

706

Supplementary Information

707

Statistical learning of bacterial growth in combinatorially constructed environments

708

Andrea Arrabal, Magdalena San Román, Juan Díaz-Colunga, Álvaro Sánchez

709

710

711

712

713

714

715

716

717

718

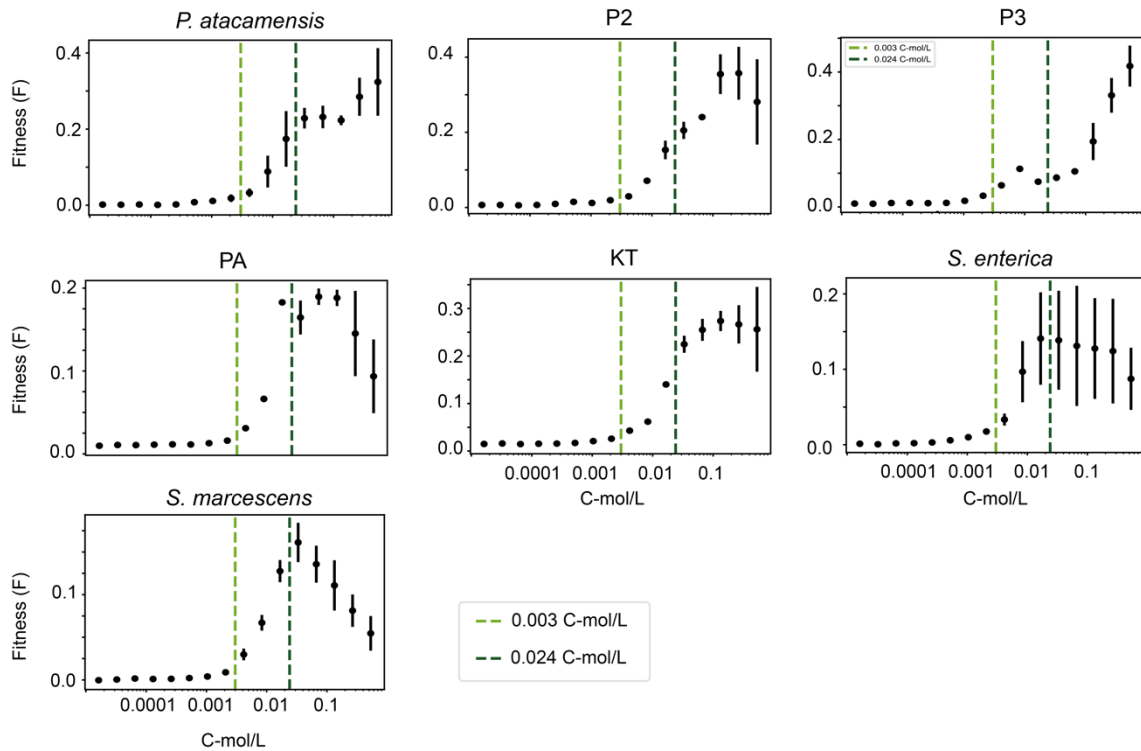


Figure S1. Fitness after 24 hours of growth in M9 supplemented with a glucose concentration gradient. To identify the carbon concentration range at which growth remains carbon-limited, seven bacterial strains (*Pseudomonas atacamensis*, P2, P3, PA, KT, *Salmonella enterica*, and *Serratia marcescens*) were grown in M9 minimal medium (base medium) supplemented with increasing glucose concentrations (0.0001–0.1 C-mol/L). Points represent biological replicate measurements and black error bars indicate mean $\pm$ standard deviation. Light green and dark green dashed lines indicate the carbon molarity in media when a single resource or all eight resources are added, respectively.

