

Comprehensive evaluation of the capacities of microbial cell factories

Gi Bae Kim, Ha Rim Kim and Sang Yup Lee

Correspondence should be addressed to S.Y.L. (leesy@kaist.ac.kr).

This file includes:

Supplementary Note 1. Sensitivity analysis for calculation of maximum yields
Supplementary Note 2. Improving metabolic capacities for chemical production by addition of heterologous reactions.

Supplementary Fig. 1. Maximum theoretical yields (Y_T) and maximum achievable yields (Y_A) of 235 bio-based chemicals.

Supplementary Fig. 2. Y_T and Y_A of bio-based chemicals derived from D-glucose 6-phosphate, D-glyceraldehyde 3-phosphate, and 3-phospho-D-glycerate.

Supplementary Fig. 3. Y_T and Y_A of bio-based chemicals derived from D-erythrose 4-phosphate.

Supplementary Fig. 4. Y_T and Y_A of bio-based chemicals derived from pyruvate.

Supplementary Fig. 5. Y_T and Y_A of bio-based chemicals derived from acetyl-CoA.

Supplementary Fig. 6. Y_T and Y_A of bio-based chemicals derived from 2-oxoglutarate.

Supplementary Fig. 7. Y_T and Y_A of bio-based chemicals derived from oxaloacetate.

Supplementary Fig. 8. The number of added heterologous reactions for a target chemical for each host strain.

Supplementary Fig. 9. Hierarchical clustering of 272 metabolic pathways based on Y_T ranks across host strains under aerobic condition using D-glucose as the sole carbon source.

Supplementary Fig. 10. Hierarchical clustering of 272 metabolic pathways based on Y_A ranks across host strains under aerobic condition using D-glucose as the sole carbon source.

Supplementary Fig. 11. L-lysine biosynthetic pathways for host strains.

Supplementary Fig. 12. Hierarchical clustering of predicted heterologous reaction sets across host strains.

Supplementary Fig. 13. Hierarchical clustering of predicted heterologous reaction sets using iML1515 across chemicals.

Supplementary Fig. 14. Flux-sum differences of metabolites between native pathway and NADH-dependent homoserine dehydrogenase, homoserine deaminase, and pyruvate carboxylase introduced pathway for Y_T calculation.

Supplementary Fig. 15. The distribution of percentage increases Y_T with the number of added heterologous reactions.

Supplementary Fig. 16. Hierarchical clustering of NAD(P)H cofactor usage ratios in *E. coli* across different conditions.

Supplementary Fig. 17. Metabolic pathways for chemicals in clusters I and J of the FVSEOF-based cluster map.

Supplementary Fig. 18. Hierarchical clustering of 272 metabolic pathways using metabolic reaction profiles predicted by FVSEOF analysis.

Supplementary Fig. 19. Metabolic pathways for chemicals in the cluster A of the FVSEOF-based cluster map.

Supplementary Fig. 20. Metabolic pathways for chemicals in the cluster B of the FVSEOF-based cluster map.

Supplementary Fig. 21. Metabolic pathways for chemicals in the cluster C of the FVSEOF-based cluster map.

Supplementary Fig. 22. Metabolic pathways for chemicals in the cluster D of the FVSEOF-based cluster map.

Supplementary Fig. 23. Metabolic pathways for chemicals in the cluster E of the FVSEOF-based cluster map.

Supplementary Fig. 24. Metabolic pathways for chemicals in the cluster F of the FVSEOF-based cluster map.

Supplementary Fig. 25. Metabolic pathways for chemicals in the cluster G of the FVSEOF-based cluster map.

Supplementary Fig. 26. Metabolic pathways for chemicals in the cluster H of the FVSEOF-based cluster map.

Supplementary Fig. 27. Metabolic pathways for chemicals in the cluster I of the FVSEOF-based cluster map.

Supplementary Fig. 28. Metabolic pathways for chemicals in the cluster J of the FVSEOF-based cluster map.

Supplementary Fig. 29. Hierarchical clustering of 272 metabolic pathways using metabolic reaction profiles predicted by iBridge analysis.

Supplementary Fig. 30. Metabolic pathways for chemicals in the cluster A of the iBridge-based cluster map.

Supplementary Fig. 31. Metabolic pathways for chemicals in the cluster B of the iBridge-based cluster map.

Supplementary Fig. 32. Metabolic pathways for chemicals in the cluster C of the iBridge-based cluster map.

Supplementary Fig. 33. Metabolic pathways for chemicals in the cluster D of the iBridge-based cluster map.

Supplementary Fig. 34. Metabolic pathways for chemicals in the cluster E of the iBridge-based cluster map.

Supplementary Fig. 35. Metabolic pathways for chemicals in the cluster F of the iBridge-based cluster map.

Other supplementary material for this manuscript includes the following:

Supplementary Data 1. Maximum yields for bio-based chemicals in *Bacillus subtilis* (iYO844).

Supplementary Data 2. Maximum yields for bio-based chemicals in *Corynebacterium glutamicum* (iCW773).

Supplementary Data 3. Maximum yields for bio-based chemicals in *Escherichia coli* (iML1515).

Supplementary Data 4. Maximum yields for bio-based chemicals in *Pseudomonas putida* (iJN1463).

Supplementary Data 5. Maximum yields for bio-based chemicals in *Saccharomyces cerevisiae* (Yeast8).

Supplementary Data 6. Heterologous reactions which improve maximum theoretical yield for bio-based chemical production in *Bacillus subtilis* (iYO844).

Supplementary Data 7. Heterologous reactions which improve maximum theoretical yield for bio-based chemical production in *Corynebacterium glutamicum* (iCW773).

Supplementary Data 8. Heterologous reactions which improve maximum theoretical yield for bio-based chemical production in *Escherichia coli* (iML1515).

Supplementary Data 9. Heterologous reactions which improve maximum theoretical yield for bio-based chemical production in *Pseudomonas putida* (iJN1463).

Supplementary Data 10. Heterologous reactions which improve maximum theoretical yield for bio-based chemical production in *Saccharomyces cerevisiae* (Yeast8).

Supplementary Data 11. Cofactor swapping target reactions which improve maximum theoretical yield for bio-based chemical production in *Bacillus subtilis* (iYO844).

Supplementary Data 12. Cofactor swapping target reactions which improve maximum theoretical yield for bio-based chemical production in *Corynebacterium glutamicum* (iCW773).

Supplementary Data 13. Cofactor swapping target reactions which improve maximum theoretical yield for bio-based chemical production in *Escherichia coli* (iML1515).

Supplementary Data 14. FVSEOF simulation results for *Bacillus subtilis* (iYO844).

Supplementary Data 15. FVSEOF simulation results for *Corynebacterium glutamicum* (iCW773).

Supplementary Data 16. FVSEOF simulation results for *Escherichia coli* (iML1515).

Supplementary Data 17. FVSEOF simulation results for *Pseudomonas putida* (iJN1463).

Supplementary Data 18. FVSEOF simulation results for *Saccharomyces cerevisiae* (Yeast8).

Supplementary Data 19. iBridge simulation results for *Bacillus subtilis* (iYO844).

Supplementary Data 20. iBridge simulation results for *Corynebacterium glutamicum* (iCW773).

Supplementary Data 21. iBridge simulation results for *Escherichia coli* (iML1515).

Supplementary Data 22. iBridge simulation results for *Pseudomonas putida* (iJN1463).

Supplementary Data 23. iBridge simulation results for *Saccharomyces cerevisiae* (Yeast8).

Supplementary Data 24. Abbreviations for metabolites and reactions.

Supplementary Data 25. Cosine similarities of metabolic reaction profiles calculated by FVSEOF simulation performed for *E. coli* under aerobic conditions with D-glucose as the sole carbon source.

Supplementary Data 26. Cosine similarities of metabolic reaction profiles calculated by iBridge simulation performed for *E. coli* under aerobic conditions with D-glucose as the sole carbon source.

Supplementary Data 27. Maximum achievable yields for bio-based chemicals in *Escherichia coli* with 5% threshold.

Supplementary Data 28. Maximum achievable yields for bio-based chemicals in *Escherichia coli* with 15% threshold

Supplementary Data 29. Maximum achievable yields for bio-based chemicals in *Escherichia coli* using eciML1515.

Supplementary Data 30. Maximum achievable yields for bio-based chemicals in *Escherichia coli* using iJO1366.

Supplementary Data 31. Heterologous reactions which improve maximum achievable yield for bio-based chemical production in *Bacillus subtilis* (iYO844).

Supplementary Data 32. Heterologous reactions which improve maximum achievable yield for bio-based chemical production in *Corynebacterium glutamicum* (iCW773).

Supplementary Data 33. Heterologous reactions which improve maximum achievable yield for bio-based chemical production in *Escherichia coli* (iML1515).

Supplementary Data 34. Heterologous reactions which improve maximum achievable yield for bio-based chemical production in *Pseudomonas putida* (iJN1463).

Supplementary Data 35. Heterologous reactions which improve maximum achievable yield for bio-based chemical production in *Saccharomyces cerevisiae* (Yeast8).

Supplementary Data 36. Cofactor swapping target reactions which improve maximum achievable yield for bio-based chemical production in *Bacillus subtilis* (iYO844).

Supplementary Data 37. Cofactor swapping target reactions which improve maximum achievable yield for bio-based chemical production in *Corynebacterium glutamicum* (iCW773).

Supplementary Data 38. Cofactor swapping target reactions which improve maximum achievable yield for bio-based chemical production in *Escherichia coli* (iML1515).

Supplementary Data 39. Hierarchical clustering of 272 metabolic pathways based on Y_T ranks across host strains under aerobic condition using D-glucose as the sole carbon source.

Supplementary Data 40. Hierarchical clustering of 272 metabolic pathways based on Y_A ranks across host strains under aerobic condition using D-glucose as the sole carbon source.

Supplementary Data 41. Hierarchical clustering of predicted heterologous reaction sets across host strains.

Supplementary Data 42. Hierarchical clustering of predicted heterologous reaction sets using iML1515 across chemicals.

Supplementary Data 43. Hierarchical clustering of NAD(P)H cofactor usage ratios in *E. coli* across different conditions.

Supplementary Notes

Supplementary Note 1. Sensitivity analysis for calculation of maximum yields

To evaluate the robustness of calculated maximum achievable yields (Y_{AS}), we performed a sensitivity analysis by varying the lower bound of the specific growth rate used to account for minimal growth. Using iML1515, a genome-scale metabolic model (GEM) of *Escherichia coli*, we calculated Y_A for chemical production under aerobic conditions using D-glucose as the carbon source, setting the lower bound at 5%, 10%, and 15% of the maximum biomass production rate (Supplementary Data 27 and 28). Statistical analyses of the calculated Y_{AS} showed that, although the absolute Y_A values shift with the choice of threshold, the overall distributions remain largely unchanged. For example, Y_{AS} of chemical production calculated at the 10% growth rate threshold exhibit a Pearson correlation coefficient of 0.9991 (p-value < 0.001, $n = 272$) with those calculated at a 5% threshold. A two-sided Kolmogorov-Smirnov test indicates that the distributions are statistically indistinguishable (p-value of 0.2010, $n = 272$). Additionally, a one-sided Wilcoxon signed-rank test showed that the Y_{AS} of chemical production calculated at the 10% threshold are generally lower than those calculated with a 5% threshold (p-value < 0.001, $n = 272$). A similar pattern was observed when comparing the 15% threshold to the 10% threshold, with high correlation (Pearson correlation coefficient of 0.9991, p-value < 0.001, $n = 272$), indistinguishable distributions (two-sided Kolmogorov-Smirnov test, p-value of 0.2010, $n = 272$), and a statistically significant reduction in absolute values (one-sided Wilcoxon signed-rank test, p-value < 0.001, $n = 272$). These results suggest that while increasing the minimum growth requirement leads to a systematic decrease in absolute Y_A , the overall trends remain consistent.

To assess the influence of enzyme kinetics and proteome allocation on maximum yield calculations, we compared results from iML1515 with those from eciML1515, an enzyme-

constrained counterpart of iML1515. For the proteome-integrated model, we used the ecModel_batch version, wherein enzyme usage reactions are connected to a common protein pool. The model parameters were calibrated for aerobic conditions using D-glucose as the carbon source based on experimental measurements. Because the target chemical production flux does not exhibit a strict linear relationship with the carbon source uptake flux, we simulated the minimum required carbon source uptake fluxes across a range of target chemical production fluxes. Maximum yields were then calculated as the highest ratio of target chemical production flux to carbon source uptake flux, rather than relying on a single loopless flux balance analysis run (Supplementary Data 29). Additionally, to evaluate the sensitivity of predictions to the choice of GEM, we calculated maximum yields using iJO1366, a precedent GEM of *E. coli* (Supplementary Data 30). This allowed us to examine the effects of incorporating additional enzyme kinetics and proteome allocation information, as well as the impact of potentially missing information in older models, on maximum yield calculations.

The analyses show that Y_{TS} calculated by eciML1515 are generally higher than those calculated by iML1515 (one-sided Wilcoxon signed-rank test; p -value < 0.001 , $n = 272$), while Y_{AS} from eciML1515 are generally lower than those from iML1515 (one-sided Wilcoxon signed-rank test; p -value < 0.001 , $n = 272$). Despite these shifts in absolute values, maximum yields from eciML1515 are highly correlated with those from iML1515 (Pearson correlation coefficient of 0.8955 for Y_T and 0.8742 for Y_A ; both p -value < 0.001 and $n = 272$), and the overall distributions of maximum yields are statistically indistinguishable between the two models (two-sided Kolmogorov-Smirnov tests, p -value of 1.0000 and 0.5916 for Y_T and Y_A , respectively; $n = 272$ for both cases). Likewise, a similar trend was observed when comparing maximum yields between iML1515 and iJO1366. Although both Y_T and Y_A from iJO1366 are generally higher, their distributions are well correlated (Pearson correlation coefficients of 0.9971 for Y_T and 0.9961 for Y_A ; both p -value < 0.001 and $n = 272$) and statistically

indistinguishable (two-sided Kolmogorov-Smirnov tests, p-value of 1.0000 and 0.9545 for Y_T and Y_A , respectively; $n = 272$ for both cases). Collectively, these sensitivity analyses demonstrate that our conclusions regarding comparative metabolic capacities are reproducible and not highly sensitive to the specific choice of growth rate threshold or to the use of proteome-constrained models.

Supplementary Note 2. Improving metabolic capacities for chemical production by addition of heterologous reactions and modifying the use of cofactors in metabolic reactions

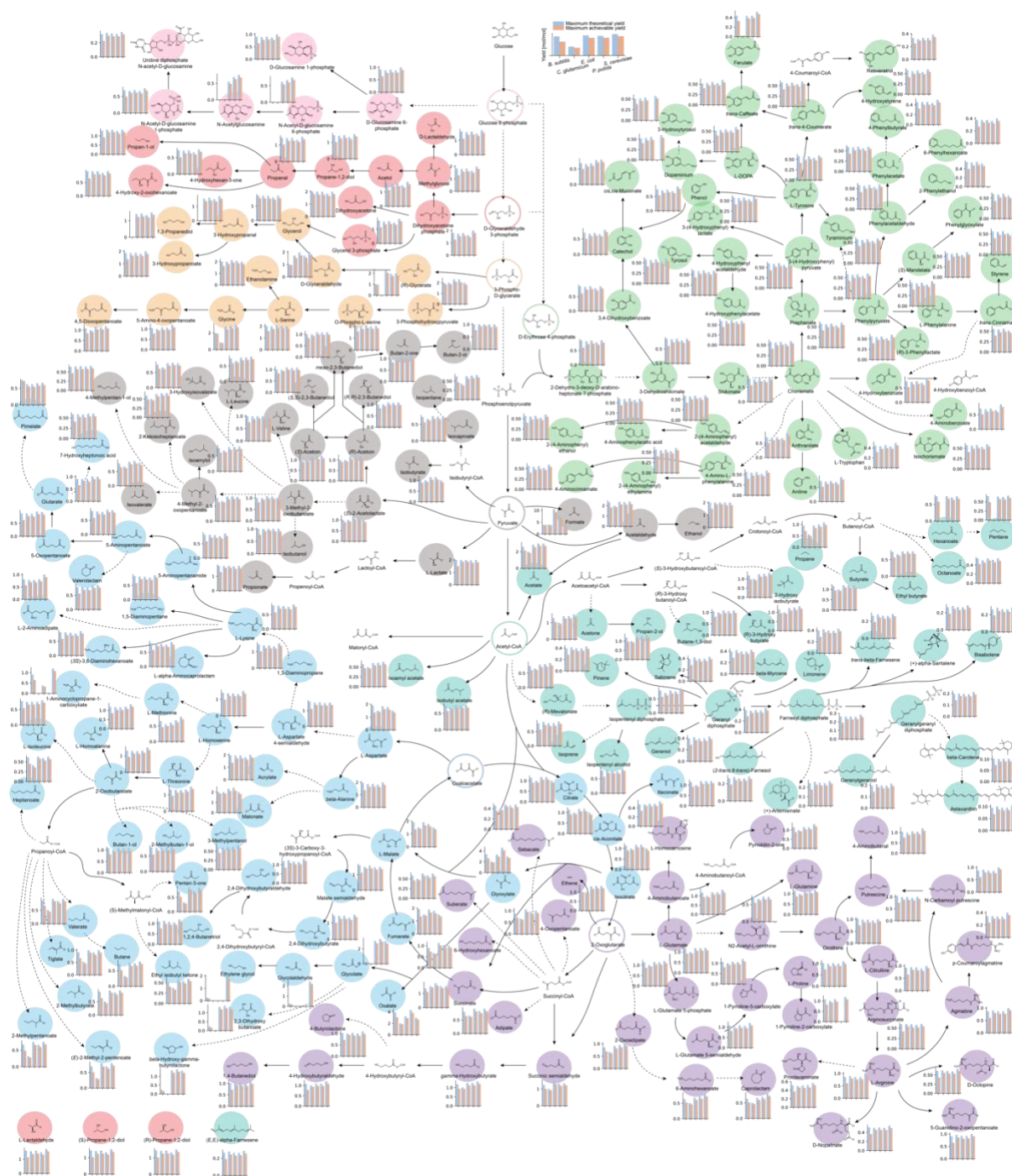
We performed *in silico* simulations to find yield-improving heterologous reactions for all 1,360 constructed GEMs. Although the simulation was performed to predict sets of heterologous reactions that improve maximum theoretical yield (Y_T), the predicted reactions also improved both Y_T and maximum achievable yield (Y_A) of chemical production. For example, the addition of ornithine transacetylase, which catalyzes the conversion of L-glutamate to ornithine, improved Y_T of ornithine production in *E. coli*, increasing it from 0.9817 mol/mol D-glucose to 1.0142 mol/mol D-glucose (3.31% increase) while Y_A was increased from 0.8632 mol/mol D-glucose to 0.8962 mol/mol D-glucose (3.82% increase) under aerobic conditions. Similarly, the addition of NADH-dependent homoserine dehydrogenase, homoserine deaminase, and pyruvate carboxylase improved Y_T of propan-1-ol production in *E. coli*, increasing it from 0.7059 mol/mol D-glucose to 1.0909 mol/mol D-glucose (54.55% increase) while Y_A was increased from 0.5003 mol/mol D-glucose to 0.7736 mol/mol D-glucose (54.62% increase) under anaerobic conditions.

