Supplementary Materials for

Network Context and Selection in the Evolution to Enzyme Specificity

Hojung Nam, Nathan E. Lewis, Joshua A. Lerman, Dae-Hee Lee, Roger L. Chang, Donghyuk Kim, Bernhard O. Palsson*

*To whom correspondence should be addressed. E-mail: palsson@ucsd.edu

Published 31 August 2012, *Science* **337**, 1101 (2012)
DOI: 10.1126/science.1216861

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/337/6098/1101/DC1)

Database S1 as a MS Excel file: (A) *E. coli* enzyme table, (B) *E. coli* reaction table, (C) *M. barkeri*, (D) *C. reinhardtii*, and (F) *S. cerevisiae*.

Table of Contents

Material and Methods

E. coli culturing and phenotyping

Phenotype information, including metabolite uptake rates and growth rates were previously obtained for aerobic growth on glucose, glycerol, and propylene glycol and anaerobic growth on glucose (*30-32*). We augmented these data with additional growth phenotyping data for anaerobic growth on glucose M9 minimal medium with and without nitrate (Table S1). Glycerol stocks of *E. coli* K-12 MG1655 were inoculated into 2 g/L glucose M9 minimal media and grown at 37°C overnight. Freshly sparged media was inoculated with the overnight culture to the optical density at 600nm of 0.02. To measure growth on nitrate, the protocol was repeated, with the exception that the anaerobic culture was supplemented with potassium nitrate (20 mM) at 37°C. Cells were grown exponentially while sampling growth rates and media multiple times. Growth rates were determined by measuring the optical density of cultures at 600nm. Glucose uptake and byproduct secretion were also measured by HPLC. The range of feasible nitrate uptake rates was approximated from the *i*AF1260 model of *E. coli* by constraining the model growth rate and glucose uptake rate to the measured values (0.34 ± 0.007 h^{-1}, and 4.05 ± 0.5 mmol gDW^{-1} h^{-1}, respectively) and minimizing/maximizing nitrate update using flux balance analysis (*33*). However, as shown in fig. S16, our results were qualitatively robust with variations in nitrate uptake rate within the computed range.

Gene-expression profiling

To provide support for the MCMC sampling-based computational flux change predictions, we compared these with gene expression changes. This was done using expression data sets generated for this study and published microarray data in which cultures of exponentially-growing *E. coli* were profiled under aerobic conditions on glucose, glycerol, or propylene glycol M9 minimal media and anaerobically on glucose M9 minimal media with and without added nitrate (*17, 31, 32*). The glycerol-glucose shift provides support for the changes between substrates that are metabolized similarly, while the glycerol-propylene glycol shift data we present here support our results for substrates that differ substantially in how they are metabolized, despite only differing by one hydroxyl group in the molecular structure.

Since *E. coli* cannot normally grow on propylene glycol, a strain of *E. coli* K-12 MG1655, adapted for glycerol growth was evolved to also metabolize propylene glycol (*30, 34*). This strain was subsequently grown and expression profiled on both 2g/L glycerol M9 minimal media and 2g/L propylene glycol M9 minimal media at 37°C. Affymetrix *E. coli* Antisense Genome Arrays were used for all transcriptional analyses. Each experimental condition was tested in triplicate in the respective carbon sources (i.e., glycerol or propylene glycol) using independent cultures and processed following the manufacturer-recommended protocols. Cultures were grown to mid-exponential growth phase aerobically (OD_{600} = 0.3) in minimal media, supplemented with the appropriate carbon source. Three ml of culture was added to 2 volumes of RNAprotect Bacteria Reagent (Qiagen) and total RNA was then isolated using RNeasy columns (Qiagen) with DNase I treatment. Total RNA yields and quality were measured using a Nanodrop 1000

(Thermo Scientific) and agarose gels. cDNA synthesis, fragmentation, end-terminus biotin labeling, and array hybridization were performed as recommended by the Affymetrix standard protocol.

The Affymetrix CEL files were normalized using GCRMA (version 2.20.0) implemented in R (version 2.11.1). Genes were considered not expressed if their median expression level across replicates was lower than than the median value of intergenic (IG) probes. Genes were subsequently removed from further analysis if they were not expressed in any conditions. Differentially expressed genes were determined using a two-tailed *t*-test followed by false discovery rate (FDR) *P*-value adjustment (FDR ≤ 0.01).

Designation of specialist and generalist enzymes

For this study we classified 1,147 enzymes from the *E. coli* genome-scale model (*i*AF1260) (*5*) following the detailed process shown in Fig. S17. First, we identified 1,081 proteins with enzymatic activity, as reported in the EcoCyc Database (*35*). The 66 proteins that were removed did not have experimentally-validated catalytic activities. These included non-catalytic members of enzyme complexes (e.g., the electron transferring protein flavodoxin (b0684)) and predicted enzymes (e.g., predicted carbamate kinase (b0521)). Among these 1,081 enzymes, through careful manual curation, 677 enzymes were classified as specialists, which only catalyze one specific chemical reaction, and 404 enzymes were classified as generalist enzymes since they catalyze more than one reaction (Fig. S17A).

Following enzyme classification, specialist enzyme reactions (SERxns) and generalist enzyme reactions (GERxns) were identified and classified (Fig. S17B). The reaction lists were also filtered to remove reactions with ambiguous classification. We also note that transport reactions were removed as they usually do not represent canonical metabolic catalysis beyond, for example, ATP hydrolysis in ABC transporters. However, the presence of transporters did not qualitatively change the results in this work (Fig. S3).

Markov chain Monte Carlo sampling

The distribution of feasible fluxes for each reaction in the models used here were determined using Markov chain Monte Carlo (MCMC) sampling (*36*), as previously described (*37, 38*), and was implemented with the COBRA Toolbox v2.0 (*39*). Published uptake rates were used to constrain the models. To model more realistic growth conditions (*40*), sub-optimal growth was modeled. Specifically, the biomass objective function, a proxy for growth rate (*41*), was provided a lower bound of 90% of the optimal growth rate as computed by flux balance analysis (*33*). Thus, the sampled flux distributions represented sub-optimal flux-distributions, but still simulated fluxes relevant to cell growth and maintenance.

MCMC sampling was used to obtain thousands of feasible flux distributions (referred to here as “points”) using the artificially centered hit-and-run algorithm with slight modifications, as described previously (*37, 38*). Briefly, a set of non-uniform points was generated. Each point was subsequently moved randomly, while remaining within the feasible flux space. To do this, a random direction was first chosen. Second, the limit for how far the point can travel in the randomly-chosen direction was calculated. Lastly, a

new random point on this line was selected. This process was repeated until the set of points approached a uniform sample of the solution space, as measured using the mixed fraction metric described previously (*42*). A mixed fraction of approximately 0.50 was obtained, suggesting that the space of all possible flux distributions is nearly uniformly sampled.

For each reaction, a distribution of feasible steady-state flux values was acquired from the uniformly sampled points, subject to the network topology and model constraints. For the *E. coli* model such distributions of feasible flux values could be determined for 2,314 of the 2,382 reactions. The remaining 68 reactions were involved in loops (*43*) and therefore reliable flux estimates were not available. Thus, sampling distributions for these 68 reactions were removed from all analysis in this work. Similar measures were taken for all other models in this work.

Model parameterization

In general, metabolic models were used in their published format with published uptake and secretion rates (*5, 20-22, 31, 44*). For the 174 simulated media formulations in *E. coli*, uptake rates were estimated as follows. After setting the glucose uptake to zero in the *i*AF1260 model, flux balance analysis was used to find which of all other carbon sources could support growth. For each of the 174 growth-supporting carbon sources, an uptake rate was set that was consistent with the uptake rate of glucose in the published *i*AF1260 model (i.e., 8 mmol gDW^{-1} hr^{-1}), normalized by the number of carbons in the metabolite. For example, since glucose has 6 carbons, the uptake rate of lactate, with 3 carbons, was set as 16 mmol gDW^{-1} hr^{-1} (which is similar to the actual reported lactate uptake rate in M9 minimal media (*32*)). While this was used to standardize the media conditions, variations in carbon uptake rates did not significantly impact the results presented in this work.

All models were selected based on the availability of carefully curated genome-scale metabolic network reconstructions with measured metabolite uptake rates. Specific media conditions for the eukaryotic and archaea models included the following. *M. barkeri* growth was simulated on minimal media containing methanol, acetate, pyruvate, or H_2 and CO_2. *S. cerevisiae* growth was simulated with glucose, acetate, ethanol, and maltose minimal media. For *C. reinhardtii*, three growth conditions were used: light with no acetate, light with acetate, and dark with acetate. Further details on media formulations are provided elsewhere (*20-22, 44*).

Ranking of flux magnitude

We compared the flux magnitudes of SERxns and GERxns. In order to avoid biases resulting from variations between growth conditions, we used a rank-based metric to compare flux between conditions. This was done as follows. The median flux magnitude values for each reaction were calculated from the MCMC-sampled flux loads for each of the 174 media formulations. For each condition, reactions were filtered out if they were transporter related, involved in loops, non-enzymatic, or unable to carry flux. The median fluxes for each reaction were then rank-ordered and the distributions of ranks for SERxns and GERxns were compared. In comparing the relative flux between reactions,

higher flux magnitudes correspond to higher rank in this study. The significance of higher flux magnitudes in SERxns were statistically evaluated by using one-tailed *t*-tests and Fisher's method.

Essentiality

Gene essentiality was assessed in comparison to a previously published set of 300 experimentally-determined essential genes in *E. coli* (*11*). In addition, an *in silico* analysis was used to assess reaction essentiality with respect to the synthesis of biomass precursors, since we hypothesize that the selective pressure would exert its influence through the reactions themselves. The *in silico* approach used MCMC sampling to simulate growth (>95% of the *in silico*-predicted optimal growth rate). The distributions of feasible flux values of each reaction was used to assess the correlation of flux between it and the biomass reaction (a pseudo-reaction that simulates the consumption of all biomass precursor metabolites in order to produce biomass) (*41*). Reactions that significantly contribute to or are essential for growth are identified by having a significant *P*-value from the computation of the Pearson's correlation coefficient. While we selected a *P*-value cutoff of $1x10^{-10}$, the results were consistent for any reasonable *P*-value cutoff (Fig. S6). These correlated reactions contain no redundant pathways, and would therefore provide the most stringent selective pressures since they are the most essential reactions.

Flux change predictions in media shifts

To simulate changes in reaction flux occurring in a shift between two conditions, the sampled fluxes for each reaction were compared between two media conditions as follows. First, reactions that carried no flux in both conditions or that were involved in loops (*43*) were removed and not used in further analysis. Next, flux magnitudes were normalized between each pair of media conditions. To do this, the flux value of each sample point was divided by the sum of all flux magnitudes for the sample point.

$$normed\_flux_{ij} = flux_{ij} / \sum_{i=1}^{n} abs(flux_{ij}) \quad , n = \text{number of reactions}$$

Once the flux values were normalized, the changes of fluxes between two conditions were determined as previously described (*37, 45*). Briefly, differential flux for each reaction was determined by assuming that a reaction is differentially activated if the distributions of feasible flux states (obtained from MCMC sampling) under two different conditions do not significantly overlap. For each metabolic reaction, a *P*-value was obtained by computing the probability of finding a flux value for a reaction in one condition that is equal to or more extreme than a given flux value in the second condition. Significance of *P*-values were adjusted for multiple hypotheses (FDR $\leq$ 0.01).

Clustering of reaction changes

An *m* x *n* binary matrix with *m* reactions and *n* media shifts (n = 15,051) was made, detailing in which shifts each reaction significantly changed flux (FDR $\leq$ 0.01). The reactions were subjected to *K*-means clustering (K = 3). Clustering was repeated 100

times with different seed values to find consensus clusters. Enrichment tests in the clusters were done using the hypergeometric test.

Enrichment of metabolic regulation

Lists of metabolic proteins with post-translational modifications (PTMs) were obtained from studies that identified sites of protein acetylation, phosphorylation, and succinylation in *E. coli* (*18, 46-48*). All reported occurrences of non-covalent metabolite-mediated metabolic regulation were obtained from EcoCyc (*35*). Metabolic regulation events labeled in EcoCyc as allosteric, noncompetitive, and uncompetitive were used. Competitive inhibition was also analyzed and compared with the others to distinguish between different regulatory properties of these enzymes (Fig. S11). Enrichment and depletion of PTMs and metabolite-mediated metabolic regulation events in the gene lists and reaction clusters were determined using the hypergeometric test.

Cosine similarity

The patterns of how reaction flux changes when two or more reactions share the same generalist enzyme were estimated by using the cosine similarity metric. For each shift, the median flux magnitudes of a reaction in conditions *x* and *y* were represented as a vector ($R_\alpha(f_x, f_y)$). The similarity score of two reactions, α and β, was then measured by the cosine similarity of the two vectors, R_α and R_β.

$$\cos(R_\alpha, R_\beta) = \frac{(R_\alpha \bullet R_\beta)}{|R_\alpha| \times |R_\beta|}$$

The similarity score of reactions catalyzed by the same enzyme (e_i) was calculated as the mean value of all pair-wise cosine similarity scores for reactions catalyzed by that generalist enzyme.

$$\text{similarity}(e_i) = \sum^{n} \text{abs}(\cos(R_\alpha, R_\beta)) / n$$

For example, for an enzyme *e1* that catalyzes three reactions (*r1*, *r2*, and *r3*), the flux similarity score of *e1* is calculated as an average value of cosine distances of three reaction pairs (Fig. S10A). Similarity scores for generalist enzymes were compared to 2000 cosine similarity scores from randomly-selected reaction pairs.

Assessment of kinetic parameters

Differences in enzyme kinetic parameters were assessed using previously curated k_{cat} values for wild-type enzymes in *E. coli* K12 (*23*). Significance in comparisons were made using the Wilcoxon rank sum test.

Supplementary Text

More detailed properties of generalist enzymes and an assessment of classification robustness

Among these 1,081 enzymes, 677 enzymes were classified as specialists, which only catalyze one specific chemical reaction. The 404 enzymes catalyzing more than one reaction were classified as generalist enzymes (Fig. S17A). These specialists and generalist enzymes are encoded by 720 genes and 473 genes, respectively. We anticipate that a small fraction of enzymes may be misclassified due to incomplete study of some enzymes or incorrect interpretation of biochemical studies. However, these concerns should not substantially affect the conclusions in this work, since variations on the categorization yield qualitatively similar results (see Fig. S3). In this classification, among the canonical generalist enzymes, we found that 89% of the generalist enzymes are enzymes that exhibit substrate promiscuity, and 11% are multifunctional enzymes (Fig. S1A). Furthermore, 3% were multifunctional with promiscuous active sites. Lastly, 2% were bifunctional enzymes in which the different catalytic activities were used exclusively to catalyze two reactions in series on one substrate (e.g., in substrate channeling). More detailed explanation for combining substrate promiscuity and multifunctional enzymes is provided in the section entitled '*Justification of combining promiscuous and multifunctional enzymes into the "generalist" class*'.

Individual generalist enzymes catalyzed between 2 and 249 reactions. Once the transporters were filtered out, about 50% (189 enzymes) of generalist enzymes catalyzed two distinct reactions in the model. Furthermore, the catalytic degree of generalist enzymes closely followed power-law distribution ($P(k) \sim ck^{-1.6}$) (Fig. S1B).