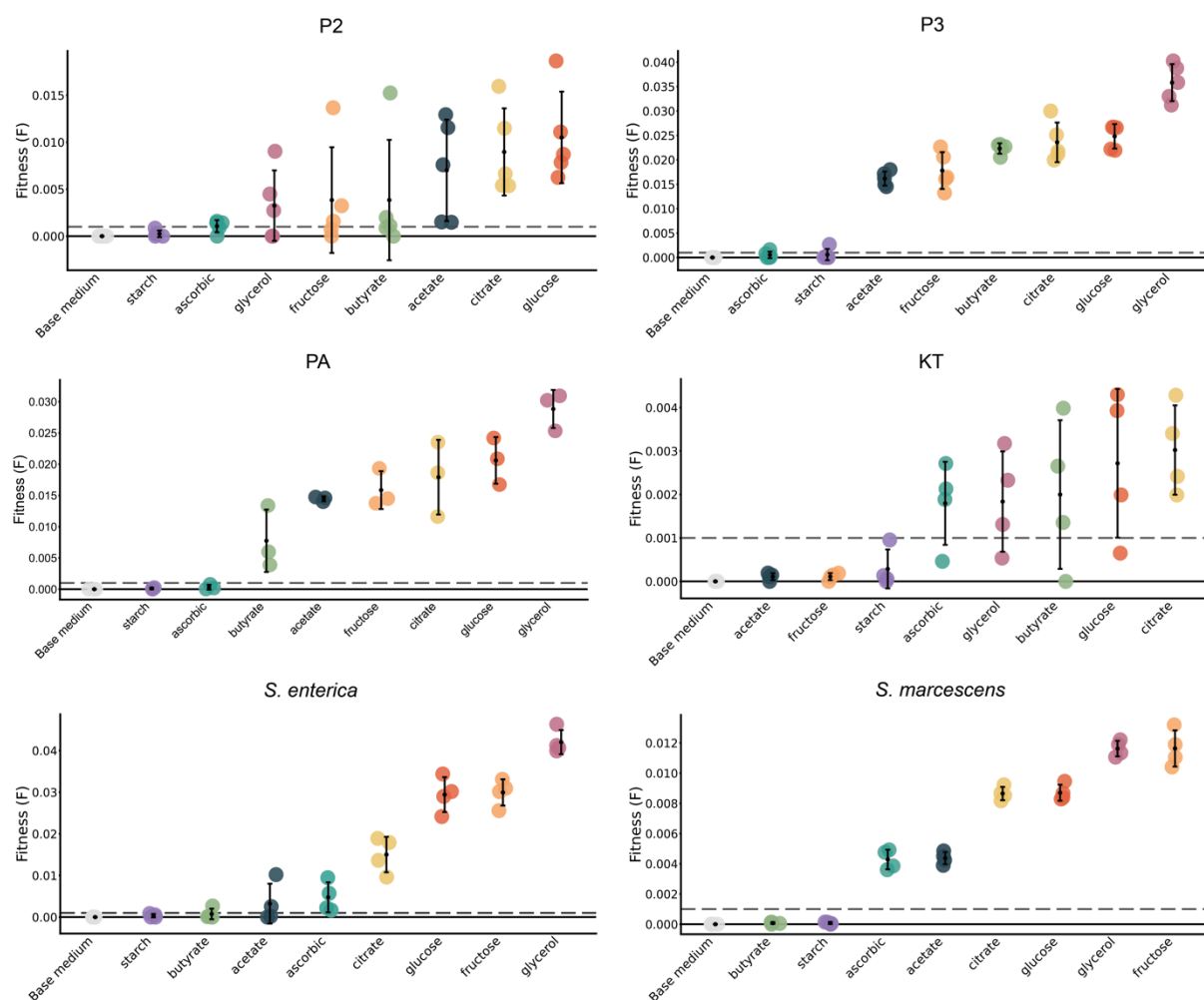


Figure S2. Individual fitness on single carbon sources. Fitness value of each strain when grown on base medium supplemented exclusively with each of the eight carbon sources tested. Error bars represent standard deviation across replicates. The dashed horizontal line indicates the detection limit (0.001). Substantial variation in strain-specific utilization efficiency is evident across carbon sources.

