Supplementary Information

Assembly of microbial communities engaging in metabolic division of labor in a diffusion-limited environment is governed by metabolic flux

Xiaoli Chen[1, 2], Miaoxiao Wang[3, 4], Yexin Xing[5], Yong Nie[1#] and Xiao-Lei Wu[1, 2, 6#]

[1] College of Engineering, Peking University, Beijing 100871, China

[2] Institute of Ocean Research, Peking University, Beijing 100871, China

[3] Microbial Systems Ecology Group, Institute of Biogeochemistry and Pollutant Dynamics, Department of Environmental Systems Sciences, ETH Zurich, 8006, Zurich, Switzerland

[4] Department of Environmental Microbiology, Eawag: Swiss Federal Institute of Aquatic Sciences, 8600, Duebendorf, Switzerland

[5] School of Resource and Environmental Engineering, Hefei University of Technology, Hefei 230000

[6] Institute of Ecology, Peking University, Beijing 100871, China

[#]Corresponding author: Research Scientist, College of Engineering, Peking University. Tel: +86 10-62759047; Fax: +86 10-62759047; E-mail: nieyong@pku.edu.cn

[#]Corresponding author: Professor, College of Engineering, Peking University. Tel: +86 10-62759047; Fax: +86 10-62759047; E-mail: xiaolei_wu@pku.edu.cn

S1 Formulation, simulation, and analyses of our individual-based model

We first built the formulations of our individual-based model describing substrate (Equation S1.1 and S1.2), intermediate (Equation S1.3 and S1.4), and product (Equation S1.5 and S1.6) concentrations in a two-step MDOL community*(1)*. Here, the diffusion of S, I, and P was modeled by the finite element method built-in *gro*. The growth dynamics of [1, 0] and [0, 1] were modeled using a product-dependent Monod-type formulation (Equation S1.7-S1.10). Detailed parameter meanings were listed in Supplementary Table 1 and Supplementary Table 2.

$$\frac{dS_1}{dt} = -\frac{k_1 E_1}{K_1 + S_1} S_1 \cdot V_1 + D_S \cdot \nabla^2 S_1 \quad [S1.1]$$

$$\frac{dS_2}{dt} = -D_S \cdot \nabla^2 S_1 \quad [S1.2]$$

$$\frac{dI_1}{dt} = \frac{k_1 E_1}{K_1 + S_1} S_1 \cdot V_1 - D_I \cdot \nabla^2 I_1 \quad [S1.3]$$

$$\frac{dI_2}{dt} = D_I \cdot \nabla^2 I_2 - \frac{k_2 E_2}{K_2 + I_2} I_2 \cdot V_2 \quad [S1.4]$$

$$\frac{dP_1}{dt} = D_P \cdot \nabla^2 P_1 - \frac{kg_1 P_1}{Kg_1 + P_1} P_1 \quad [S1.5]$$

$$\frac{dP_2}{dt} = \frac{k_2 E_2}{K_2 + I_2} I_2 \cdot V_2 - \frac{kg_2 P_2}{Kg_2 + P_2} P_2 - D_P \cdot \nabla^2 P_2 \quad [S1.6]$$

$$V_1 = \frac{4}{3}\pi \cdot (\frac{w}{2})^3 + \pi \cdot (\frac{w}{2})^2 \cdot L_1 \quad [S1.7]$$

$$V_2 = \frac{4}{3}\pi \cdot (\frac{w}{2})^3 + \pi \cdot (\frac{w}{2})^2 \cdot L_2 \quad [S1.8]$$

$$\frac{dL_1}{dt} = (\frac{kg_1}{Kg_1 + P_1} P_1 \cdot Y_1 \cdot C_1 \ - D_1) \cdot L_1 \quad [S1.9]$$

$$\frac{dL_2}{dt} = (\frac{kg_2}{Kg_2 + P_2} P_2 \cdot Y_2 \cdot C_2 \ - D_2) \cdot L_2 \quad [S1.10]$$

Our simulations were run on Windows Server 2019 system using the *gro* platform (https://github.com/liaupm/GRO-LIA) based on C++ language. To perform a series of simulations with a range of parameter sets and replicates, we edit a custom Mathematica script to control the automatic running of *gro* using the RobotTools package of Mathematica. The unconcerned parameters were assigned with the default values listed in Table S2. During simulations, images of simulated colony patterns at each time point were generated from *gro* using the built-in function ‘Snapshot’. Custom functions were

included in the *gro* codes to record the position coordinates of every cell, concentrations of the masses (i.e., substrate, intermediate, and products), as well as the cell number of each population at each time point. These simulation data were analyzed and visualized using custom Mathematica scripts (version 12.0) and R scripts (version 4.2.2).

