evolutionary feedbacks in the human gut microbiome. *Nature  
847 Communications*. 2023;14(1). doi:10.1038/s41467-023-42769-3.
- 848 [80] Tikhonov M, Monasson R. Collective Phase in Resource Competition in a Highly Diverse  
849 Ecosystem. *Physical Review Letters*. 2017;118(4). doi:10.1103/physrevlett.118.048103.
- 850 [81] Feng Z, Blumenthal E, Mehta P, Goyal A. A theory of ecological invasions and its  
851 implications for eco-evolutionary dynamics. *ArXiv*. 2025; p. arXiv–2503.
- 852 [82] Cohan FM, Perry EB. A systematics for discovering the fundamental units of bacterial  
853 diversity. *Current biology*. 2007;17(10):R373–R386.
- 854 [83] Zheng W, Zhao S, Yin Y, Zhang H, Needham DM, Evans ED, et al. High-throughput,  
855 single-microbe genomics with strain resolution, applied to a human gut microbiome.  
856 *Science*. 2022;376(6597):eabm1483.
- 857 [84] Jin X, Yu FB, Yan J, Weakley AM, Dubinkina V, Meng X, et al. Culturing of a complex gut  
858 microbial community in mucin-hydrogel carriers reveals strain- and gene-associated spatial  
859 organization. *Nature Communications*;14(1):3510. doi:10.1038/s41467-023-39121-0.
- 860 [85] Zhou W, Chow Kh, Fleming E, Oh J. Selective colonization ability of human fecal microbes  
861 in different mouse gut environments. *The ISME Journal*. 2019;13(3):805–823.

- 862 [86] Wong DPGH, Good BH. Quantifying the adaptive landscape of commensal gut bacteria  
863 using high-resolution lineage tracking. *bioRxiv*. 2023;doi:10.1101/2022.05.13.491573.
- 864 [87] Jiang H, Lei R, Ding SW, Zhu S. Skewer: a fast and accurate adapter trimmer for  
865 next-generation sequencing paired-end reads. *BMC Bioinformatics*;15(1):182.  
866 doi:10.1186/1471-2105-15-182.
- 867 [88] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*;9(4):357–359. doi:10.1038/nmeth.1923.
- 869 [89] Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM. Distribution of fixed beneficial  
870 mutations and the rate of adaptation in asexual populations. *Proceedings of the National  
871 Academy of Sciences*. 2012;109(13):4950–4955. doi:10.1073/pnas.1119910109.
- 872 [90] Lelieveld SH, Maas T, Duk TCX, Gourlé H, van den Ham HJ. InSilicoSeq 2.0: Simulating  
873 realistic amplicon-based sequence reads. *bioRxiv*. 2024;doi:10.1101/2024.02.16.580469.
- 874 [91] Bitter MC, Berardi S, Oken H, Huynh A, Lappo E, Schmidt P, et al. Continuously fluctuating  
875 selection reveals fine granularity of adaptation. *Nature*;634(8033):389–396.  
876 doi:10.1038/s41586-024-07834-x.