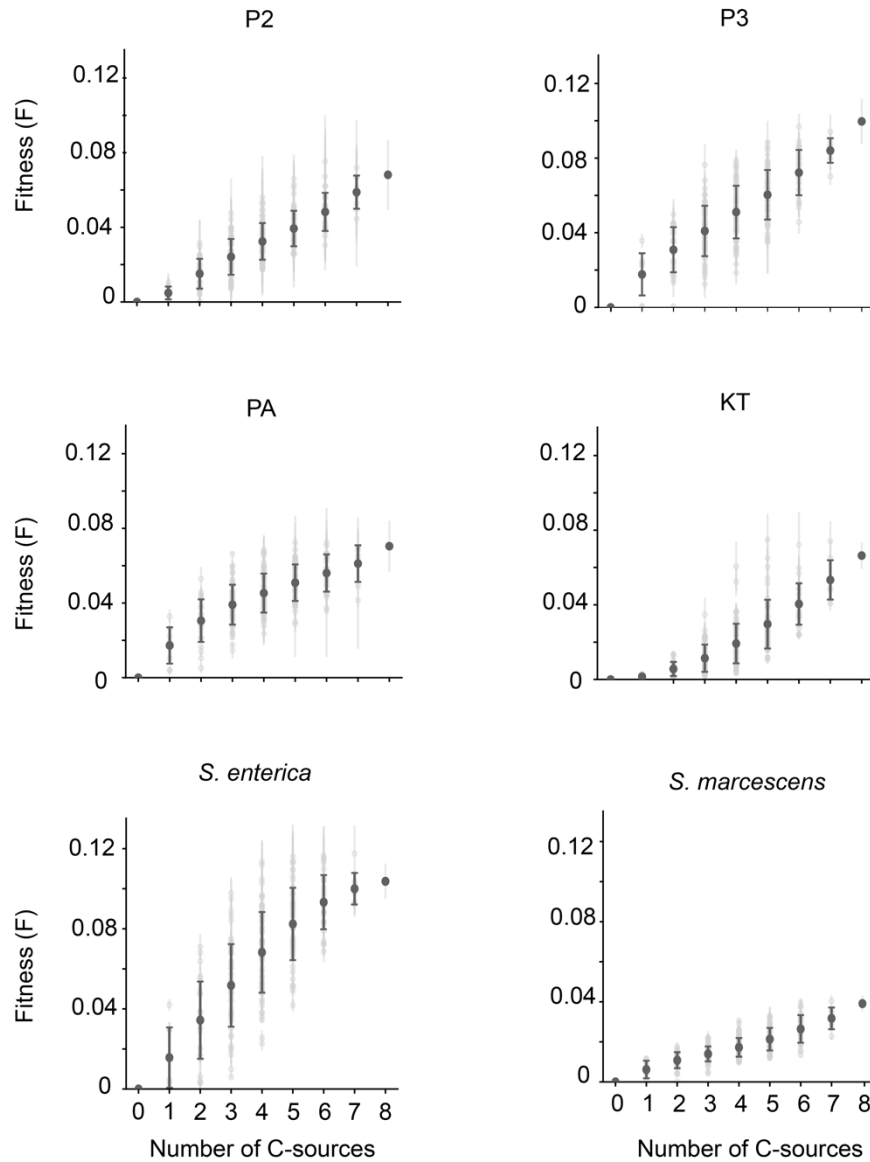


Figure S3. Resource diversity and fitness relationships across strains. Fitness (F) measured across 255 environments plotted against the number of carbon sources in the mixture. Light grey points represent individual fitness of each strain in each individual environment; dark grey points indicate the average fitness ($\pm$ standard deviation) as a function of the number of carbon sources present in the mixture.