S2 Genetic manipulations

We introduced the DNA fragments using derivatives of the mini-Tn7T-Gm transposon and the delivery plasmid pJM220 (Table S4). Specifically, we constructed the derivative transposons by first extracting the pJM220 plasmid from an overnight culture of *Escherichia coli* Trans5α. We next PCR amplified the *nahG* gene using DNA polymerase (2 × Phanta max master mix, Vazyme Biotech Co., China) and the oligonucleotide primers listed in Table S5. These primers contain the homologous sequences of 20 bp that we used to clone the PCR products into the pJM220 plasmid. The pJM220 plasmid was also amplified using DNA polymerase. The PCR products derived from both *nahG* gene and pJM220 plasmid were purified by DNA extraction kit (Solarbio, China). We then ligated the PCR product of *nahG* gene into a fragment of the amplified pJM220 plasmid (Hieff clone plus one step cloning kit, Yeasen Biotech Co., China). We designated the newly generated plasmid as pJM-nahG (Table S4). It was transformed into chemically competent *E. coli* Trans5α cells (TransGen Biotech) and plated onto Luria Bertani (LB) agar supplemented with 10 μg/mL gentamycin. The assembled derivative plasmid pJM-nahG was confirmed via Sanger sequencing. We used conjugative four-parental mating to deliver the pJM-nahG plasmid along with the helper pRK2013 and pTNS3 plasmids into *P. stutzeri* AN0000 following the standard protocol*(2)* (Table S3). The *P. stutzeri* exconjugant was selected by plating on LB agar plates containing 5 μg/mL of nalidixic acid and 25 μg/mL of gentamycin. We performed FRT excision of the gentamycin resistance marker using pFLP3 plasmid. The constructed *P. stutzeri* AN0000-*nahG* strain was validated by PCR and DNA sequencing, in which the *rhaSR-PrhaBAD-nahG* module was located 25-nucleotide downstream of the *glmS* gene in the chromosome.

S3 Supplementary Tables

Table S1 Definitions of variables in the IBM model

Variable	Description	Units
S_1	The concentration of substrate in [1, 0]	C-mmol
S_2	The concentration of substrate in [0, 1]	C-mmol
I_1	The concentration of intermediate in [1, 0]	C-mmol
I_2	The concentration of intermediate in [0, 1]	C-mmol
P_1	The concentration of the final product in [1, 0]	C-mmol
P_2	The concentration of the final product in [0, 1]	C-mmol
V_1	The volume of the [1, 0] cells	fL
V_2	The volume of the [0, 1] cells	fL
L_1	Length of the [1, 0] cells	μm
L_2	Length of the [0, 1] cells	μm

Table S2 Definition and value of parameters in IBM

Parameter	Description	Default value and units	Source
K_1	Michaelis-Menten constant first reaction.	0.096 C-mmol·L^{-1}	Beatriz Ca´mara*(3)*
E_1	Enzyme concentration of the first reaction.	0.001-0.06 C-mmol	Steven B. Zimmerman*(4)*
k_1	The specific rate of the first reaction.	46 min^{-1}	Beatriz Ca´mara *(3)*
K_2	Michaelis-Menten constant second reaction.	0.015 C-mmol·L^{-1}	Ambra Viggiani*(5)*
E_2	Enzyme concentration of the second reaction.	0.02 C-mmol	Steven B. Zimmerman*(4)*
k_2	The specific rate of the second reaction.	2000 min^{-1}	Ambra Viggiani*(5)*
D_S	Diffusivity coefficient for the substrate	2.0	The maximum value in *gro*
D_I	Diffusivity coefficient for the intermediate	2.0	The maximum value in *gro*
D_P	Diffusivity coefficient for the product	0.005-2	Miaoxiao Wang*(1)*
Kg	The half-saturation constant of Monod growth.	0.02 C-mmol·L^{-1}	Miaoxiao Wang*(1)*
kg	The maximum consumption rate of the final product for cell growth.	0.64 C-mmol·min^{-1}	Miaoxiao Wang*(1)*

d	Apparent maintenance rate	0.0003 min^{-1}	Miaoxiao Wang*(1)*
w	Width (diameter) of the capsular cells	1.0 μm	The default value in *gro*
Y	Yield coefficient for biomass production	0.1 C-$mmol^{-1}$	Miaoxiao Wang*(1)*
S	The initial concentration of the substrate	10 C-mmol·L^{-1}	