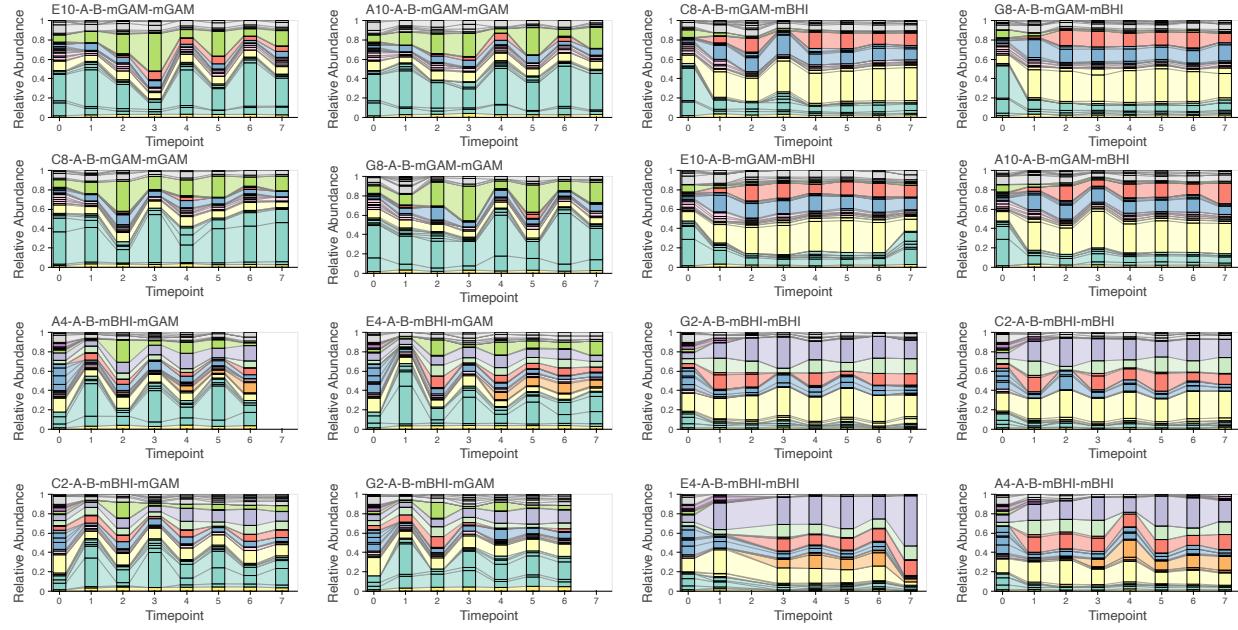
## Supplementary Figures



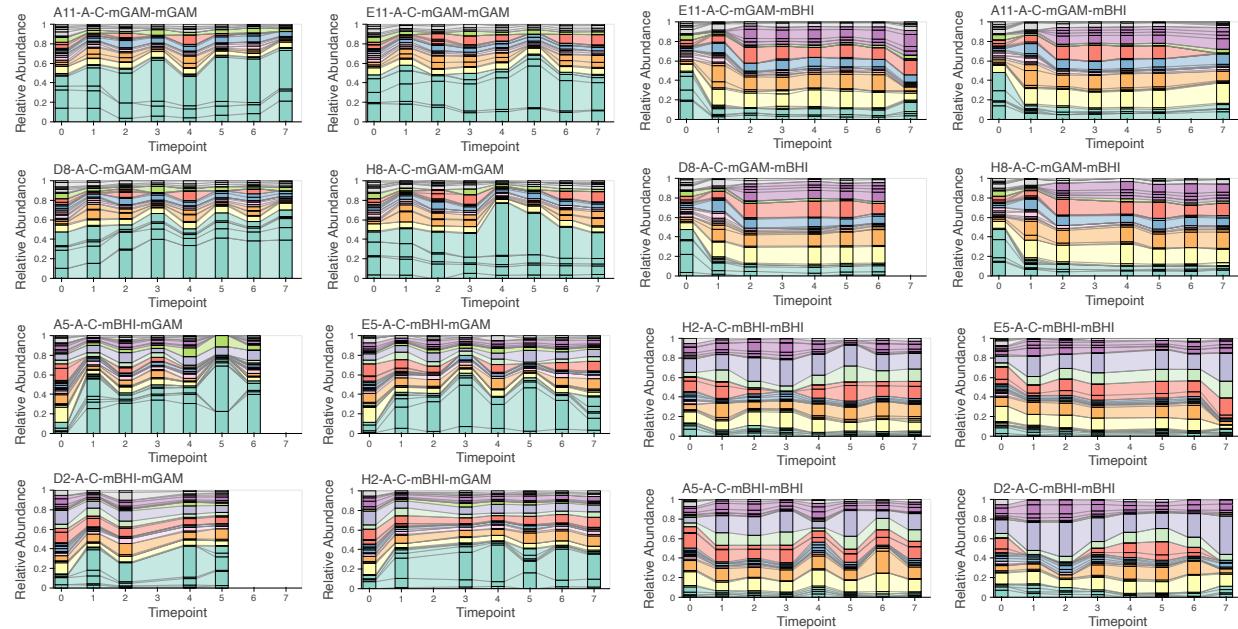
**Figure S1: Assembly species composition** Lines represent species, colors represent families. Same color scheme as in Fig 1A



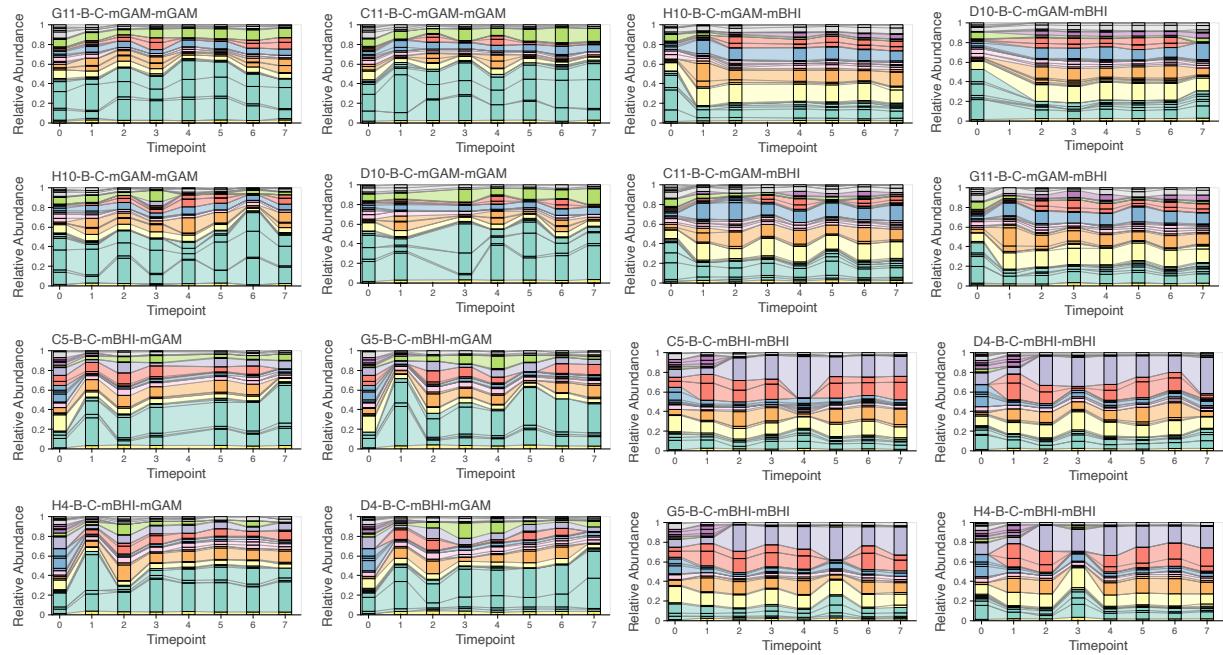
**Figure S2: Autologous controls composition** Lines represent species, colors represent families. Same color scheme as in Fig 1A



