**Cell Systems, Volume *13***

## Supplemental information

# Functional attractors

# in microbial community assembly

**Sylvie Estrela, Jean C.C. Vila, Nanxi Lu, Djordje Bajić, Maria Rebolleda-Gómez, Chang-Yu Chang, Joshua E. Goldford, Alicia Sanchez-Gorostiaga, and Álvaro Sánchez**

# Supplementary Figures and Tables

**This PDF file includes:**

- Supplementary Figures S1-S20
- Supplementary Tables S1 and S2

## Supplementary Figures

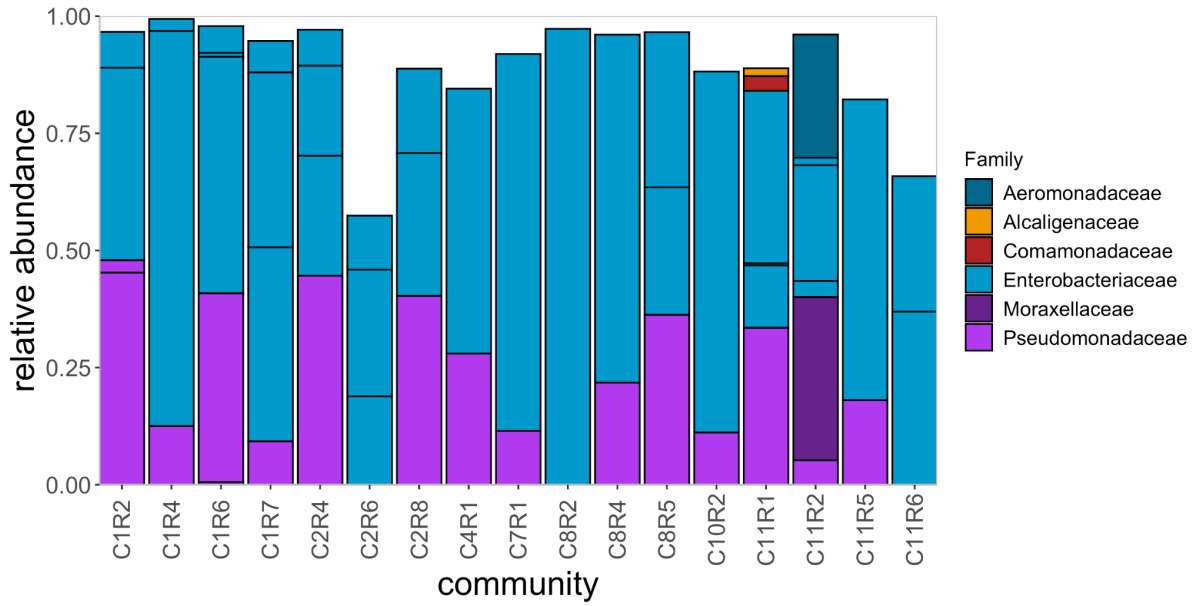


**Fig. S1. The strains isolated represented most of the taxonomic composition of the 17 communities from where they were collected.** N=73 isolates from 17 communities (CxRy) covering on average 89.4% of the taxonomic composition (Methods).

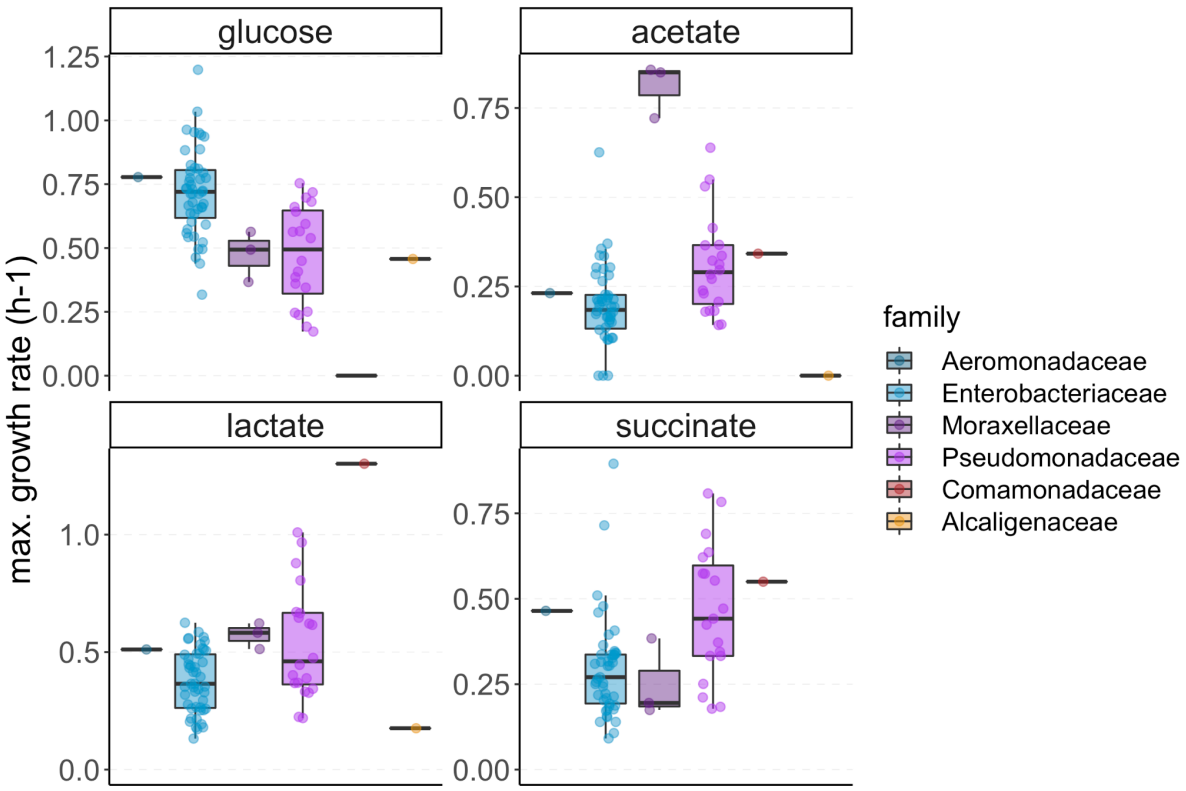


**Fig. S2. Family-level variation in maximum growth rate.** Maximum growth rate of isolates belonging to different families when grown in M9 minimal media supplemented with either glucose (N=73), acetate (N=73), lactate (N=73) or succinate (N=69) as the single carbon source.

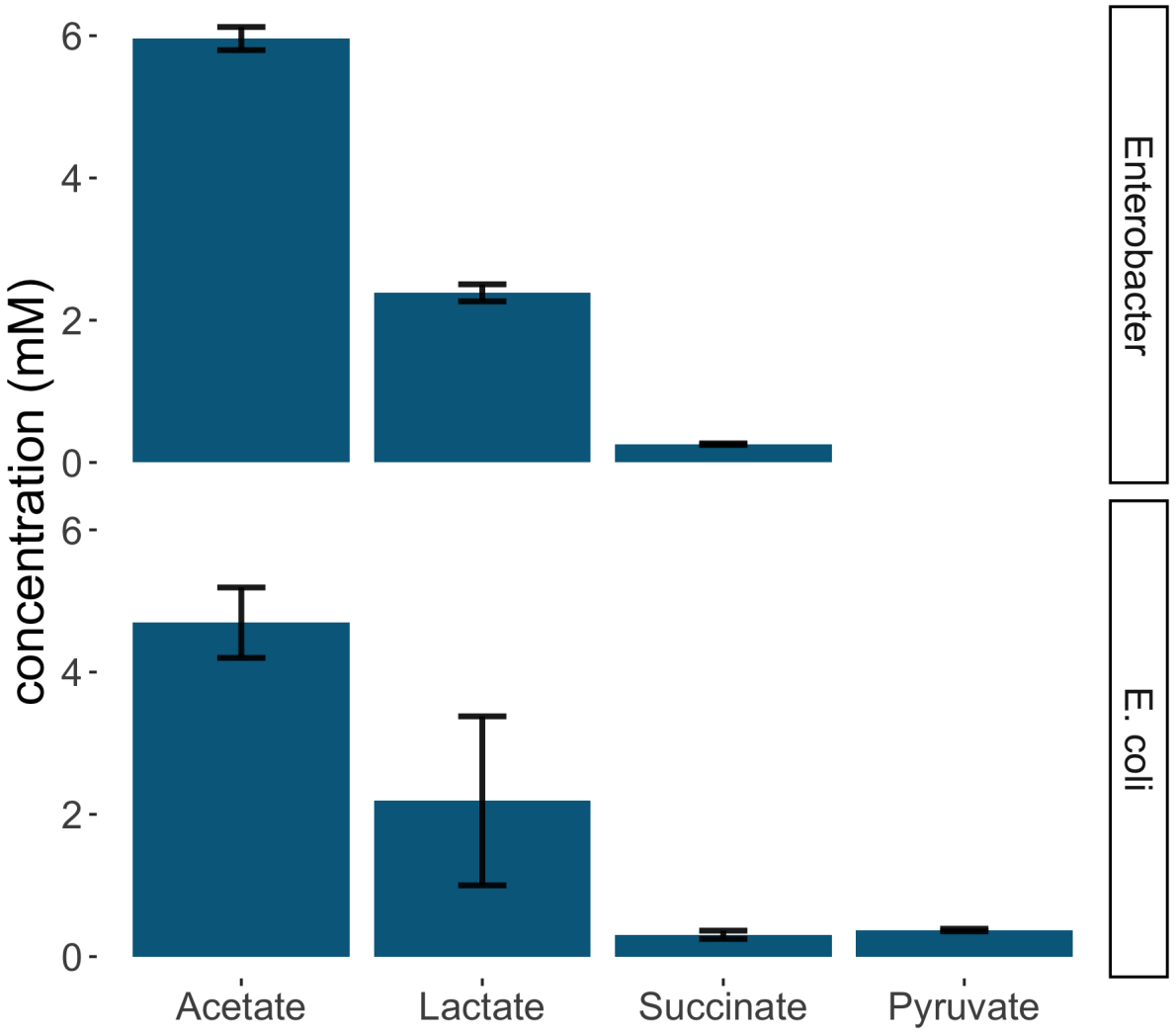


**Fig. S3. Concentration of metabolites secreted by *E. coli* and *Enterobacter* growing on glucose.** Metabolites found in the supernatant of each monoculture after 28h of growth in minimal glucose media. Metabolites were quantified through LC-MS (Methods). Shown are the metabolites with mean concentrations above 0.1 mM (mean ±range) of two replicates.