Table S3 Sources of the strains used in this study

Strain	Features	Source
E.coli Trans5α	F$^-$φ80 *lac ZΔM15Δ(lacZYA-arg F) U169 endA1 recA1 hsdR17(*r_k^-, m_k^+*) supE44λ$^-$ thi -1 gyrA96 relA1 phoA*	TransGen
E.coli HB101	F$^-$ *mcrB mrr hsdS20*(rB mB) *recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Smr) *glnV44*	Chao Gao
P. stutzeri AN0000	AN1111Δ*nahA::G*Δ*nahC::G*Δ*nahG::T*Δ*nahTH::A*	Miaoxiao Wang*(6)*
P. stutzeri AN0001	AN1111Δ*nahA::A*Δ*nahC::C*Δ*nahG::G*	Miaoxiao Wang*(6)*
P. stutzeri AN0010	AN1111Δ*nahA::T*Δ*nahC::A*Δ*nahTH::C*	Miaoxiao Wang*(6)*
P. stutzeri AN0000-*nahG*	AN0000 containning pJM-nahG	This study
P. stutzeri AN0000-*nahG* -eGFP2	AN0000-*nahG* containing pMMPc-eGFP2	This study
P. stutzeri AN0000-*nahG*-mCherry2	AN0000-*nahG* containing pMMPc-mCherry2	This study
P. stutzeri AN0001-eGFP2	AN0001 containing pMMPc-eGFP2	This study
P. stutzeri AN0001-mCherry2	AN0001 containing pMMPc-mCherry2	This study
P. stutzeri AN0001ΔpilAB	Deletion of *pilA* and *pilB* in *P. stutzeri* AN0001	Miaoxiao Wang*(6)*
P. stutzeri AN0000-*nahG*ΔpilAB	Deletion of *pilA* and *pilB* in *P. stutzeri* AN0000-*nahG*	This study
P. stutzeri	*P. stutzeri* AN0001ΔpilAB containing pMMPc-	This study

AN0001ΔpilAB-eGFP2	eGFP2	
P. stutzeri AN0001ΔpilAB-mCherry2	*P. stutzeri* AN0001ΔpilAB containing pMMPc-mcherry2	This study
P. stutzeri AN0000-*nahG*ΔpilAB-eGFP2	*P. stutzeri* AN0000-*nahG*ΔpilAB containing pMMPc-eGFP2	This study
P. stutzeri AN0000-*nahG*ΔpilAB-mCherry2	*P. stutzeri* AN0000-*nahG*ΔpilAB containing pMMPc-mcherry2	This study

Table S4 Sources of the plasmids used in this study

Plasmid	Features	Source
pRK2013	Helper plasmid for conjugation; Km^R	Lab of Professor Lin Min, CAAS
pMMPc-Gm	Shuttle vector between *E. coli* and *P. stutzeri*	Lab of Professor Ping Xu, SJTU
pMMPc-eGFP2	*eGFP* gene on pMMPc-Gm	This study
pMMPc-mCherry2	*mcherry* gene on pMMPc-Gm	This study
pJM220	pUC18T-miniTn7T-gm-rhaSR-PrhaBAD; Gm^R	Addgene
pJM-nahG	*nahG* gene on pJM220	This study
pTNS3	Helper plasmid; Amp^R	Addgene
pFLP3	Containing the recombinase structural gene *flp* and *SacB* selection marker; Tet^R	Addgene

Table S5 Primers used in this study

Gene	Primer name	Primer sequence (5'-3')	Application
nahG	*nahG*-Tn7-F	ACACGTTTCCAAGAGGTACCATGAACGACATGAACGCTAAGAAACCAGC	Amplication of *nahG*
	nahG-Tn7-R	GAGGTACCGGGCCCAAGCTTTTACCCTTGACGCAGCGCAAT	
pJM220 fragment	Tn7-nh-F	AAGCTTGGGCCCGGTACCT	Amplication of pJM220 fragment
	Tn7-nh-R	GGTACCTCTTGGAAACGTGTATCCTGCAG	
pJM-nahG	Tn7-G-F	GCGGTGAGCATCACATCACCACACAAAATAGTTGGGAACTGGGAGGGGT	pJM-nahG construction
	Tn7-G-R		
glmS-nahG	PglmS-up	ATGCCGATCAATGTCGGACT	pJM-nahG integration
	PglmS-down	ACTGTCTTTGCCGATCGTGAAG	