Recently, Khersonsky *et al.* introduced the concept of degree of promiscuity, which measures the functional diversity of enzymes (*49*). Here, we assessed the degrees of functional diversity for generalist enzymes using the unique Enzyme Commission (EC) numbers. The EC numbers of enzymes were collected from the KEGG database (*50*). As Fig. S1C depicts, the majority of generalist enzymes had a single or two unique EC numbers, suggesting that reactions catalyzed by many generalist enzymes in *E. coli* often involve similar substrates.

One of the most trivial explanations for the categorization of an enzyme as a specialist or as a generalist is the biased depth of knowledge in enzymes. We attempted to rule out this possibility with respect to *E. coli* metabolism, as carefully detailed and curated in the *i*AF1260 reconstruction. Specifically, we quantified the number of abstracts per gene of *E. coli* K-12 MG1655 in MEDLINE to assess if the accumulated knowledge is uniformly represented in the classes of specialist and generalist enzymes. The number of MEDLINE abstracts was calculated by using the genes2pubmed file from NCBI FTP repository (downloaded October 20, 2011). The 1,260 genes in *i*AF1260 had total number of 61,727 citations (Fig. S1D). Genes associated with specialist and generalist enzymes had similar numbers of MEDLINE abstracts (Fig. S1E). That is, we found no correlation between our classification and knowledge depth. Neither specialist nor generalist enzymes had been studied in greater depth, which leads to the conclusion that the catalytic degree of generalists in *i*AF1260 does not likely result from increased study of either group.

We further investigated this hypothesis that specificity assignments might solely reflect the level of study by looking to see if generalists occurred more in the well-studied central metabolic pathways or less studied periphery of the metabolic network. The specialists and generalists are not uniformly distributed among metabolic subsystems in *E. coli* (Fig. S1F). Specialist enzymes are more frequently associated with core processes such as tRNA charging and central metabolism (including glycolysis, TCA cycle, and the pentose phosphate pathway), and generalist enzymes are more frequently associated with more peripheral pathways, such as murein biosynthesis and glycerophospholipid metabolism.

One may argue that the activities of generalists in these peripheral pathways may be latent promiscuous enzyme activities measured *in vitro*, and that these activities are likely not functional *in vivo*. In this study we focus on these *in vivo* enzymatic activities, since only these will be under evolutionary selection. Furthermore, the classification of generalists and specialists assumes that all of the associated reactions can be active *in vivo* and do not represent latent promiscuous activities. Thus, it is necessary that the reactions be assessed as to if they can take place in the context of the metabolic network (i.e., if pathways providing reaction precursors are present and carry flux, and if the products can be metabolized or secreted). To do this, we simulated flux through GERxns using experimentally-measured phenotype data from six media conditions (i.e., substrate uptake rates). These six conditions included M9 minimal media supplemented with (i) glucose, (ii) glycerol, (iii) propylene glycol (evolved, see (*30*)), (iv) glycerol (evolved on propylene glycol, see (*30*)), (v) glucose under anaerobic conditions, and (vi) glucose with nitrate under anaerobic conditions. Flux distributions were simulated using MCMC sampling implemented in the COBRA Toolbox v.2.0 (*39*).

After filtering out thermodynamically-infeasible loops (*43*) and transport reactions, the ratio of reactions that carry non-zero predicted flux was computed. From this analysis, about 80% of GERxns carried flux in any one growth condition (Fig. S2). When this analysis was repeated for the 174 possible carbon sources (aerobic and anerobic), 85% of the GERxns could carry flux in at least one growth condition. We note that transporters and tRNA charging reactions were removed from this analysis. The high percentage of active GERxns, and the fact that the percentage was not lower than SERxns, suggests that the GERxns are more likely to be reactions that occur *in vivo* and not latent promiscuous activities that are only relevant *in vitro*.

While we are confident in the classification system, we note that it is entirely possible that the classification of individual enzymes may change with further study. However, it seems that specialist enzymes tend to be well-studied and overrepresented in core metabolic processes, and that generalist enzyme functions mostly do not represent latent promiscuous activities with only *in vitro* relevance. Furthermore, we have tested the robustness of many results with respect to variations in classification and found qualitatively similar results (Fig. S3). Therefore, we are confident that the classification is robust enough to be amenable to variations and updates and still provide qualitatively similar results, compared to what was presented in the main text of this study.

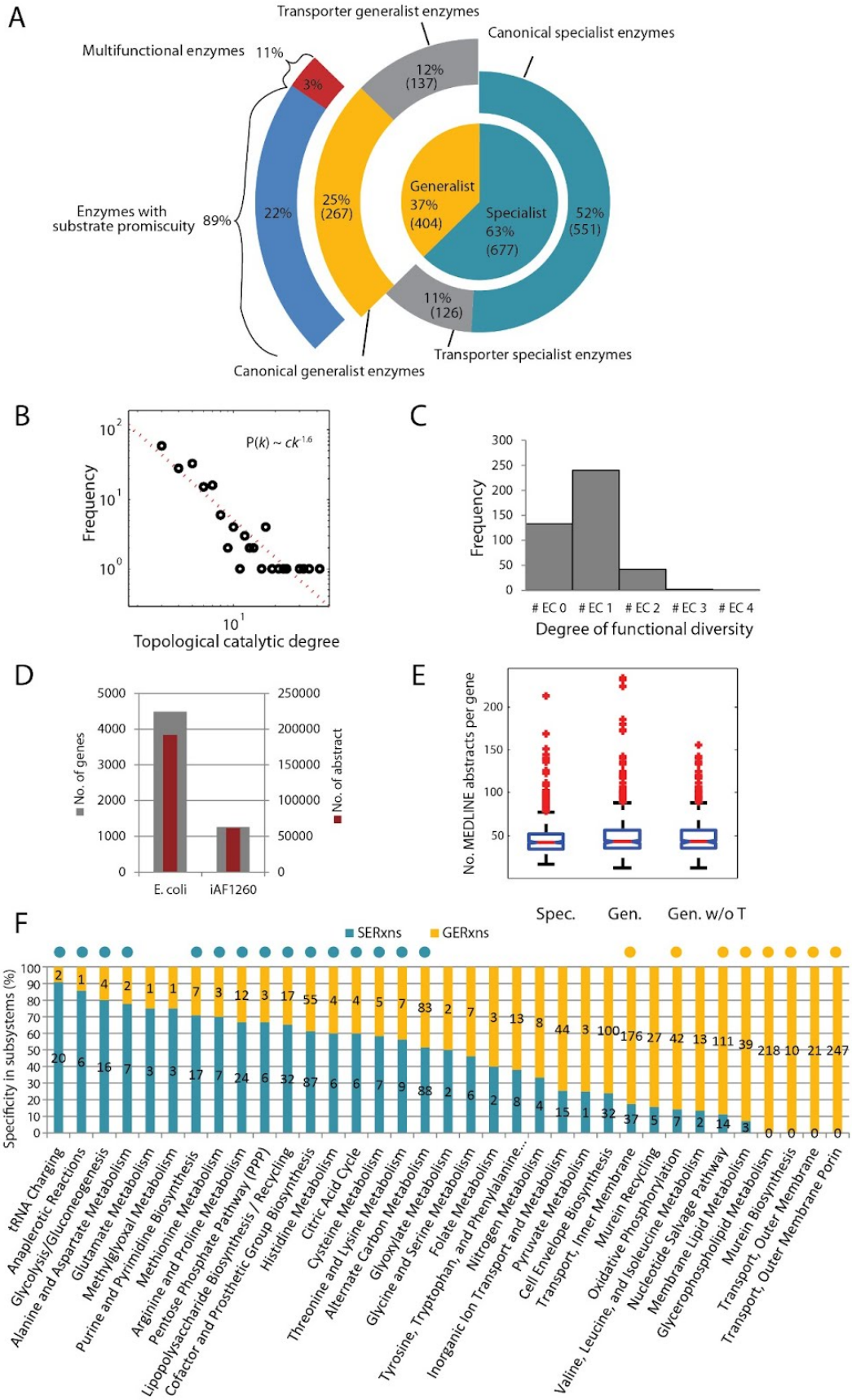
A
Transporter generalist enzymes
Canonical specialist enzymes
Multifunctional enzymes
11%
3%
12%
(137)
Enzymes with
substrate promiscuity
89%
22%
25%
(267)
Generalist
37%
(404)
Specialist
63%
(677)
52%
(551)
11%
(126)
Canonical generalist enzymes
Transporter specialist enzymes
B
P(k) ~ ck^1.6
Frequency
Topological catalytic degree
C
Frequency
EC 0
EC 1
EC 2
EC 3
EC 4
Degree of functional diversity
D
No. of genes
No. of abstract
E. coli
iAF1260
E
No. MEDLINE abstracts per gene
Spec.
Gen.
Gen. w/o T
F
SERxns
GERxns
Specificity in subsystems (%)
tRNA Charging
Anaplerotic Reactions
Glycolysis/Gluconeogenesis
Alanine and Aspartate Metabolism
Glutamate Metabolism
Methylglyoxal Metabolism
Purine and Pyrimidine Biosynthesis
Arginine and Proline Metabolism
Methionine Metabolism
Pentose Phosphate Pathway (PPP)
Lipopolysaccharide Biosynthesis / Recycling
Cofactor and Prosthetic Group Biosynthesis
Histidine Metabolism
Citric Acid Cycle
Cysteine Metabolism
Threonine and Lysine Metabolism
Alternate Carbon Metabolism
Glyoxylate Metabolism
Glycine and Serine Metabolism
Folate Metabolism
Tyrosine, Tryptophan, and Phenylalanine...
Inorganic Ion Transport and Metabolism
Nitrogen Metabolism
Pyruvate Metabolism
Cell Envelope Biosynthesis
Transport, Inner Membrane
Murein Recycling
Oxidative Phosphorylation
Valine, Leucine, and Isoleucine Metabolism
Nucleotide Salvage Pathway
Membrane Lipid Metabolism
Glycerophospholipid Metabolism
Murein Biosynthesis
Transport, Outer Membrane
Transport, Outer Membrane Porin

Fig. S1.

Properties and knowledge depth of specialist and generalist enzymes. (A) Generalist enzymes can be further partitioned into subclasses. The majority of the non-transporter generalist enzymes in the *E. coli* metabolic network exhibit substrate promiscuity. Only 11% had more than one active site. (B) The network topology degree of generalist enzymes is represented in log-scale, with the X-axis showing the number reactions, and y-axis showing the frequency. (C) The distribution of degrees of promiscuity for generalist enzymes shows that a large percentage of generalist enzymes catalyze multiple similar reactions. The X-axis shows the number of unique EC numbers for an enzyme, with their frequency represented by the y-axis. (D) The total number of MEDLINE abstracts were counted in *E. coli* K-12 MG1655 and *i*AF1260. (E) The classification is likely not biased by the lack of study of either class of enzymes. Box plots show the number of MEDLINE abstract for each gene for specialists (Spec.), generalists with transporters (Gen.), and generalists without transporters (Gen. w/o T). (F) The numbers of SERxns and GERxns are shown for all 35 *E. coli* metabolic functional subsystems in *i*AF1260. The ratio of SERxns (green) and GERxns (yellow) in the functional subsystems is depicted on each bar. The circles above functional subsystem bars represent which subsystems have a significant over-representation of SERxns or GERxns against the background distribution (Green dot: enrichment of SERxns, yellow dot: enrichment of GERxns, hypergeometric test, $P \leq 0.05$).

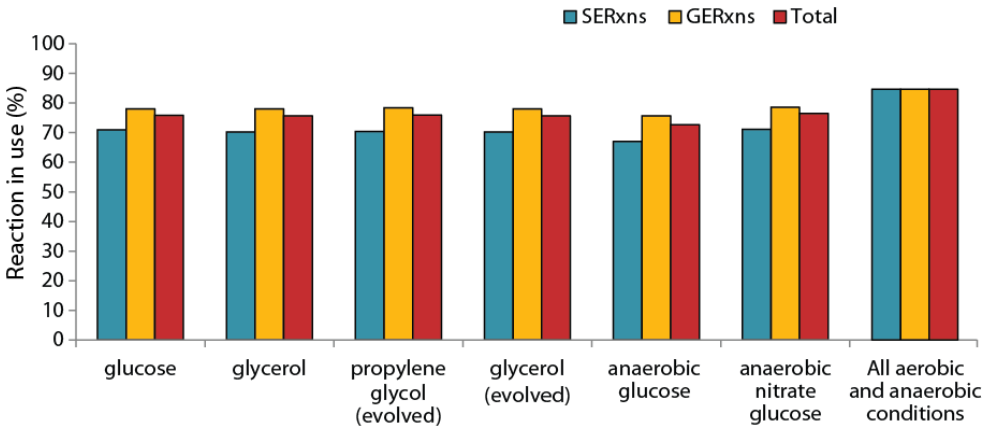

Fig. S2.

Most reactions associated with specialist and generalist enzymes can be active *in silico*. The percentage of reactions carrying flux for each growth condition was computed for reactions associated with specialist enzymes (SERxns) and generalist enzymes (GERxns). For any given condition, ~80% of the GERxns (yellow) were active, and 85% of the GERxns were active in at least one of all simulated conditions (174 carbon sources under aerobic and anaerobic conditions).

More stringent classifications of generalists

In order to demonstrate the robustness of our results with variations in the classification of SERxns and GERxns, we further examined the percentage of reactions that change in environmental shifts for (i) GERxns with transport reactions added back into the set, (ii) conditional GERxns (cases in which more than one reaction is active for a generalist enzyme under the given condition) and (iii) subclasses of GERxns (reactions associated with promiscuous and multifunctional enzymes).

The assignment of conditional GERxns followed the schema in Fig. S3A. Since the magnitude of flux varies in the environmental changes, a generalist enzyme might only catalyze only a single reaction under a specific condition since substrates for its other reactions might not be available under that condition. For example, suppose there is a generalist enzyme that catalyzes reaction r_1 and reaction r_2 (Fig. S3A). If, in a given growth condition A, both reactions r_1 and r_2 carry flux, the reactions are classified as type I conditional GERxns for the given condition. However, if in another condition B, the enzyme only catalyzes one reaction (r_1) (since reaction r_2 cannot carry flux), the reactions are classified as type II conditional GERxns in condition B. For the assessment of conditional GERxns, type II conditional GERxns are not considered in our shift analysis here. The total number of type I conditional GERxns in five representative perturbations are listed in the Table S2.

When we assessed the effects of GERxns with transporting reactions, conditional GERxns and subclasses of GERxns on the fraction of reactions that change, it is clear that none of these changes qualitatively change the results of this study. Specificially, the fraction of reactions with significantly different flux values in shifts continues to vary greatly by enzyme class (Fig. S3B). This analysis was also repeated on a recently published update to the *E. coli* K12 metabolic reconstruction, *i*JO1366 (*51*), which showed qualitatively similar results (Fig. S3C).