We also analyzed the effect of cofactor exchanges in the 1,360 constructed GEMs to find yield-improving target reactions. As before, although the simulation was performed to predict cofactor exchanges that improve Y_T , the predicted reactions also improved both Y_T

and Y_A of chemical production. For example, when the cofactor (NAD) of the glyceraldehyde 3-phosphate dehydrogenase was swapped for NADP, Y_T of mevalonic acid production in *E. coli* was increased from 0.8000 mol/mol D-glucose to 0.8229 mol/mol D-glucose (2.86% increase) while Y_A of mevalonic acid production was increased from 0.7143 mol/mol D-glucose to 0.7369 mol/mol D-glucose (3.17% increase) under aerobic conditions. Similarly, when the cofactor (NADPH) of the glutamate dehydrogenase was exchanged for NADH, Y_T of putrescine production in *E. coli* was increased from 0.9907 mol/mol D-glucose to 1.0221 mol/mol D-glucose (3.16% increase) while Y_A of putrescine production was increased from 0.8712 mol/mol D-glucose to 0.9030 mol/mol D-glucose (3.64% increase) under aerobic conditions.

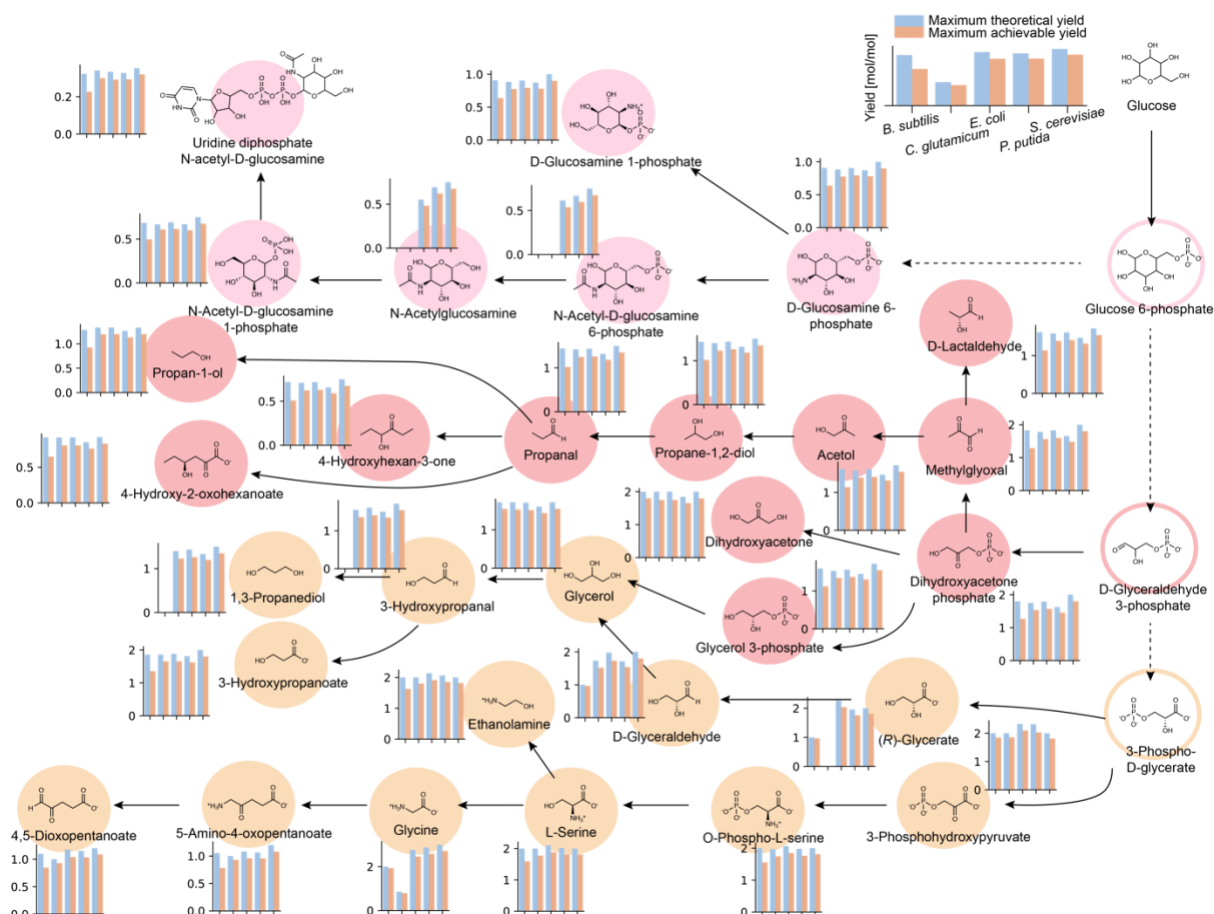
To provide metabolic engineering strategies for improving Y_A of chemicals for all 1,360 constructed GEMs, *in silico* simulations were performed again to find Y_A -improving heterologous reactions and cofactor exchange targets (Supplementary Data 31-38). The identified engineering strategies include 614,438 heterologous reaction targets and 21,220 cofactor exchange targets, of which 460,966 heterologous reaction targets and 18,091 cofactor exchange targets also overlapped with the identified strategies for improving Y_T .

Supplementary Figures

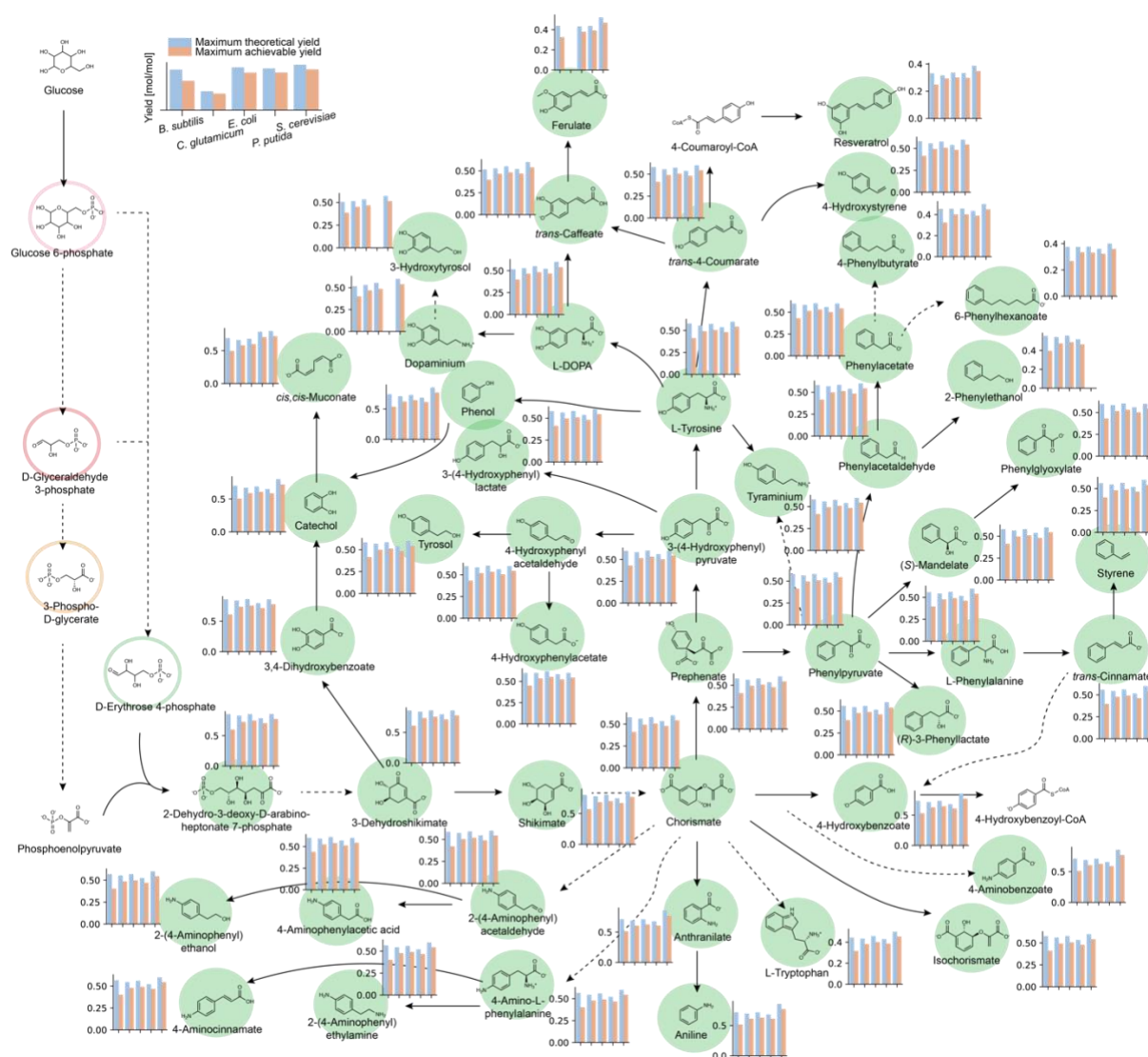


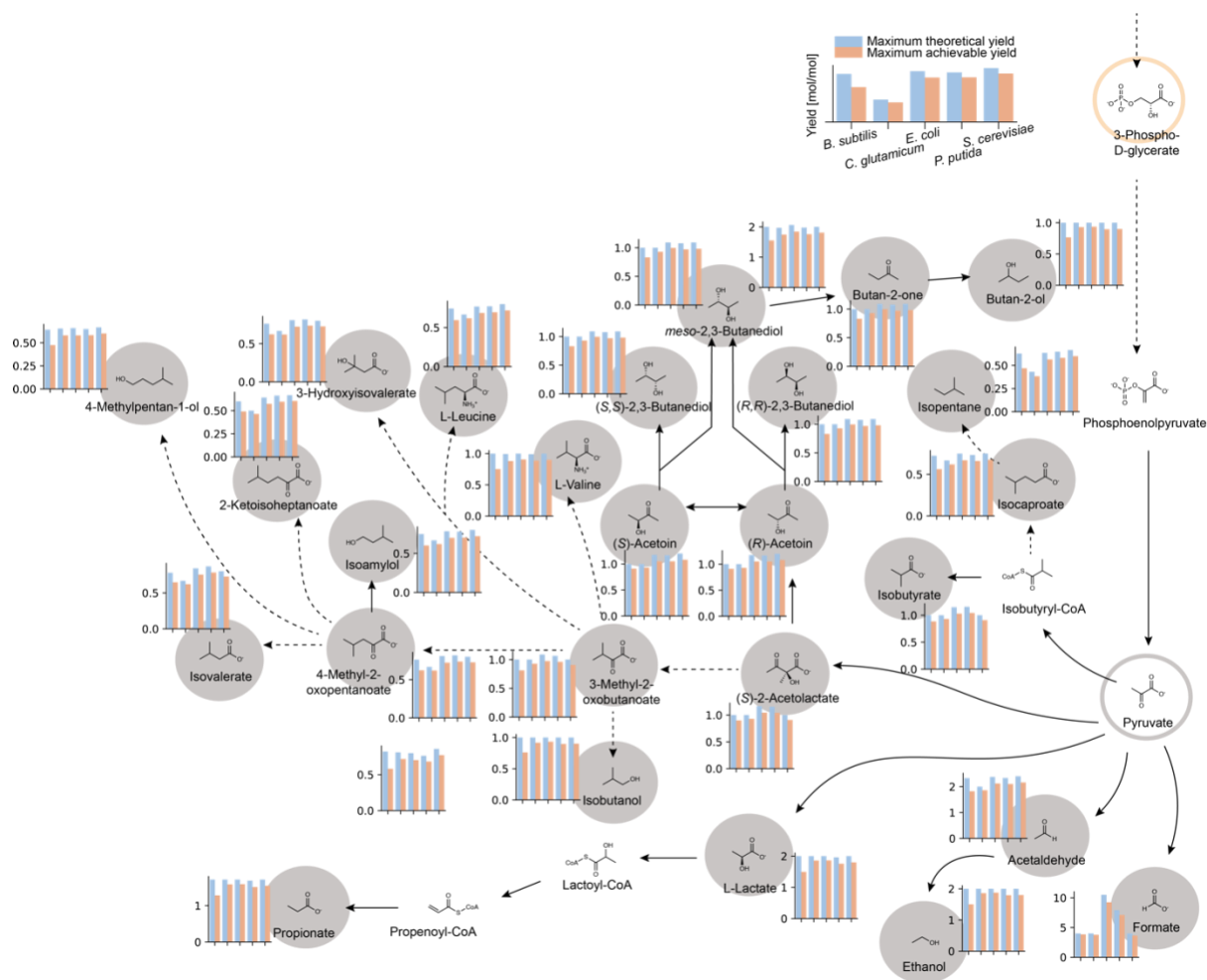
Supplementary Fig. 1. Maximum theoretical yields (Y_T) and maximum achievable yields (Y_A) of 235 bio-based chemicals. The maximum yields were calculated under aerobic conditions with D-glucose as the sole carbon source. Chemicals derived from 2-oxoglutarate, 3-phospho-D-glycerate, acetyl-CoA, D-erythrose 4-phosphate, D-glucose 6-phosphate, D-glyceraldehyde

3-phosphate, oxaloacetate, and pyruvate were colored in purple, orange, cyan, green, pink, red, blue, and gray, respectively. The bar graph adjacent to each chemical represents the Y_T (blue bar) and Y_A (orange bar) of the chemical. Metabolic pathways for each chemical group and their maximum yields are available in Supplementary Figs. 12-17.