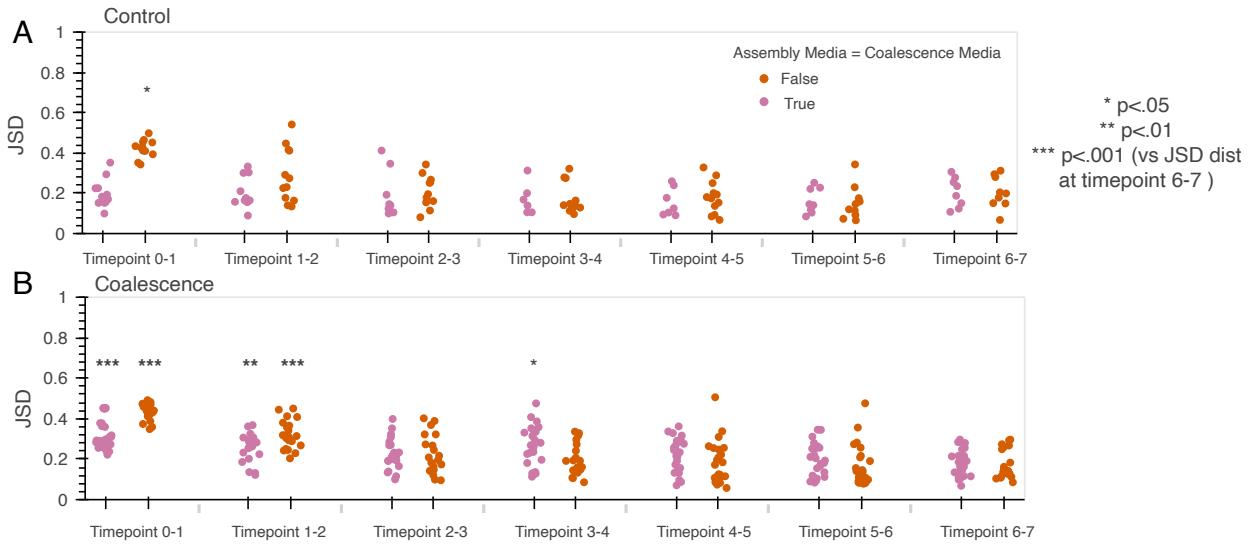
**Figure S3: A and B coalescence composition** Lines represent species, colors represent families. Same color scheme as in Fig 1A



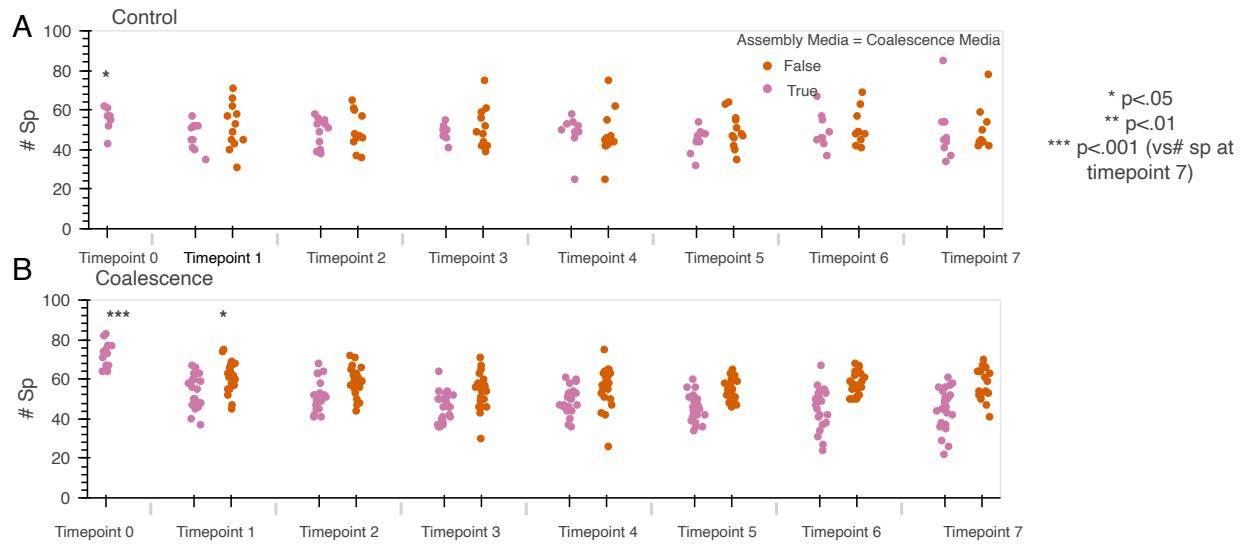
**Figure S4: A and C coalescence composition** Lines represent species, colors represent families. Same color scheme as in Fig 1A