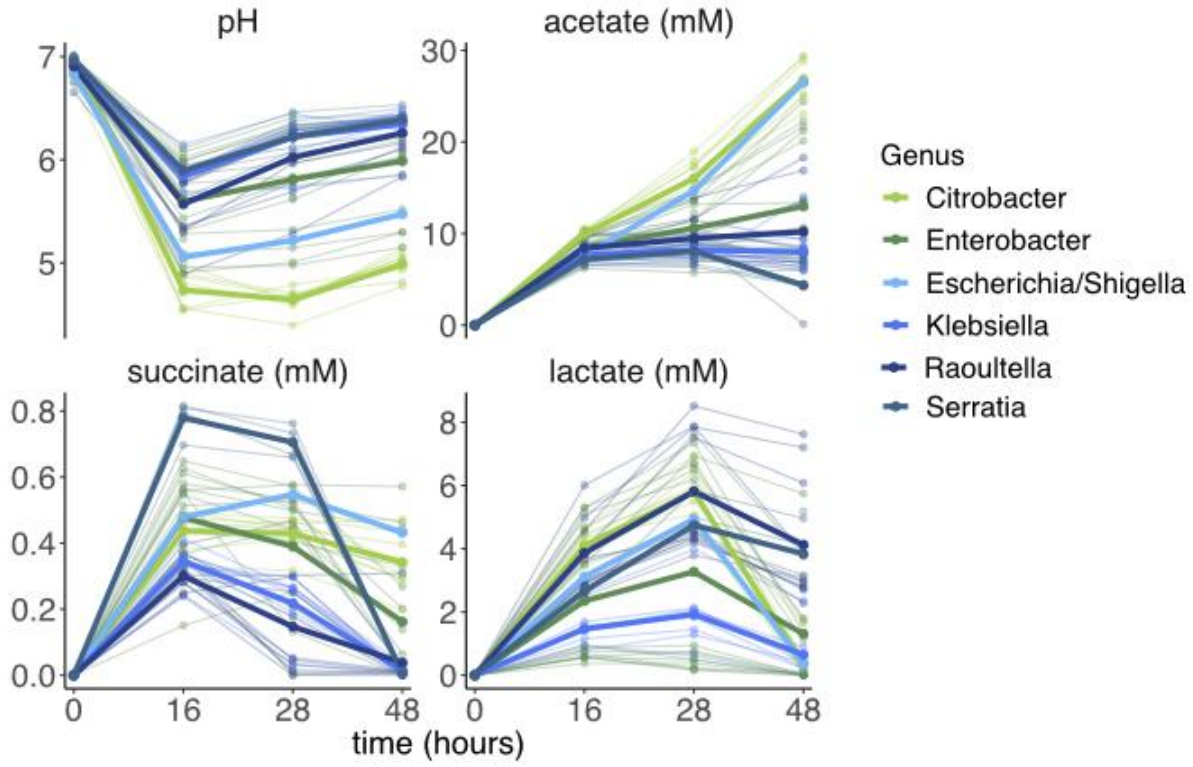


**Fig. S4. Genus-level variation in the amount of organic acids excreted by Enterobacteriaceae.** pH and concentrations of acetate, succinate, and lactate in the supernatant of isolates grown in minimal media with glucose. Supernatant collected at different timepoints during a 48h growth cycle. The thick line shows the mean of each genus (N=47).

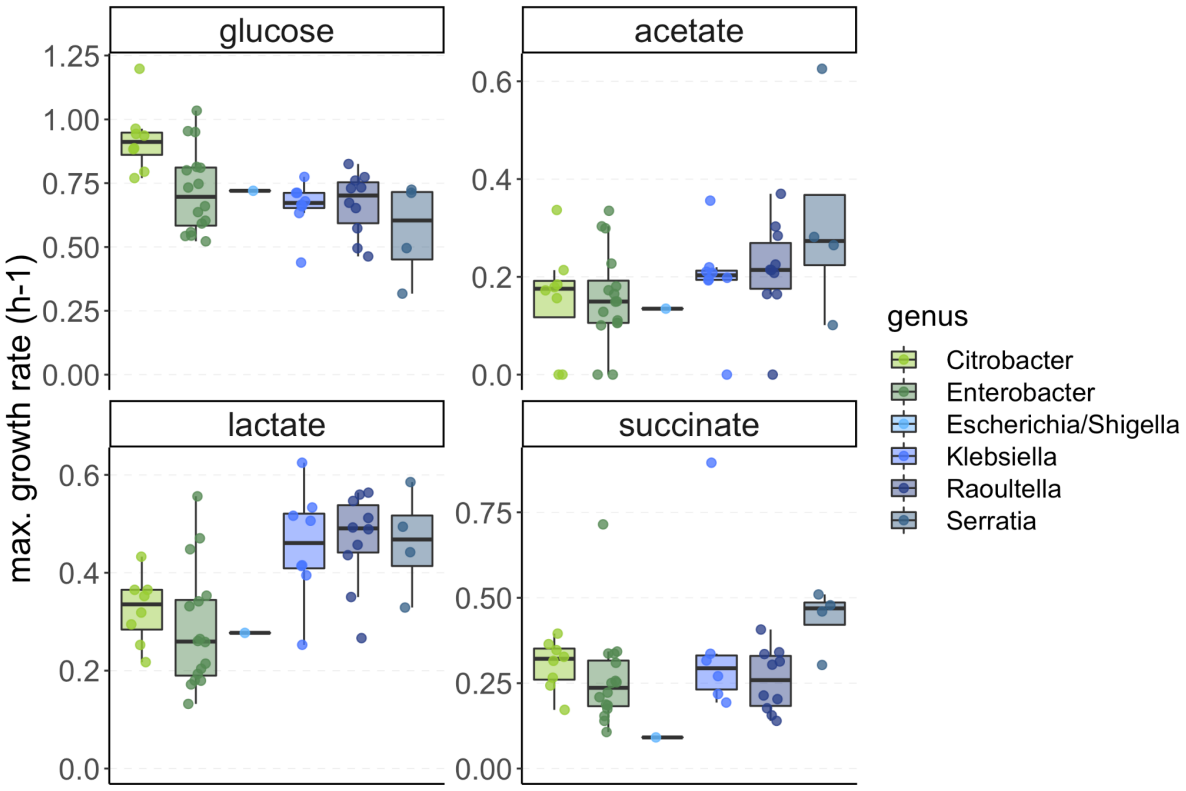


**Fig. S5. Genus-level variation in maximum growth rate.** Maximum growth rate of Enterobacteriaceae isolates (N=47, grouped by genus) when grown in M9 minimal media supplemented with either glucose, acetate, succinate, or lactate as single carbon source.