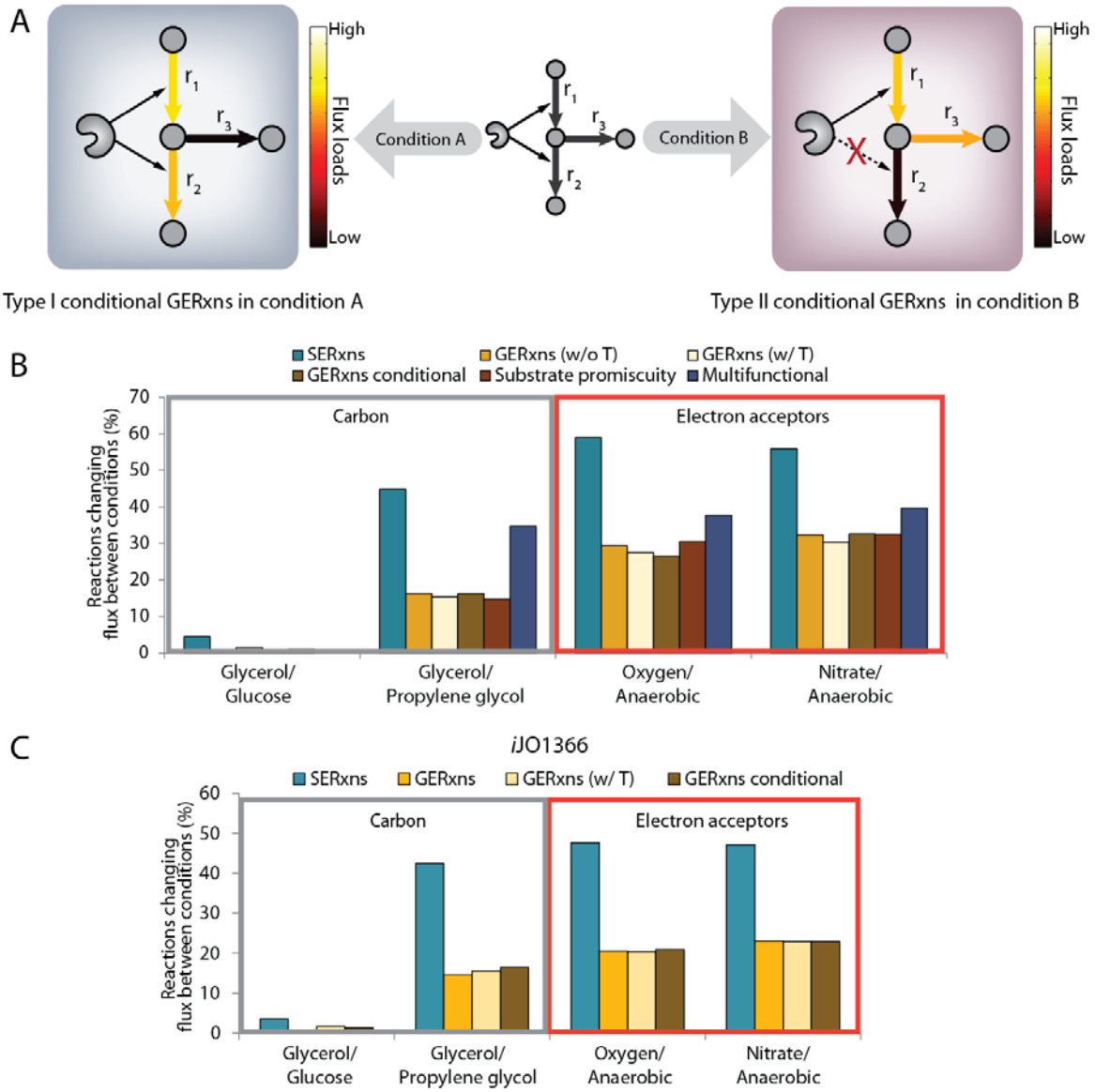

Fig. S3.

Robustness of result showing the association of specialist enzymes with growth condition shifts. (A) A conceptual illustration is provided, describing conditional GERxns. Left: For type I conditional GERxns, their associated generalist enzyme catalyzes more than one active reaction under a given condition. Right: For type II conditional GERxns, their associated generalist enzyme catalyzes only one active reaction under the given condition. Only type I conditional GERxns were considered in our shift analysis here. (B) The percentage of SERxns (green), GERxns without transporters (GERxns w/o T; yellow), GERxns including transporters (GERxns w/ T; light-yellow), conditional GERxns (brown), substrate promiscuity GERxns (tan), and multifunctional GERxns (dark-blue) that change were computed using the *i*AF1260 model. (C) The percentage of SERxns (green), GERxns without transporters (yellow),

GERxns including transporters (light-yellow), and conditional GERxns (brown) that change were also computed using the *i*JO1366 model.

Strengths and limitations of the model employed in this analysis

The results of this work depends in part upon the accuracy of the models employed. The models here are "genome-scale" with respect to metabolism, but do not include non-metabolic processes. This may be a modest concern since dynamic natural environments exert their influence through complex regulation, signaling, and protein-protein interactions. However, even though this work does not explicitly account for these mechanisms, we feel that the study of properties of enzyme specificity can be done reliably here for a few reasons we list below.

First, the definition of specialist and generalist enzymes is the basis of this work. These designations are based on the curation of decades of careful biochemical studies. This curation process has been handled by several researchers over the course of about 20 years (*5, 51-56*). There may be some inaccuracies in the *E. coli* model from mistakes in previous biochemical studies or in the interpretation of these studies. However, it is believed that these are mostly correct because of the success of these models in predicting growth phenotypes (*57*). We also note that the authors of this work have manually curated all generalist enzymes and classified these as multifunctional or exhibiting substrate promiscuity (see Database S1).

Second, the model used accounts for almost a third of all genes in *E. coli*. Furthermore, the detailed biochemical functions are included, manually curated, and all chemical reactions are mass balanced. While it may be desirable to include processes that are not as well characterized, such as transcription regulation, protein-protein interactions, etc., efforts are still underway to find modeling methods that best account for these and more accurately simulate the effects of these interactions. However, since the question we have approached in this study addresses the evolution of metabolic enzymes, the genome-scale metabolic network accounts rather well for the scope and functions of this system. It allows the assessment of selective pressures on it.

Third, the accuracy of the *E. coli* metabolic model is apparent in its ability to accurately predict growth phenotypes on measured growth conditions for 76% of different minimal media formulations and 92% of tested single gene deletion mutants (*5*). While the predictions are less reliable for organisms that have not been characterized biochemically as carefully, these models provide reliable predictions for how all of the metabolic enzymes contribute to the cell phenotype. Thus, since we are directly investigating the evolution of metabolic enzymes, the metabolic network accounts for all of these enzymes and many of the factors shaping enzyme specificity.

Fourth, we show in the supporting materials here that differential flux under substrate shifts correlated strongly with differential gene expression (Fig. S8). This is also true for differential expression of protein, as measured using quantitative proteomics (Lewis, *et al.*, in preparation). Other studies have also shown that the expression of metabolic genes and proteins is optimized in laboratory evolution (*58*). Therefore, it seems that in metabolism, gene and protein expression may change to follow the metabolic needs of the cell, as predicted *in silico*.

Fifth, the constraints imposed on the metabolic network by mechanisms such as signaling and transcription regulation are malleable. These mechanisms evolve more

rapidly than does metabolism (*59*), and so if they are temporarily constraining the evolution of enzymes in a sense that is non-optimal, evolution can ease these constraints (see for example (*60, 61*)). Thus, it is likely that they do not exert a substantial influence on the time scale in which enzymes evolve.

In conclusion, we do not expect that all flux predictions will be completely correct, since this model suffers a small rate of incorrect predictions when genes are systematically removed (suggesting that pathway usage predictions are not completely correct or that transcription regulation might be temporarily affecting phenotype). However, the fact that most growth and gene deletion predictions are correct provides some confidence in the assessment of general network properties as we present here (e.g., flux of two general classes of enzymes, essentiality of reaction deletion, and differential reaction flux under varying environments). In addition, the comparison of our results with known enzyme regulation and post-translational modifications only provide further experimental evidence that the results presented here are supported by real data.

In the future, we anticipate that the results presented here may be enhanced as models begin to include other cell processes. While much of the literature on enzyme specificity focuses on metabolic reactions, there are many examples of specificity of protein-protein interactions, kinase-target interactions, codon-tRNA interactions, etc. that may be addressed in the future. Current efforts are now expanding the scope of these models to account for additional systems such as the transcription and translation machinery (*62*) and also protein structures (*63, 64*). While the model used here can still account for the biochemical functions of ~1/3 of all proteins in *E. coli*, the new larger models will facilitate more broad studies since they will account for mechanisms including transcription, translation, and transcription regulation. Since these will add ~600 more genes in *E. coli*, we anticipate that they will provide additional insight but that the results should not deviate far from the results we present here. Future work, with improved genome-scale models, will be important to build upon the results in this work.

Specificity is selected and maintained in high-flux metabolic pathways

In the main text, it was shown that SERxns maintain higher flux than do GERxns (Fig. 1C). While this was demonstrated by looking at the median value of all feasible flux levels for each reaction, this result was also obtained when using the average or mode of flux for each reaction (Fig. S4A). The rank distribution of median flux through SERxns is significantly higher than flux through all GERxns, GERxns catalyzed by enzymes exhibiting substrate promiscuity, and GERxns catalyzed by multifunctional enzymes (Fig. S4B-C). High-flux SERxns (reactions contributing to the top 10% flux values in at least one condition) were enriched in central and anabolic metabolism (Fig. S4C) (hypergeometric $P < 0.01$). However, a fraction of SERxns consistently demonstrated a low flux. These lower-flux SERxns are enriched in 'Cofactor and Prosthetic Group Biosynthesis', 'Lipopolysaccharide Biosynthesis/Recycling', and 'Cell Envelope Biosynthesis' pathways (hypergeometric $P < 0.01$). When comparing flux magnitudes of each pathway to the other 35 pathways, 'Cofactor and Prosthetic Group Biosynthesis' and 'Lipopolysaccharide Biosynthesis / Recycling' pathways clearly carried low flux in all 174 conditions (Fig. S4D). The genes associated with these pathways were also essential for growth (Fig. S6A). We note, however, that flux estimates for some cofactor

biosynthesis pathways might be lower than expected *in vivo* because they synthesize the few biomass components for which the synthesis rates are unknown and were thus previously provided low estimates (*5*). However, on average, specialist enzymes are selected for pathways that must maintain a higher flux to aid in the processing of nutrients to make the major biomass components.

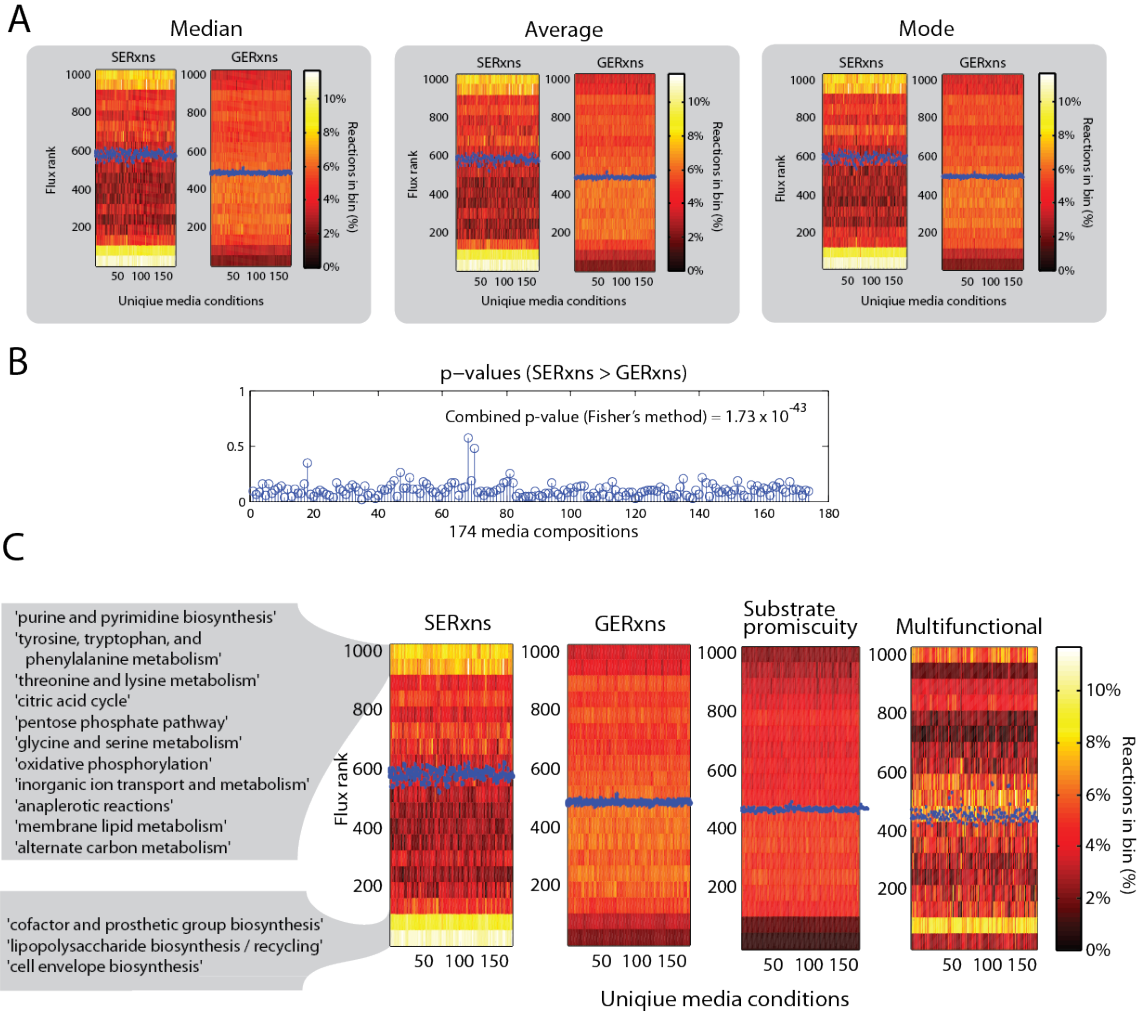
A
Median
Average
Mode
SERxns
GERxns
Flux rank
Reactions in bin (%)
Uniqiue media conditions
B
p–values (SERxns > GERxns)
Combined p-value (Fisher's method) = 1.73 x 10^-43
174 media compositions
C
'purine and pyrimidine biosynthesis'
'tyrosine, tryptophan, and phenylalanine metabolism'
'threonine and lysine metabolism'
'citric acid cycle'
'pentose phosphate pathway'
'glycine and serine metabolism'
'oxidative phosphorylation'
'inorganic ion transport and metabolism'
'anaplerotic reactions'
'membrane lipid metabolism'
'alternate carbon metabolism'
'cofactor and prosthetic group biosynthesis'
'lipopolysaccharide biosynthesis / recycling'
'cell envelope biosynthesis'
SERxns
GERxns
Substrate promiscuity
Multifunctional
Flux rank
Reactions in bin (%)
Uniqiue media conditions
D
alanine and aspartate metabolism
alternate carbon metabolism
anaplerotic reactions
arginine and proline metabolism
cell envelope biosynthesis
citric acid cycle
cofactor and prosthetic group biosynthesis
cysteine metabolism
folate metabolism
glutamate metabolism
glycerophospholipid metabolism
glycine and serine metabolism
glycolysis/gluconeogenesis
glyoxylate metabolism
histidine metabolism
inorganic ion transport and metabolism
lipopolysaccharide biosynthesis / recycling
membrane lipid metabolism
methionine metabolism
methylglyoxal metabolism
murein biosynthesis
murein recycling
nitrogen metabolism
nucleotide salvage pathway
oxidative phosphorylation
pentose phosphate pathway
purine and pyrimidine biosynthesis
pyruvate metabolism
threonine and lysine metabolism
transport, inner membrane
transport, outer membrane
transport, outer membrane porin
tRNA charging
tyrosine, tryptophan, and phenylalanine metabolism
unassigned
valine, leucine, and isoleucine metabolism
log(A/B)
A = median(pathway of interest), B = median(others)
'cofactor and prosthetic group biosynthesis' vs. others
frequency
rank order
other subsystems
'cofactor....'
A = median rank order('cofactor and...') = 125
B = median rank order(others) = 666.5
log(A/B) = -1.673

Fig. S4.