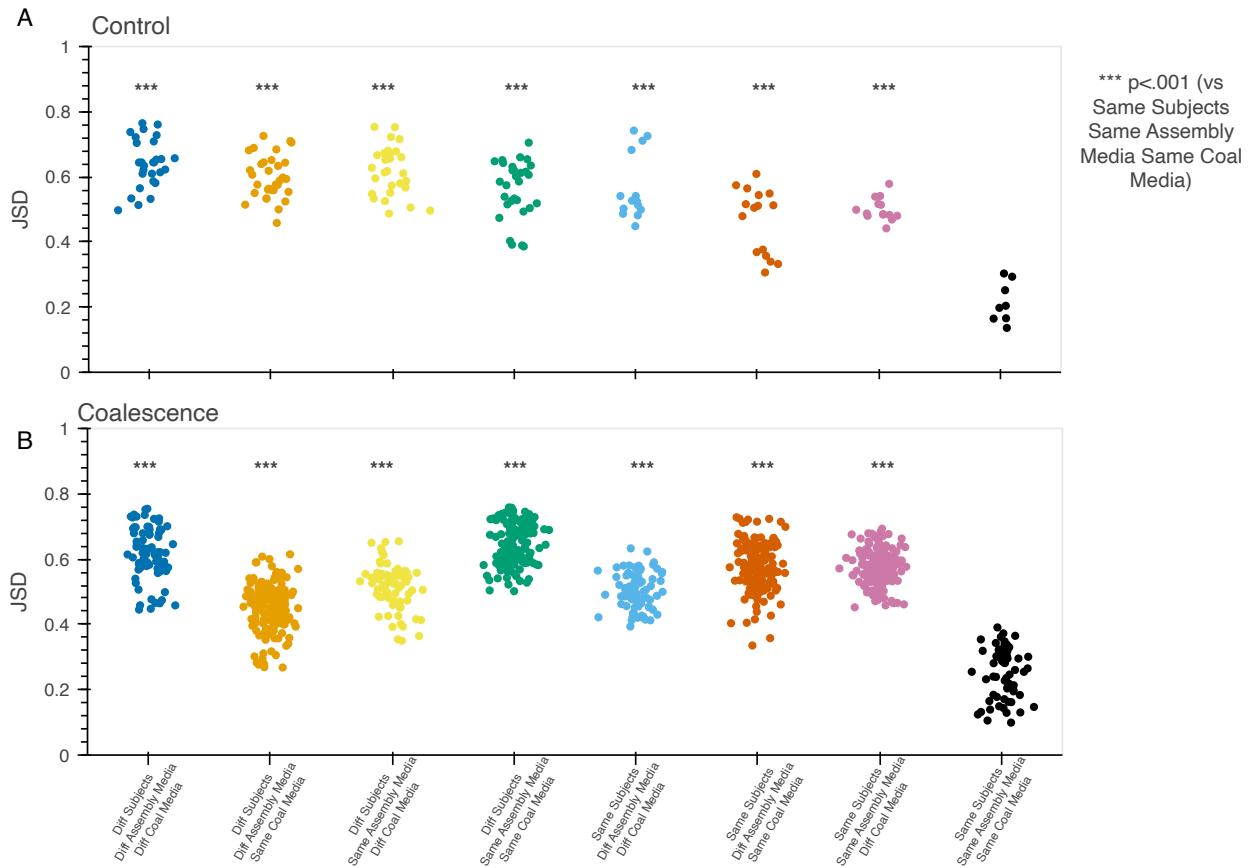
**Figure S5: B and C coalescence composition** Lines represent species, colors represent families. Same color scheme as in Fig 1A