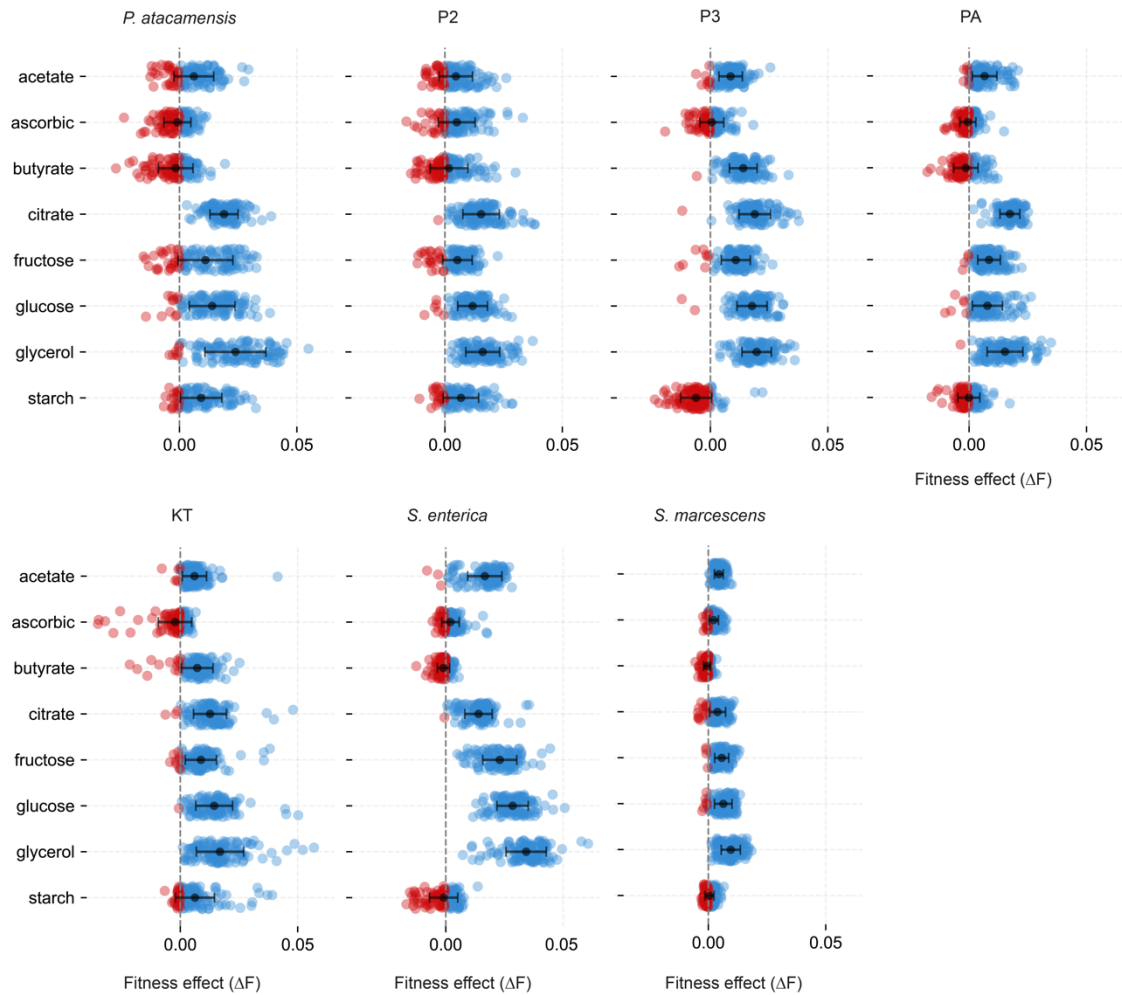


Figure S4. Context-dependent dual behavior of carbon sources across strains. Fitness effects distribution for each carbon source (given on the y-axis) added to every possible background environment across strains. Blue and red dots indicate environments in which the resource acts as a fitness enhancer or fitness inhibitor, respectively.