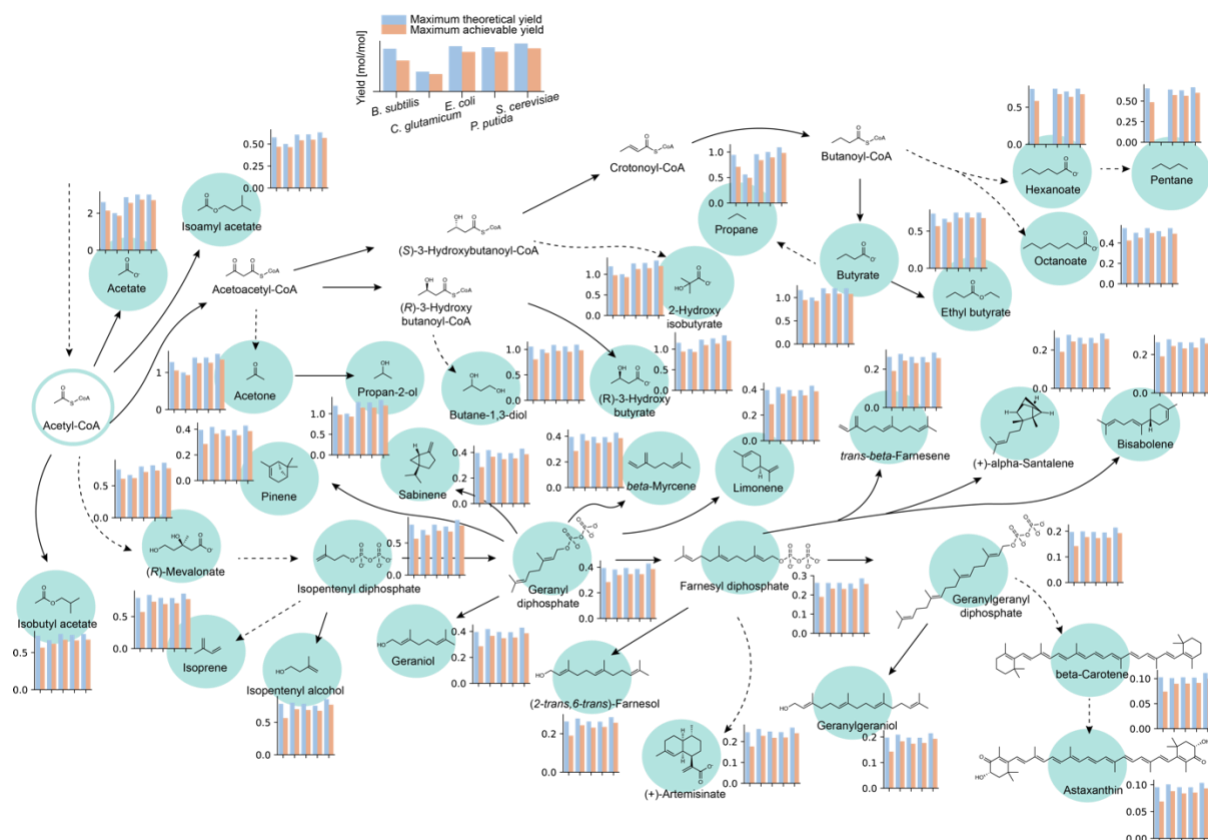


Supplementary Fig. 2. Y_T and Y_A of bio-based chemicals derived from D-glucose 6-phosphate, D-glyceraldehyde 3-phosphate, and 3-phospho-D-glycerate. Chemicals derived from D-glucose 6-phosphate, D-glyceraldehyde 3-phosphate, and 3-phospho-D-glycerate were colored in pink, red, and orange, respectively. The bar graph adjacent to each chemical represents the Y_T (blue bar) and Y_A (orange bar) of the chemical.

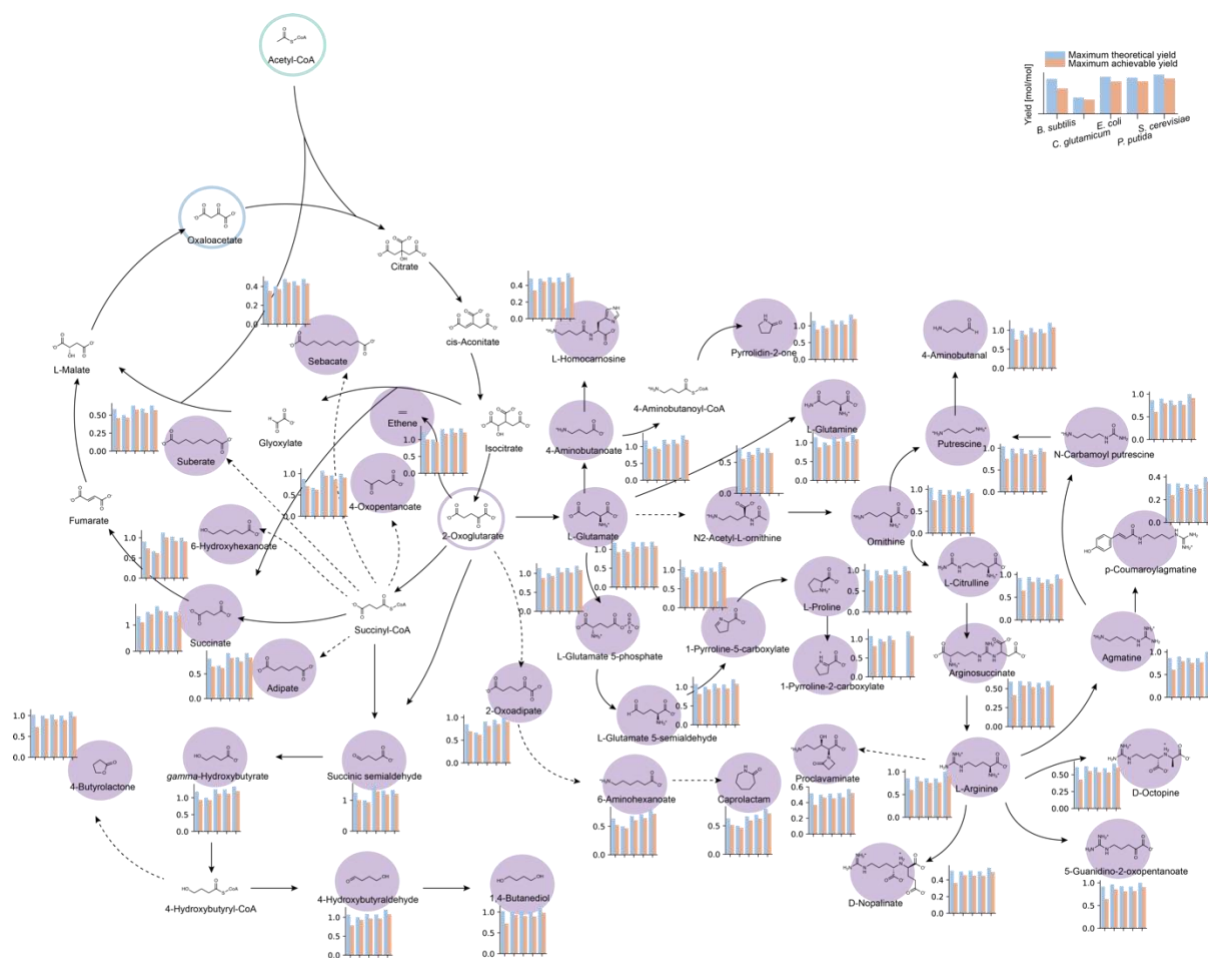




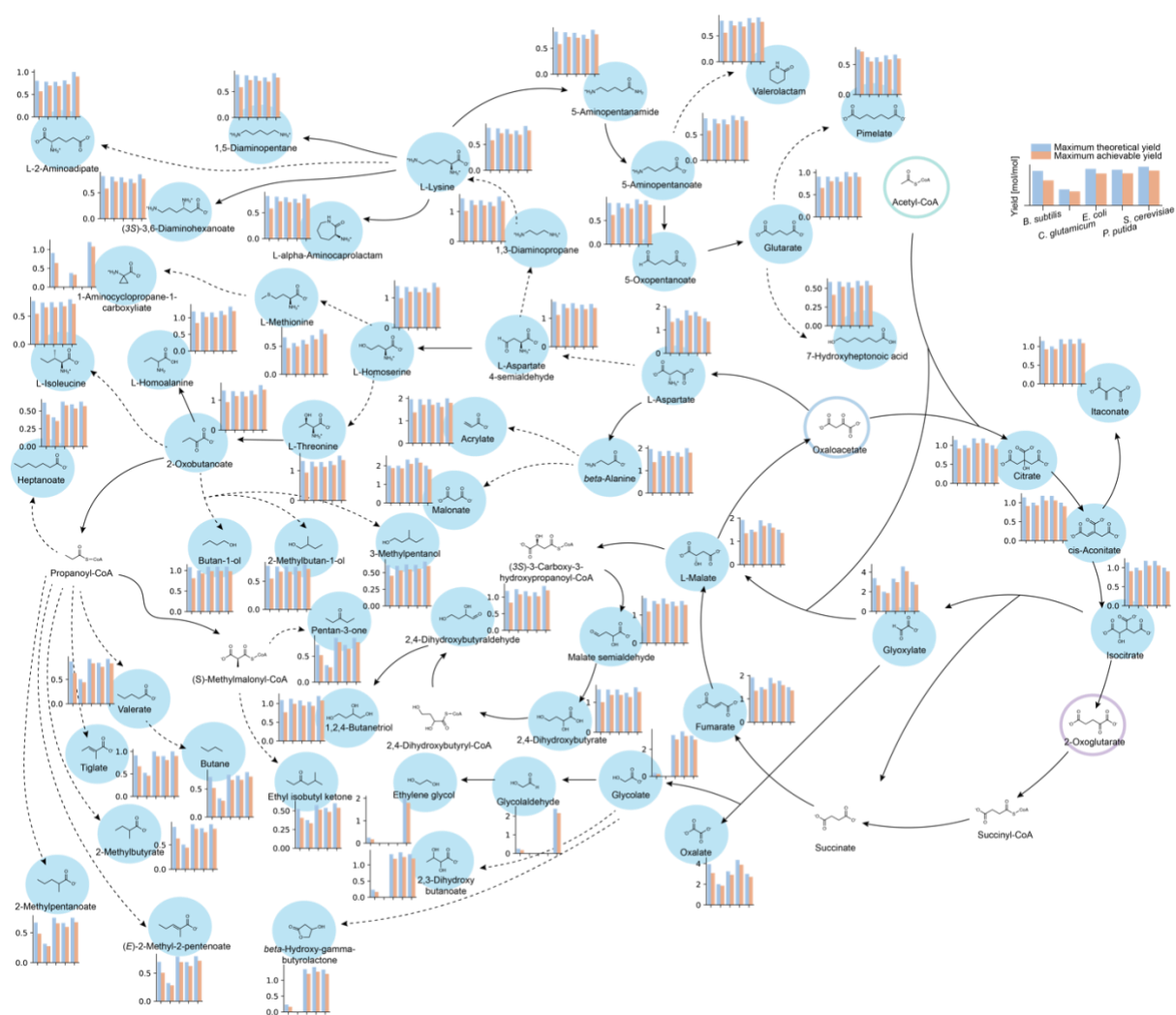
Supplementary Fig. 4. Y_T and Y_A of bio-based chemicals derived from pyruvate. The bar graph adjacent to each chemical represents the Y_T (blue bar) and Y_A (orange bar) of the chemical.



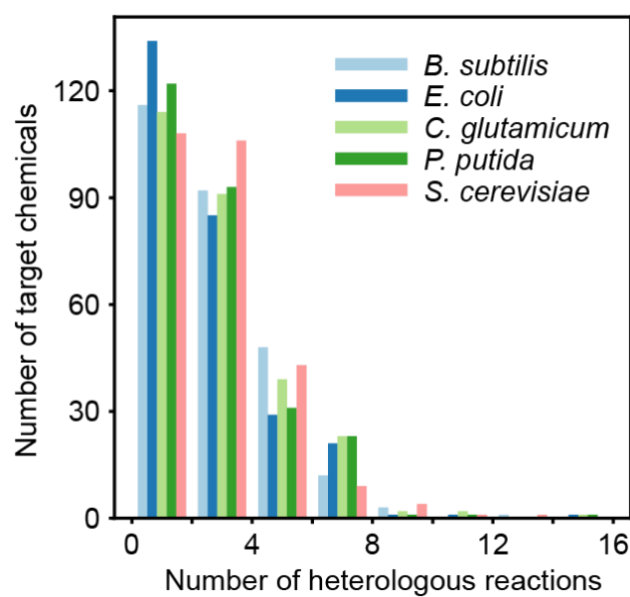
Supplementary Fig. 5. Y_T and Y_A of bio-based chemicals derived from acetyl-CoA. The bar graph adjacent to each chemical represents the Y_T (blue bar) and Y_A (orange bar) of the chemical.



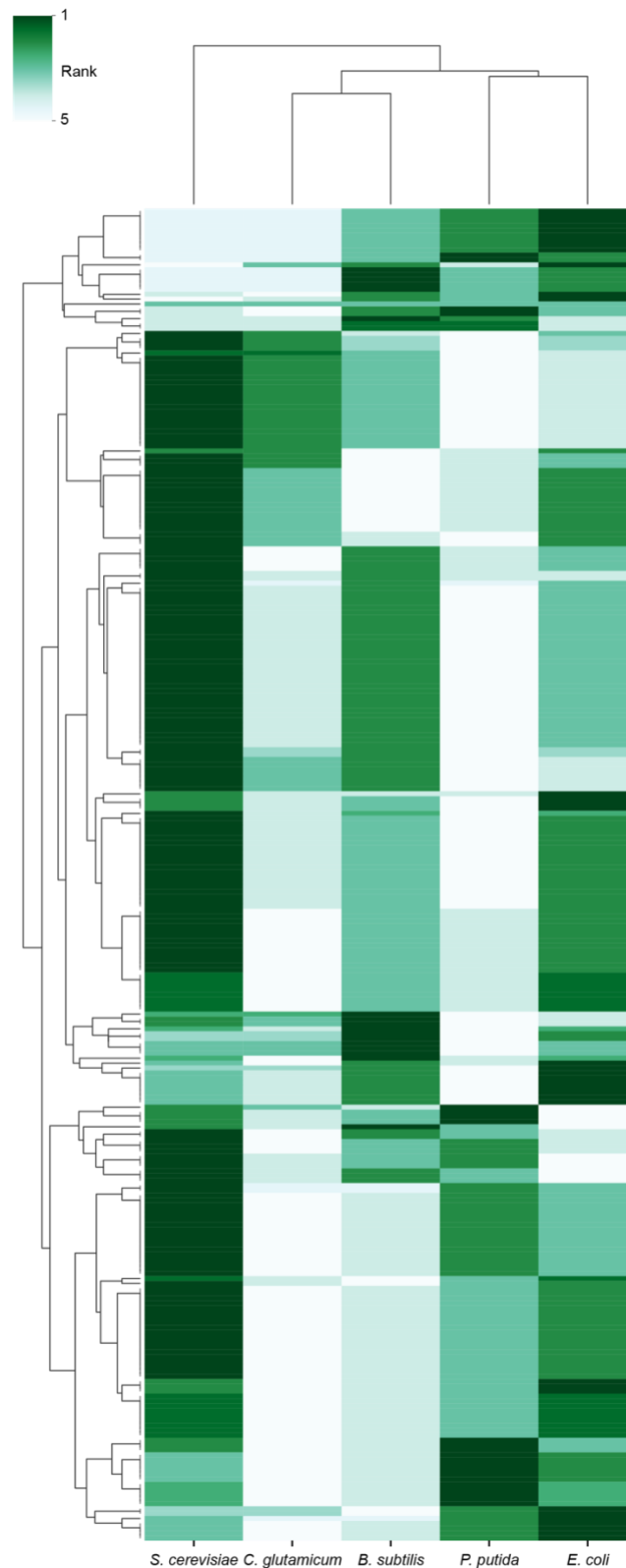
Supplementary Fig. 6. Y_T and Y_A of bio-based chemicals derived from 2-oxoglutarate. The bar graph adjacent to each chemical represents the Y_T (blue bar) and Y_A (orange bar) of the chemical.



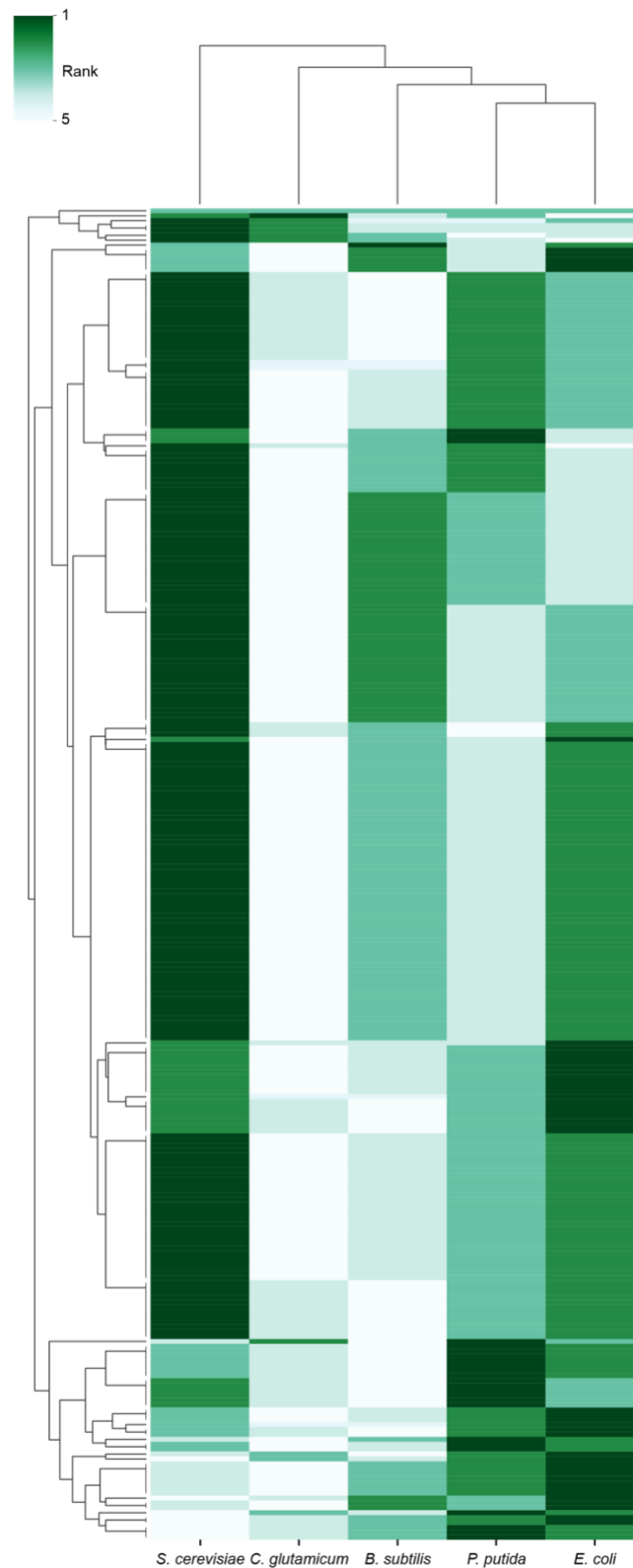
Supplementary Fig. 7. Y_T and Y_A of bio-based chemicals derived from oxaloacetate. The bar graph adjacent to each chemical represents the Y_T (blue bar) and Y_A (orange bar) of the chemical.