A detailed assessment of flux vs. specificity. (A) The assessment of flux vs. specificity was conducted using the median, average, and mode flux values for all reaction flux estimates on all 174 media conditions. The magnitude of flux for each reaction was rank ordered, with a rank of 1 being the smallest non-zero flux. On average, flux through SERxns was consistently higher than flux through GERxns. (B) *P*-values were computed against the null hypothesis that the flux through SERxns were not higher than the flux through GERxns for each of the 174 *in silico* media formulations using the one-tailed *t*-test. Each column of the heatmap represents the binned flux magnitudes for one of the 174 carbon substrate conditions for SERxns (left) and GERxns (right). The color of each cell shows the percentage of reactions that are within the range of flux ranks. The blue circle represents the median rank of reactions in each media formulation. (C) SERxns also have a higher flux then GERxns catalyzed by multifunctional enzymes and enzymes exhibiting substrate promiscuity. Metabolic subsystems were identified that were enriched in the highest and lowest 10% of flux for SERxns. (D) The flux magnitudes were further analyzed for each metabolic subsystem across all 174 growth conditions. The heatmap depicts flux magnitudes between a subsystem of interest and the other pathways. Black cells show that the subsystem of interest has a lower flux magnitude than the other pathways, and white shows higher flux magnitudes than the other pathways.

Enzyme kinetic parameters and specificity

Using a curated set of kinetic parameters (*23*) derived from the BRENDA database, we assessed kinetic properties of enzyme specificity. We note that relatively few studies had both k_{cat} and K_m values in the curated kinetic parameters for *E. coli*. However, this is not too much of a concern because steady-state metabolic flux is more dependent on k_{cat} values since many metabolites are found in concentrations near their K_m values (*65*). Since the cells have managed to adjust their *in vivo* metabolite levels to their needed values, higher k_{cat} values should be more important for higher flux reactions (consistent with the positive correlations between simulated flux and measured k_{cat}; fig. S13C), while lower K_m values should be more important for lower flux essential enzymes that need to avoid substrate competition. Since a recent study (*23*) suggests that there may be some of a trade-off between k_{cat} values and K_m values (i.e., turnover rate improvement may occur at the cost of substrate binding affinity), the K_m values might also increase as high flux enzymes evolve to increase their turnover rate.

As demonstrated in Fig. 1C and Fig. S4, many of the SERxns consistently demonstrated a much higher flux than most GERxns. We hypothesized that the the high flux would influence the evolutionary enhancement of enzyme activity in order to increase the turnover rate of the enzymes and thereby reduce the amount of required enzyme mass. Consistent with this hypothesis we found that the k_{cat} values of the high flux SERxns (reactions with a rank greater than 900 in at least one growth condition) were significantly higher (Wilcoxon rank-sum $P < 2.91 \times 10^{-8}$) than the rest of SERxns (Fig. S5A). Moreover, the k_{cat} values for these high-flux SERxns were also higher than those of all other enzymes (Wilcoxon rank-sum $P < 2.84 \times 10^{-7}$; Fig. S5C). Thus, these high-flux enzymes have possibly evolved to become specialists to allow the attainment of

more optimal k_{cat} values. With this higher k_{cat}, the enzyme concentration may decrease while maintaining the high flux, thereby decreasing the cost of the duplication event and possibly freeing up space for other enzymes that enhance growth. We note that while the K_m values also increase (Fig. 5B,D), the magnitude of the median k_{cat} values tends to increase more than the magnitude of the median K_m value for these high-flux SERxns. The increased K_m may result as a trade-off for achieving a higher k_{cat} (*23*), but that this is likely not a concern since the k_{cat} seems to increase more and the metabolite concentrations seem to compensate since they are often are near their associated enzyme K_m values.

Lower K_m values for an enzyme suggests that the enzyme has a particularly higher affinity for the metabolite in question. A higher affinity may be beneficial to avoid substrate competition at an enzyme's active site, which may be particularly important for enzymes that synthesize essential biomass components. While some essential enzymes maintain a higher flux, it is interesting to note that many of these are specialist enzymes that maintain a lower flux (compare figs. S4C and S6A). This is consistent with the observation that essential enzymes, on average, have slightly lower k_{cat} values (Fig. S5E). However, consistent with the hypothesis that they might have significantly lower than average K_m values to decrease the chance of substrate competition in the synthesis of the essential biomass precursors, we found the K_m values for essential enzymes indeed tends to be lower than the remaining enzymes (one-sided Wilcoxon $P < 1.13 \times 10^{-11}$;Fig. S5F).

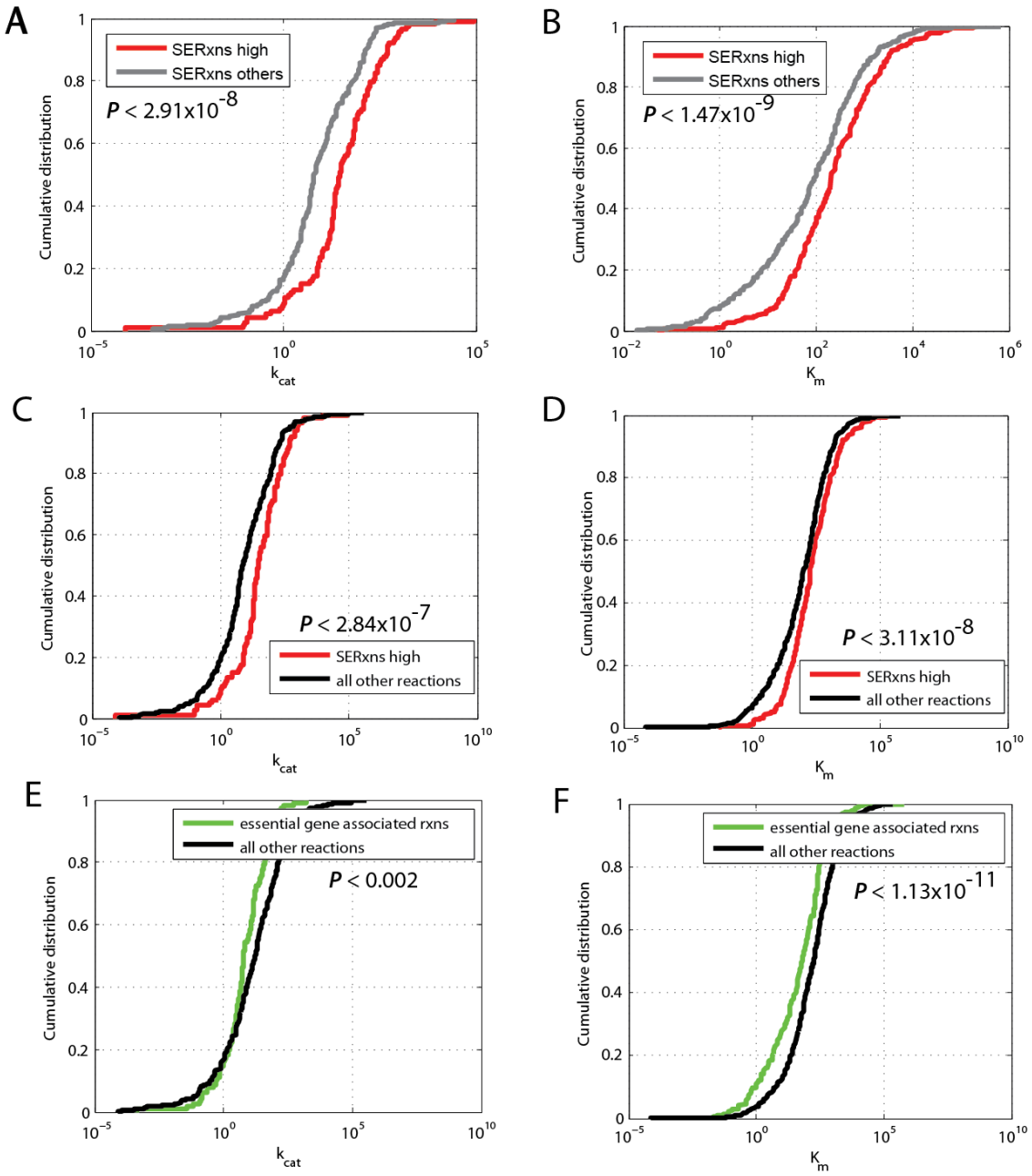

Fig. S5.

Kinetic parameters of specialist and generalist enzymes in BRENDA. The distribution of k_{cat} and K_m values in high flux SERxns are compared against the values associated with (A-B) all other specialist enzymes and (C-D) all other enzymes in *E. coli*. The k_{cat} values seem to increase more in magnitude than do K_m values, which may just increase as a trade-off to attain higher k_{cat} values. (E-F) Essential enzymes, which tend to have lower flux, also have slightly lower k_{cat} values. However, they exhibit much lower K_m values, which suggests that they may have evolved higher affinity for their substrates to avoid substrate competition.

Essentiality

A number of enzymes are essential for the synthesis of several biomass precursors and other cell processes. Since these must be carefully regulated for growth, the associated enzymes may have evolved to become specialists to avoid substrate competition. Consistent with this, we found that specialist enzymes are more essential (Fig. 1D). Furthermore, many of the experimentally validated essential genes are found in low-flux pathways synthesizing essential cofactors, prostectic groups, and the cell envelope, or involved in tRNA charging (Fig. S6A). Most of these are ancient processes upon which other enzymes depend.

To test if the model supports this hypothesis that enzyme specificity is associated with essential processes for growth, we used MCMC sampling to simulate growth. Since MCMC sampling provides a number of randomly selected feasible steady state flux distributions, these distributions can be used to test the dependencies between any given pair of reactions, i.e., a correlation coefficient can be computed to test if two reactions depend on each other.

For the 174 sampled media conditions, the correlation coefficients between each reaction and the biomass reaction were computed. Reactions that significantly contribute to or are essential for growth are identified by having a significant *P*-value from the computation of the Pearson's correlation coefficient. These correlated reactions represent the reactions for which there are no redundant pathways, and would therefore provide the most stringent selective pressures since they are the most essential reactions.

Interestingly, as the *P*-value gets more significant, it is clear that GERxns are less frequently correlated with growth rate (Fig. S6B). In glucose minimal media, GERxns only account for 19% of the essential reactions with a correlation significance cutoff of $P = 1x10^{-10}$ (see Fig. S6B, inset). This trend for few GERxns is seen for any reasonable correlation *P*-value cutoff (Fig. S6B,C). In like manner, within all 174 sampled media conditions, specialist enzymes are significantly enriched in the group of reactions that contribute most strongly with growth (Fig. S6D). SERxns account for 56% of all reactions with flux that correlates with the growth rate (Fig. S6D, inset), and this trend is seen in all growth conditions (Fig. S6E). Thus, since SERxns dominate the set of enzyme-catalyzed reactions necessary for optimal growth, it seems that the demand for biomass synthesis may exert a selective pressure that causes some enzymes to become specialists, while enzymes upon which growth is less dependent are under less pressure to specialize.

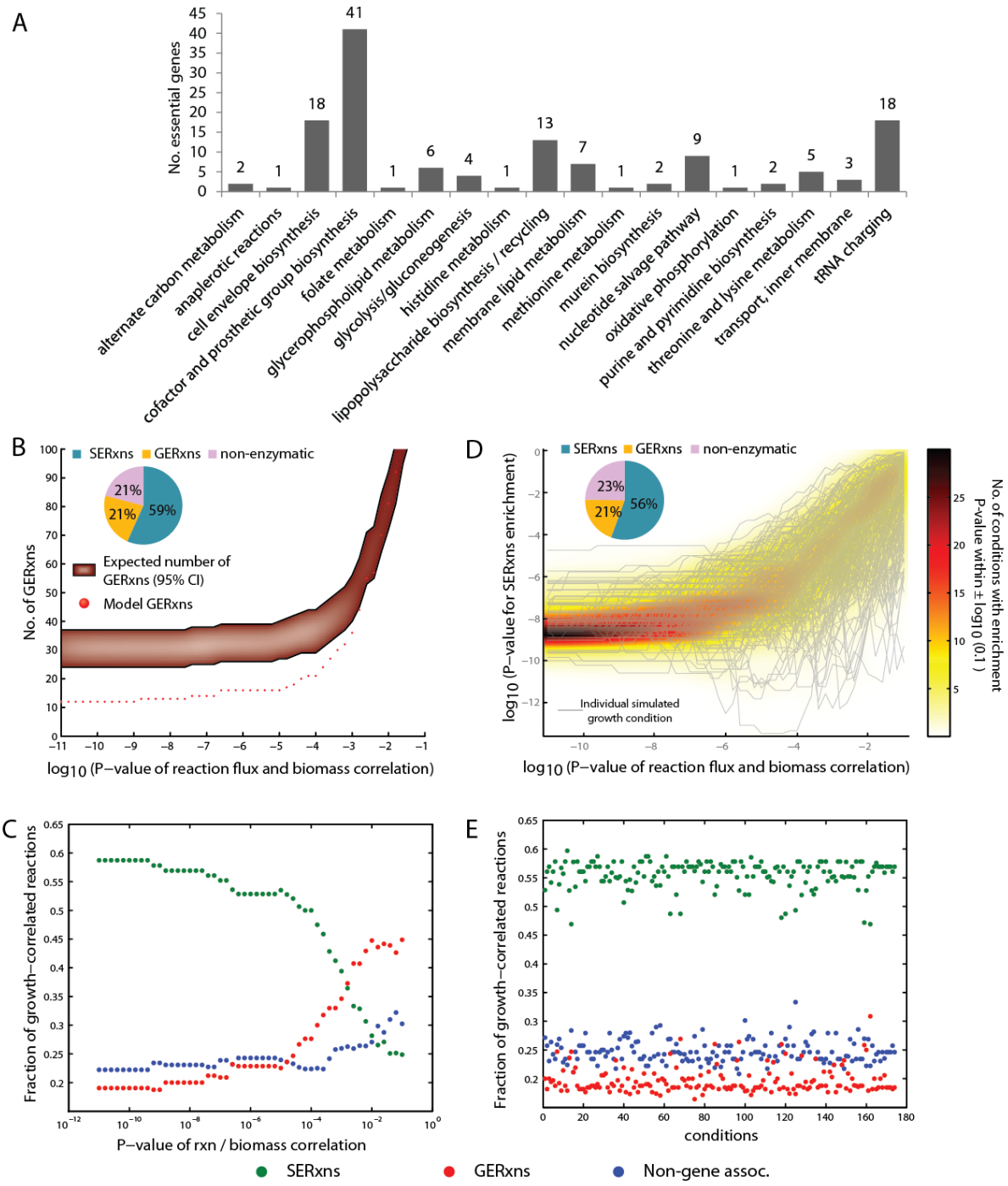

Fig. S6.