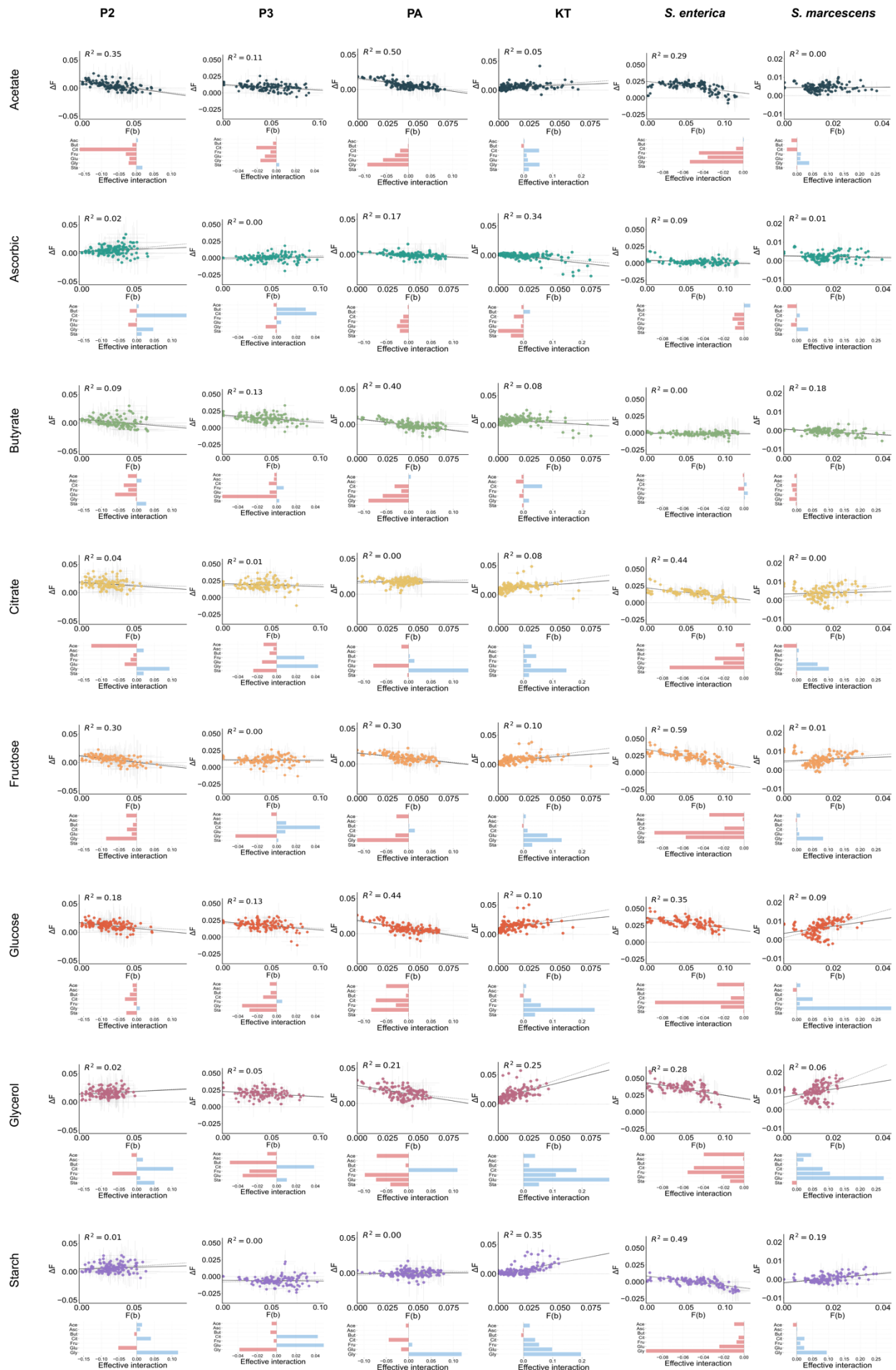


Figure S5. Fitness effects across strains. Fitness effects of adding a carbon source (color-coded and positioned on the y-axis) to a background environment are well predicted by linear models linking fitness effects to the fitness of the background environment. Panels show results for P2, P3, PA, KT, *S. enterica* and *S. marcescens* (indicated at the bottom). Data represent mean $\pm$ standard deviation across replicates.

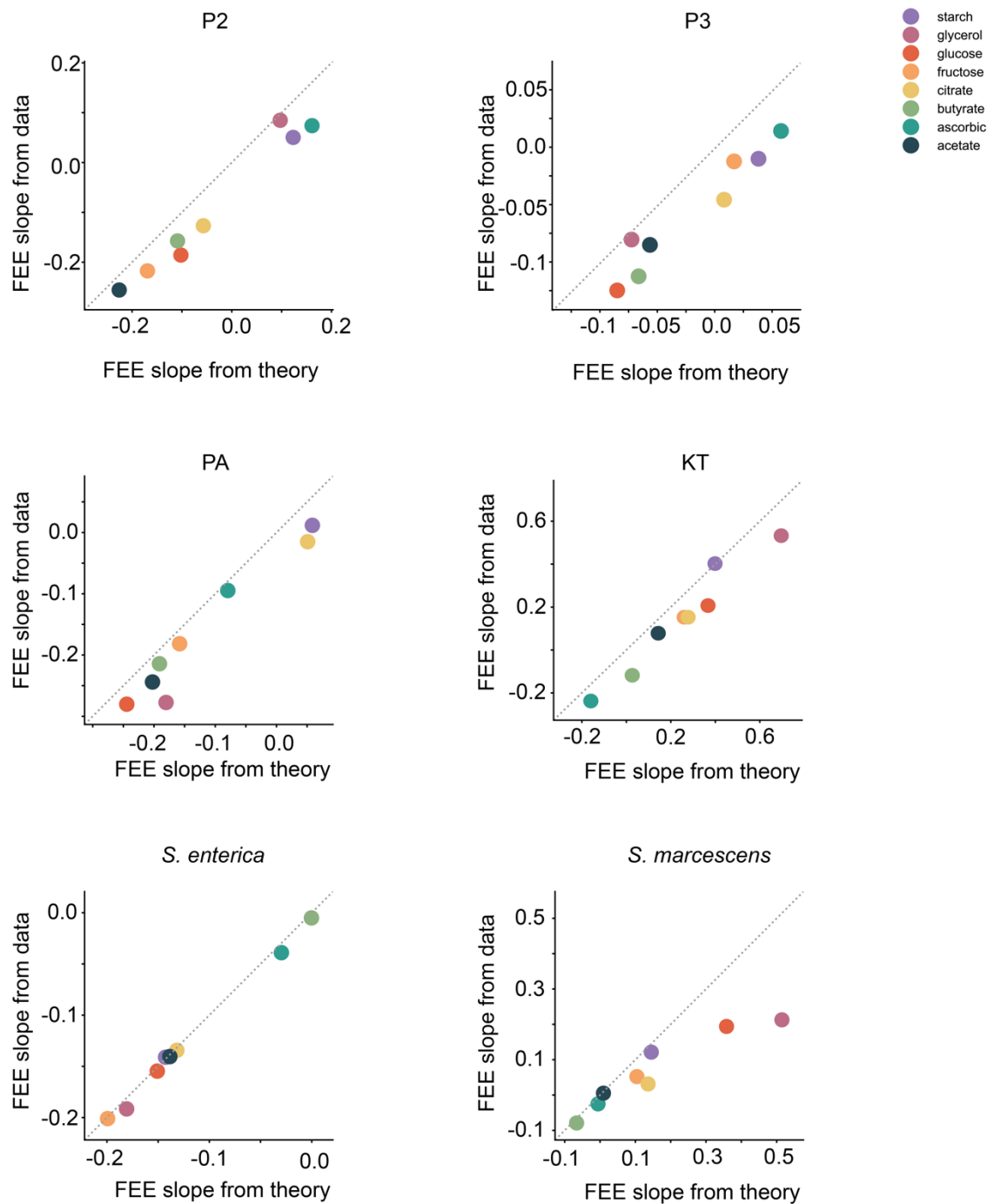


Figure S6. Empirical vs theoretical FEE slopes. FEE slopes from data are compared with theoretical predictions across strains.