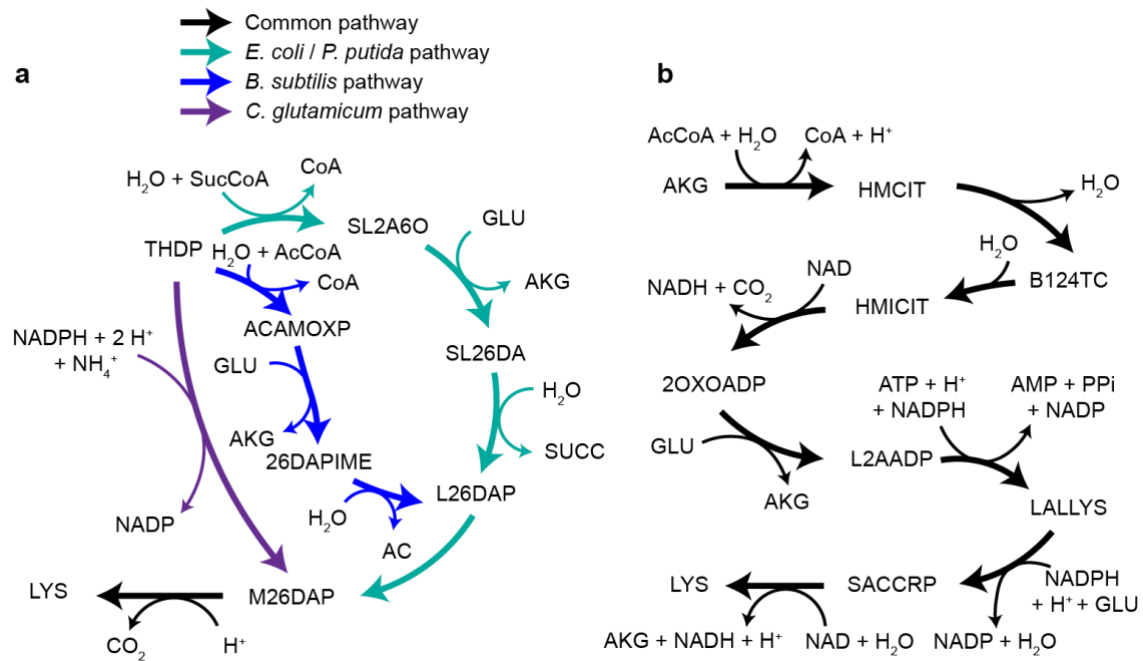
Supplementary Fig. 8. The number of added heterologous reactions for a target chemical for each host strain. Source data are provided as a Source Data file.



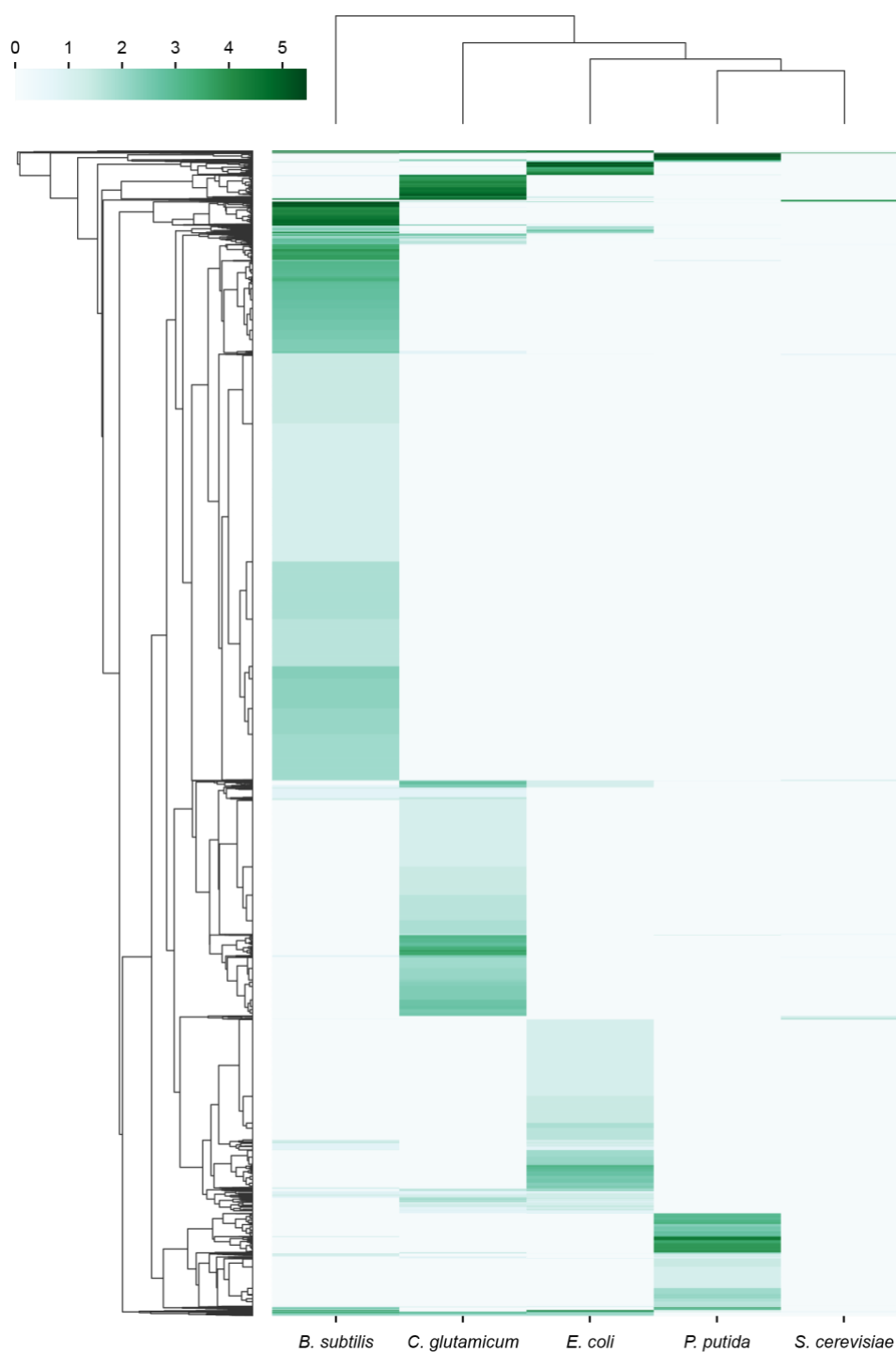
Supplementary Fig. 9. Hierarchical clustering of 272 metabolic pathways based on Y_T ranks across host strains under aerobic condition using D-glucose as the sole carbon source. The detailed rank and metabolic pathway lists are available in Supplementary Data 39.



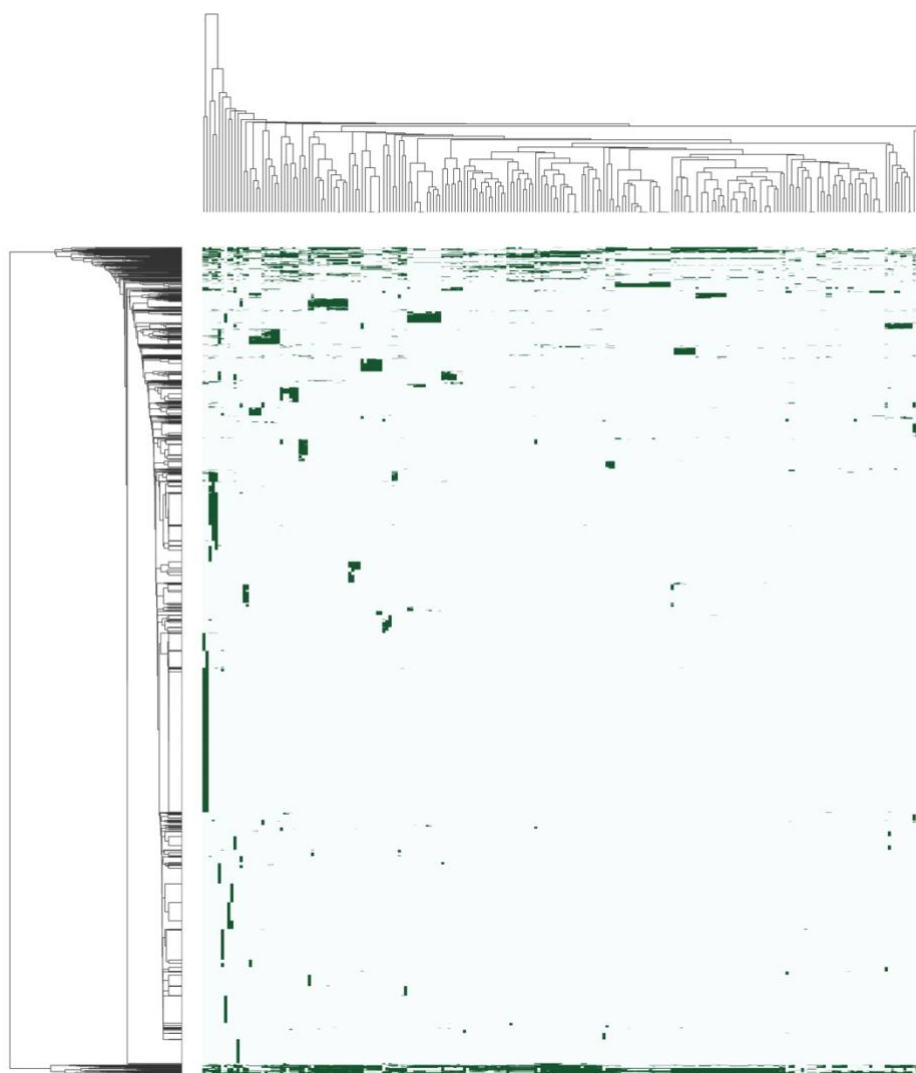
Supplementary Fig. 10. Hierarchical clustering of 272 metabolic pathways based on Y_A ranks across host strains under aerobic condition using D-glucose as the sole carbon source. The detailed rank and metabolic pathway lists are available in Supplementary Data 40.



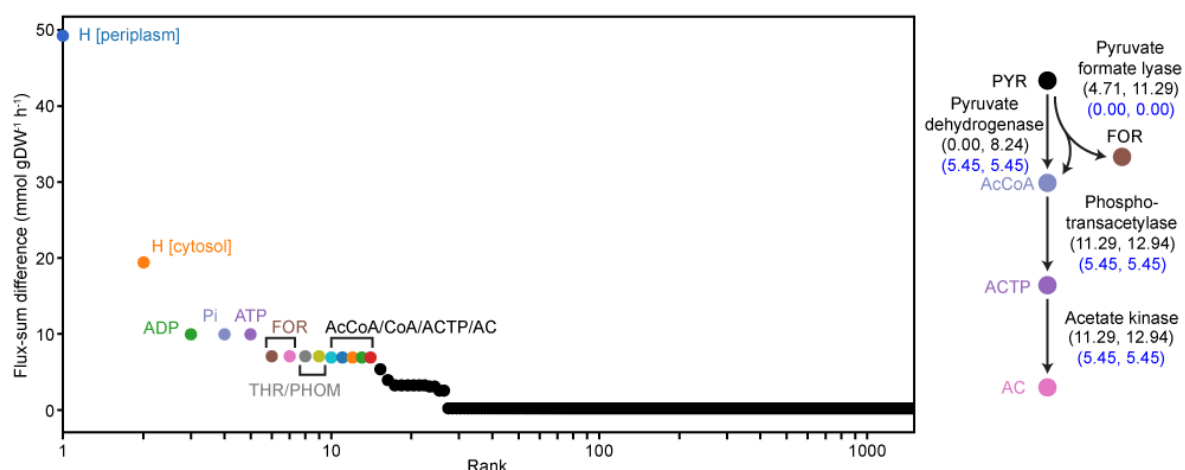
Supplementary Fig. 11. L-lysine biosynthetic pathways for host strains. (a) Diaminopimelate pathway for the biosynthesis of L-lysine. Black arrows are the common pathway for *B. subtilis*, *C. glutamicum*, *E. coli*, and *P. putida* while mint arrows are for *E. coli* and *P. putida*, blue arrows are for *B. subtilis*, and purple arrows are for *C. glutamicum*, respectively. (b) L-2-Aminoadipate pathway for the biosynthesis of L-lysine used in *S. cerevisiae*.



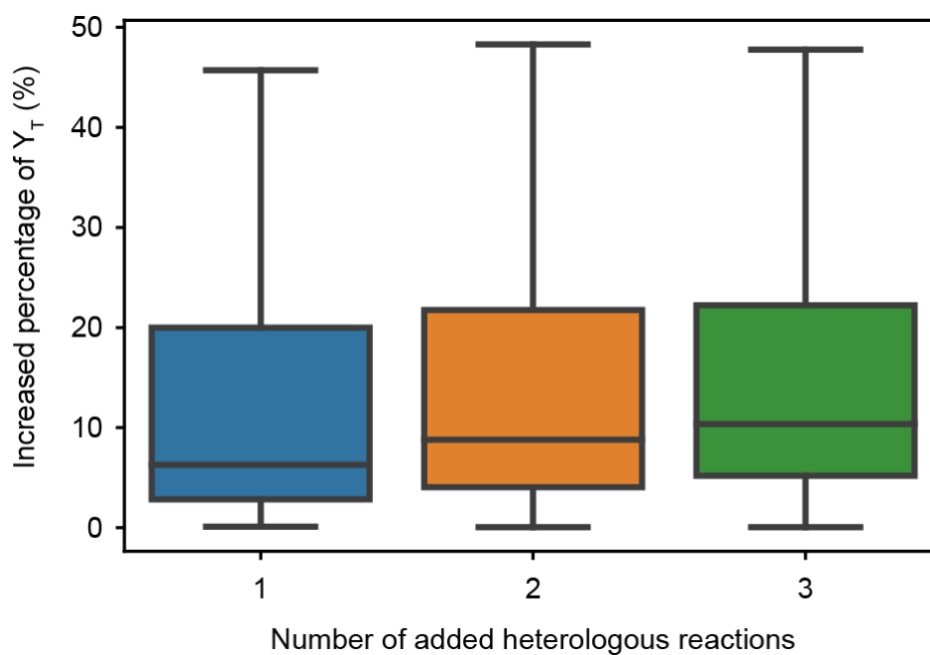
Supplementary Fig. 12. Hierarchical clustering of predicted heterologous reaction sets across host strains. Among 26,690 heterologous reaction sets, 17,317 reaction sets that were predicted at least twice were analyzed, with frequency normalized by log1p transformation. Detailed frequencies of heterologous reaction sets are available in Supplementary Data 41.