Specificity and its association with *in vivo* and *in silico* essentiality. (A) The number of essential genes (measured *in vivo*) was determined for each metabolic subsystem in the *i*AF1260 reconstruction. (B) Far fewer GERxns fluxes are significantly correlated with growth (i.e., flux through the biomass objective) than expected for growth on glucose minimal medium. The distribution of SERxns, GERxns, and non-enzyme-catalyzed reactions with a *P*-value < $1x10^{-10}$ is shown (inset), and (C) this dominance of SERxns

among the essential reactions increases with increasing correlation between reaction flux and *in silico*-predicted biomass production. (D) For all 174 simulated growth conditions SERxns are significantly enriched among reactions showing a correlation with growth. The mean distribution for all 174 conditions is shown for reactions with a *P*-value < $1x10^{-10}$ (inset), and (E) individual fractions are shown for all 174 media formulations.

Reasoning why dynamic environments may select for enzyme specificity: serine hydroxymethyltransferase as a case study

The cellular microenvironment is inherently dynamic. Metabolites can rapidly change concentration and some microbes see regular shifts in their environmental conditions (e.g., microbes in the gut), and can therefore change gene and protein expression in anticipation (*66*). When nutrients change in concentration, these changes directly affect the levels of metabolic flux throughout the metabolic network. If a flux needs to change, proper metabolic regulation can rapidly change the flux to avoid catalyzing undesirable reactions, and force the flux to adopt the needed steady-state flux distribution. Transcriptional regulation can also aid in adaptation. While these sorts of responses would be simple and allow the cell to easily control enzymes that catalyze one particular reaction, the combinatorial complexity significantly increases for generalist enzymes. Take, for example, serine hydroxymethyltransferase (GlyA), an enzyme that can catalyze a few reactions in *E. coli*. GlyA converts serine to glycine and transfers a methyl group to tetrahydrofolate. GlyA also catalyzes the hydrolysis of 5,10-methenylTHF to 5-formylTHF (*67*), and reversibly cleaves 3-hydroxy amino acids, such as 3-phenylserine, threonine, and allothreonine, to form glycine and an aldehyde (*68, 69*). In the model, all five reactions that are catalyzed by GlyA can be used in the known metabolic network under most simulated growth conditions. If, for a hypothetical shift, the serine hydroxymethyltransferase and L-allo-threonine aldolase fluxes had to increase, stay the same or decrease, there would be nine qualitative flux states that could be assumed (Fig. S7A-B). This number rapidly expands when one accounts for all five of the reactions catalyzed by this enzyme in the *i*AF1260 model. Thus, the complex regulation for all of these reactions may pose a disadvantage if the fluxes do not often co-vary (as seen for many reactions associated with generalist enzymes; see Fig. S10), especially if the reactions carry a high flux in many conditions.

For GlyA in particular, there are five reactions included in the *i*AF1260 model that it is known to catalyze. While claims have been made that the serine hydroxymethyltransferase reaction may be the primary reaction that occurs *in vivo* (*70*), we note that all five can be used in most simulated growth conditions, i.e., all of the substrates are available, and so it is likely that GlyA will also catalyze all five reactions *in vivo*. The flux through the D-alanine transaminase, L-alanine transaminase, and threonine aldolase reactions significantly changes in 3.7%, 3.9%, and 10.0% of the simulated substrate shifts, respectively, and they have lower flux than average, ranking at the 40^{th}, 30^{th}, and 47^{th} percentile for flux among all active reactions, respectively. Meanwhile, serine hydroxymethyltransferase and L-allo-threonine aldolase change flux more frequently. Their flux changes in 58% and 37% of the simulated media shifts, respectively. In addition, they maintain much higher flux values, ranking at the 95^{th} and 85^{th} percentiles. Thus given the explanation above, the serine hydroxymethyltransferase

and L-allo-threonine aldolase activities would be more sensitive to selective pressures for enzyme evolution because their flux frequently changes when the nutritional environment changes. In addition, they maintain a much higher flux on average than then other three enzymatic activities, and the flux between these two reactions does not co-vary more than randomly-selected reactions (P = 0.9).

Interestingly, GlyA actually has an evolutionarily-related isozyme in *E. coli*, LtaE (Fig. S7C). GlyA shows a higher preference to catalyze the serine hydroxymethyltransferase reaction, while still weakly catalyzing the L-allo-threonine aldolase reaction. However, LtaE more efficiently catalyzes L-allo-threonine aldolase, and only has a weak serine hydroxymethyltransferase activity (*70*). Both of these isozymes, however, are still able to catalyze the other reactions that maintain a lower flux and are less subject to changing their flux when the nutritional environment changes. In addition, acetylation and succinylation sites have been found on GlyA, and many of these sites are near the serine hydroxymethyltransferase active site (*18, 46, 47*). Thus, while these enzymes have remained generalists through evolution, it seems that mutations have diversified the enzymes to control flux through the two reactions that will require more regulation. Since the enzymes have yet to abolish all other catalytic activities, the question remains as to if the current generalist functions of GlyA and LtaE represent a continuing evolutionary process or if the specificity they have achieved represents the upper limit due to structural needs at the active sites (*70*). However, it is clear that the partial divergence of these catalytic functions agrees with the findings of this work in which high flux and a need to change flux in dynamic nutritional environments tends to evolve enzymes away from being generalists.

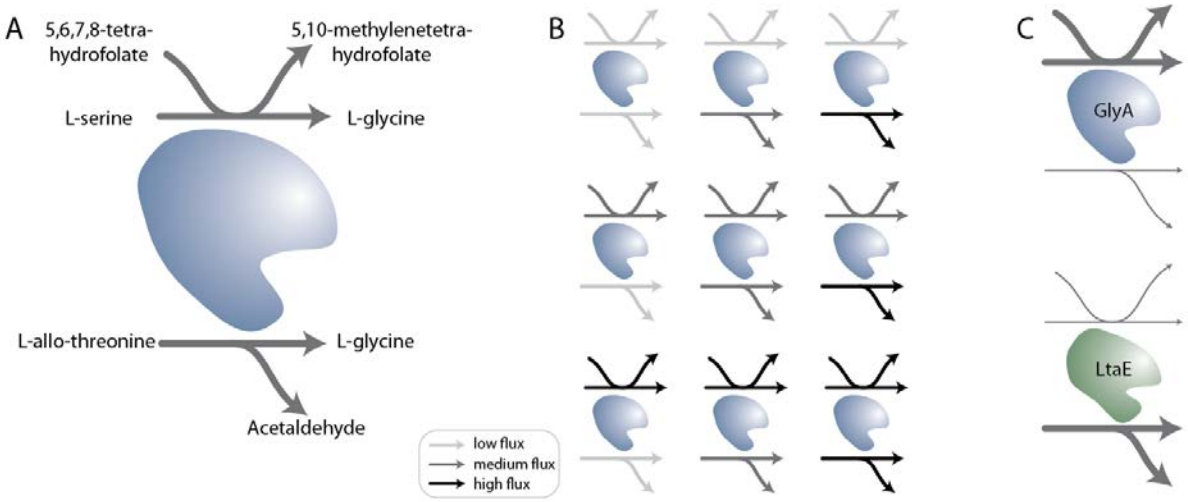

Fig. S7.

Changes in flux in dynamic environments increases the complexity required to regulate flux on a generalist enzyme. (A) GlyA is a generalist enzyme. (B) In any given substrate shift, the flux through its two most prominent reactions can undergo nine qualitative changes (e.g, one reaction can increase in flux and the other can decrease). The complex regulatory mechanisms required to control this in any number of changes in the nutritional microenvironment around the cell may encourage gene duplication and specialization of the different copies. (C) Consistent with this idea, GlyA has an

evolutionarily-related isozyme, LtaE, that catalyzes the L-allo-threonine aldolase activity more efficiently, while GlyA is more efficient at catalyzing the serine hydroxymethyltransferase activity.

Validity of predicted flux changes in this study

Various studies have provided some support for predicted pathway usage from of Flux Balance Analysis (FBA) and related methods (*14, 58*). With the reliability of pathway usage established, the immediate following question is if differential gene expression can be predicted when *E. coli* is grown on two different growth conditions. While a recent published study seems to support this idea (*14*), we provide further support here.

To address the question as to if experimental differential gene expression is consistent with model predictions, a method is presented here which is a slight variant on a previously published MCMC sampling-based method (*37, 45*). The approach presented here extends the published method by associating reaction flux with gene expression change. Herein we call this method Flux Space Shift analysis (FSS). FSS utilizes MCMC sampling of the metabolic solution space to compute the distribution of all possible steady-state fluxes an enzymatic reaction can carry in a cell in a given growth condition.

The aim of constraint-based modeling is to define a space of possible phenotypes by adding a series of known biologically-relevant governing constraints (*7*). Assuming the constraints are accurate, the true steady state flux through the network should be within the *in silico* solution space (Fig. S8A). The range and distribution of reaction fluxes within these solution spaces are dependent on the constraints, such as reaction thermodynamics, metabolite uptake rates, etc. Therefore, the space is condition-specific, i.e., the various dimensions of the space might move when the model is simulated under two different growth conditions. Specifically, as shown in Fig. S8B, the flux may be significantly higher in the second growth condition (reaction 2), significantly lower (reaction 1), or show no significant change between the two growth conditions (reaction 3). Thus, for each reaction, a *P*-value can be assigned to measure the difference between the distributions of fluxes for the two growth conditions.

Since properly constrained reactions do not demonstrate uniform distributions of feasible steady-state fluxes, the distribution is determined by using MCMC sampling of the solution space. To do this, a large number of samples within the solution space are randomly moved until they are well mixed, thereby sampling the entire solution space (see (*42*) for a more detailed description). This sampling process yields a distribution of feasible steady-state fluxes for each reaction. This process is then repeated for the second growth condition. For each reaction, a *P*-value is computed from the distributions of possible fluxes for the reaction under the two conditions. This *P*-value represents the probability of choosing a flux value from a reaction in condition 1 that is also within the distribution for that reaction under condition 2. The *P*-values are then corrected for multiple hypotheses, and the list of reactions that show significantly different fluxes for the two conditions is returned, along with the direction of the change in magnitude. All significantly changed fluxes are then decomposed into a list of genes that help to catalyze the reactions using the gene-protein-reaction associations in the model. Through this, one can obtain lists of genes that are expected to be up-regulated or down-regulated (genes

that are associated both with reactions that increase and other reactions that decrease are removed from the analysis since the kinetics for the different reactions are often not known).

While it may seem unreasonable for such a high level of tuned condition-specific gene expression to occur for a large number of gene expression states, it is hypothesized that if the flux through a network is predicted to increase (or decrease) in magnitude, that the expression level and/or abundance of active enzyme will increase (or decrease) to meet the change in flux. These model-predicted up- and down-regulated genes can be compared to gene and protein expression data for different growth conditions.

This process was conducted for growth conditions used in this study and compared to the new and previously-published experimental data. Specifically, to assess the accuracy for different nutritional shifts, FSS was tested on changes in carbon source and electron acceptor source in M9 minimal media (Fig. S8D-E). In the models, many genes are predicted to change their expression level in each shift, and these changes are growth-condition specific; therefore, one might not expect that there should be significant agreement between the model predictions and data from a short time following the shift, since it would require highly specific regulatory mechanisms. Surprisingly, there was significant agreement for the tested environmental conditions (in which only one nutritional component was changed). There is a significant overlap between model-predicted up- (down-) regulated genes, and the actual up- (down-) regulation seen in the microarrays (Fig. S8D-E) and semi-quantitative proteomic data (unpublished).

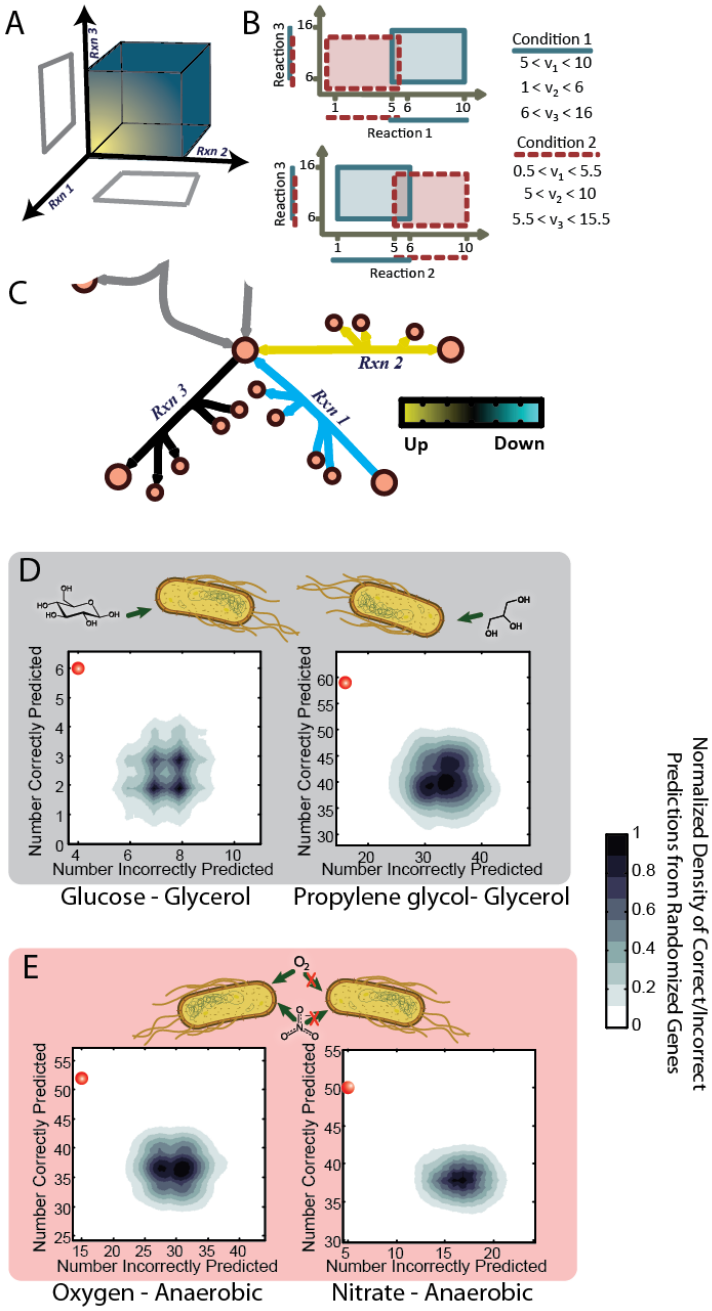

Fig. S8.

Model-predicted differential gene expression of metabolic genes is consistent with measured gene expression changes for various shifts in growth condition. (A) Constraint-based modeling employs governing constraints to define a space of possible phenotypes (represented by feasible steady-state fluxes for each reaction). (B) When growth conditions change (e.g., a change in carbon source, or aerobicity), the space of feasible fluxes can change, as is reflected in the computed flux levels shown here for reactions 1 and 2 (C). These changes are subsequently mapped back to the genes and proteins associated with these reactions to predict differential gene and protein expression. (D)

The prediction of gene expression changes from changes in flux was tested against microarray data for different changes in carbon source, and (E) changes in electron acceptor (i.e., oxygen and nitrate). All of these demonstrated significant agreement between model predictions and the actual direction of expression change (i.e., up/down regulation) from the microarray data (red) when compared to randomized expression change.