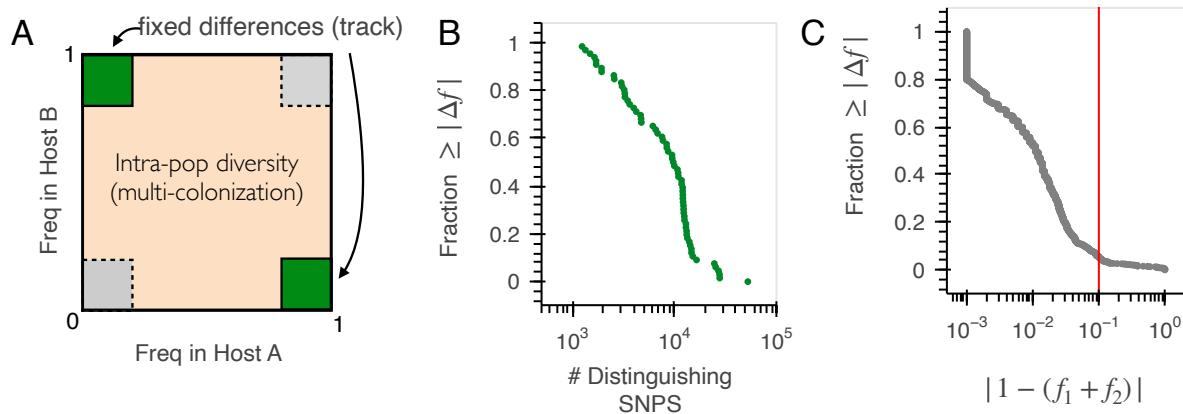
**Figure S6: Community coalescence yields stable communities at the species level (JSD quantification).** Jensen Shannon Distance between pairwise timepoints in control **A** and coalescence **B** communities. Points are colored by whether they underwent a media shift at timepoint 0. Timepoint 0 indicates the passage in which coalescence was performed, which is 7 dilution cycles after communities were derived from fecal samples. Distributions of JSD values between consecutive periods (i.e. 0-1) were compared to JSD comparisons between timepoints 6-7 to evaluate community stability (permutation test, significance threshold .05). Significant comparisons are indicated.



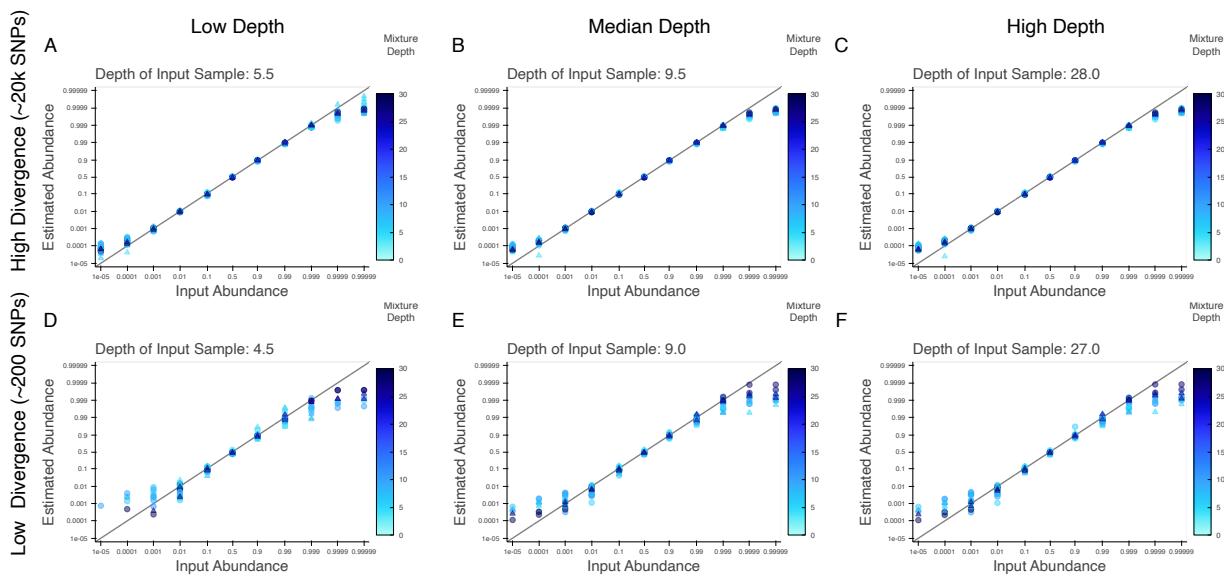
**Figure S7: Community coalescence yields stable communities at the species level (number of species quantification).** Number of species detected at relative abundance  $> .1\%$  for control A and coalescence B communities. Points are colored by whether they underwent a media shift at timepoint 0. Timepoint 0 indicates the passage in which coalescence was performed, which is 7 dilution cycles after communities were derived from fecal samples. Distributions of species diversities at a given timepoint were compared to that at timepoint 7 evaluate community stability (permutation test, significance threshold .05).



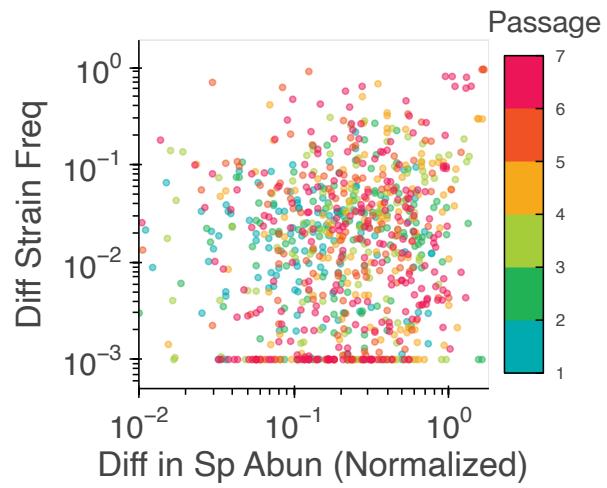
**Figure S8: Community coalescence yields reproducible communities at the species level.** Jensen shannon distance between pairs of communities at timepoint 7 in control A and coalescence B communities. For these comparisons samples with less than 20 species were discarded, for this was a sign of low depth sequencing. All comparisons to 'Same subjects Same Parent Media Same Coal Media' for both coalescence and control communities are statistically significant, illustrating that communities are reproducible, and subject and media specific ( $p < .001$ , permutation test).