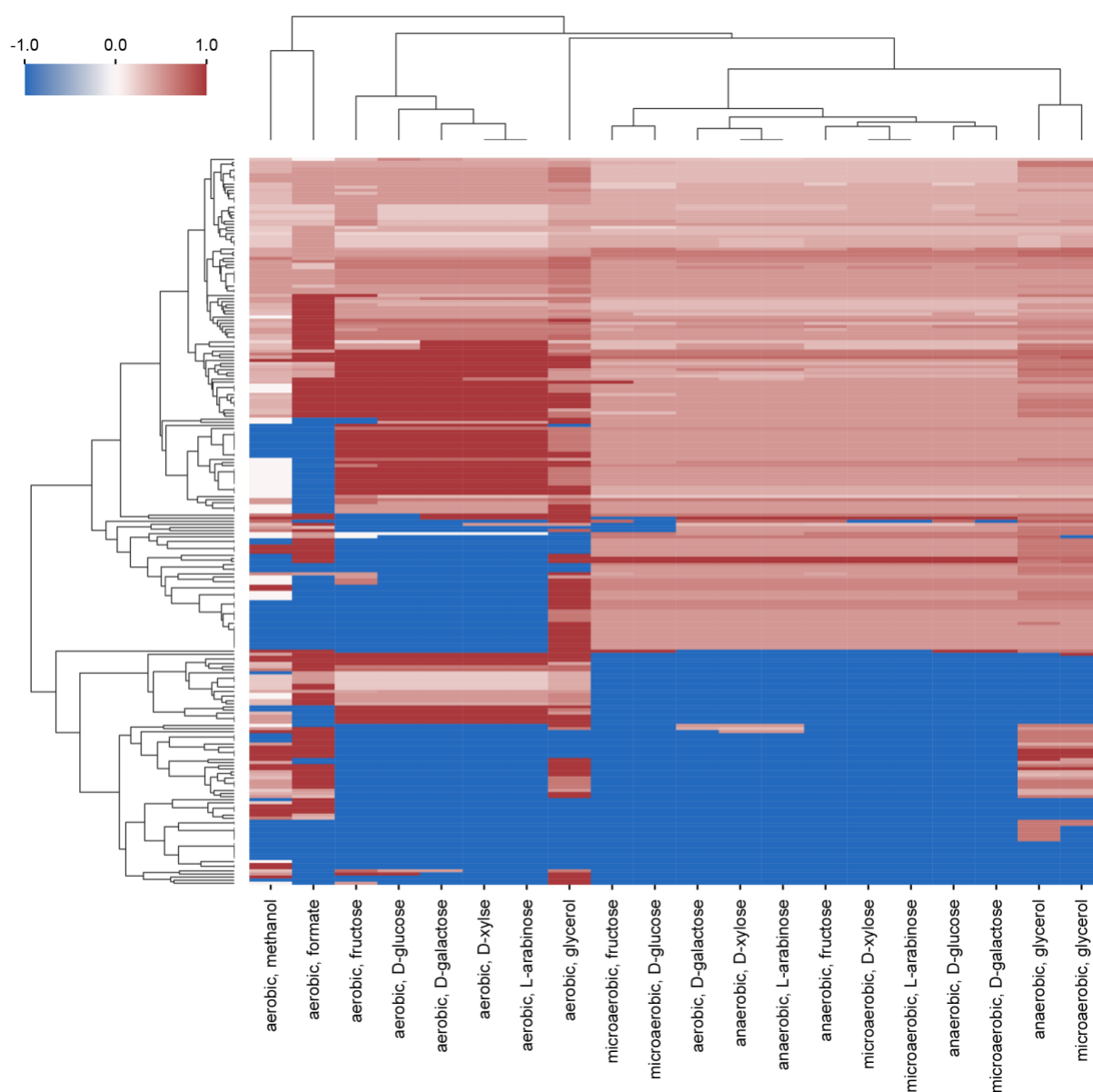
Supplementary Fig. 13. Hierarchical clustering of predicted heterologous reaction sets using iML1515 across chemicals. Binary indicators of heterologous reaction set predictions were used to cluster bio-based chemicals and heterologous reactions. Detailed binary values are available in Supplementary Data 42.



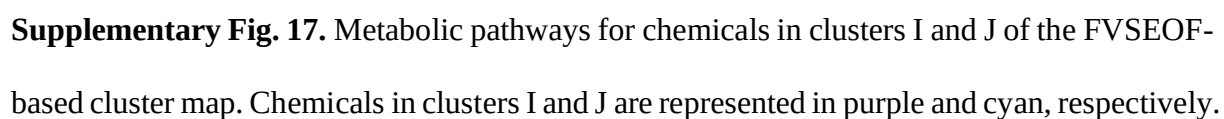
Supplementary Fig. 14. Flux-sum differences of metabolites between native pathway and NADH-dependent homoserine dehydrogenase, homoserine deaminase, and pyruvate carboxylase introduced pathway for Y_T calculation. Metabolites are sorted in descending order of flux-sum decrease. Flux values in the parentheses are the lower bound and upper bound of FVA result when maximizing the target chemical production flux. FVA results for native pathways and heterologous reaction introduced pathways are in black and blue, respectively. Source data are provided as a Source Data file.

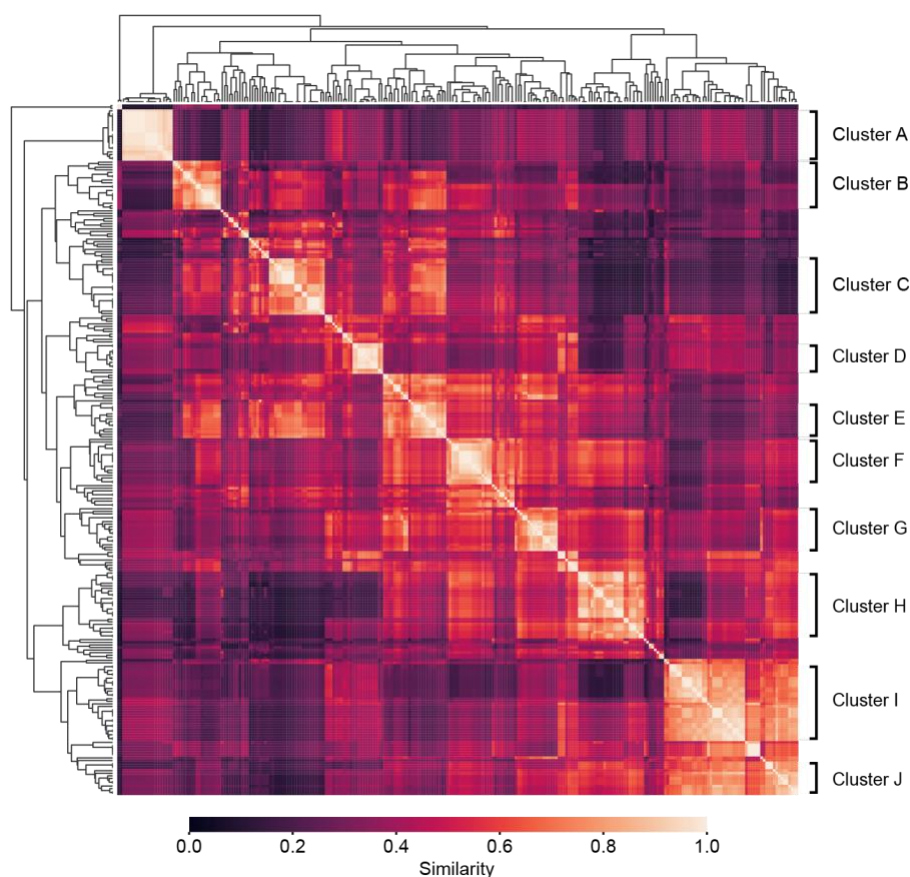


Supplementary Fig. 15. The distribution of percentage increases Y_T with the number of added heterologous reactions. Center line, box limits, and whiskers, represent median, upper quartiles, lower quartiles, and $1.5 \times$ interquartile range, respectively. Source data are provided as a Source Data file.

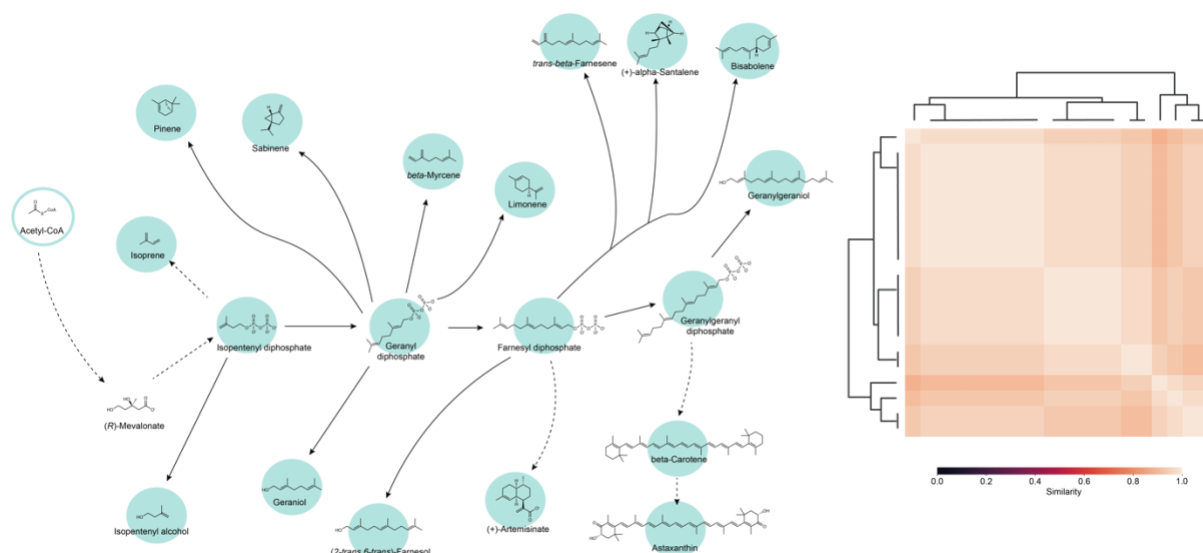


Supplementary Fig. 16. Hierarchical clustering of NAD(P)H cofactor usage ratios in *E. coli* across different conditions. The rows and columns represent bio-based chemicals and cultivation conditions. The fraction of NADPH usage, that is $\text{NADPH} / (\text{NADPH} + \text{NADH})$, was used for the analysis. If no cofactor exchange target was predicted, the -1 was assigned for the value. Detailed fractions of NADPH usage for each chemical are provided in Supplementary Data 43.

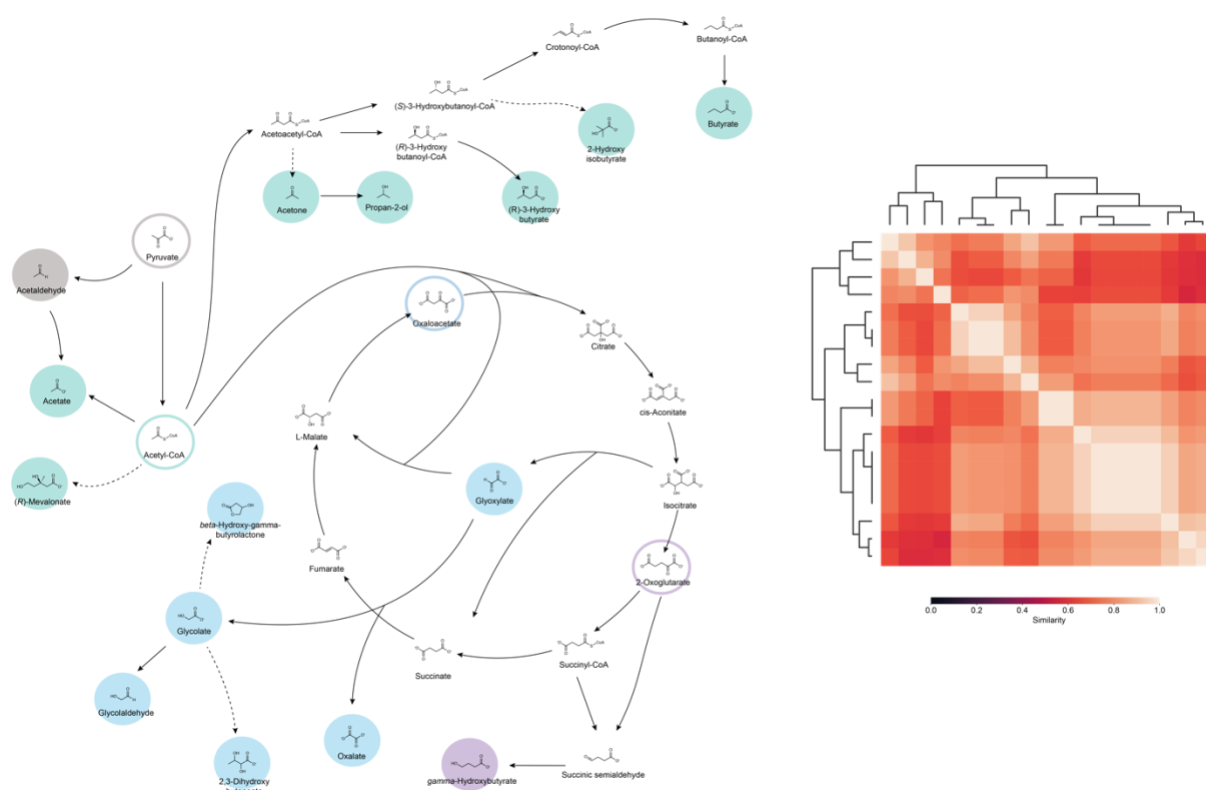




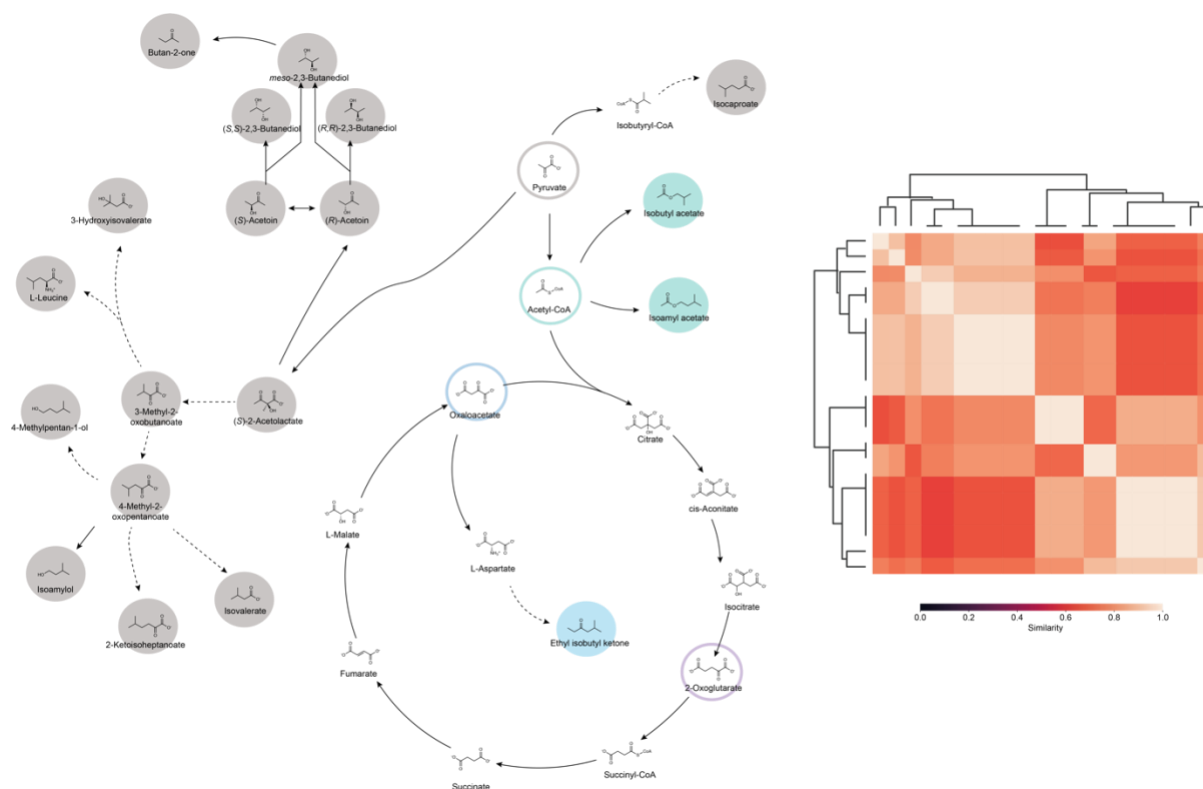
Supplementary Fig. 18. Hierarchical clustering of 272 metabolic pathways using metabolic reaction profiles predicted by FVSEOF analysis. Metabolic reaction profiles, representing the types of predicted metabolic reactions to be up-regulated for improved chemical production, were analyzed across 272 metabolic pathways which lead to the production of 235 chemicals. Pairwise similarities of metabolic reaction profiles were calculated using cosine similarity. High similarity between metabolic reaction profiles is indicated by bright red, while low similarity is indicated by dark red. Chemicals with similar metabolic reaction profiles were grouped into distinct clusters (A-J). The simulation was performed for *E. coli* under aerobic conditions with D-glucose as the sole carbon source. Source data are provided as a Source Data file.