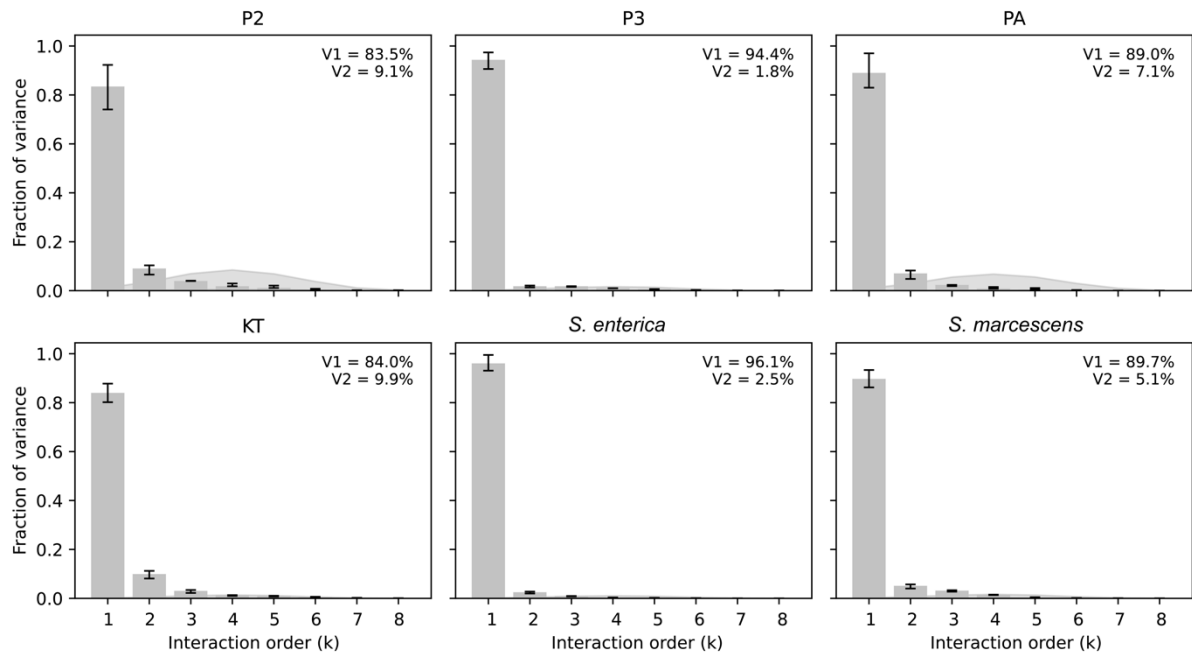


Figure S7. Contribution of interactions at all orders to fitness variation across strains. Fitness variance is predominantly explained by additive and pairwise effects across all environments across strains. Consistent with the findings for *P. atacamensis* (Fig. 3C), the results are quantitatively and qualitatively similar, demonstrating that higher-than-pairwise interactions have a negligible effect on the fitness impact of new resources under these conditions.