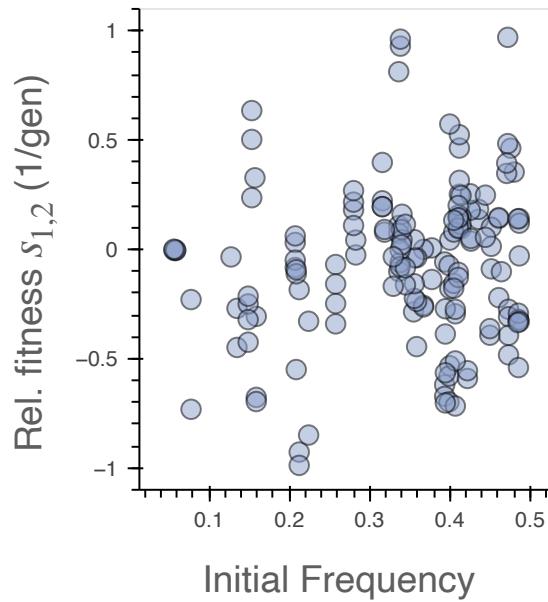
**Figure S9: Inferring sets of distinguishing SNPs to track conspecific strains in metagenomic data.** **A** Schematic of how distinguishing SNPs are inferred. SNPs in the dark green corners, which are fixed differences between samples, are used to create sets of distinguishing SNPs. **B** Distribution of the number of distinguishing SNPs for pairs of conspecific strains. **C** When inferred separately, the frequencies of strains from each parent add up to close to 1. Here is the distribution of the deviations of the frequencies of strains from both parents from 1.



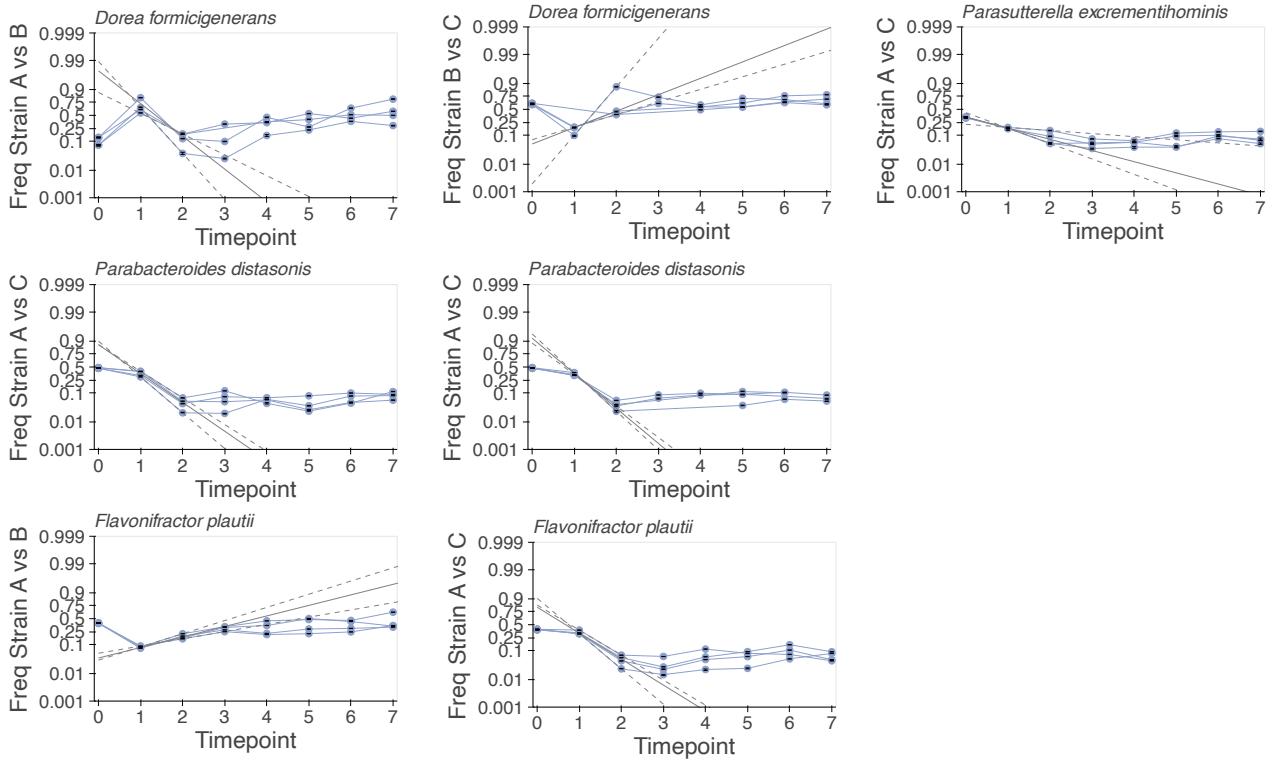
**Figure S10: Validation strain inference using simulated metagenomic data.** **A-C** Inferring the abundance of GUT\_GENOME000472 in mixtures of GUT\_GENOME000472 and the MIDAS2 reference genome for *B. theta*. **D-F** Inferring the abundance of GUT\_GENOME001637 in mixtures of GUT\_GENOME001637 and the MIDAS2 reference genome for *B. theta*. Points are colored by the depth of *B. theta* the mixture sample and differences in point shape indicate that different relative abundance profiles were used to generate samples for marker SNP inference (circle for A10-e003Coalescence-mBHI-p7, triangle for G8-e003Coalescence-mBHI-p7\_S125). Plots are titled by the depth of samples used to infer marker SNPs.