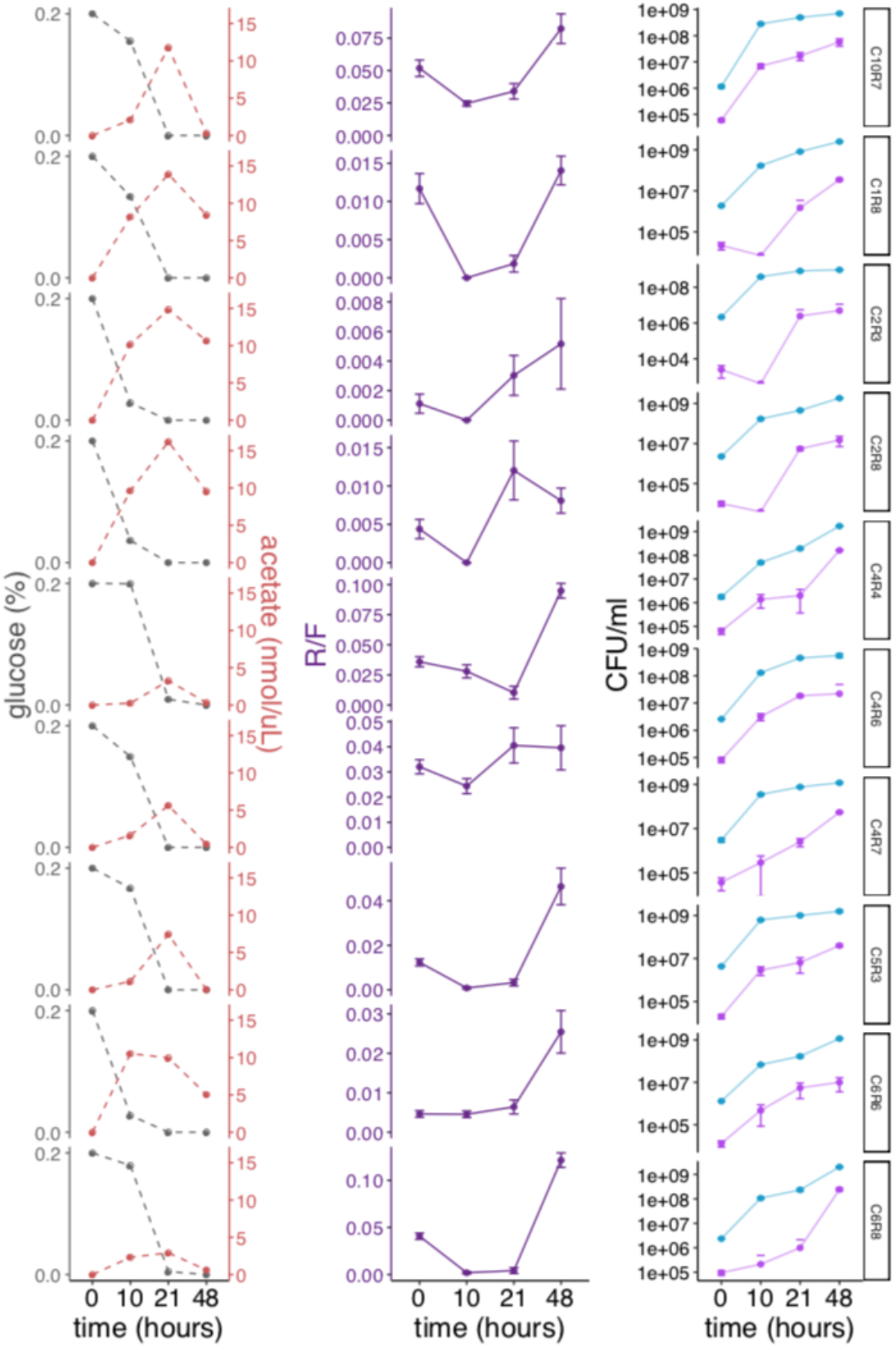
glucose (%)
acetate (nmol/uL)
R/F
CFU/ml
time (hours)
time (hours)
time (hours)
0 10 21 48
C10R7
C1R8
C2R3
C2R8
C4R4
C4R6
C4R7
C5R3
C6R6
C6R8

**Fig. S6. Growth of fermenters is associated with glucose depletion, whereas growth of respirers is associated with acetate depletion.** We thawed whole communities and revived them in glucose minimal media for three passages. We selected a subset of communities where fermenters (F) and respirers (R) coexist after three serial transfers (N=9 communities from 6 different inocula where CxRy corresponds to inoculum x, replicate y) and measured their R/F ratio, and concentrations of glucose and acetate at different timepoints during a 48h growth cycle (0h, 10h, 21h and 48h). The R/F ratio represents the mean$\pm$sd of the CFU ratios (bootstrap, N=1000). For three of the communities (C1R8, C2R3, and C2R8), there were no detectable R colonies on the plates, so we set R/F to 0. To make sure that the observed decrease in R/F ratio from 0h to 10h is not due to differences in the total number of CFUs across timepoints, we confirmed that the total CFU count at 10h is larger than the total CFU count at 0h. C10R7 corresponds to the community shown in **Fig 1C**.

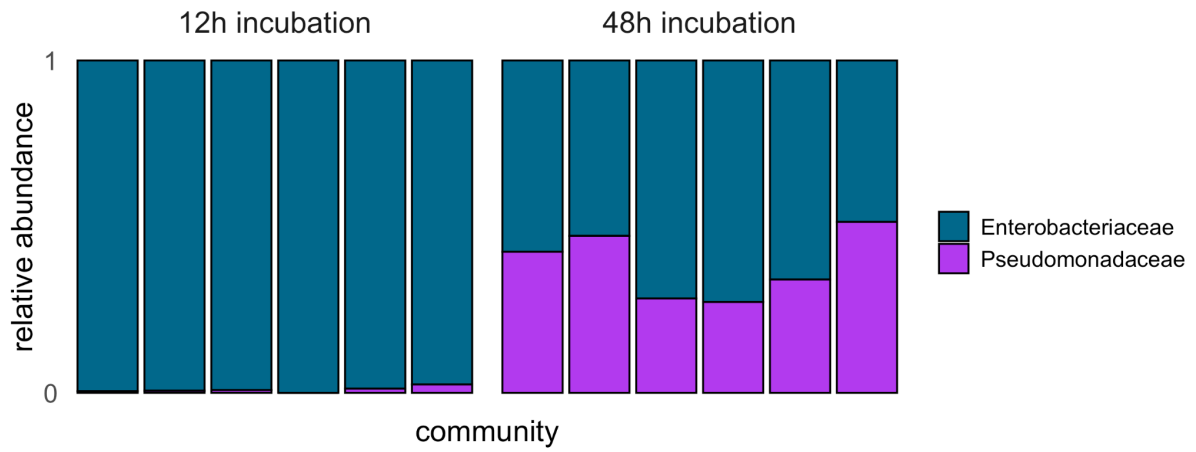


**Fig. S7. Shorter incubation time favours Enterobacteriaceae and disfavours *Pseudomonas*.** A representative community composed of four different strains (three strains belonging to the Enterobacteriaceae family-- Klebsiella, Raoultella, Citrobacter- and one *Pseudomonas* strain) was reconstituted by mixing the isolates 1:1 (Methods). The reconstituted community (N=6 replicates) was then serially-transferred to fresh m9-glucose medium every 48h for 5 transfers. After the 5th transfer, the communities were propagated for 5 more transfers of 48h (control) (N=6) or 12h each (N=6). Shown is the relative abundance of each family at Transfer 10.

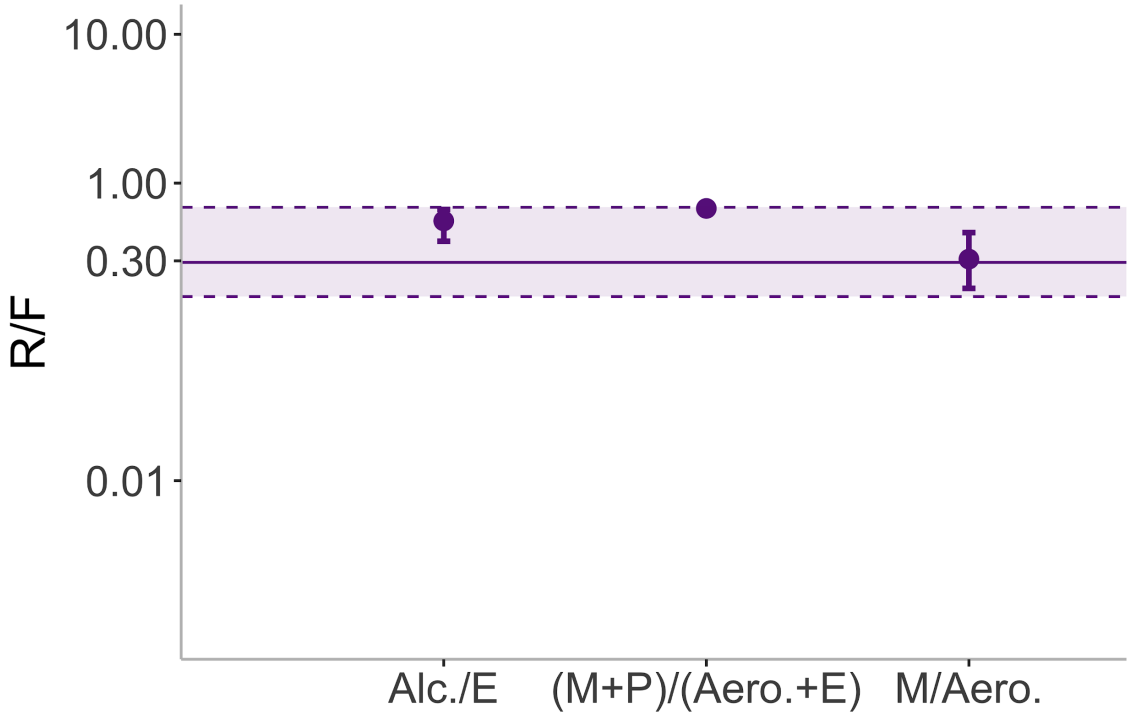


**Fig. S8. Communities dominated by other respirator and fermenter families exhibit similar R/F ratios.** Shown is the median R/F ratio for communities where Pseudomonadaceae, Enterobacteriaceae or both had been totally or partially replaced by other families with the same functional role (i.e. respirators or fermenters, respectively). Error bars represent IQR. The Alcaligenaceae/ Enterobacteriaceae communities were taken from **Fig. 2B** (state dominated by Enterobacteriaceae and Alcaligenaceae as dominant F and R respectively; N=65, median =0.56, Q1=0.41, Q3=0.66); the (Moraxellaceae + Pseudomonadaceae/ Aeromonadaceae + Enterobacteriaceae) community corresponds to one community (C11R2) in Goldford et al. (2018) that is not dominated by *Pseudomonas* and Enterobacteriaceae (N=1, R/F=0.68), and the Moraxellaceae/ Aeromonadaceae communities were from an experiment where one community (12 replicates) was passaged for 11 transfers in 5mL of glucose M9 minimal media inside a 50mL Falcon tube under the same other experimental conditions as before (N=12, median=0.31, Q1=0.20, Q3=0.46). The horizontal line shows the median R/F ratio of the glucose communities presented in **Fig. 1A** (R/F=0.29) and the dashed lines show the interquartile range (Q1=0.17, Q3=0.69).