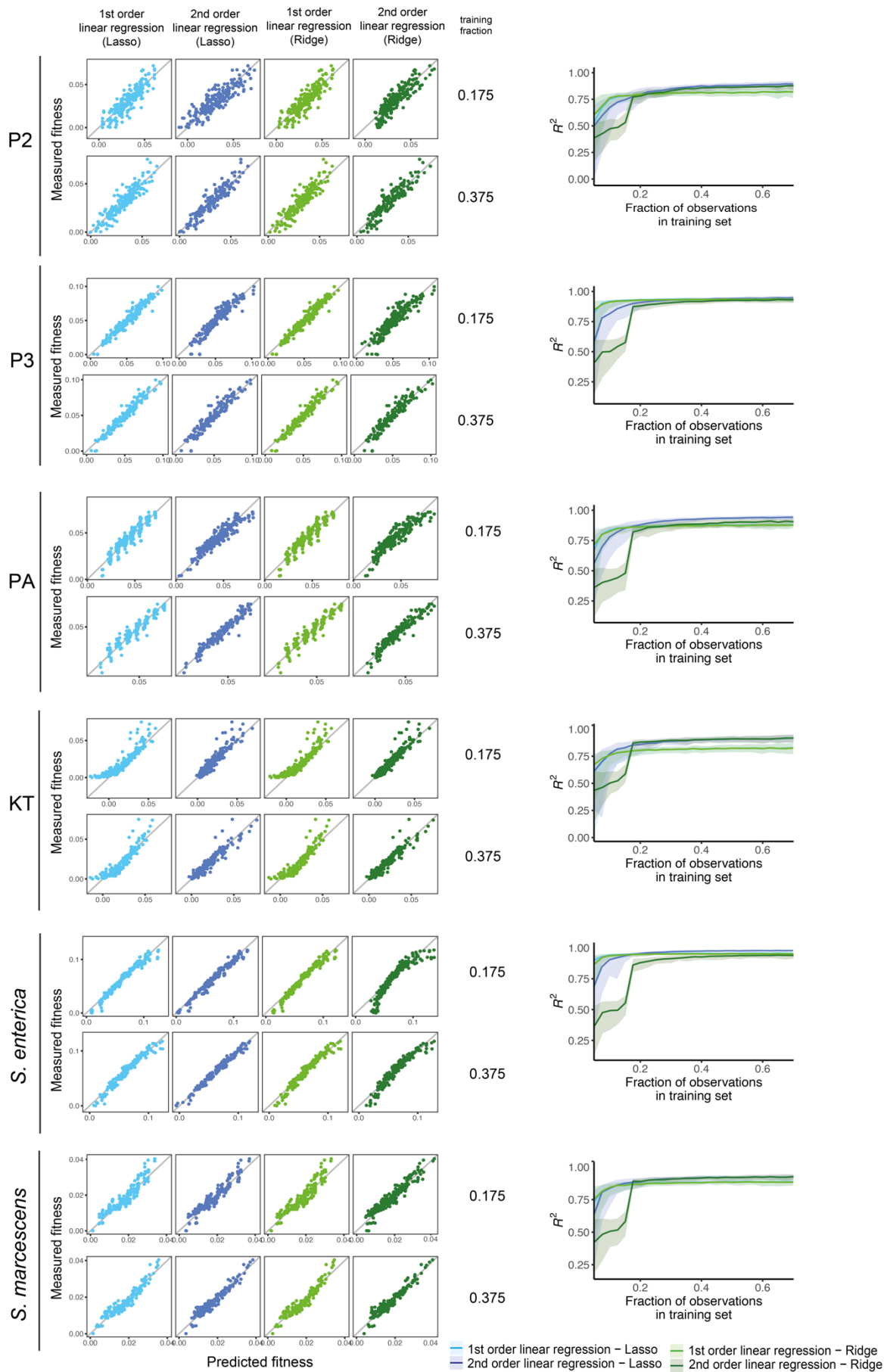


Figure S8. Model prediction accuracy across bacterial isolates and species. (Left panels) Measured versus predicted fitness values for first-order and second-order linear regression models (Lasso and Ridge) at training set fractions of 0.175 and 0.375 across strains. (Right panels) Relationship between training set fraction and prediction accuracy (R^2) with shaded interquartile ranges. Second-order models (dark colors) consistently outperform first-order models (light colors) across all species. Both Lasso and Ridge regularization show similar performance with Ridge achieving slightly higher R^2 values at higher training fractions.