A comparison of shifts between structurally similar carbon substrates

When the differences between changes in SERxns and GERxns were assessed for all 15,051 media shifts, a change-difference parameter, termed CD here, was computed to assess if SERxn or GERxn changes dominated. The CD was computed as:

$$CD = \frac{\#\text{ of SERxns changing}}{\#\text{ of SERxns with flux}} - \frac{\#\text{ of GERxns changing}}{\#\text{ of GERxns with flux}}$$

The CD parameter exhibits a bimodal distribution, in which the environmental shifts causing more than 8% of the model reactions to change demonstrated a more dramatic difference between SERxns and GERxns (Fig. S9). Shifts showing a weaker preference to change SERxns (lower CD) tended to have fewer than 8% of the enzyme-catalyzed reactions actually changing.

To test if this is due in part to structural similarity of the limiting metabolites in the media, pairwise Tanimoto coefficients were computed for 154 of the 174 consumable carbon substrates in the *E. coli i*AF1260 model. The remaining 20 metabolites did not have appropriate identifiers to unambiguously identify a metabolite structure. Pairwise Tanimoto coefficients were calculated using the software tool Pipeline Pilot (Accelrys Software Inc.: http://www.accelrys.com/) with the FCFP_6 fingerprint descriptor. When Tanimoto coefficients were compared with the CD values, there was a significant negative correlation (Spearman rank correlation, $P < 1x10^{-307}$). Thus, while almost all shifts show that SERxn flux is more sensitive to nutritional shifts, the shifts that show a weaker preference toward changing SERxn flux are due to the fact that the main carbon substrates are structurally similar and therefore cause few flux changes throughout the network.

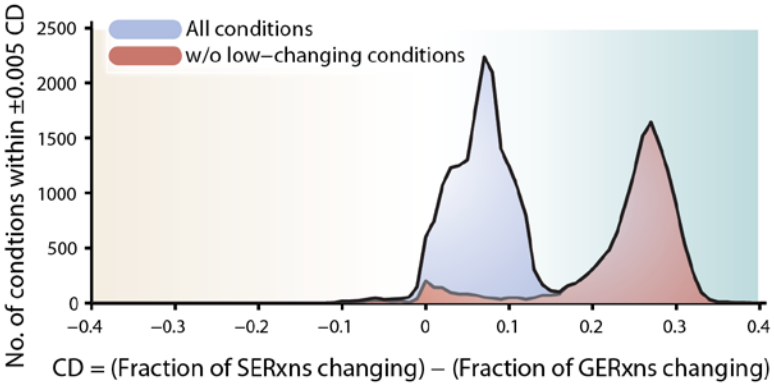

Fig. S9.

Flux through specialist enzymes changes more than flux through generalist enzymes. The difference between the fraction of SERxns that change flux and the fraction of GERxns that change was computed for each media shift. The distribution for all conditions exhibited a bimodal characteristic. When condition shifts that change fewer than 8% of the enzyme-catalyzed reactions in the metabolic network are removed (blue), almost all condition shifts show a dominance of SERxns changing (red). CD = change-difference parameter.

Flux through reactions using the same generalist enzyme is often concerted

Flux through generalist enzymes tends to be less sensitive to environmental changes and subject to less metabolic regulation. However, when the fluxes through GERxns change, do the fluxes catalyzed by the same enzyme co-vary under nutritional environment shifts? Such a scenario would require less complex regulation since the flux for reactions sharing an enzyme would change together for most or all nutritional shifts. To assess this, a cosine similarity metric was used to quantify the amount of co-variance between fluxes through generalist enzymes. For example, for an enzyme e_1 that catalyzes three reactions (r_1, r_2, and r_3), the flux similarity score of e_1 is calculated as an average value of cosine distances of three reaction pairs (Fig. S10A). This metric is applied to 15,051 carbon shift conditions. In each shift condition, the median value of averaged similarity scores of generalist enzymes are evaluated. For the control experiment, cosine similarity scores of the random number of reactions that correspond to the generalist's catalytic degree are used as a control distribution. This metric demonstrates that the individual reactions catalyzed by the same generalist enzyme tend to be regulated in same direction (i.e., the magnitude of their flux either increases or decreases together), more so than seen for randomly chosen pairs of reactions (Fig. S10B). This suggests that generalist enzymes may not require more complex regulation to balance conflicting changes in flux between multiple reactions.

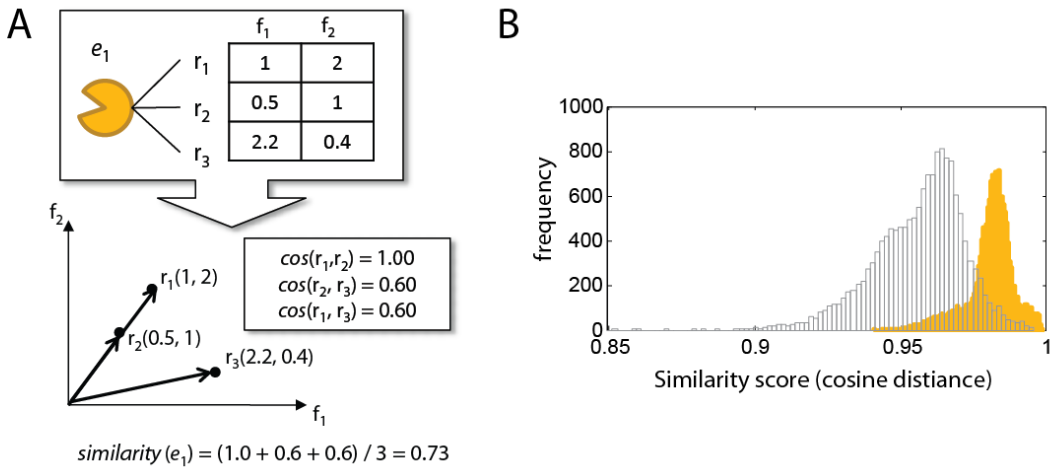

Fig. S10.

Similarity scores for flux changes suggest that generalist enzyme flux often co-varies. (A) An example of calculating the similarity score of a generalist enzyme is provided. (B) Across the 15,051 media shifts, reactions associated with the same generalist enzyme (yellow) tend to change flux in the same direction more frequently than randomly selected reactions (gray), as reflected by a higher similarity score.

Enrichment of metabolic regulation on specialist enzymes

An increased role for metabolic regulation may be necessary for enzymes that are more sensitive to fluctuations in the nutritional environment. Thus, is there evidence for

increased metabolic regulation on specialist enzymes? Metabolic regulation in microbes involves the modulation of enzyme activity under different stimuli, and is often mediated by allosteric binding of small molecules or post-translational enzyme modification (*18, 19*). Thus, to quantify the level of metabolic regulation, the abundance of known allosteric regulatory interactions and enzyme PTMs associated with specialist and generalist enzymes was assessed. Specifically, genes associated with small molecule mediated activation/inhibition were identified from EcoCyc (*35*). Of these, 64 were known to undergo allosteric, uncompetitive, or noncompetitive regulation, and these were enriched in SERxns ($P = 9x10^{-4}$). To complement this, enzymes with PTMs were identified from several proteomic data sets assessing phosphorylation (*48*), acetylation (*46, 47*), and succinylation (*18*). Consistent with small-molecule-mediated regulation, specialist enzymes are more frequently post-translationally modified ($P = 5x10^{-3}$), and enzymes associated with reactions that more frequently change flux in more dynamic environments tend to also have more PTMs associated with them (Fig. 3D).

When compared to the clusters reported in Fig. 3A, regulated enzymes were particularly enriched within the cluster of enzymes associated with reactions that most frequently change flux (dominated by specialist enzymes), while small molecule-mediated regulation and PTMs were depleted from the cluster in which reactions rarely changed flux (dominated by generalist enzymes) (Fig. S11). Only enzymes with known competitive inhibitors did not show a significant preference for any one cluster, which is not surprising. Unlike other modes of regulation, competitive inhibition mostly temporarily slows the reaction rate until the natural substrate concentration increases to counter-balance the inhibition. Thus it is less expected that competitive mechanisms would be employed in nutritional shifts where regulatory mechanisms need to change reaction flux to a new steady state level. Thus, competitive mechanisms may not be selected for on specialist enzymes, since they seem to evolve due to a need to control flux levels under dynamic nutritional conditions.

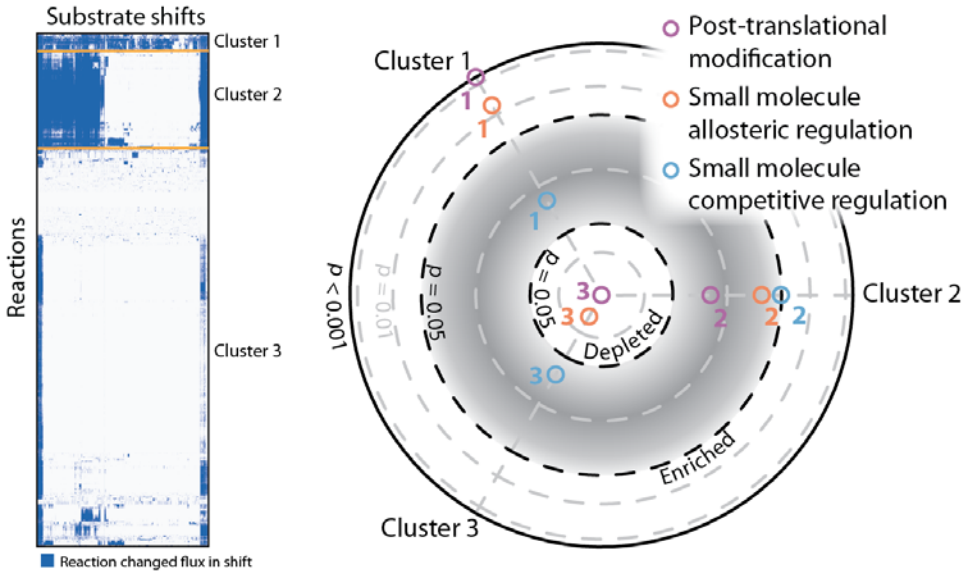

Fig. S11.

Post-translational modifications and metabolite-mediated allosteric regulation are enriched among reactions that are predicted to change more frequently. Specifically, three main clusters of reactions appear in a binary matrix that details which reactions change flux in each of 15,051 media shifts. A modified rose plot shows the enrichment of regulatory mechanisms in these three clusters. In this plot, enrichment *P*-values are plotted in the outer ring, and depletion *P*-values are plotted in the inner ring, with the grey band in the middle representing *P*-values greater than 0.05. In this plot it is clear to see that allosteric regulation and PTMs are enriched in cluster 1, while cluster 3 is depleted in these modes of regulation. It is also clear that small-molecule-mediated competitive regulation is not enriched in any cluster, which is not surprising since this mode of regulation would be less effective in substrate shift conditions.

Network properties associated with enzyme specificity are conserved across micobes

The properties of network context discussed in this work show how enzyme specificity correlates with the holistic functions of the *E. coli* metabolic network. However, if these properties influence selection of enzyme specificity in protein evolution, one may expect the properties to be conserved. Thus, we examined conservation of these properties using genome-scale metabolic models of microbes from the other domains of life: the archeon *Methanosarcina barkeri* (*20*), and the eukaryotes *Saccharomyces cerevisiae* (*21*) and *Chlamydomonas reinhardtii* (*22*). Similar to *E. coli*, the three organisms contain numerous generalist enzymes, ranging from 25%-45% of the known metabolic enzymes in these organisms (Fig. S12). Common growth conditions were simulated for each organism using MCMC sampling to estimate metabolic flux. The median flux value for each reaction was obtained and these were ranked to compare the relative flux levels for specialist and generalist enzymes. In each organism, specialist

enzymes maintained a higher flux on average than generalist enzymes. We further simulated environmental shifts for each organism and found that generalist enzymes were less likely to change flux between growth conditions for most conditions. The only shifts for which differences were not seen were for shifts between highly similar substrates (e.g., shift from maltose to glucose in *S. cerevisiae*). Thus, through the diversification of microbes, higher flux and a need for regulation in varying environments remain as general features of specialist enzymes.

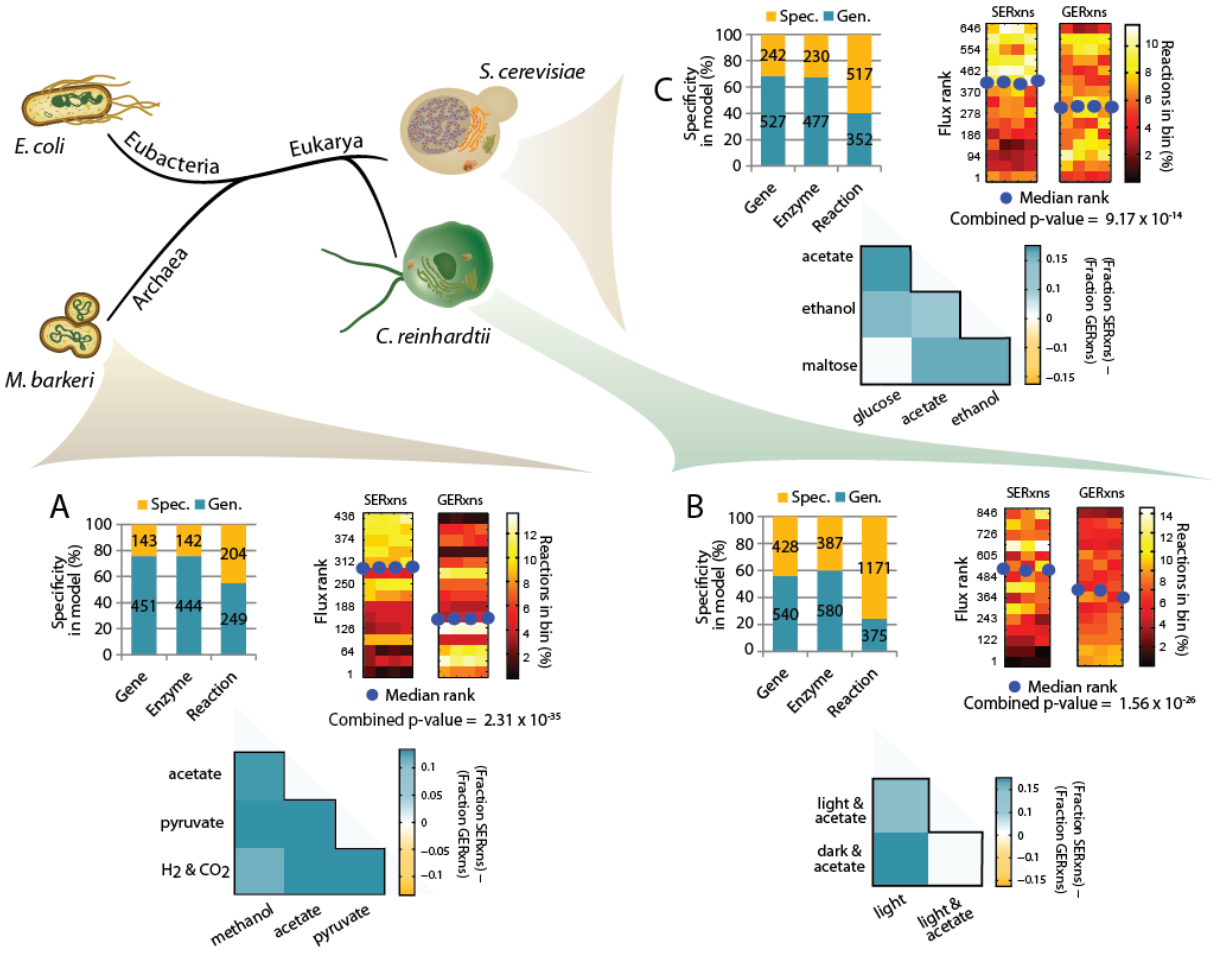

Fig. S12.