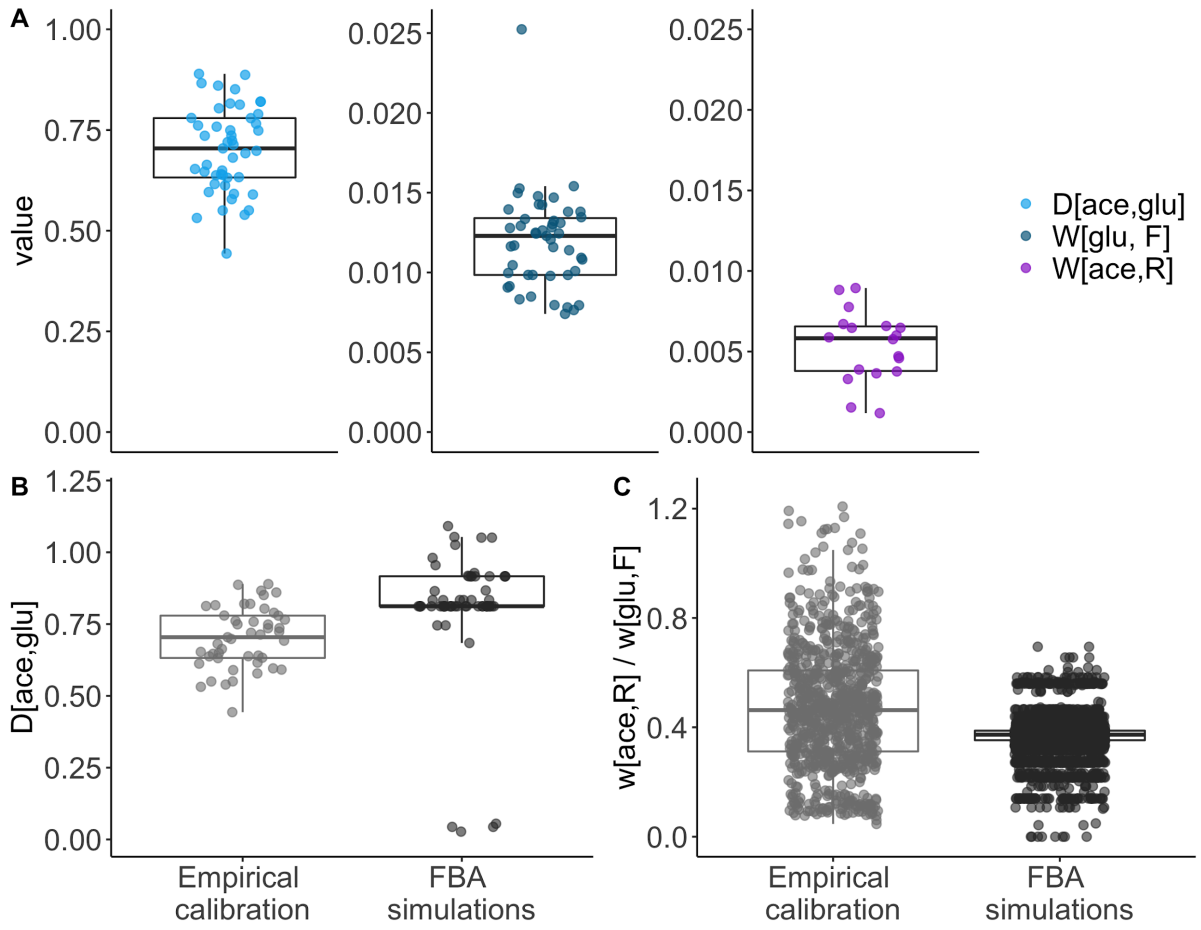


**Fig. S9. FBA predicts well the empirically measured ratio of acetate to glucose in our Enterobacteriaceae isolates. A.** The values of the parameters shown in **Eq. 1** ($D_{ace,glu}$, $w_{glu}^{F}$, and $w_{ace}^{R}$) were estimated empirically for a collection of taxa isolated from our communities - 47 Enterobacteriaceae isolates (F) and 18 Pseudomonadaceae isolates (R) (**Fig. 1A** and **Fig. 1B**). $D_{ace,glu}$ represents the amount of acetate released per glucose molecule consumed after 16h of monoculture growth by F in glucose minimal medium. $w_{glu}^{F}$ and $w_{ace}^{R}$ were estimated for each F and R isolate by growing them in monoculture in glucose and acetate minimal media, respectively (Methods and Supplementary Methods). **B, C.** The values of the parameters were also predicted by CAFBA for 59 Enterobacteriaceae genome-scale models and 74 Pseudomonas metabolic models (Methods and Supplementary Methods).

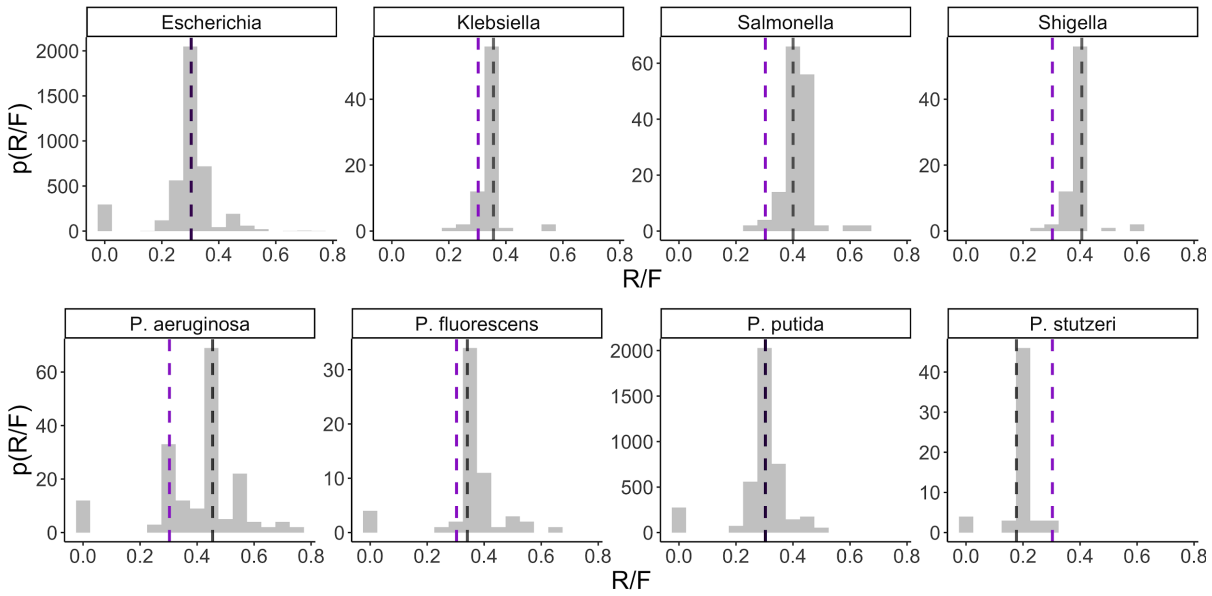


**Fig. S10. Probability distribution of the FBA predicted R/F ratio for different pairs of Enterobacteriaceae -Pseudomonas models.** Distributions are shown separately for each genus of Enterobacteriaceae (N=59), or species of *Pseudomonas* (N=74). The purple line shows the median of all *Pseudomonas*/ Enterobacteriaceae pairs (R/F=0.3, N=4366) and the black line shows the median for the pairs shown in the respective facet.

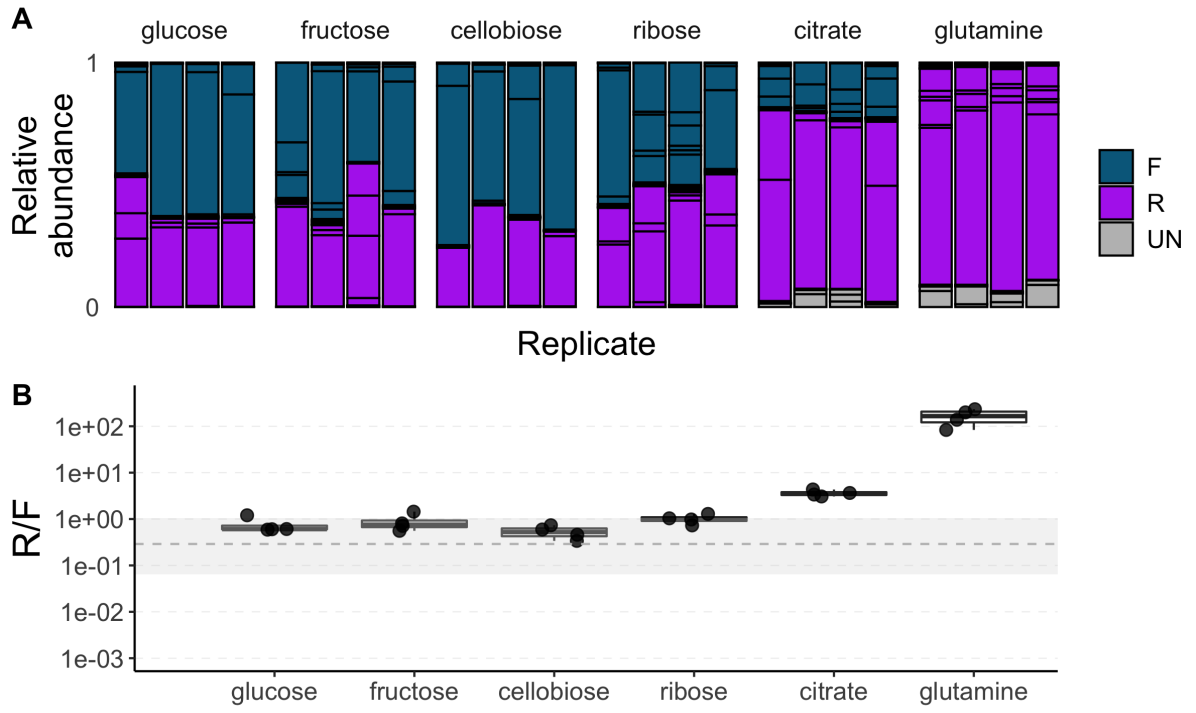


**Fig. S11. The R/F ratio is carbon source dependent.** Four replicate communities, all started from the same inoculum, were serially transferred every 48h in minimal medium supplemented with a single carbon source (glucose, fructose, cellobiose, ribose, citrate, and glutamine) for a total of 10 transfers (Methods). **A.** Community composition coloured by R and F type. UN means that the fermentation profile is not assigned. **B.** The dashed line indicates the median and the gray shading area indicates the region between the 2.5% and 97.5% percentiles of the communities shown in **Fig. 1A**.