Table S6 Sensitivity analyses by Spearman rank correlation

Parameter	Value setting	*R*	*P*
Y	0.025	0.756	0.001
Y	0.05	0.874	2.09×10^{-5}
Y	0.1 (initial setting)	0.983	6.26×10^{-11}
Y	0.2	0.921	1.07×10^{-6}
Y	0.4	0.997	4.60×10^{-16}
k_2	2	0.526	0.044
k_2	20	0.932	4.46×10^{-7}
k_2	40	0.986	1.41×10^{-11}
k_2	200	0.892	7.98×10^{-6}
k_2	2000 (initial setting)	0.983	6.26×10^{-11}
kg	0.04	0.635	0.011
kg	0.16	0.667	0.007
kg	0.64 (initial setting)	0.983	6.26×10^{-11}
kg	2.56	0.984	4.46×10^{-11}

S2 Supplementary Figures

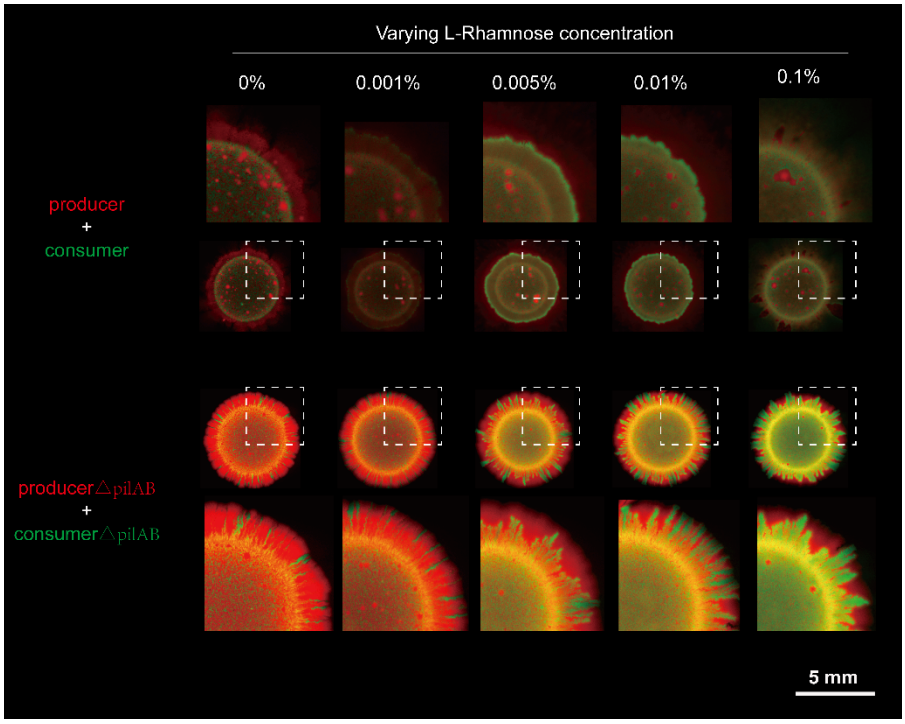

Figure S1 Expansion of co-cultures [1, 0] and [0, 1] in the absence or presence of *pilAB*. Deletion of the *pilAB* gene in *P. stutzeri* results in an increase in well-defined spatial structures. To focus on the direct effects of substrate consumption rate on spatial organization, our experiments were henceforth conducted with *P. stutzeri ΔpilAB*.