Enzyme-specificity characteristics hold for microbes in all domains of life, as shown here for (A) *M. barkeri*, (B) *C. reinhardtii*, and (C) *S. cerevisiae*. Generalist enzymes are abundant, as shown by the gene, enzyme, and reaction (G/E/R) composition for each species. In addition, SERxns maintain higher magnitudes of flux than do GERxns. The shift plots also show that a higher fraction of SERxns change flux levels in shifts between two different growth conditions.

Correlation of flux vs. gene expression, protein concentration, and enzyme efficiency

Metabolic flux is determined by a number of factors (e.g., transcription and translation rates, metabolic regulation, kinetic properties of enzymes, etc.). Therefore, it should not necessarily correlate well with any one factor on its own. However, we assessed how flux values from our model predictions correlate with measured data to gain insight into how each of the various properties may influence flux, and also to provide some confidence in our simulated flux rankings. Specifically, we evaluated model predicted flux values with respect to transcript levels, enzyme concentration and enzyme kinetic parameters. The data we used included 42 previously published microarrays (*32, 38*), quantitative protein abundance data (*71*), and k_{cat} values obtained from BRENDA. The median values of gene expression and protein abundance measurements for each mRNA/protein were compared to the rank-ordered flux, as were the k_{cat} values. For each of these, we found a correlation between the parameters and flux. Consistent with the idea that k_{cat} values and mRNA/protein levels contribute directly to flux, we saw the most significant correlation of these parameters with the predicted flux (Fig. S13A-C, Spearman correlation coefficient: 0.383, 0.411, and 0.381 respectively).

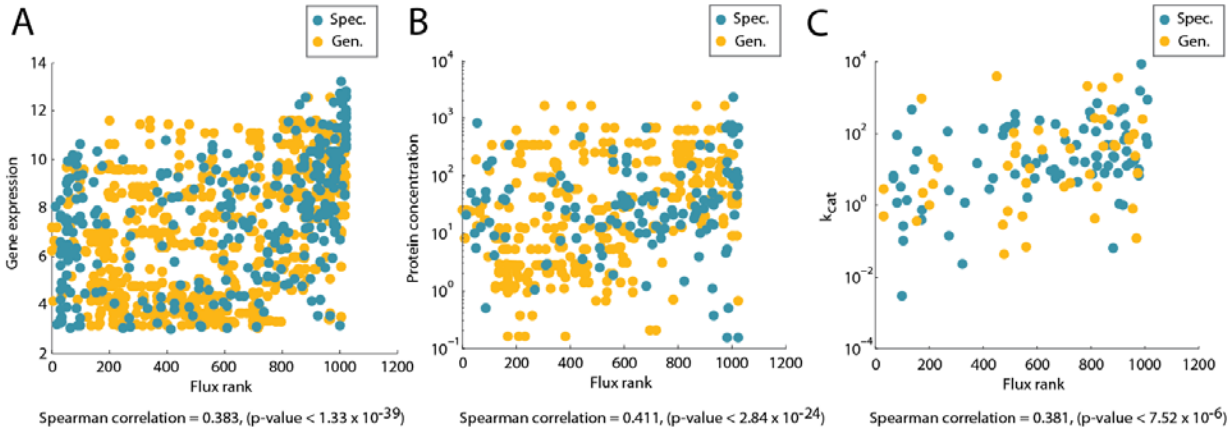

Fig. S13.

Comparisons were made between model-predicted flux and various types of experimental data, such as (A) microarray data, (B) protein concentration, and (C) $k_{cat.}$ values.

Substrate/product-mediated regulation of enzymes

Here we provide a more detailed assessment of cases when enzymes are regulated by their own substrates. It has been previously suggested that this might be important to prevent an active site from catalyzing undesirable promiscuous reactions (*2*). The only case of regulation that would seem to facilitate this is that of allosteric activation. However, in *E. coli* there are few cases of substrate-mediated allosteric activation (Fig. S14). These include the specialist enzymes GltA (b0720), AnsA (b1767), and RffE (b3786), which are regulated by their substrates Acetyl-CoA, L-Asparagine, and UDP-N-acetyl-D-glucosamine, respectively. The only known generalist enzyme that is activated by its substrate, 'pyruvate formate lyase I (PflB)' is positively regulated by one of its substrates, pyruvate. Interestingly, this regulation is indirect. Pyruvate binds to PflA, which then activates PflB's anaerobic pyruvate cleavage activity. Further discussion of this enzyme can be found below.

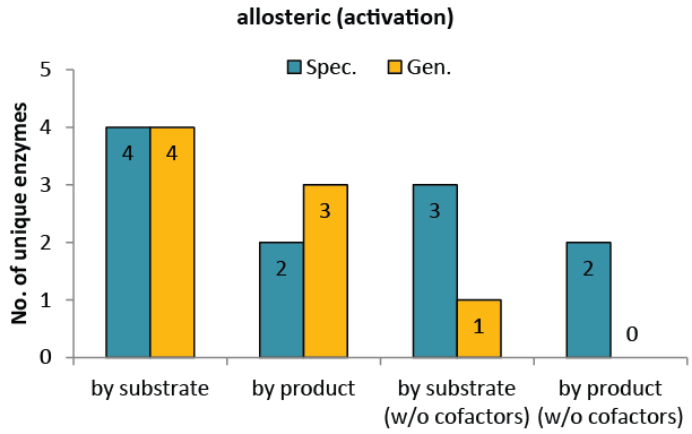

Fig. S14.

The Number of enzymes that are regulated by their own substrates or products in allosteric activation. Spec.: Specialist, Gen: Generalist.

Isoenzymes in generalist enzymes

Since gene duplication is a common component in models of enzyme evolution (*3*), is there evidence of greater catalytic redundancy among generalist enzymes? Here we provide an initial analysis of isoenzymes and their relation to enzyme specificity. In the *E. coli* model, 1,045 reactions are catalyzed a single enzyme, while 270 reactions have known isoenzymes (Fig. S15A). In *E. coli* we found that reactions catalyzed by multiple isoenzymes are more likely to be catalyzed by generalist enzymes than expected by chance (binomial $P = 6.4 \times 10^{-8}$). We looked into the flux magnitudes of reactions catalyzed by multiple generalist enzymes and found that they exhibited a bimodal distribution of flux levels (Fig. S15B). Furthermore, many of the GERxns that maintain higher flux levels could be catalyzed by more than one enzyme. These higher flux generalist enzymes include important biosynthetic pathways that need to synthesize components of the cell envelope and membrane (hypergeometric $P=4\times10^{-4}$ and $P=3\times10^{-12}$, respectively), which are needed in bulk quantities as cells grow. In addition, enzymes needed to recycle and modify nucleotides were also enriched (hypergeometric $P=7\times10^{-8}$). Both of these processes benefit from being able to do similar reaction chemistries on a variety of substrates. In addition, the isoenzymes have overlapping substrate specificities, such as seen for FabA and FabZ, which are both beta-hydroxyacyl-ACP dehydratases required for cell envelope synthesis. While their specificities significantly overlap, FabA is more highly active on intermediate length β-hydroxyacyl-ACPs, while FabZ has a preference for short and long chain β-hydroxyacyl-ACPs. FabA can also use short and long chain β-hydroxyacyl-ACPs, but only the saturated forms (*72*). It is anticipated that having generalist isoenzymes such as these provide a fitness advantage since they can act on a broad range of substrates to synthesize a diverse range of structural components for the cell, and the presence of different isozymes provides additional diversity to molecules that are synthesized.

Among the high flux GERxns that can be catalyzed by isoenzymes, we found a few additional proteins of interest. For example, pyruvate formate lyase (PflB) is an important enzyme in glucose metabolism under anaerobic conditions, and a homologous isozyme has been found (TdcE). PflB is known to catalyze the non-oxidative cleavage of pyruvate to make formate and acetyl-CoA (*73*). Under anaerobic conditions, it is the primary enzyme that produces acetyl-CoA from pyruvate, since pyruvate dehydrogenase is inactive under anaerobic conditions. PflA is an activase that is necessary for PflB activity, and pyruvate allosterically activates PflA (*74*). It is believed that pyruvate cleavage is the primary function of PlfB, and this enzyme maintains a high flux in anaerobic conditions in our model (97^{th} percentile), which significantly decreases when grown aerobically or with nitrate. In addition to the cleavage of pyruvate, PflB has a secondary metabolic function in which it catalyzes the cleavage of 2-ketobutyrate to make propionyl-CoA and formate, which is one step in L-threonine metabolism (*75*).

However, a homolog to pflB has been discovered. The TdcE enzyme has 82% homology in protein sequence to PflB. As seen with PflB, TdcE is also expressed anaerobically. However, initial studies suggested that its pyruvate formate lyase activity is lower than that seen by PflB, while it has a greater preference for the cleavage of 2-ketobutyrate (*75*). It is believed that the physiological role of this enzyme is the anaerobic degradation of L-threonine. However, the flux for this enzyme ranks in the 18^{th} percentile under anaerobic conditions.

A few questions arise when one assesses the generalist PflB and TdcE isozymes. Both of the enzyme activities seem to have important physiological roles under anaerobic conditions. While they have both evolved greater specificity for their substrate of interest, they both can replace each other's activity to some extent (*73*, *75*). Do they maintain their generalist properties because they cannot evolve greater specificity? Or are they in the process of diverging, but are doing so more slowly since the 2-ketobutyrate cleavage activity is under a weaker selective pressure since the reaction is lower flux? Similar hypotheses exist for other generalist isozyme pairs (e.g., GlyA and LtaE). Further detailed study and experimentation (e.g., using directed evolution) is required to test these hypotheses.

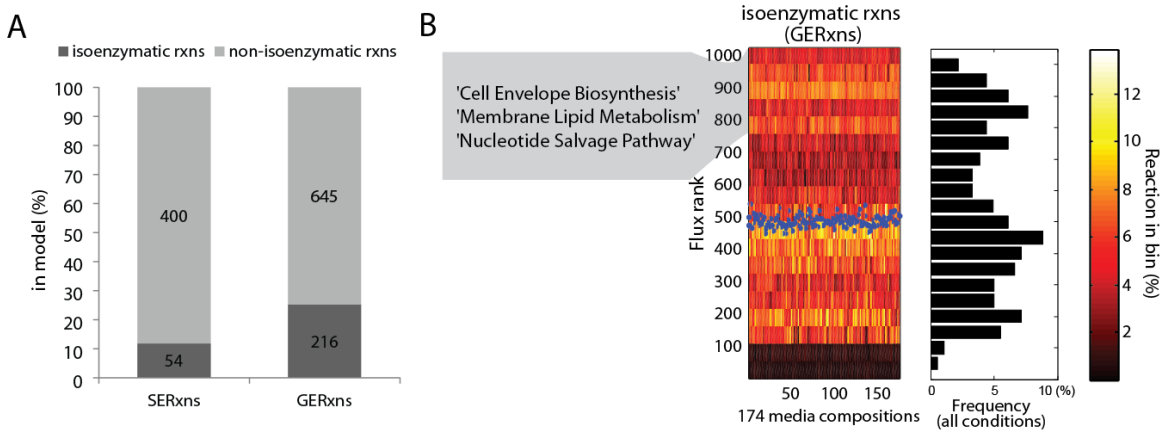

Fig. S15.

(A) Isozymes contribute to a sizable number of reactions and are more prevalent among GERxns in *E. coli* metabolism. (B) The rankings of isozyme-related GERxn flux showed a bimodal distribution of flux values, including an enrichment high-flux GERxns associated with several metabolic subsystems.

Assessment of nitrate uptake rate

The range of feasible nitrate uptake rates was approximated from the *i*AF1260 model of *E. coli* by constraining the model growth rate and glucose uptake rate to the measured values (0.34±0.007 h^{-1}, and 4.05±0.5 mmol gDW^{-1} h^{-1}, respectively). Once the model is constrained with the measured growth rate and the glucose uptake rate, the feasible nitrate uptake range was determined by Flux Variability Analysis (FVA) (*76*). Within the given feasible range, we used uptake rates in the range of 25th, 50th, and 75th percentile, and compared the qualitative difference between specialist and generalist enzymes in the nitrate-anaerobic shift. However, results presented in Fig. 2A were qualitatively robust with variations in nitrate uptake rate within the computed range (Fig. S16).

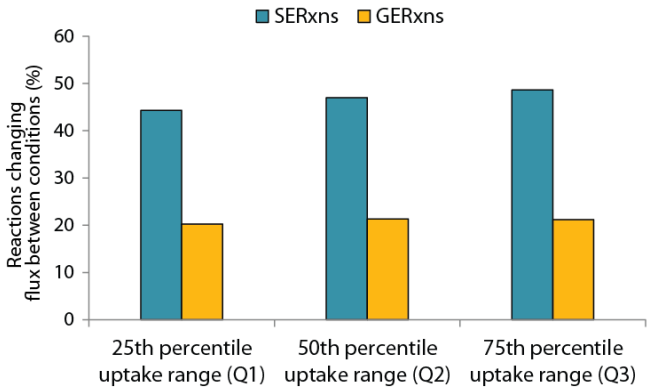

Fig. S16.

Qualitative properties of specialist and generalist enzymes were robust to variations in estimated values of nitrate metabolism. Due to the difficulty of measuring the rate of nitrate metabolism, the uptake rate was computed from the model by using experimentally-measured glucose uptake rates and growth rates. This calculation provided a range of feasible nitrate metabolic rates. Across this range, SERxns change flux more frequently. Thus, this result is qualitatively robust, within the range of feasible nitrate uptake rates (between approximately 15-25 mmol gDW^{-1} h^{-1}).

Details of the classification of specialist and generalist enzymes

For this study we classified 1,147 enzymes from the *E. coli* genome-scale model (*i*AF1260) (*5*) following the detailed process shown in Fig. S17. First, we identified 1,081 proteins with enzymatic activity, as reported in the EcoCyc Database (*35*). The 66 proteins that were removed did not have experimentally-validated catalytic activities. These included non-catalytic members of enzyme complexes (e.g., the electron transferring protein flavodoxin (b0684)) and predicted enzymes (e.g., predicted carbamate kinase (b0521)). Among these 1,081 enzymes, 677 enzymes were classified as specialists, which only catalyze one specific chemical reaction on a single set of substrates. The 404 enzymes catalyzing more than one reaction were classified as generalist enzymes (Fig. S17A). These specialists and generalist enzymes are encoded by

720 genes and 473 genes, respectively. We anticipate that a small fraction of enzymes may be misclassified due to incomplete study of some enzymes or incorrect interpretation of biochemical studies. However, these concerns should not substantially affect the conclusions in this work, since variations on the categorization yield qualitatively similar results (see Fig. S3).

Generalist enzymes were further curated to verify that they indeed were associated with reactions on multiple substrates. These were also partitioned into subclasses according to if they were multifunctional or enzymes that exhibit substrate promiscuity.