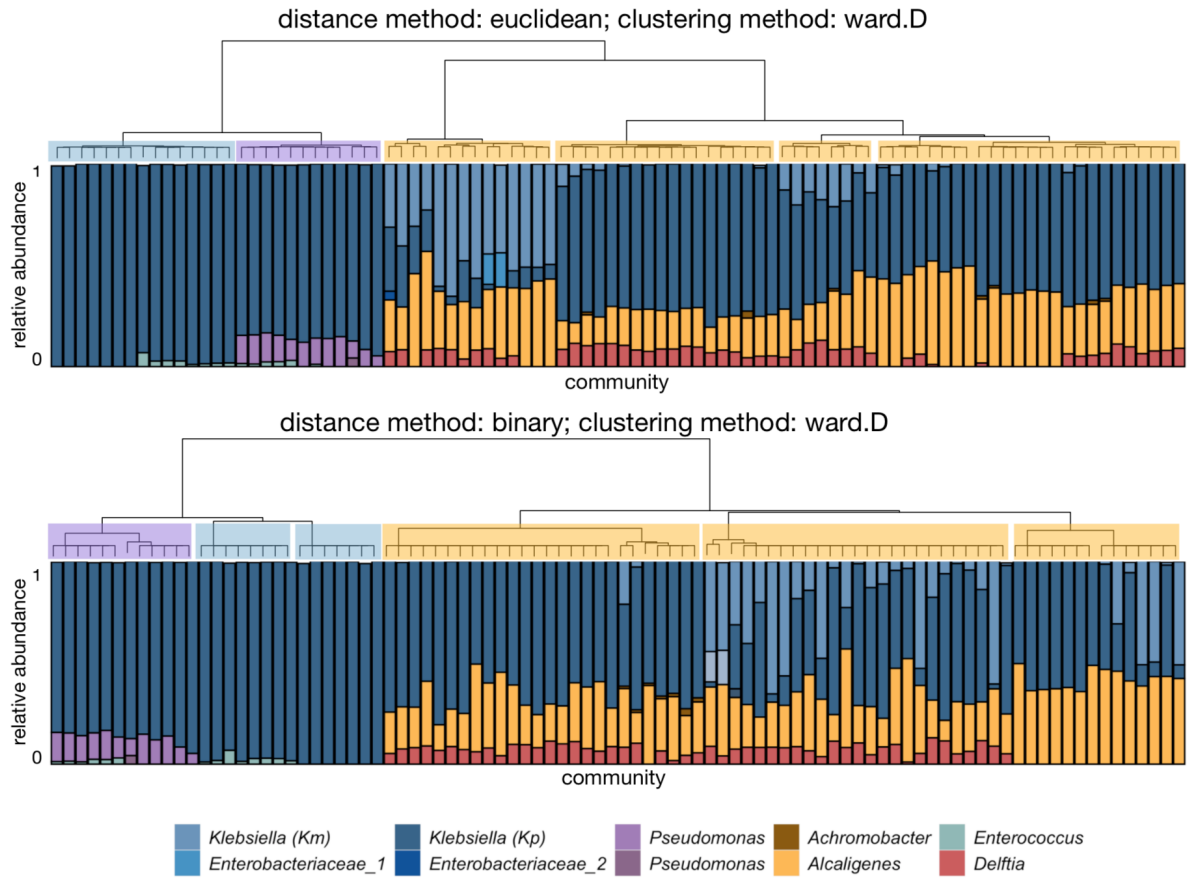


**Fig. S12. Detection of alternative states in community composition.** Taxonomic profile of communities shown at the exact sequence variant (ESV) level (one color per ESV) with corresponding genus and family level assignments. Only the ESVs with a minimum relative abundance > 0.01 are shown. The communities are sorted according to the order obtained from the hierarchical tree (dendrogram) of the relative abundance of all ESVs using hierarchical clustering (Ward's minimum variance method) and either the "euclidean" (top) or "binary" (bottom) distance (implemented in the *hclust* R function). Six alternative states in community composition (clusters) obtained using the *cutree* function in R to cut the tree generated by *hclust* are coloured as in **Fig. 2B**.

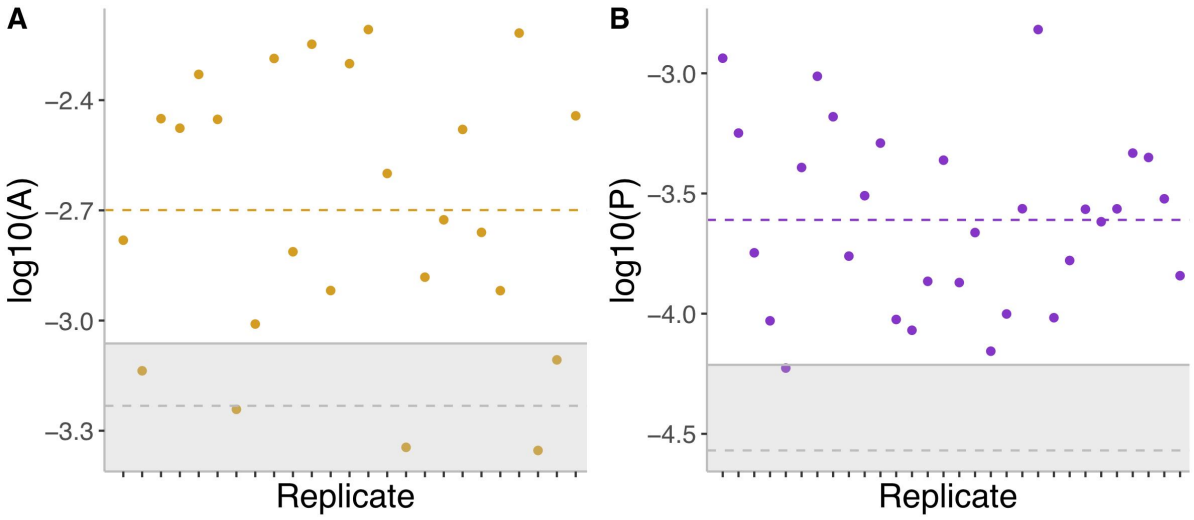


**Fig. S13. Low *Alcaligenes* and low *Pseudomonas* are above the error threshold in most communities.** To correct for spurious detection of *Alcaligenes* and *Pseudomonas* during amplicon sequencing, we determine an error threshold for the dominant ESV of *Alcaligenes* and for the dominant ESV of *Pseudomonas* using two positive controls as reference (Methods). The points show the relative abundance of *Alcaligenes* in communities with low *Alcaligenes* **(A)** and the relative abundance of *Pseudomonas* in communities with low *Pseudomonas* **(B)**. The dashed coloured line shows the mean relative abundance across all the communities shown. The dashed gray line shows the detected relative abundance of *Alcaligenes* **(A)** and of *Pseudomonas* **(B)** across all positive controls (N=22, Methods) and the gray shaded area indicates the 95% confidence interval. The communities that fail the 95% confidence interval test (that is, within the shaded area) were excluded from the calculation of the number of communities that contain both *Alcaligenes* and *Pseudomonas* after 18 transfers (main text). Most communities, however, pass the 95% confidence interval test (that is, above the shaded area), demonstrating that rare *Alcaligenes* and rare *Pseudomonas* are indeed present in the communities despite rare. The fact that *Pseudomonas* is still present but rare (0.01) in *Alcaligenes* dominated communities, and similarly, that *Alcaligenes* is still present but rare (0.01) in *Pseudomonas* dominated communities, strongly suggests that the alternative states we observed were generally not caused by random sampling.