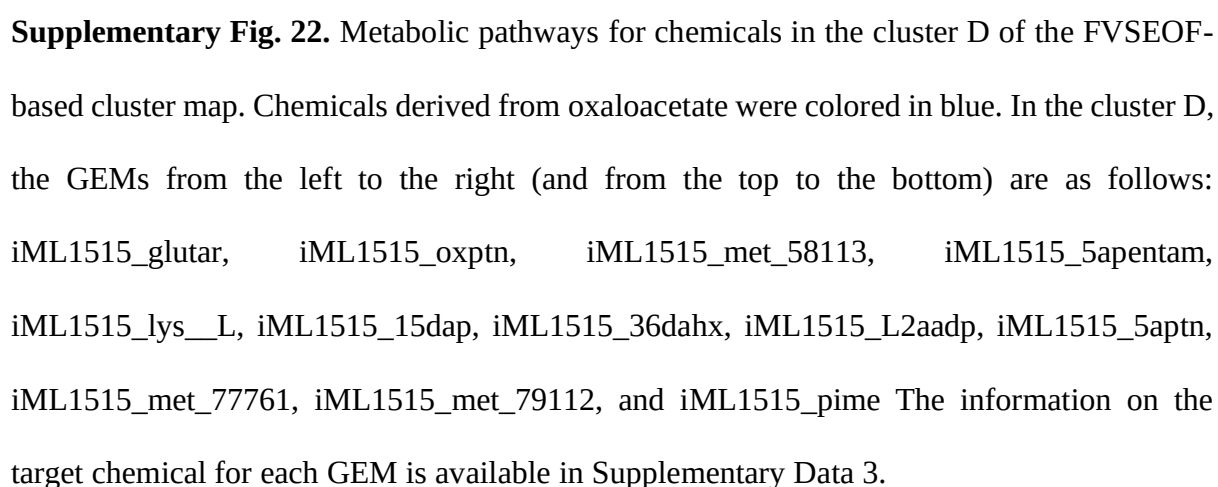
Supplementary Fig. 19. Metabolic pathways for chemicals in the cluster A of the FVSEOF-based cluster map. Chemicals derived from acetyl-CoA were colored in cyan. In the cluster A, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_met_64782, iML1515_met_61677, iML1515_met_49243, iML1515_met_46762, iML1515_met_40968, iML1515_met_16619, iML1515_met_10418, iML1515_caro, iML1515_met_10280, iML1515_met_50029, iML1515_met_28660, iML1515_met_17447, iML1515_limnen, iML1515_met_17221, iML1515_met_35194, iML1515_met_62898, iML1515_ipdp, iML1515_grdp, iML1515_frdp, and iML1515_ggdp. The information on the target chemical for each GEM is available in Supplementary Data 3.

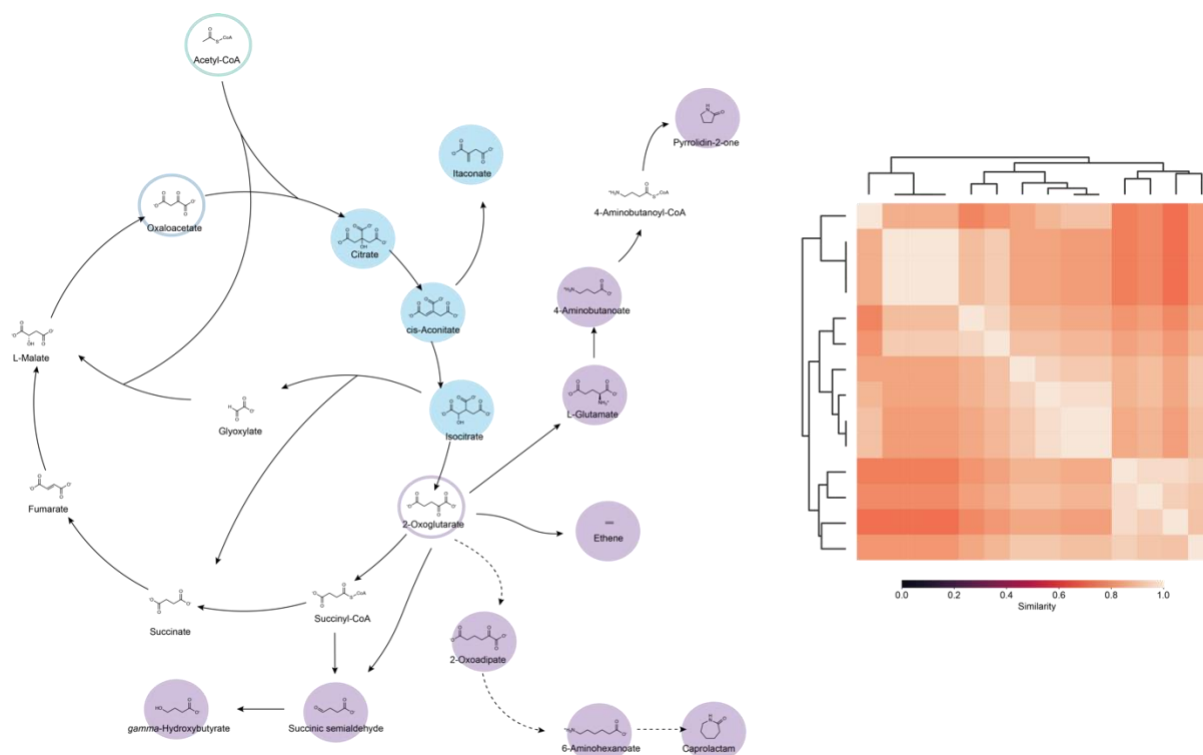


Supplementary Fig. 20. Metabolic pathways for chemicals in the cluster B of the FVSEOF-based cluster map. Chemicals derived from 2-oxoglutarate, acetyl-CoA, oxaloacetate, and pyruvate were colored in purple, cyan, blue, and gray, respectively. In the cluster B, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_bhb, iML1515_mev_R, iML1515_acald, iML1515_but, iML1515_ac, iML1515_acetone, iML1515_acetone_path_1, iML1515_2ppoh, iML1515_met_19641, iML1515_ghb_path_1, iML1515_ghb_path_2, iML1515_gcald, iML1515_met_HBA_path_1, iML1515_met_HBA, iML1515_met_DHB23, iML1515_met_DHB23_path_1, iML1515_glyclt, iML1515_glx, and iML1515_oxa. The information on the target chemical for each GEM is available in Supplementary Data 3.

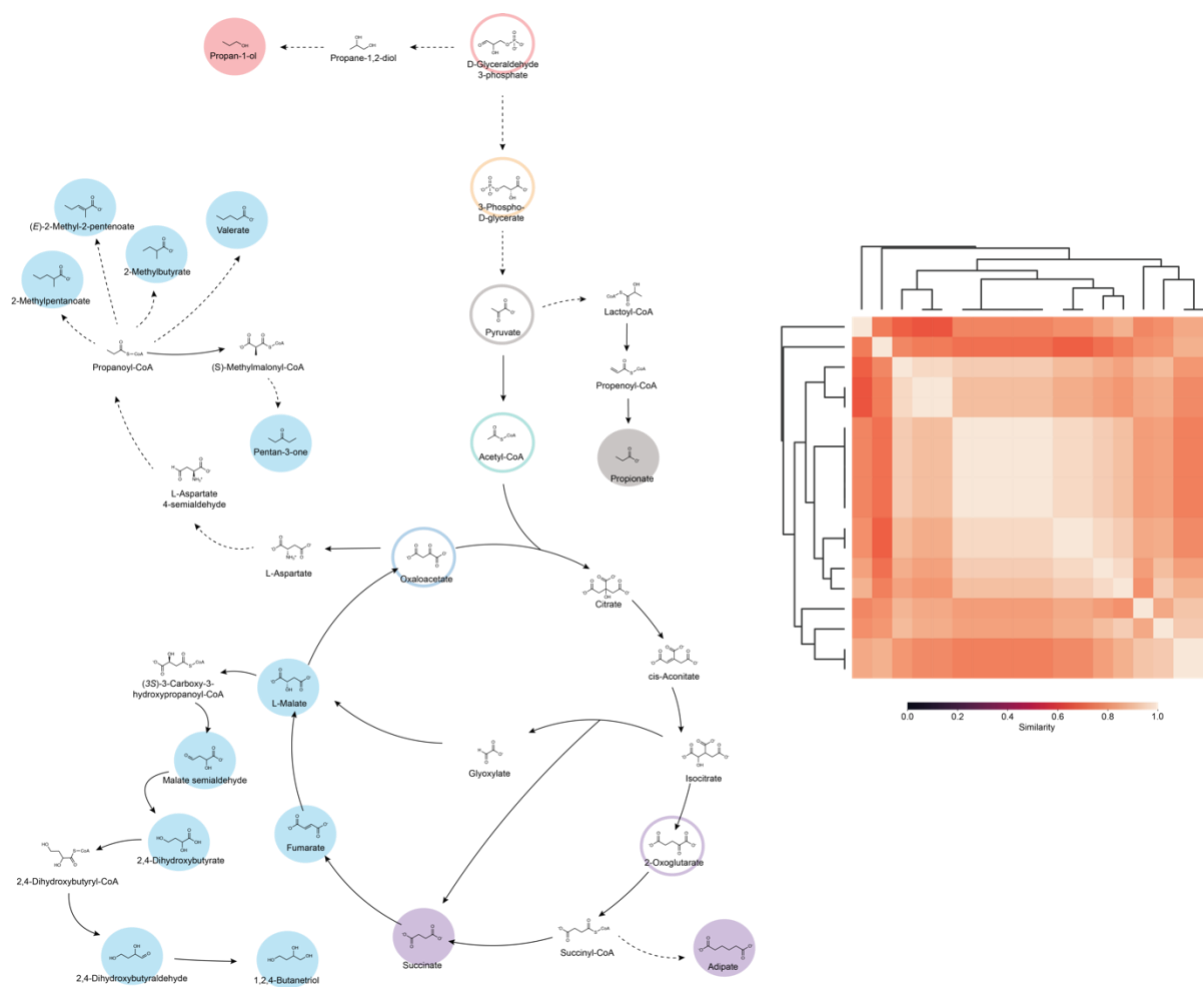


Supplementary Fig. 21. Metabolic pathways for chemicals in the cluster C of the FVSEOF-based cluster map. Chemicals derived from acetyl-CoA, oxaloacetate, and pyruvate were colored in cyan, blue, and gray, respectively. In the cluster C, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_leu__L, iML1515_met_89209, iML1515_3mb, iML1515_met_5M2OPT, iML1515_met_74904_path_1, iML1515_met_63910, iML1515_iamoh, iML1515_iamac, iML1515_4mop, iML1515_CE2028, iML1515_met_15687, iML1515_actn__R, iML1515_alac__S, iML1515_ibutac, iML1515_met_74904, iML1515_met_75460_path_1, iML1515_met_75460, iML1515_met_28398_path_1, iML1515_btd_RR, iML1515_met_16812, and iML1515_3mob. The information on the target chemical for each GEM is available in Supplementary Data 3.

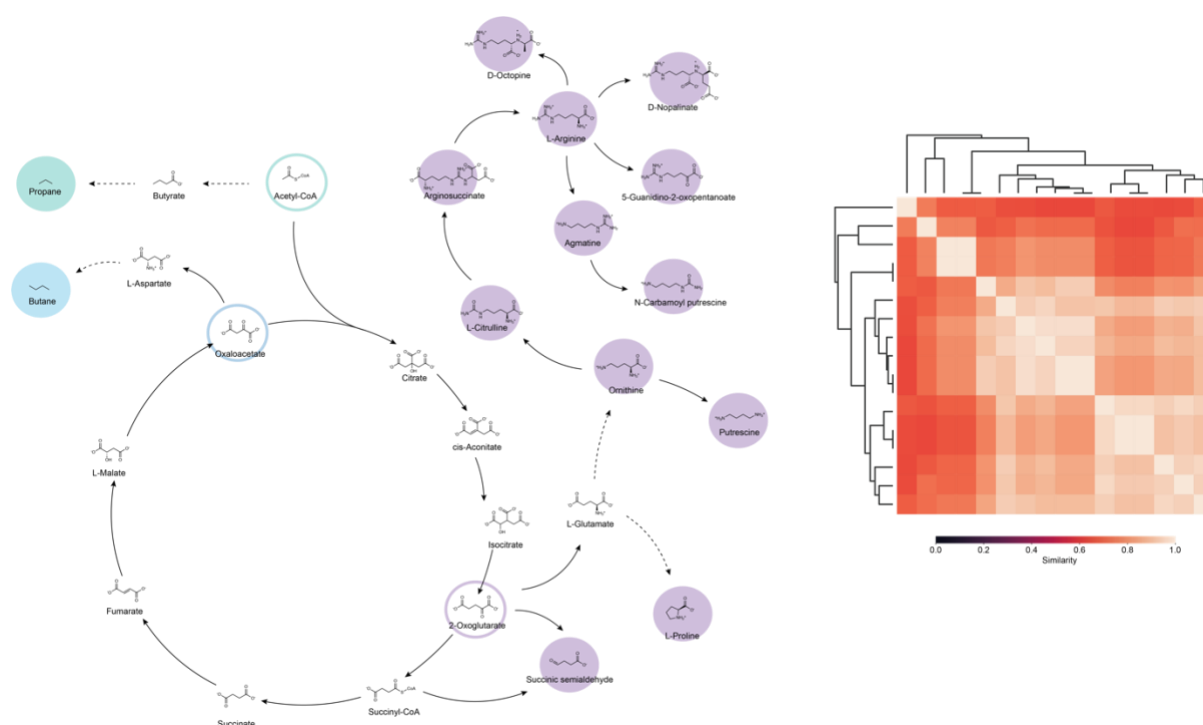




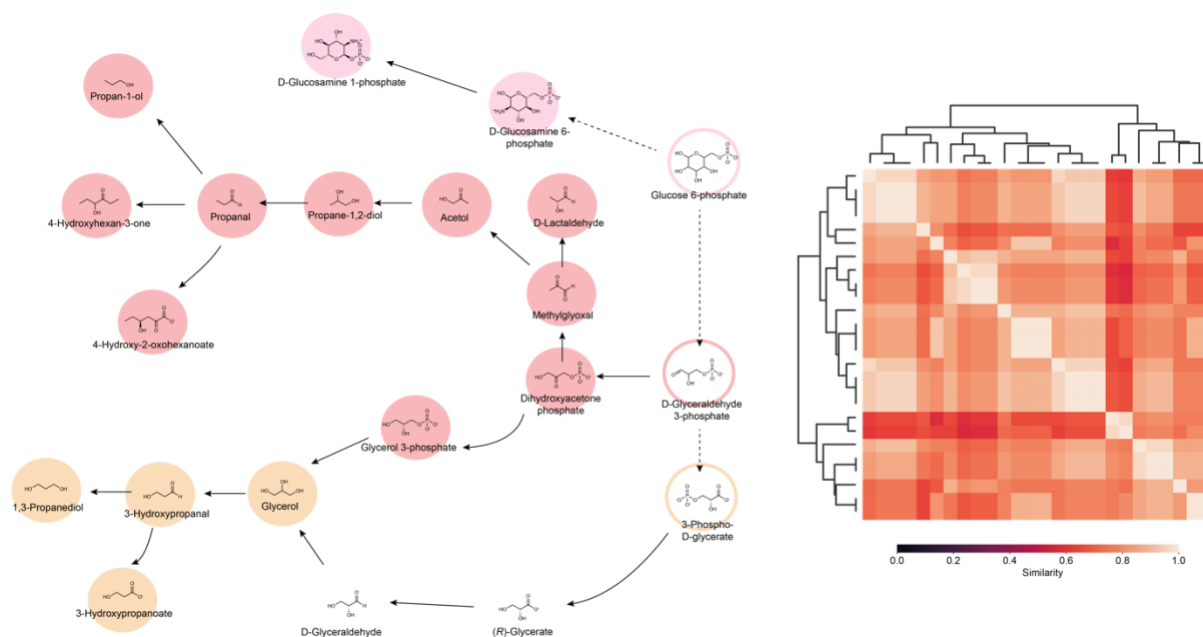
Supplementary Fig. 23. Metabolic pathways for chemicals in the cluster E of the FVSEOF-based cluster map. Chemicals derived from 2-oxoglutarate and oxaloacetate were colored in purple and blue, respectively. In the cluster E, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_met_36592, iML1515_met_28579, iML1515_2oxoadp, iML1515_6ax, iML1515_met_18153_path_1, iML1515_sucsal_path_1, iML1515_ghb, iML1515_glu__L, iML1515_4abut, iML1515_4abut_path_1, iML1515_acon_C, iML1515_icit, iML1515_cit, and iML1515_itacon. The information on the target chemical for each GEM is available in Supplementary Data 3.



Supplementary Fig. 24. Metabolic pathways for chemicals in the cluster F of the FVSEOF-based cluster map. Chemicals derived from 2-oxoglutarate, D-glyceraldehyde 3-phosphate, oxaloacetate, and pyruvate were colored in purple, red, blue, and gray, respectively. In the cluster F, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_ppa_path_1, iML1515_ppoh_path_2, iML1515_met_DHB, iML1515_met_88063, iML1515_met_DHBA, iML1515_met_MalSA, iML1515_mal_L_path_1, iML1515_mal_L, iML1515_fum, iML1515_fum_path_1, iML1515_succ, iML1515_succ_path_1, iML1515_met_87755, iML1515_met_M2P2TEN, iML1515_met_17128_path_1, iML1515_met_MP2TN, iML1515_met_48946, and iML1515_pta. The information on the target chemical for each GEM is available in Supplementary Data 3.



Supplementary Fig. 25. Metabolic pathways for chemicals in the cluster G of the FVSEOF-based cluster map. Chemicals derived from 2-oxoglutarate, acetyl-CoA, and oxaloacetate were colored in purple, cyan, and blue, respectively. In the cluster G, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_met_37808, iML1515_sucsal, iML1515_pro__L, iML1515_pro__L_path_1, iML1515_met_32879, iML1515_argsuc, iML1515_orn, iML1515_citr__L, iML1515_ptrc, iML1515_ptrc_path_1, iML1515_arg__L, iML1515_Ncbmpts, iML1515_agm, iML1515_octp, iML1515_5g2oxpt, and iML1515_nop. The information on the target chemical for each GEM is available in Supplementary Data 3.