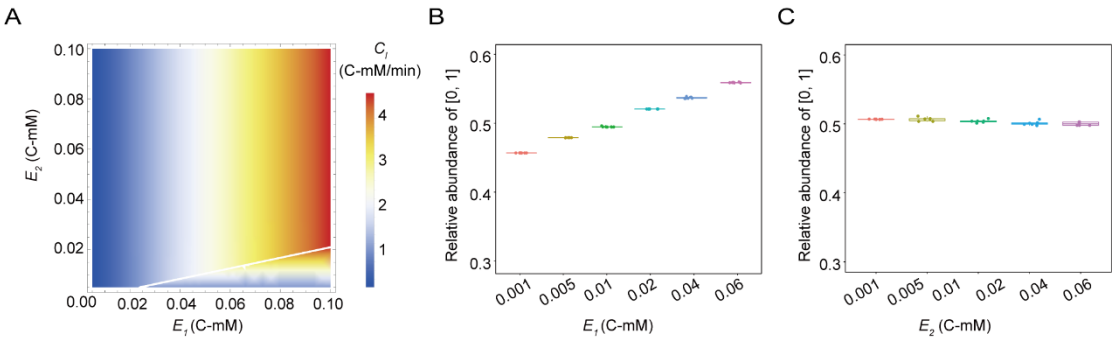

Figure S2 Effects of enzyme concentrations on the actual conversion concentration from the substrate to product per unit of time C_I and structure of MDOL communities. (A) Effects of E_1 and E_2 on the conversion rate from the substrate to product. (B) The relative abundance of [0, 1] upon the variations of E_1. Parameter values used in this simulation: $k_1 = 46$, $k_2 = 2000$, $K_1 = 0.096$, $K_2 = 0.015$, $S = 10$, $E_2 = 0.02$. (C) The relative abundance of [0, 1] upon the variations of E_2. Parameter values used in this simulation: $k_1 = 46$, $k_2 = 2000$, $K_1 = 0.096$, $K_2 = 0.015$, $S = 10$, $E_1 = 0.02$.

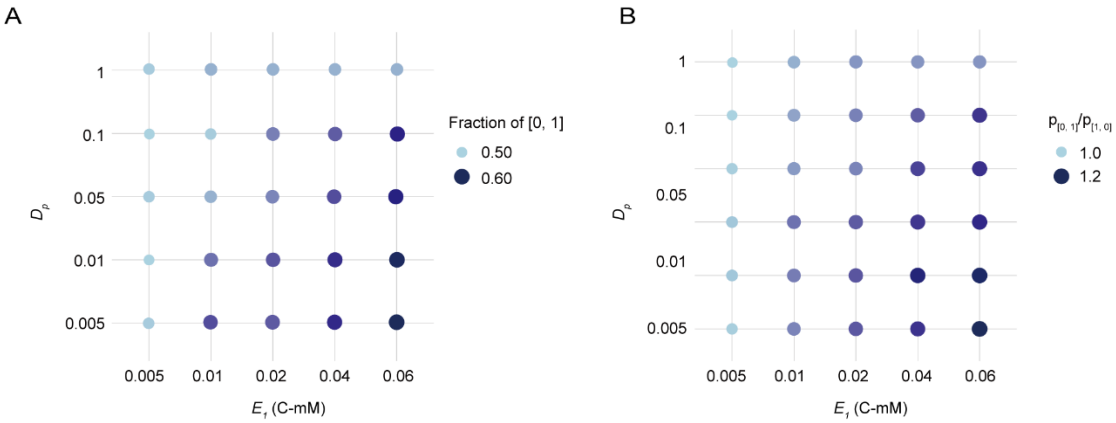

Figure S3 Analysis of community composition in the colonies from our individual-based model (IBM) using different E_1 and diffusion coefficients of the product. (A) Plot showing how both the nahG concentration and diffusion coefficients of the product affect the relative abundance of [0, 1]. (B) Plot showing how both the E_1 concentration and diffusion coefficient of the product affect the ratio of the product concentration between [0, 1] and [1, 0].

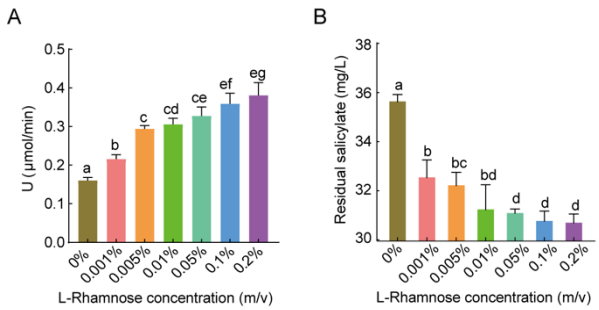

Figure S4 Enzymic activity of salicylic acid 1-hydroxylase induced by different concentrations of L-rhamnose. Sodium pyruvate was added to the MF medium as the sole carbon source. (A) The enzyme activity measurement. (B) Residual sodium salicylate concentration after the reaction of salicylic acid 1-hydroxylase with the substrate. Sodium pyruvate was added to the MF medium as the sole carbon source. n=3 biological replicates. Letters represent significant difference comparisons using the two-tailed Student's t-test.