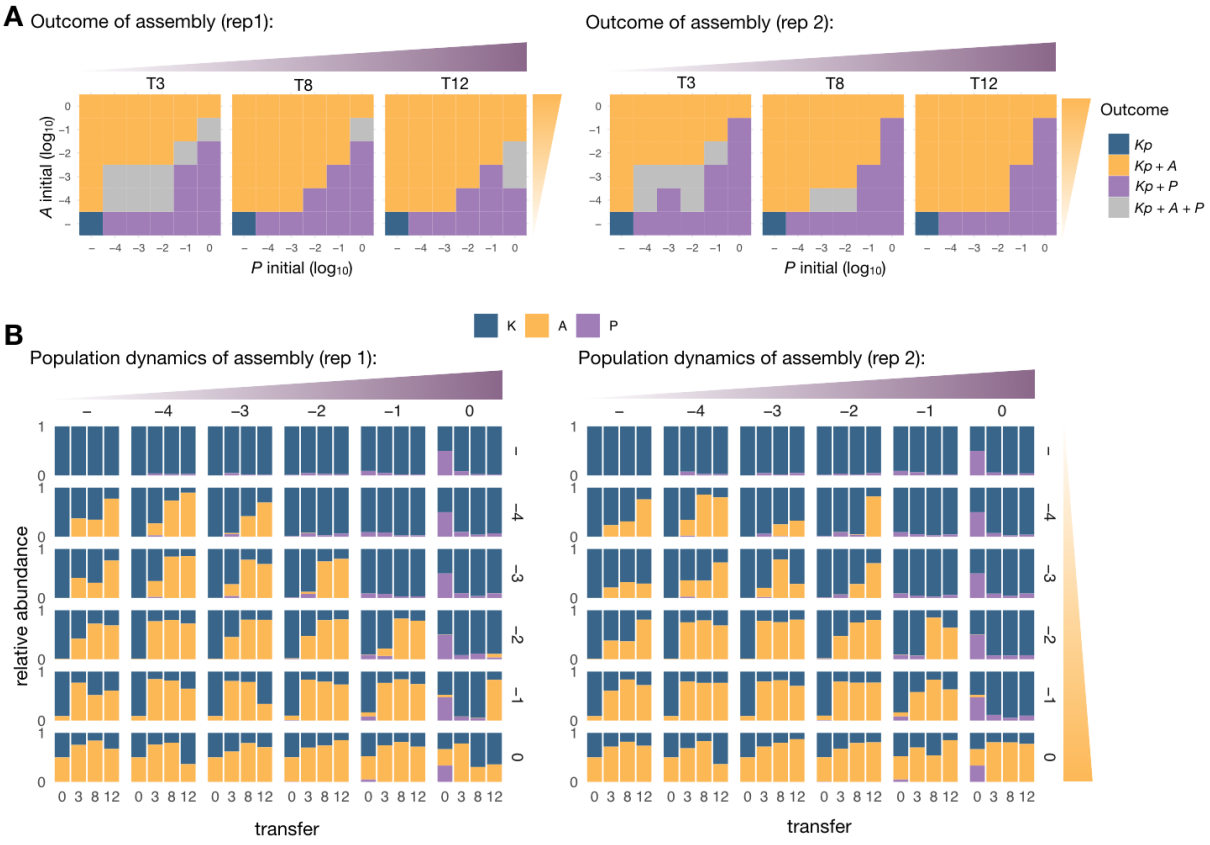


**Fig. S14. Density-dependent multistable population dynamics.** We isolated the three dominant strains - *Klebsiella* (*Kp*), *Alcaligenes* (*A*), and *Pseudomonas* (*P*) that make up the two major alternative attractors, and grew them in pairwise coculture (*Kp*+*A* or *Kp*+*P*) or in three species consortia (*Kp*+*A*+*P*) by mixing *Kp* with different initial densities of *A* and/or *P* (see Methods). These reconstituted communities were passaged for 12 transfers in M9+glucose in two biological replicates. **A.** The phase portraits show the state of the community at T= 3, 8, and 12 transfers for two biological replicates. A square is colored yellow if a community that was started there contained *A* but not *P* at time T, and it is purple if it contained *P* but not *A*. It is gray if both *A* and *P* were visibly present. We can see that the phase portrait is divided in two regions: The upper-left diagonal is made up by the basin of attraction of *Alcaligenes* dominated communities, whereas the bottom-right diagonal contains the basin of attraction for *Pseudomonas* dominated communities. *Alcaligenes* and *Pseudomonas* mutually exclude each other depending on their starting densities. **B.** Shown are the population dynamics of each species in both replicates (left and right panel) when started under different initial densities of *A* and/or *P*. The amount of *P* inoculated increases from left to right whereas *A* increases from bottom to top.

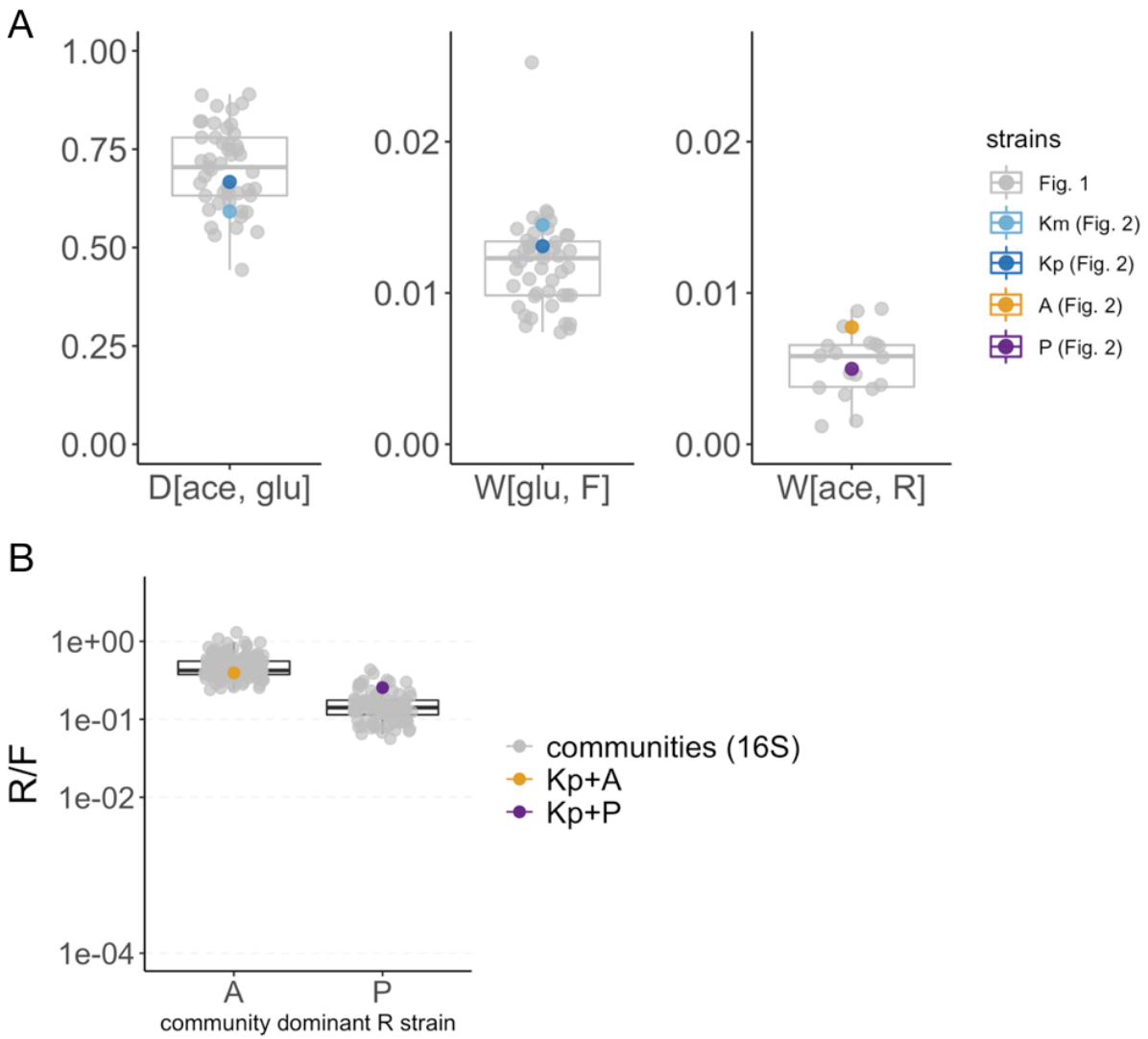


**Fig. S15. Model predictions for the community assembly experiment shown in Figure 2.** **A.** The three parameters used in our resource-partitioning model (Eq. 1; $D_{ace,glu}$, $w_{glu}^{F}$, and $w_{ace}^{R}$) were estimated empirically for the dominant strains of the community assembly experiment shown in Fig. 2. Specifically, we estimated $D_{ace,glu}$ and $w_{glu}^{F}$ for the two dominant F strains (*Kp* and *Km*), and estimated $w_{ace}^{R}$ for the two dominant R strains (*A* and *P*). For comparison, in gray we plot the values estimated for the isolates shown in Fig. 1 (same as shown in Fig. S9). **B.** R/F ratio predicted by the empirical model for the KpA (R/F=0.39) and KpP (R/F=0.25) community states. Shown in gray are the R/F ratios of the communities shown in Fig. 5C and grouped by community state outcome (i.e. whether *A* or *P* is the dominant R strain).