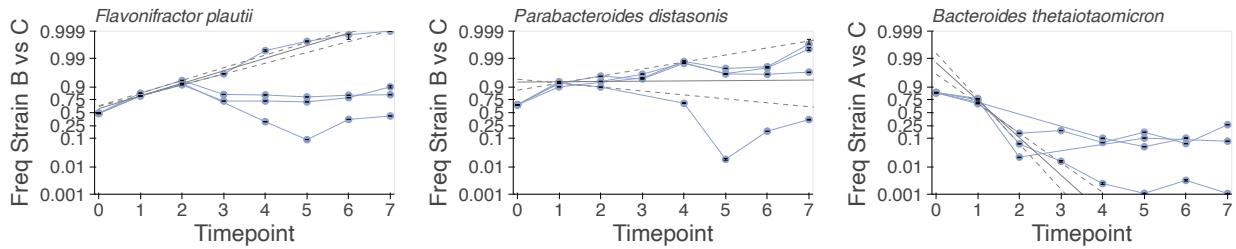
**Figure S11: Examining final strain frequencies.** Relationship between pairwise absolute difference in focal species relative abundance and strain relative abundance at timepoint 7 in replicate collisions. Differences in species abundance were normalized to the mean relative abundance of the focal species in two communities. Relationship is statistically significant (PearsonR,  $p < .001$ ,  $r^2 = .287$ ).



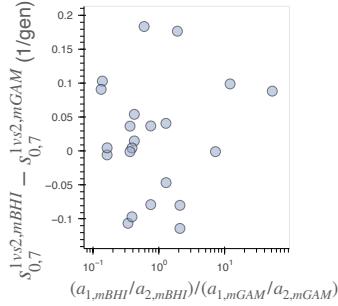
**Figure S12: Initial abundance does not predict early phase selection**  $r^2 = .113$ ,  $p = .097$



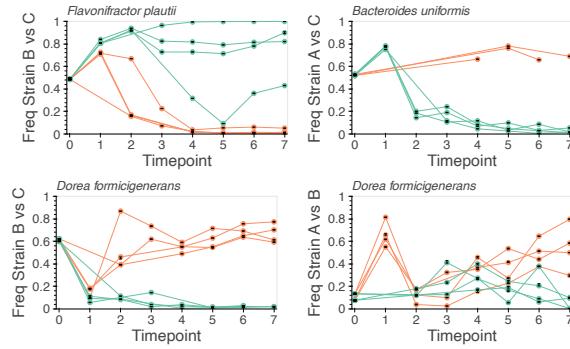
**Figure S13: Additional examples of time varying selection across replicates** Identical format as Figures 3A-C. Colored lines indicate data, and grey lines represent the median (solid) and highest and lowest (dashed) predicted strain trajectories under a model of logistic growth fit using data from timepoints 1 and 2. Examples with data on 2 or more replicates are included.



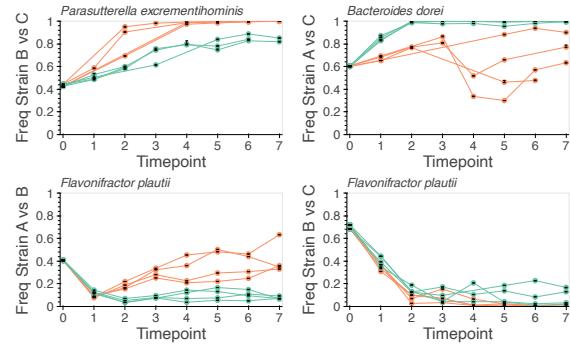
**Figure S14: Examples of inconsistent time varying selection across replicates** Identical format as Figures 3A-C. Colored lines indicate data, and grey lines represent the median (solid) and highest and lowest (dashed) predicted strain trajectories under a model of logistic growth fit using data from timepoints 1 and 2.