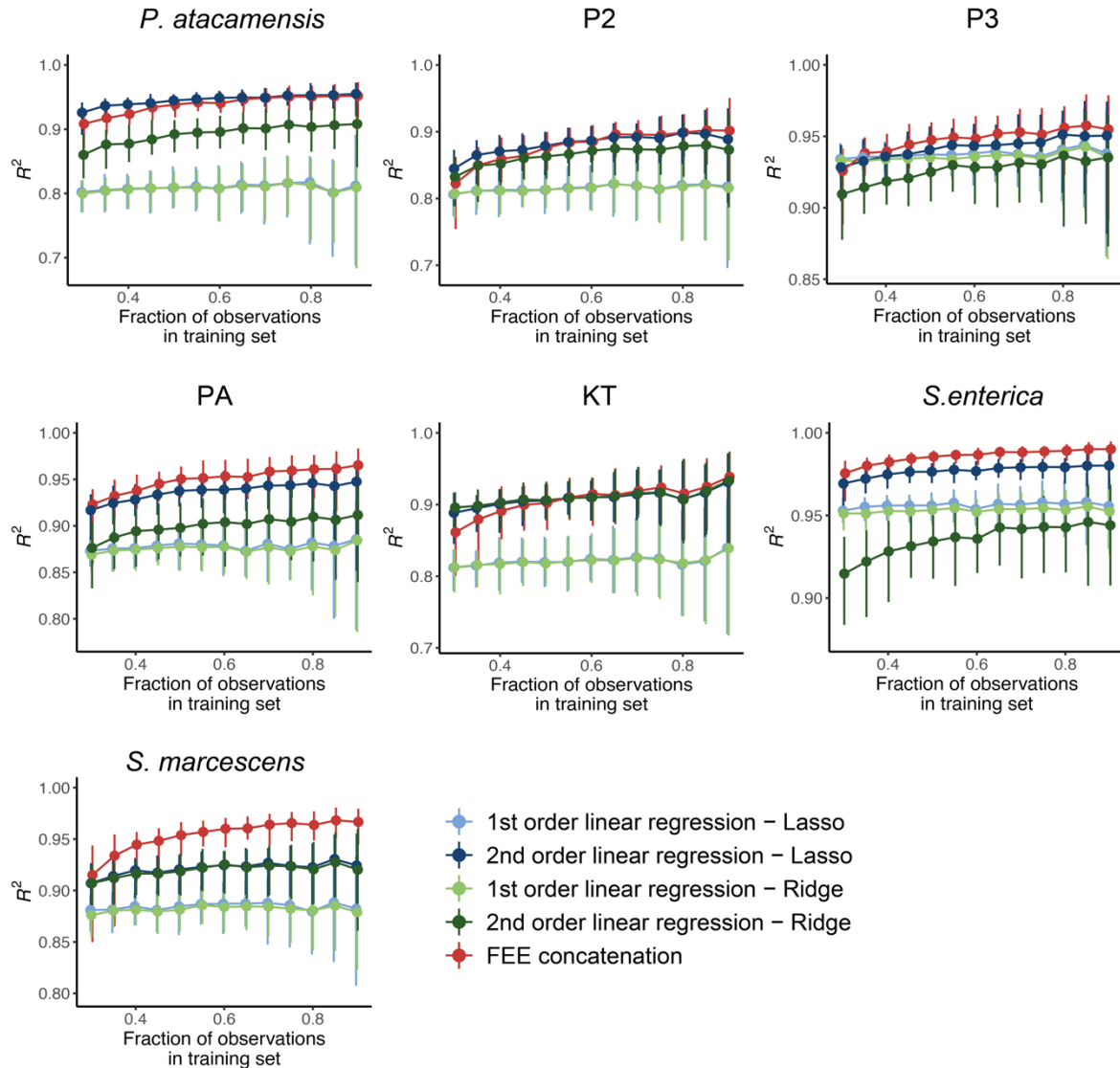


Figure S9. Linear regression models and FEE concatenation method accurately predict microbial growth out of sample across bacterial species. Relationship between training set fraction and model prediction accuracy (R^2) across strains. Each panel shows median R^2 values across 100 independent random partitions with shaded error bars representing the interquartile range. Models compared include first-order and second-order linear regression with Lasso (light and dark blue, respectively) and Ridge regularization (light and dark green, respectively), as well as an alternative FEE concatenation method (red). Results are qualitatively and quantitatively similar across all species, demonstrating the robustness and generalizability of the predictive modeling approach. Second-order models consistently outperform first-order models across all species and training fractions, while FEE concatenation achieves comparable or superior performance to linear regression, particularly at higher training set fractions.

Supplementary material

Table 1: Composition and concentrations of carbon sources used in the study, including their chemical nature, supplier information, and reference identifiers.

Reagent type	Carbon source (CS)	C-mol/L	1X	C-mol/ L (8X)	8X	Supplier	Reference
Chemical compound, drug	Sodium acetate anhydrous	0.003	0.003 C-mol/L	0.024	0.024 C-mol/L	Sigma Aldrich	W302406
Chemical compound, drug	L-ascorbic acid		0.088 g/L		0.7 g/L	Sigma Aldrich	A92902
Chemical compound, drug	Sodium butyrate		0.083 g/L		0.664 g/L	Sigma Aldrich	B5887
Chemical compound, drug	Sodium citrate monobasic		0.003 C-mol/L		0.024 C-mol/L	Sigma Aldrich	71498
Chemical compound, drug	D-(-)-Fructose		0.009%		0.072%	Sigma Aldrich	F0127
Chemical compound, drug	D-(+)-Glucose		0.009%		0.072%	Sigma Aldrich	G7021
Chemical compound, drug	Glycerol		0.009%		0.072%	Sigma Aldrich	G9012
Chemical compound, drug	Starch, soluble		0.002%		0.016%	Sigma Aldrich	S9765