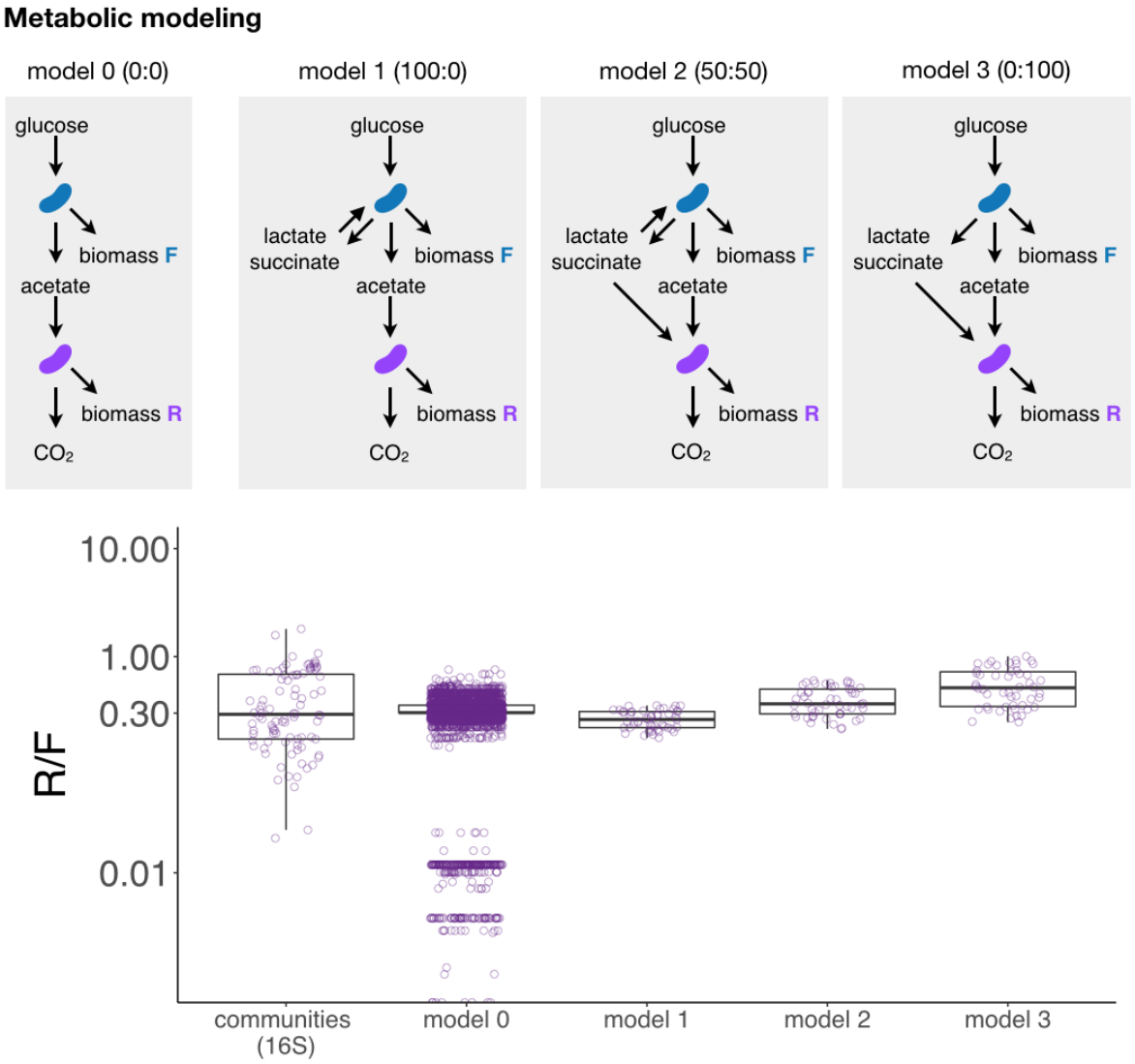


**Fig. S16. Less abundant organic acids have a relatively small effect on the R/F ratio.** Empirically observed and predicted R/F ratio. The *Experiment* and *Simulations* (model 0) show the same data as in **Fig. 1D** and are shown here to serve as a baseline. Unlike in *Model 0* which uses CAFBA and where acetate secretion is a prediction of the model, *Model 1-3* use FBA where acetate, lactate and succinate secretions are forced on the simulations based on the empirically measured median concentration of acetate, lactate and succinate by fermenter isolates from our communities (Supplementary Methods). *Model 1-3* were done using the *E. coli* (iJO1366) and *P. putida* (iJN1463) metabolic models. Each dot shows the R/F ratio estimated for an *E. coli* - *P. putida* pair where the *E. coli* is constrained to produce the empirically measured amount of acetate, lactate and succinate by one of the fermenter isolates (N=45 pairs) after 16h of monoculture growth in M9-glucose (**Fig. 1B**). Three consumption scenarios are shown: all lactate and succinate produced is consumed by F only (*Model 1*), by F and R 50:50 (*Model 2*) and by R only (*Model 3*).

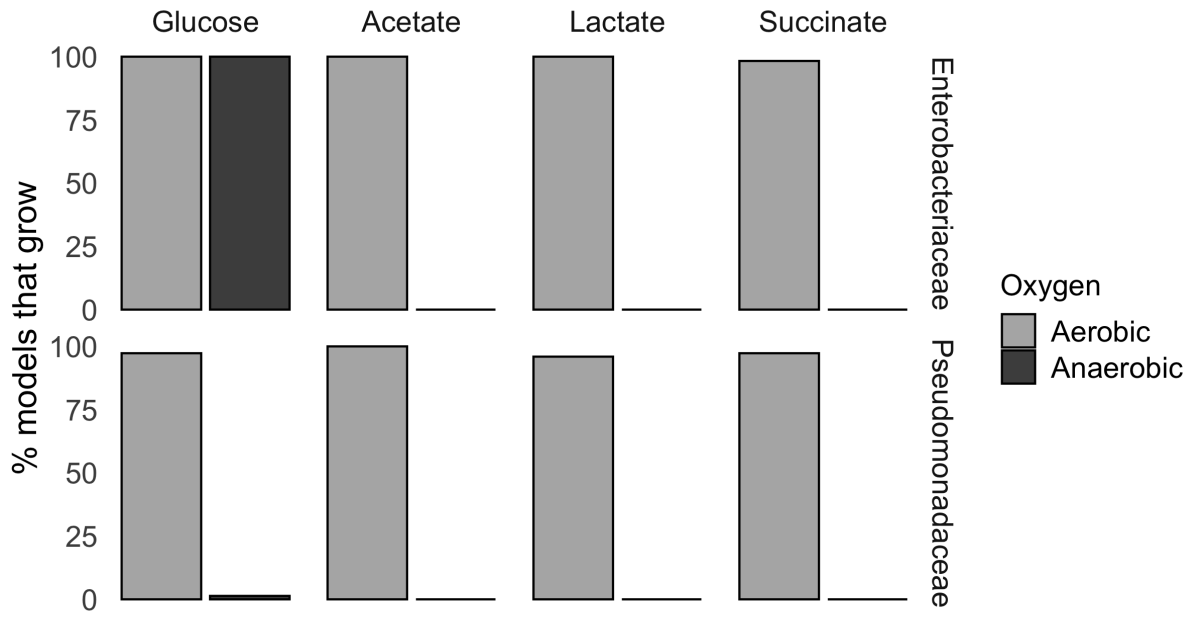


**Fig. S17. Oxygen is required for growth on acetate, lactate and succinate.** Proportion of metabolic models (for 74 *Pseudomonas* and 59 Enterobacteriaceae strains) that can grow on glucose, acetate, lactate, and succinate under aerobic and anaerobic conditions (Methods). None of the strains can grow on acetate, lactate, or succinate in the absence of oxygen. All Enterobacteriaceae strains but only one of the Pseudomonas strains can grow on glucose in the absence of oxygen.

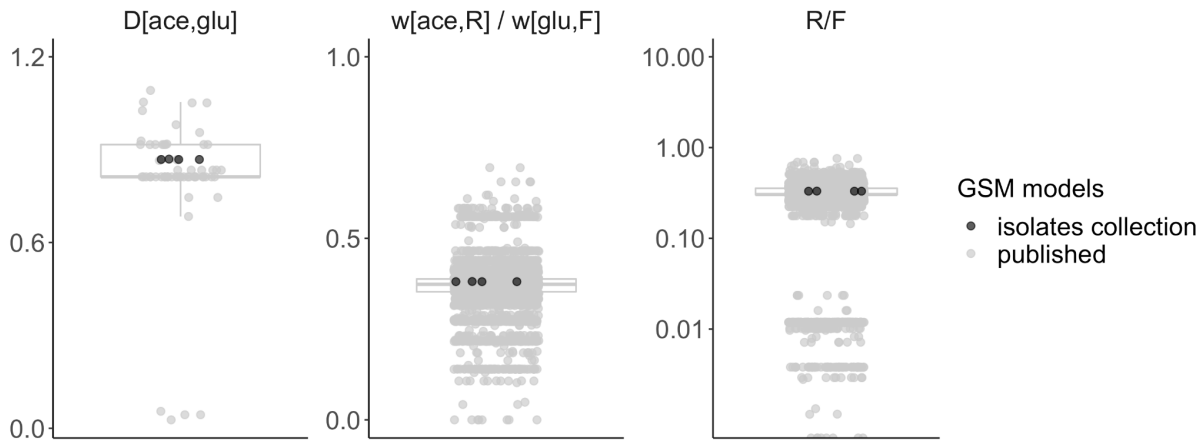


**Fig. S18. The predictions observed using a library of published metabolic models and metabolic models built with strains from our isolate collection are in close agreement.** In addition to the library of published metabolic models, we analyzed 5 genome-scale metabolic models which we built using whole genome-sequences of 5 strains from our isolate collection (Methods). These strains correspond to isolates from the communities shown in **Fig. 1A-B** and cover 5 different genera (*Klebsiella*, *Citrobacter*, *Raoultella*, *Enterobacter*, and *Pseudomonas*).