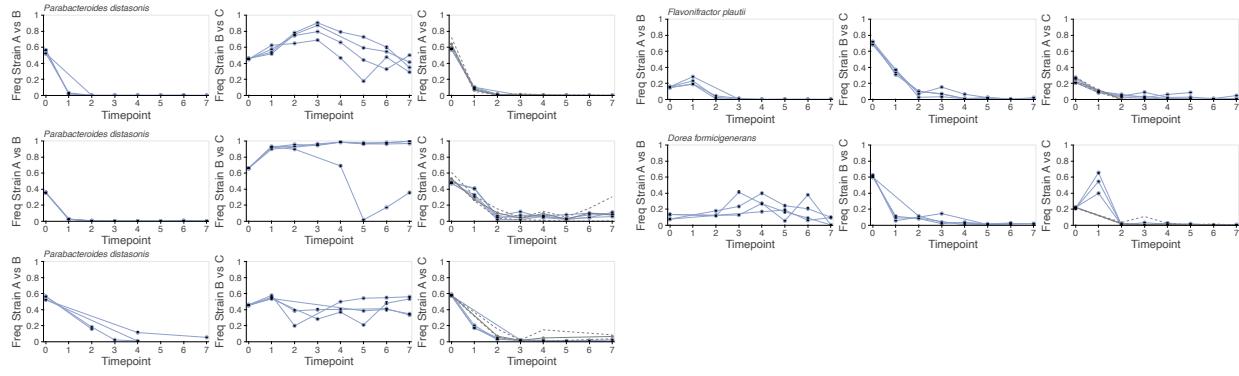
**Figure S15: The ratio of species abundances in different media does not predict differences in selection across media.**  $r^2 = .213, p = .341$



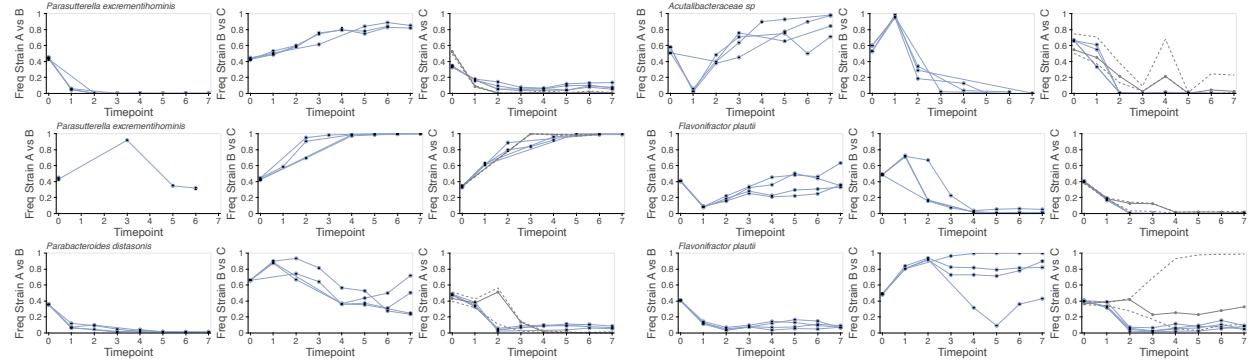
**Figure S16: Additional examples of inconsistent selection across environments, qualitative shifts** Same format as examples in Fig 4D, with replicates in mBHI colored orange and replicates in mGAM colored teal.



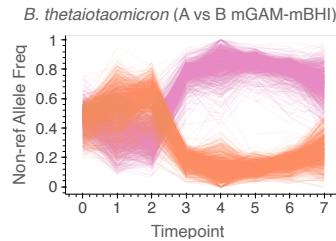
**Figure S17: Additional examples of inconsistent selection across environments, quantitative shifts** Same format as examples in Fig 4E with replicates in mBHI colored orange and replicates in mGAM colored teal.



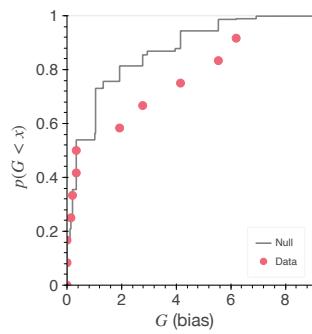
**Figure S18: Additional examples of transitive competition trios.** Each panel is identical to Fig. 5A. Each column shows the dynamics of conspecific strains in one competition for a trio of competitions. The third column also includes the predicted dynamics under the transitive model (grey). The higher and lower lines dashed represents upper and lower limits of predictions generated through bootstrapping.



**Figure S19: Additional examples of non-transitive competition trios.** Each panel is identical to Fig. 5B. Each column shows the dynamics of conspecific strains in one competition for a trio of competitions. The third column also includes the predicted dynamics under the transitive model (grey). The higher and lower lines dashed represents upper and lower limits of predictions generated through bootstrapping.



**Figure S20: Example snps plot for tracking major strains from multi-colonized populations.**



**Figure S21: Inferring community bias using only competitions from initially mono-colonized populations.** Identical to Fig E except only using competitions from initially mono-colonized populations