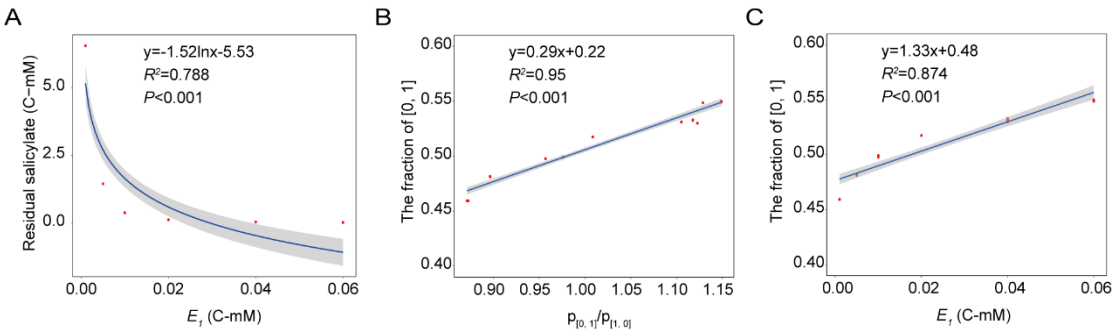

Figure S5 Simulation of the individual-based model with salicylate as the sole carbon source. (A) Residual substrate concentration on the plates decreases with increasing E_1 levels in [1, 0]. (B) Fraction of [0, 1] exhibits a positive correlation with privatization benefits of [0, 1]. The x-axis represents the ratio of the product concentration around [0, 1] to the product concentration around [1, 0]. Linear regression was fitted to the data (blue line, grey area, 95% confidence interval). (C) The relative abundance of [0, 1] is positively correlated with E_1. Linear regression was fitted to the data (blue line, grey area, 95% confidence interval). Sample size, n=6.