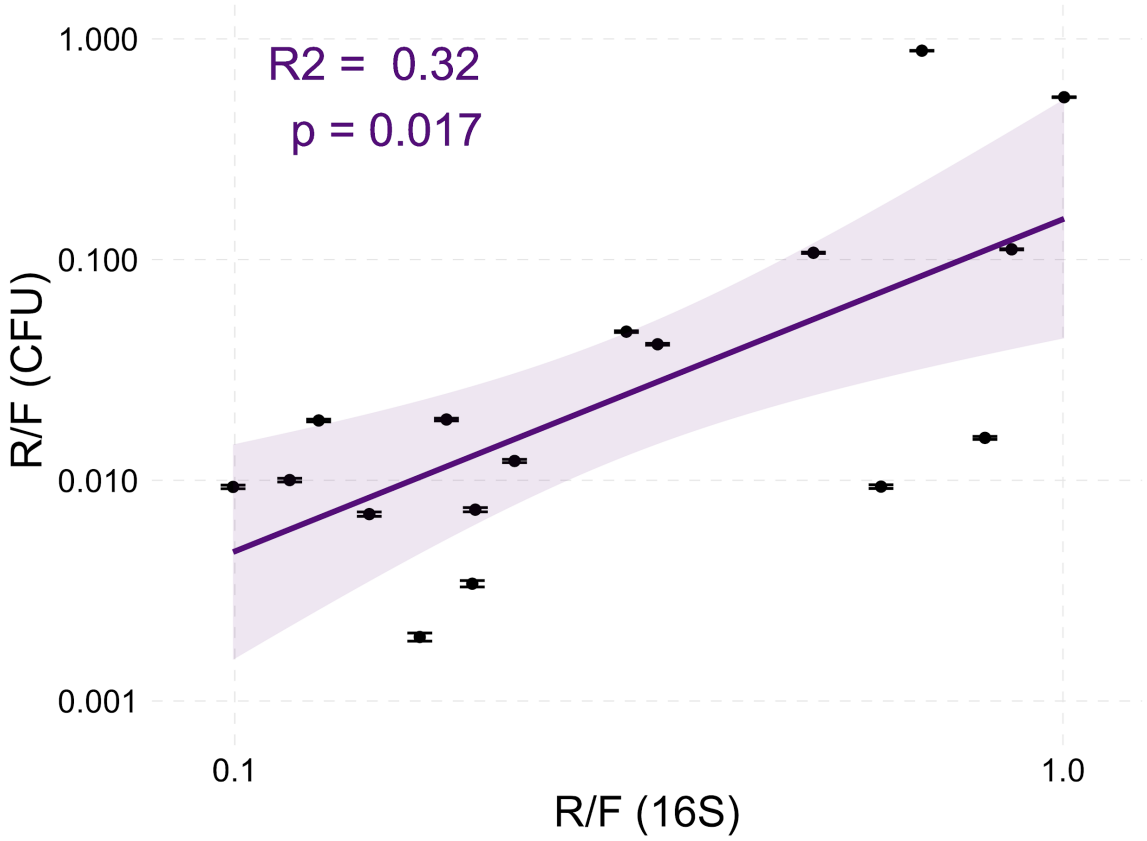


**Fig. S19. Positive correlation between the R/F ratio measured by CFU and by 16S sequencing.** Respirator:Fermenter ratio in glucose communities (T12) from Goldford et al. (2018). 16S data from Goldford et al (2018). To calculate CFUs ratio, 24 communities were thawed from frozen stock and CFUs were counted on chromogenic agar. White colonies were counted as R and blue/purple colonies were counted as F. Shown are communities with at least one colony of R and F (N=17). For each community, dots show the mean±sem (bootstrap, N=1000). Shown is a linear regression (orange line) with slope= 0.46 ($R^2$=0.32).

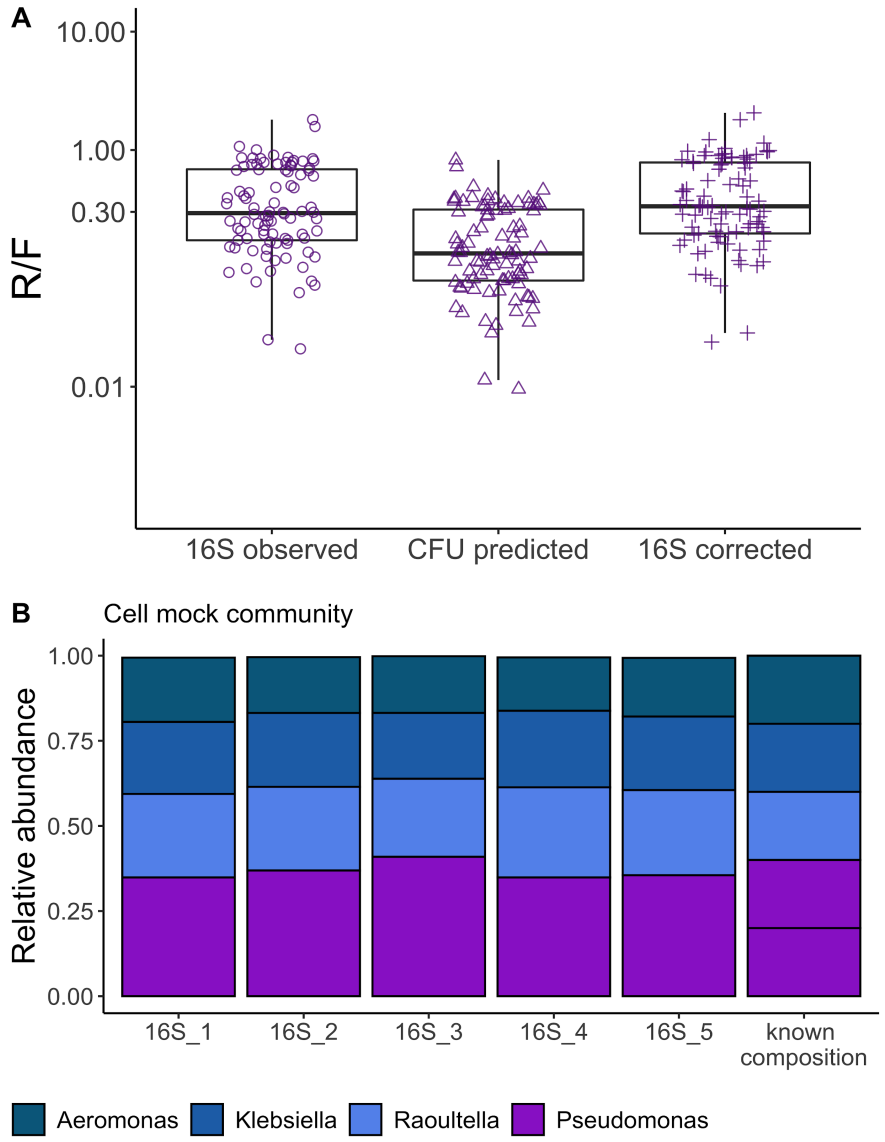


**Fig. S20. Observed and predicted R/F ratios. A.** *16S observed*: R/F ratio observed experimentally in the glucose communities described in **Fig. 1A** (median=0.29, Q1=0.17, Q3=0.69, N=92). *CFU predicted*: for each community, their observed R/F ratio was corrected by multiplying the slope of the linear regression in **Fig. S19** (Methods) (median= 0.13, Q1=0.079, Q3=0.31, N=92). *16S corrected*: estimated R/F ratio after correcting for amplicon abundance bias using as reference a cell mock community (shown in **B**) consisting of a mix of fermenters and respirators (median= 0.33, Q1=0.20, Q3=0.78, N=92). **B.** The mock community consisted of 5 isolates (1 *Klebsiella*, 1 *Raoultella*, 1 *Aeromonas*, and 2 *Pseudomonas*) all mixed 1:1 by OD, thus giving a theoretical (expected) R/F ratio of 0.6. Shown are the relative abundances of each taxa in the mock communities (5 replicates) obtained by 16S sequencing and expected (known composition) (Methods). Although a single *Pseudomonas* ESV was detected, the theoretical fraction of *Pseudomonas* is close to the detected fraction, suggesting that the two *Pseudomonas* were not discriminated through sequencing.

## Supplementary Tables

**Table S1. Glucose fermentation profile of common families**

| Family | Fermentation profile |
|---|---|
| Enterobacteriaceae | Fermenter |
| Enterococcaceae | Fermenter |
| Aeromonadaceae | Fermenter |
| Lachnospiraceae | Fermenter |
| Pseudomonadaceae | Respirator |
| Alcaligenaceae | Respirator |
| Moraxellaceae | Respirator |
| Xanthomonadaceae | Respirator |
| Comamonadaceae | Respirator |
| Oxalobacteraceae | Respirator |

**Table S2. Sensitivity analysis on the parameters used for the CAFBA simulations for the *E. coli* (iJO1366) and *P. putida* (iJN1463) metabolic models**

| Parameter | Fold change | Value | P/E |
|---|---|---|---|
| $\phi_{max}$ | Double | 0.968 | 0.350 |
| $\phi_{max}$ | Default | 0.484 | 0.357 |
| $\phi_{max}$ | Half | 0.242 | 0.370 |
| $w_E$ | Double | 0.001770326 | 0.365 |
| $w_E$ | Default | 0.000885163 | 0.357 |
| $w_E$ | Half | 0.000442581 | 0.353 |
| $w_R$ | Double | 0.338 | 0.361 |
| $w_R$ | Default | 0.169 | 0.357 |
| $w_R$ | Half | 0.0845 | 0.355 |
| *P. putida* limitation | TRUE | N/A | 0.355 |
| *P. putida* limitation | FALSE | N/A | 0.357 |