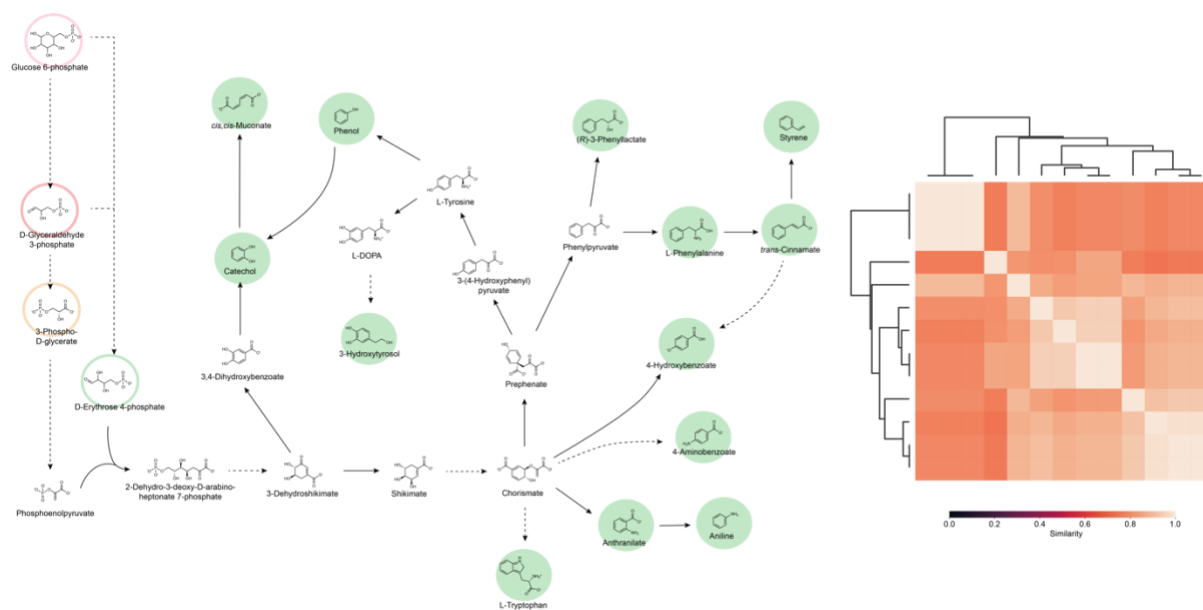


Supplementary Fig. 26. Metabolic pathways for chemicals in the cluster H of the FVSEOF-based cluster map. Chemicals derived from 3-phospho-D-glycerate, D-glucose 6-phosphate, and D-glyceraldehyde 3-phosphate were colored in orange, pink, and red, respectively. In the cluster H, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_met_16997, iML1515_ppal, iML1515_12ppd__R, iML1515_met_18351, iML1515_4hoxoh, iML1515_glyc, iML1515_dhap, iML1515_gam1p, iML1515_gam6p, iML1515_gam6p_path_1, iML1515_13ppd_path_1, iML1515_glyc_path_1, iML1515_3hpp_path_2, iML1515_3hppnl, iML1515_acetol, iML1515_mthgxl, iML1515_acetol_path_1, iML1515_lald__D, iML1515_glyc_path_2, iML1515_glyc_path_3, iML1515_ppoh_path_3, iML1515_12ppd__S, iML1515_lald__L, iML1515_glyc3p, iML1515_13ppd, and iML1515_13ppd_path_2. The information on the target chemical for each GEM is available in Supplementary Data 3.

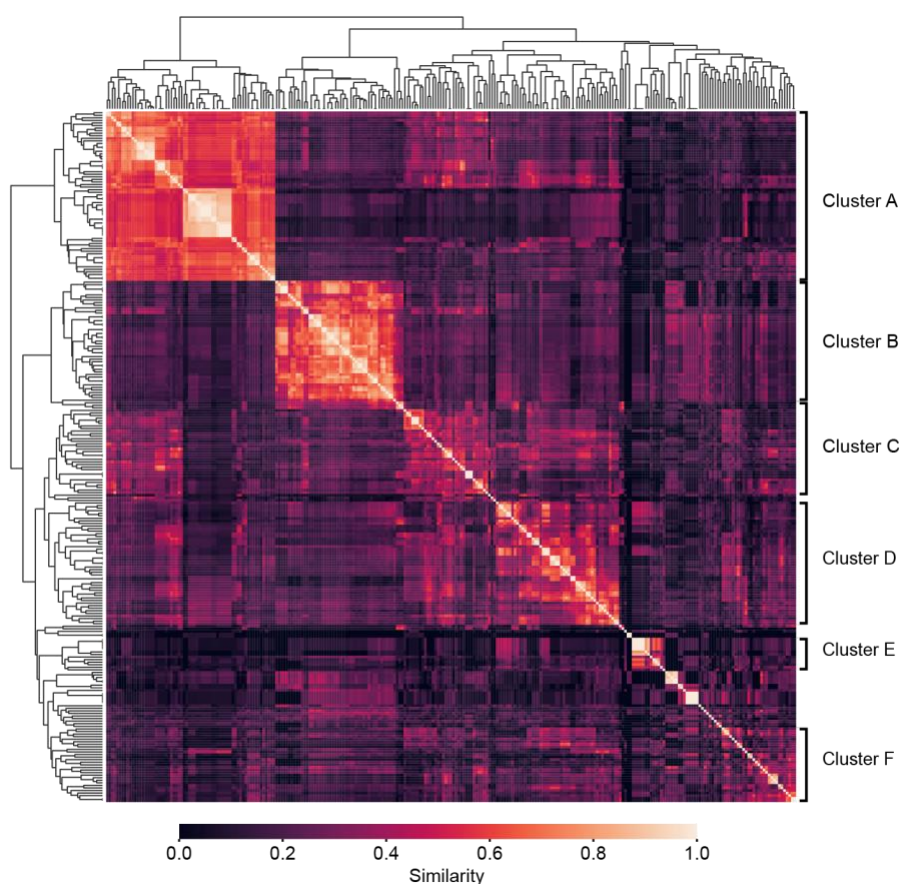
Supplementary Fig. 27. Metabolic pathways for chemicals in the cluster I of the FVSEOF-based cluster map. Chemicals derived from D-erythrose 4-phosphate were colored in green. In the cluster I, the GEMs from the left to the right (and from the top to the bottom) are as follows:

iML1515_met_143072, iML1515_met_4APEAM, iML1515_met_4ACAA, iML1515_met_4APAA, iML1515_met_4APAADH, iML1515_tyr__L, iML1515_tym, iML1515_T4hcinm, iML1515_met_1883, iML1515_pac, iML1515_met_36656, iML1515_34hpp, iML1515_4hoxpacd, iML1515_met_6PHX, iML1515_met_75317, iML1515_skm, iML1515_met_4APETH, iML1515_met_57504, iML1515_dopa, iML1515_34dhcinm, iML1515_34dhcinm_path_1, iML1515_chor, iML1515_ichor, iML1515_met_2PETO, iML1515_pacald, iML1515_34hpl, iML1515_met_1879, iML1515_pphn, iML1515_met_17756, and iML1515_phpyr

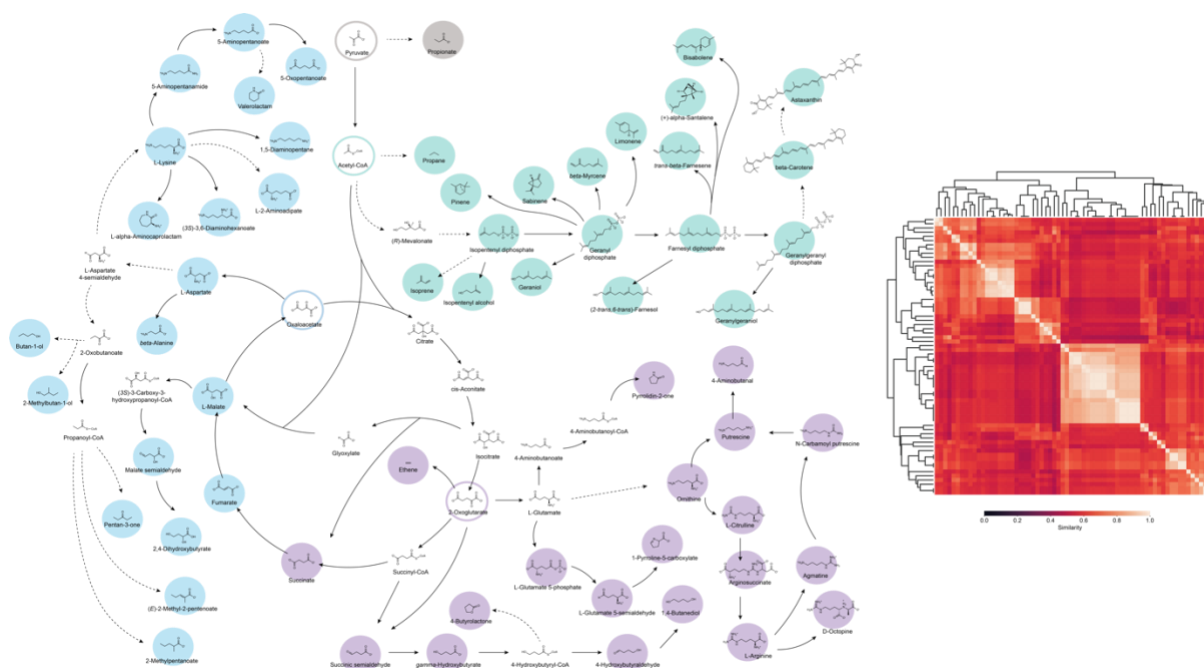
The information on the target chemical for each GEM is available in Supplementary Data 3.



Supplementary Fig. 28. Metabolic pathways for chemicals in the cluster J of the FVSEOF-based cluster map. Chemicals derived from D-erythrose 4-phosphate were colored in green. In the cluster J, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_phe__L, iML1515_cinnm, iML1515_met_27452, iML1515_trp__L, iML1515_met_11009, iML1515_4hbz, iML1515_4abz, iML1515_anth, iML1515_met_17296, iML1515_met_3HXTYR, iML1515_ccmuac, iML1515_catechol, and iML1515_phenol. The information on the target chemical for each GEM is available in Supplementary Data 3.

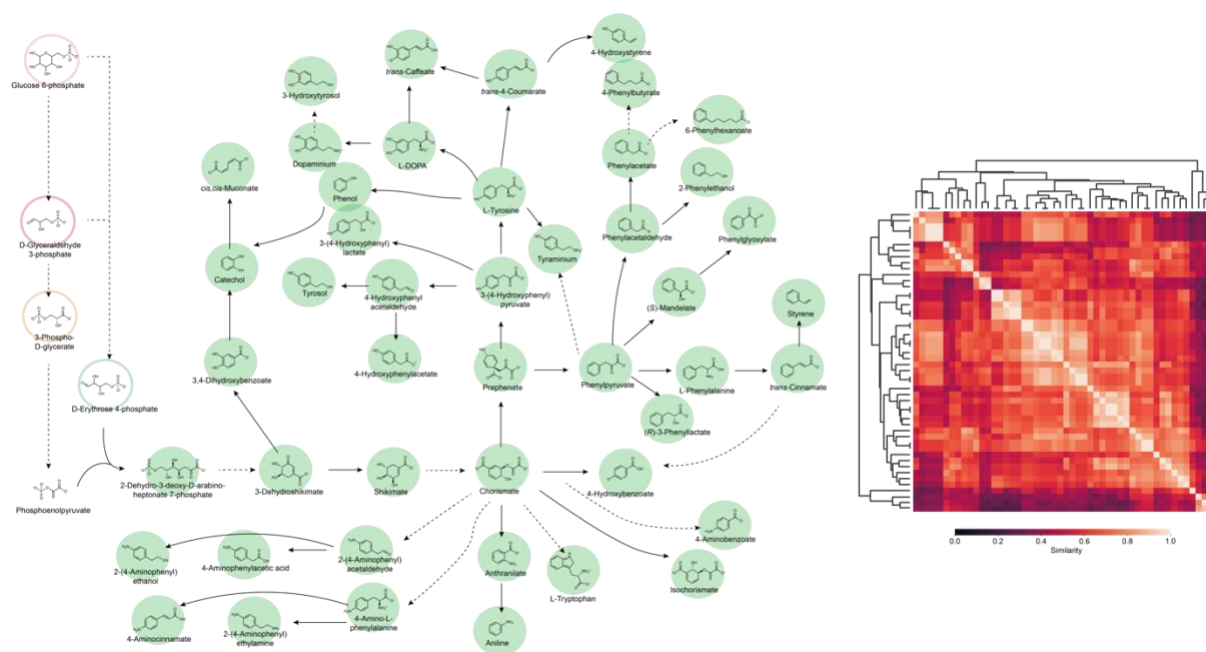


Supplementary Fig. 29. Hierarchical clustering of 272 metabolic pathways using metabolic reaction profiles predicted by iBridge analysis. Metabolic reaction profiles, representing the types of predicted metabolic reactions to be up- and down-regulated for improved chemical production, were analyzed across 272 metabolic pathways which lead to the production of 235 chemicals. Pairwise similarities of metabolic reaction profiles were calculated using cosine similarity. High similarity between metabolic reaction profiles is indicated by bright red, while low similarity is indicated by dark red. Chemicals with similar metabolic reaction profiles were grouped into distinct clusters (A-F). The simulation was performed for *E. coli* under aerobic conditions with D-glucose as the sole carbon source. Source data are provided as a Source Data file.



Supplementary Fig. 30. Metabolic pathways for chemicals in the cluster A of the iBridge-based cluster map. Chemicals derived from 2-oxoglutarate, acetyl-CoA, oxaloacetate, and pyruvate were colored in purple, cyan, blue, and gray, respectively. In the cluster A, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_ala_B, iML1515_asp_L, iML1515_4abutn, iML1515_glu5sa, iML1515_met_4HBAD, iML1515_met_GBL, iML1515_ptrc, iML1515_ptrc_path_1, iML1515_met_32879, iML1515_orn, iML1515_met_77761, iML1515_oxptn, iML1515_15dap, iML1515_5aptn, iML1515_lys_L, iML1515_L2aadp, iML1515_5apentam, iML1515_36dahx, iML1515_met_58113, iML1515_argsuc, iML1515_arg_L, iML1515_Ncbmpts, iML1515_agm, iML1515_citr_L, iML1515_octp, iML1515_ghb, iML1515_met_18153_path_1, iML1515_sucsal_path_1, iML1515_glu5p, iML1515_met_14BDO, iML1515_caro, iML1515_met_40968, iML1515_ipdp, iML1515_met_35194, iML1515_met_62898, iML1515_met_17447, iML1515_grdp, iML1515_met_50029, iML1515_met_28660, iML1515_limnen, iML1515_met_17221, iML1515_frdp, iML1515_met_16619, iML1515_met_46762, iML1515_met_61677, iML1515_met_49243, iML1515_met_10418, iML1515_ggdp, iML1515_met_10280,

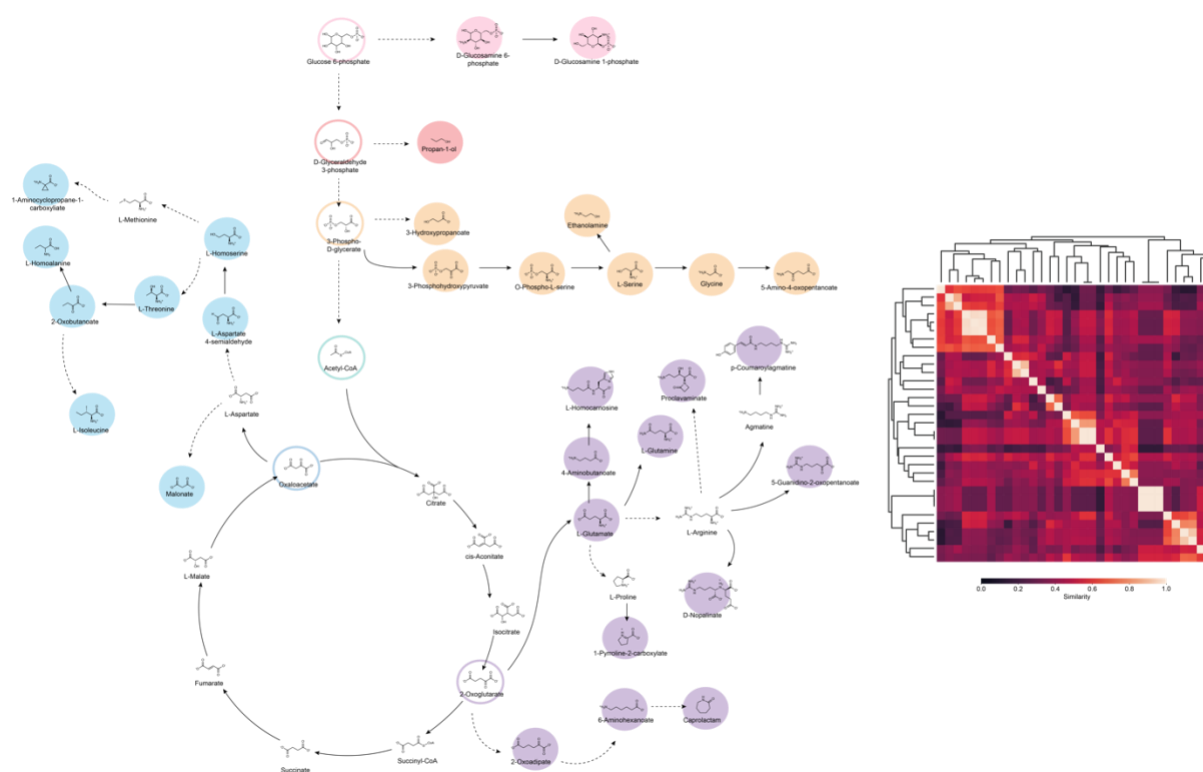
iML1515_2mbtoh, iML1515_btoh_path_1, iML1515_met_DHB, iML1515_met_MP2TN, iML1515_1pyr5c, iML1515_met_36592, iML1515_met_MalSA, iML1515_fum, iML1515_fum_path_1, iML1515_mal__L, iML1515_mal__L_path_1, iML1515_sucsal, iML1515_ppa_path_1, iML1515_met_M2P2TEN, iML1515_met_87755, iML1515_succ, and iML1515_succ_path_1 The information on the target chemical for each GEM is available in Supplementary Data 3.