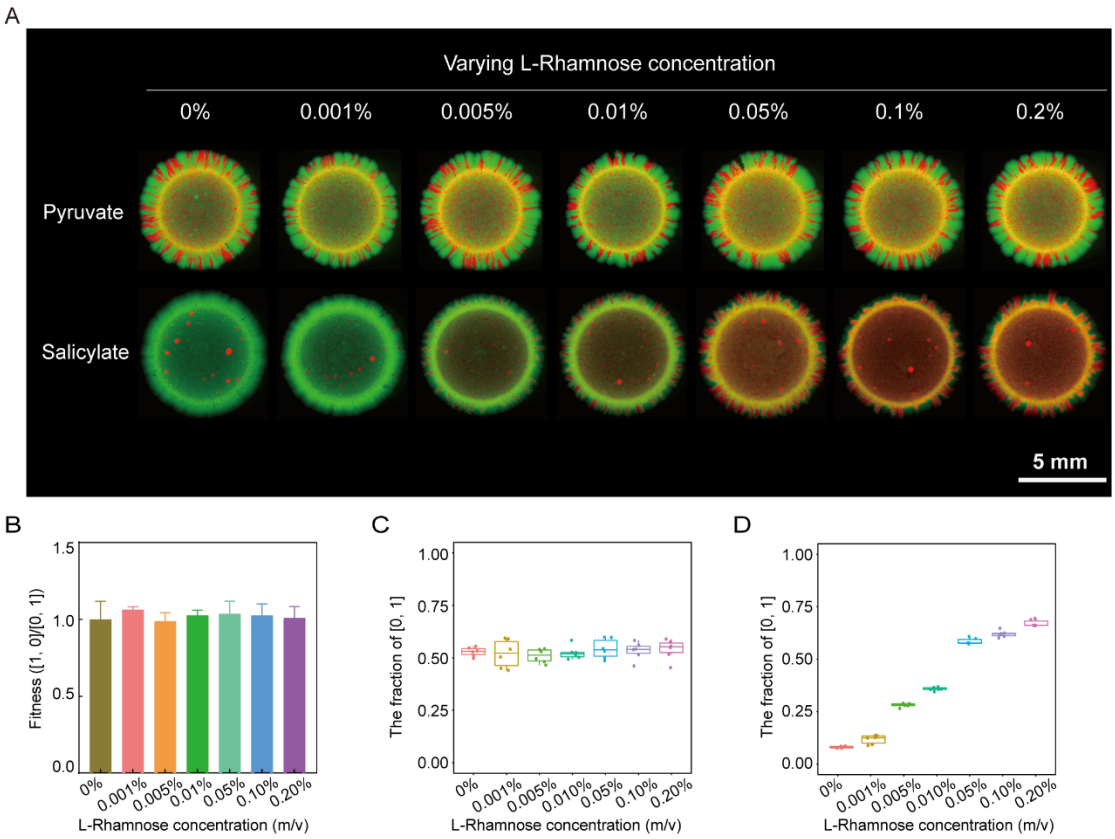

Figure S6 Effect of nahG expression levels on the community structure. Strains [1, 0] and [0, 1] carried plasmids with eGFP and mCherry, respectively. (A) Bacterial colonies were founded using a 50:50 mixture of [1, 0] and [0, 1] (10^5 *P. stutzeri* cells) and grown on minimal medium-agarose plates (1.5% agarose) with sodium salicylate (first row) or sodium pyruvate (second row) as the sole carbon source at 30°C for up to 5 days before microscopy. (B) The fitness of [1, 0] relative to [0, 1]. Statistics of the relative abundance of [0, 1] upon the variations of L-rhamnose concentration with sodium salicylate (C) or sodium pyruvate (D) as the sole carbon source. Sample size, n=6.

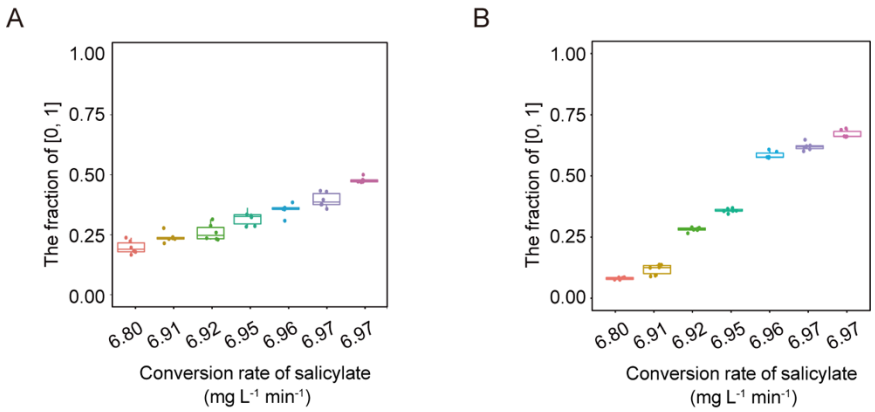

Figure S7 Effect of substrate conversion rate on the community structure in experiments. (A) Strains [1, 0] and [0, 1] carried plasmids with mCherry and eGFP, respectively. (B) Strains [1, 0] and [0, 1] carried plasmids with eGFP and mCherry, respectively. Sample size, n=6.

Reference

1. Wang, M., Chen, X., Tang, Y. Q., Nie, Y., and Wu, X. L. (2022) Substrate availability and toxicity shape the structure of microbial communities engaged in metabolic division of labor, *mLife 1*, 131-145.
2. Choi, K. H., and Schweizer, H. P. (2006) mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa, *Nat Protoc 1*, 153-161.
3. Camara, B., Bielecki, P., Kaminski, F., dos Santos, V. M., Plumeier, I., Nikodem, P., and Pieper, D. H. (2007) A gene cluster involved in degradation of substituted salicylates via ortho cleavage in Pseudomonas sp strain MT1 encodes enzymes specifically adapted for transformation of 4-methylcatechol and 3-methylmuconate, *J Bacteriol 189*, 1664-1674.
4. Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli, *J Mol Biol 222*, 599-620.
5. Viggiani, A., Siani, L., Notomista, E., Birolo, L., Pucci, P., and Di Donato, A. (2004) The role of the conserved residues His-246, His-199, and Tyr-255 in the catalysis of catechol 2,3-dioxygenase from Pseudomonas stutzeri OX1, *Journal of Biological Chemistry 279*, 48630-48639.
6. Wang, M., Chen, X., Ma, Y., Tang, Y. Q., Johnson, D. R., Nie, Y., and Wu, X. L. (2022) Type IV Pilus Shapes a 'Bubble-Burst' Pattern Opposing Spatial Intermixing of Two Interacting Bacterial Populations, *Microbiol Spectr 10*, e0194421.