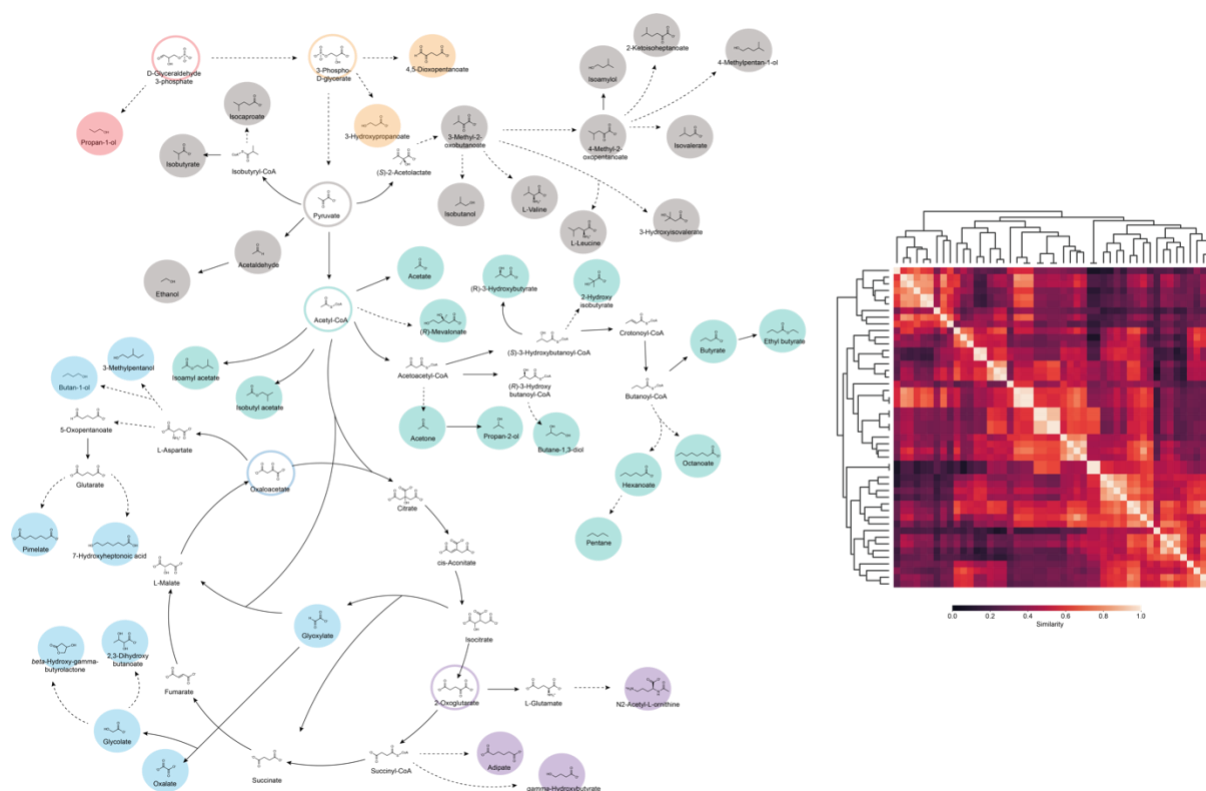
Supplementary Fig. 31. Metabolic pathways for chemicals in the cluster B of the iBridge-based cluster map. Chemicals derived from D-erythrose 4-phosphate were colored in green. In the cluster B, the GEMs from the left to the right (and from the top to the bottom) are as follows:

iML1515_met_27452, iML1515_met_3HXTYR, iML1515_met_57504,
iML1515_34dhcinm_path_1, iML1515_dopa, iML1515_34dhcinm, iML1515_cinnm,
iML1515_met_75317, iML1515_catechol, iML1515_ccmuac, iML1515_4abz,
iML1515_met_4APAA, iML1515_met_4APETH, iML1515_34dhbz,
iML1515_34dhbz_path_3, iML1515_34dhbz_path_2, iML1515_2dda7p, iML1515_3dhsk,
iML1515_met_4APAADH, iML1515_pphn, iML1515_T4hcinm, iML1515_34hpl,
iML1515_met_1879, iML1515_tym, iML1515_tyr__L, iML1515_pac, iML1515_ichor,
iML1515_34hpp, iML1515_4hoxpacd, iML1515_34dhbz_path_1, iML1515_met_36656,
iML1515_met_11009, iML1515_met_17756, iML1515_pacald, iML1515_met_2PETO,
iML1515_phpyr, iML1515_phenol, iML1515_met_1883, iML1515_anth,
iML1515_met_17296, iML1515_4hoxpac, iML1515_met_6PHX, iML1515_phe__L,
iML1515_chor, iML1515_4hbz, iML1515_skm, iML1515_trp__L, iML1515_met_143072,

iML1515_met_4ACAA, and iML1515_met_4APEAM The information on the target chemical for each GEM is available in Supplementary Data 3.

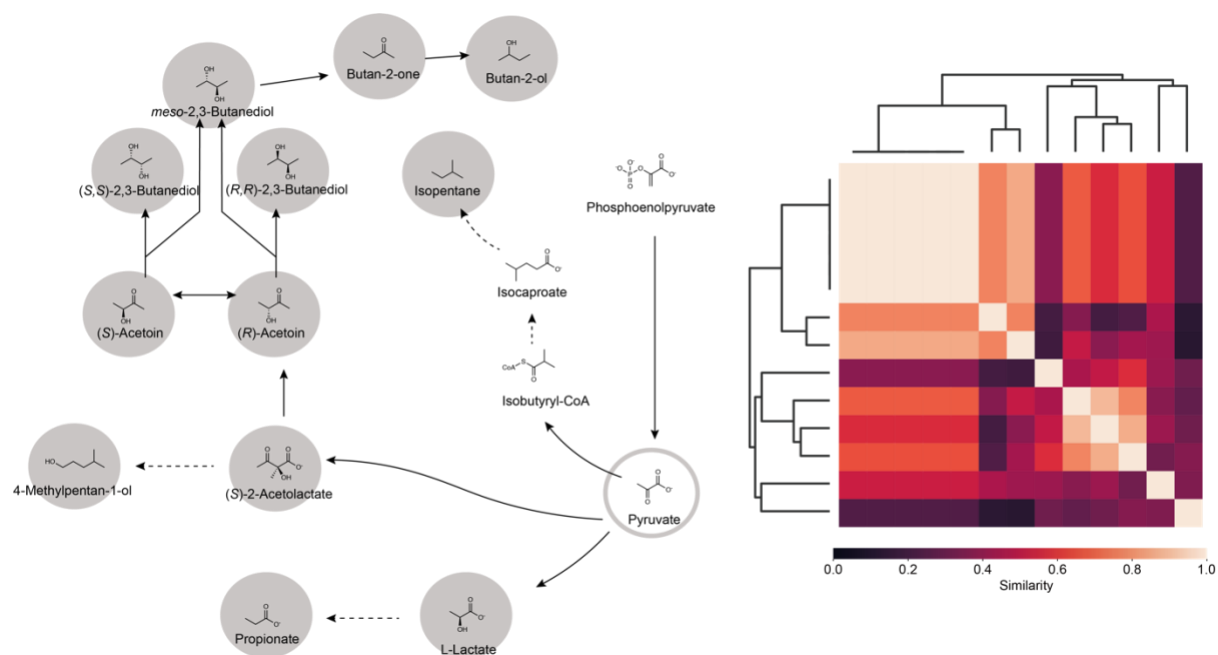


Supplementary Fig. 32. Metabolic pathways for chemicals in the cluster C of the iBridge-based cluster map. Chemicals derived from 2-oxoglutarate, 3-phospho-D-glycerate, D-glucose 6-phosphate, D-glyceraldehyde 3-phosphate, and oxaloacetate were colored in purple, orange, pink, red, and blue, respectively. In the cluster C, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_met_LHMA, iML1515_aspsa, iML1515_ppoh, iML1515_2obut, iML1515_hom__L, iML1515_thr__L, iML1515_ile__L, iML1515_met_57484, iML1515_2oxoadp, iML1515_6ax, iML1515_met_28579, iML1515_1acpc, iML1515_met_143075, iML1515_3hpp_path_1, iML1515_malon, iML1515_gln__L, iML1515_glu__L, iML1515_4abut, iML1515_4abut_path_1, iML1515_ppoh_path_1, iML1515_nop, iML1515_met_58644, iML1515_5g2oxpt, iML1515_met_57302, iML1515_gam6p_path_1, iML1515_gam1p, iML1515_gam6p, iML1515_gly, iML1515_pser__L, iML1515_etha, iML1515_ser__L, iML1515_5aop, and iML1515_met_39785 The information on the target chemical for each GEM is available in Supplementary Data 3.

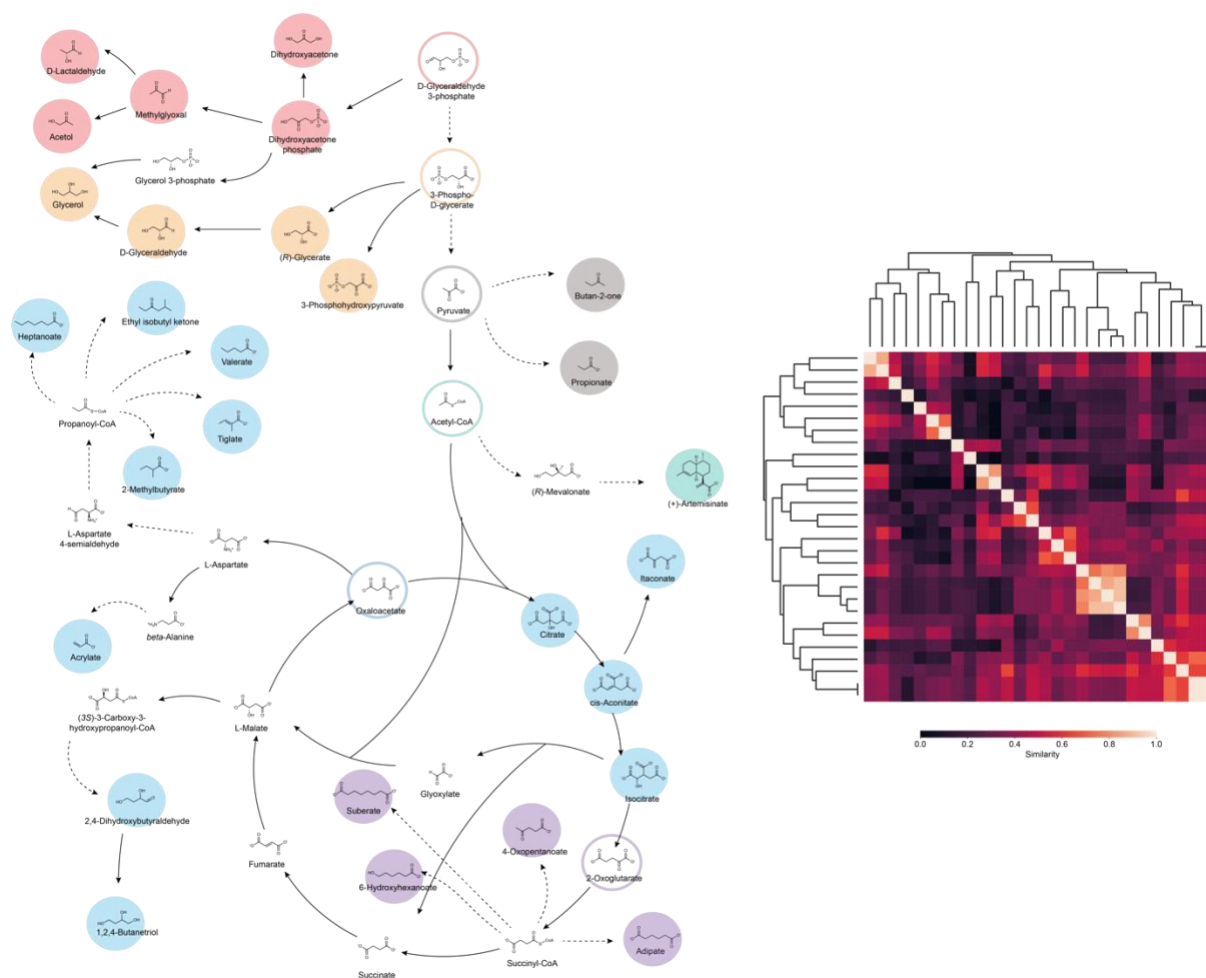


Supplementary Fig. 33. Metabolic pathways for chemicals in the cluster D of the iBridge-based cluster map. Chemicals derived from 2-oxoglutarate, 3-phospho-D-glycerate, acetyl-CoA, D-glyceraldehyde 3-phosphate, oxaloacetate, and pyruvate were colored in purple, orange, cyan, red, blue and gray, respectively. In the cluster D, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_ibutoh, iML1515_iamoh, iML1515_ibt, iML1515_3mob, iML1515_ibutac, iML1515_met_74904, iML1515_met_58055, iML1515_leu__L, iML1515_val__L, iML1515_acald, iML1515_etoh, iML1515_ppoh_path_2, iML1515_met_79112, iML1515_pime, iML1515_glx, iML1515_glyclt, iML1515_oxa, iML1515_met_87381, iML1515_CE2028, iML1515_3mb, iML1515_4mop, iML1515_met_DHB23, iML1515_met_DHB23_path_1, iML1515_met_HBA, iML1515_met_HBA_path_1, iML1515_ac, iML1515_acetone_path_1, iML1515_2ppoh, iML1515_acetone, iML1515_ghb_path_1, iML1515_ghb_path_2, iML1515_iamac, iML1515_met_63910, iML1515_met_5M2OPT, iML1515_met_74904_path_1, iML1515_bhb, iML1515_met_19641, iML1515_but,

iML1515_mev__R, iML1515_met_37830, iML1515_hxa, iML1515_met_17128, iML1515_met_88764, iML1515_octa, iML1515_acorn, iML1515_3hpp, iML1515_btoh, and iML1515_met_R13BDO The information on the target chemical for each GEM is available in Supplementary Data 3.



Supplementary Fig. 34. Metabolic pathways for chemicals in the cluster E of the iBridge-based cluster map. Chemicals derived from pyruvate were colored in gray. In the cluster E, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_met_75460_path_1, iML1515_met_75460, iML1515_met_28398_path_1, iML1515_btd_RR, iML1515_met_16812, iML1515_lac_L, iML1515_met_35687, iML1515_met_63910_path_1, iML1515_alac_S, iML1515_actn_R, iML1515_met_15687, iML1515_met_30362, and iML1515_ppa_path_2. The information on the target chemical for each GEM is available in Supplementary Data 3.



Supplementary Fig. 35. Metabolic pathways for chemicals in the cluster F of the iBridge-based cluster map. Chemicals derived from 2-oxoglutarate, 3-phospho-D-glycerate, acetyl-CoA, D-glyceraldehyde 3-phosphate, oxaloacetate, and pyruvate were colored in purple, orange, cyan, red, blue and gray, respectively. In the cluster F, the GEMs from the left to the right (and from the top to the bottom) are as follows: iML1515_mthgxl, iML1515_dhap, iML1515_glyc_path_2, iML1515_ppa, iML1515_acryl, iML1515_hpta, iML1515_subeac, iML1515_dha, iML1515_met_64782, iML1515_acetol_path_1, iML1515_lald_D, iML1515_met_28398, iML1515_met_39150, iML1515_met_TIGLATE, iML1515_glyald, iML1515_3php, iML1515_glyc_R, iML1515_itacon, iML1515_cit, iML1515_acon_C, iML1515_icit, iML1515_met_88063, iML1515_met_DHBA, iML1515_met_89209, iML1515_met_17128_path_1, iML1515_met_32383, iML1515_met_48946, and

iML1515_pta The information on the target chemical for each GEM is available in Supplementary Data 3.