Following enzyme classification, specialist enzyme reactions (SERxns) and generalist enzyme reactions (GERxns) were identified and classified (Fig. S17B). The reaction lists were also filtered to remove reactions with ambiguous classification. Specifically, 67 reactions associated with both of specialist and generalist isozymes were removed from the further analysis. We also note that transport reactions were removed as they usually do not represent canonical metabolic catalysis beyond, for example, ATP hydrolysis for in ABC transporters. However, the presence of transporters did not qualitatively change the results in this work (Fig. S3).

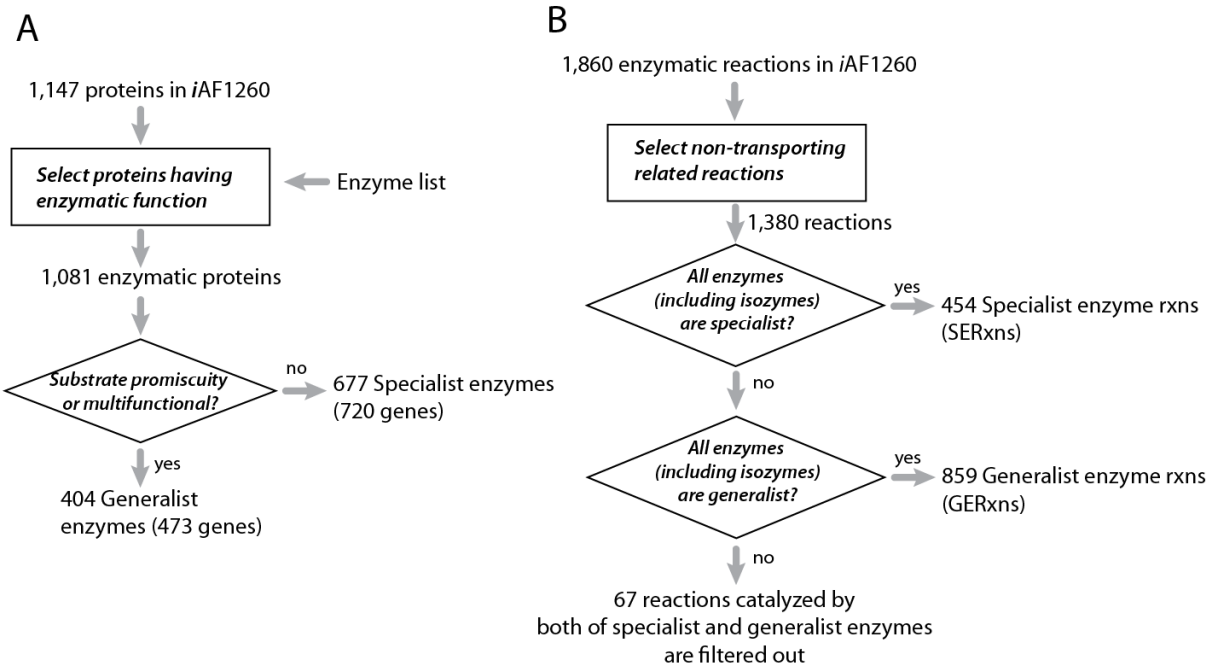

Fig. S17.

The specialist/generalist classification process employed in this study. (A) Enzymes and and their associated genes were classified based on the number of enzymatic reactions they catalyzed in the *i*AF1260 metabolic network reconstruction. (B) Reactions associated with specialist and generalist enzymes were similarly classified.

Justification of combining promiscuous and multifunctional enzymes into the "generalist" class

The evolution of enzyme specificity and gene duplication has been a topic of great interest for decades (*3*, *77*, *78*). In 1976, Roy A. Jensen published an important treatise on enzyme evolution (*1*). This seminal work included insights and theories that have since deeply influenced decades of research into the evolution of metabolic pathways and enzyme specificity. A key tenet of Jensen's work was that ancestral enzymes demonstrated broad specificity. Specifically, he stated that

"Pristine life must have been restricted to limited genetic information encoding a small number of proteins. . . . The most attractive mechanism for acquisition of additional genetic information is that of gene duplication in tandem. Divergence of the new gene copies via mutational modifications that altered enzymatic reactivities presumably allowed the expansion of metabolic capabilities and the evolution of new biochemical pathways.... Subsequent elaboration of additional enzyme proteins following gene amplification would allow the luxury of increased specialization and the improved metabolic efficiency that is thus permitted."

Jensen then raised an important point briefly: "Although modern enzymes can be extraordinarily specific, substrate specificities are perhaps broader than is generally appreciated." Despite this observation, biochemistry textbooks still define enzymes as "specific catalysts".

While Jensen referred primarily to cases in which enzymes exhibit substrate promiscuity, the combination of these enzymes with multifunctional enzymes is relevant because the selection pressures that influence the evolution of broad-specificity enzymes to toward high specificity will also influence the many multifunctional proteins, causing them to duplicate and evolve to having only one active site with any notable activity.

To demonstrate that these two groups are likely subject to the same selective pressures, we classified each generalist enzyme as promiscuous or multifunctional. In this classification, among the non-transporter generalist enzymes, we found that 89% are generalist enzymes exhibiting substrate promiscuity (such as those envisioned by Jensen), and 11% are multifunctional (Fig. S1A). Furthermore, 3% were multifunctional with one or more promiscuous active sites. Lastly, 2% were bifunctional enzymes in which the different catalytic activities were used exclusively to catalyze two reactions in series on one substrate (e.g., in substrate channeling).

To show that the selective pressures are likely relevant to multifunctional enzymes and enzymes showing substrate promiscuity, the central tests of this work were repeated on both of these groups of enzymes. Through this analysis we found that flux for both classes of enzymes was lower (Fig. S4C) and less sensitive to nutritional changes (Fig. S3B). Thus it seems that the selective pressures highlighted in this work are relevant to enzymes with substrate promiscuity and also to multifunctional enzymes.

Table S1.

Experimentally measured growth phenotypes with and without nitrate.

	Growth rate (h^{-1})	Glucose uptake rate (mmol gDW^{-1} h^{-1})	Acetate secretion rate (mmol gDW^{-1} h^{-1})	Formate secretion rate (mmol gDW^{-1} h^{-1})
Anaerobic with nitrate	0.34 ± 0.007	4.05 ± 0.504	6.99 ± 0.33	5.22 ± 0.15
Anaerobic without nitrate	0.29 ± 0.007	5.79 ± 0.44	4.64 ± 0.31	14.15 ± 0.53

Table S2.

Number of GERxns in each classification.

Number of reactions	Glycerol vs. Glucose	Propylene glycol vs. glycerol	Anaerobic vs. Aerobic	Anaerobic vs. anaerobic + nitrate
GERxns	830	828	830	837
Conditional GERxns (type I)	700	708	700	707
GERxns without Transporters	660	663	660	665

Database S1.

The excel file includes full list of specialist and generalist enzymes and their associated reactions. (A) *E. coli* enzyme table : column1 - enzyme abbreviation, column2 - enzyme full name, column3 - associated genes, column4 - associated reactions, column5 - enzyme class, column6 - enzyme subclass, column7 - associated EC number, column8 - comments. (B) *E. coli* reaction table: column1 - reaction abbreviation, column2 - reaction full name, column3 - reaction formula, column4 - subsystems, column5 - associated genes, column6 - reaction class. (C), (D), and (F) depict column1 - list of specialist gene, column2 - list of generalist genes, column3 - list of SERxns, column4 - list of GERxns of *M. barkeri*, *C. reinhardtii*, and *S. cerevisiae*, respectively.

References and Notes

1. R. A. Jensen, Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409 (1976). doi:10.1146/annurev.mi.30.100176.002205 Medline

2. O. Khersonsky, D. S. Tawfik, Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471 (2010). doi:10.1146/annurev-biochem-030409-143718 Medline

3. H. Innan, F. Kondrashov, The evolution of gene duplications: Classifying and distinguishing between models. *Nat. Rev. Genet.* **11**, 97 (2010). doi:10.1038/nrg2689 Medline

4. O. Khersonsky, S. Malitsky, I. Rogachev, D. S. Tawfik, Role of chemistry versus substrate binding in recruiting promiscuous enzyme functions. *Biochemistry* **50**, 2683 (2011). doi:10.1021/bi101763c Medline

5. A. M. Feist *et al*., A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3**, 121 (2007). doi:10.1038/msb4100155 Medline

6. A. Aharoni *et al*., The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73 (2005). Medline

7. N. E. Lewis, H. Nagarajan, B. O. Palsson, Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10**, 291 (2012). Medline

8. E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, A. L. Barabási, Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427**, 839 (2004). doi:10.1038/nature02289 Medline

9. Materials and methods are available as supporting material on *Science* Online.

10. A. Wagner, Energy costs constrain the evolution of gene expression. *J. Exp. Zool. B Mol. Dev. Evol.* **308B**, 322 (2007). doi:10.1002/jez.b.21152 Medline

11. T. Baba *et al.*, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

12. B. Papp, C. Pál, L. D. Hurst, Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661 (2004). doi:10.1038/nature02636 Medline

13. D. Deutscher, I. Meilijson, M. Kupiec, E. Ruppin, Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat. Genet.* **38**, 993 (2006). doi:10.1038/ng1856 Medline

14. S. Bordel, R. Agren, J. Nielsen, Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLOS Comput. Biol.* **6**, e1000859 (2010). doi:10.1371/journal.pcbi.1000859 Medline

15. R. Schuetz, L. Kuepfer, U. Sauer, Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.* **3**, 119 (2007). doi:10.1038/msb4100162 Medline

16. E. Gur, D. Biran, E. Z. Ron, Regulated proteolysis in Gram-negative bacteria—how and when? *Nat. Rev. Microbiol.* **9**, 839 (2011).

17. N. E. Lewis, B. K. Cho, E. M. Knight, B. O. Palsson, Gene expression profiling and the use of genome-scale in silico models of *Escherichia coli* for analysis: Providing context for content. *J. Bacteriol.* **191**, 3437 (2009). doi:10.1128/JB.00034-09 Medline

18. Z. Zhang *et al*., Identification of lysine succinylation as a new post-translational modification. *Nat. Chem. Biol.* **7**, 58 (2011). doi:10.1038/nchembio.495 Medline

19. L. Gerosa, U. Sauer, Regulation and control of metabolic fluxes in microbes. *Curr. Opin. Biotechnol.* **22**, 566 (2011). doi:10.1016/j.copbio.2011.04.016 Medline

20. A. M. Feist, J. C. Scholten, B. O. Palsson, F. J. Brockman, T. Ideker, Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol. Syst. Biol.* **2**, 2006 0004 (2006).

21. M. L. Mo, B. O. Palsson, M. J. Herrgård, Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* **3**, 37 (2009). doi:10.1186/1752-0509-3-37 Medline

22. R. L. Chang *et al*., Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* **7**, 518 (2011). doi:10.1038/msb.2011.52 Medline

23. A. Bar-Even *et al*., The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402 (2011). doi:10.1021/bi2002289 Medline

24. M. Morar, G. D. Wright, The genomic enzymology of antibiotic resistance. *Annu. Rev. Genet.* **44**, 25 (2010). doi:10.1146/annurev-genet-102209-163517 Medline

25. E. Almaas, Z. N. Oltvai, A. L. Barabási, The activity reaction core and plasticity of metabolic networks. *PLOS Comput. Biol.* **1**, e68 (2005). doi:10.1371/journal.pcbi.0010068 Medline

26. S. D. Copley, Toward a systems biology perspective on enzyme evolution. *J. Biol. Chem.* **287**, 3 (2012). doi:10.1074/jbc.R111.254714 Medline

27. B. Papp, R. A. Notebaart, C. Pál, Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.* **12**, 591 (2011). doi:10.1038/nrg3033 Medline

28. H. Nam, T. M. Conrad, N. E. Lewis, The role of cellular objectives and selective pressures in metabolic pathway evolution. *Curr. Opin. Biotechnol.* **22**, 595 (2011). doi:10.1016/j.copbio.2011.03.006 Medline

29. P. Carbonell, G. Lecointre, J. L. Faulon, Origins of specificity and promiscuity in metabolic networks. *J. Biol. Chem.* **286**, 43994 (2011). doi:10.1074/jbc.M111.274050 Medline

30. D. H. Lee, B. O. Palsson, Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-propanediol. *Appl. Environ. Microbiol.* **76**, 4158 (2010). doi:10.1128/AEM.00373-10 Medline

31. M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, B. O. Palsson, Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92 (2004). doi:10.1038/nature02456 Medline

32. S. S. Fong, A. R. Joyce, B. O. Palsson, Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* **15**, 1365 (2005). doi:10.1101/gr.3832305 Medline

33. J. D. Orth, I. Thiele, B. O. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245 (2010). doi:10.1038/nbt.1614 Medline

34. T. M. Conrad, N. E. Lewis, B. O. Palsson, Microbial laboratory evolution in the era of genome-scale science. *Mol. Syst. Biol.* **7**, 509 (2011). doi:10.1038/msb.2011.42 Medline

35. I. M. Keseler *et al*., EcoCyc: A comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.* **39** (Database issue), D583 (2011). doi:10.1093/nar/gkq1143 Medline

36. J. Schellenberger, B. O. Palsson, Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.* **284**, 5457 (2009). doi:10.1074/jbc.R800048200 Medline

37. A. Bordbar, N. E. Lewis, J. Schellenberger, B. O. Palsson, N. Jamshidi, Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol. Syst. Biol.* **6**, 422 (2010). doi:10.1038/msb.2010.68 Medline

38. N. E. Lewis *et al*., Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat. Biotechnol.* **28**, 1279 (2010). doi:10.1038/nbt.1711 Medline

39. J. Schellenberger *et al*., Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nature Protocols* **6**, 1290 (2011).

40. S. Schuster, T. Pfeiffer, D. A. Fell, Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.* **252**, 497 (2008). doi:10.1016/j.jtbi.2007.12.008 Medline

41. A. M. Feist, B. O. Palsson, The biomass objective function. *Curr. Opin. Microbiol.* **13**, 344 (2010). doi:10.1016/j.mib.2010.03.003 Medline

42. J. Schellenberger, N. E. Lewis, B. O. Palsson, Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.* **100**, 544 (2011). doi:10.1016/j.bpj.2010.12.3707 Medline

43. D. A. Beard, S. D. Liang, H. Qian, Energy balance for analysis of complex metabolic networks. *Biophys. J.* **83**, 79 (2002). doi:10.1016/S0006-3495(02)75150-3 Medline

44. P. Daran-Lapujade *et al*., Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae. A chemostat culture study. *J. Biol. Chem.* **279**, 9125 (2004). doi:10.1074/jbc.M309578200 Medline

45. A. Bar-Even, E. Noor, N. E. Lewis, R. Milo, Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8889 (2010). doi:10.1073/pnas.0907176107 Medline

46. B. J. Yu, J. A. Kim, J. H. Moon, S. E. Ryu, J. G. Pan, The diversity of lysine-acetylated proteins in *Escherichia coli*. *J. Microbiol. Biotechnol.* **18**, 1529 (2008). Medline

47. J. Zhang *et al*., Lysine acetylation is a highly abundant and evolutionarily conserved modification in *Escherichia coli*. *Mol. Cell. Proteomics* **8**, 215 (2009). doi:10.1074/mcp.M800187-MCP200 Medline

48. B. Macek *et al*., Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol. Cell. Proteomics* **7**, 299 (2008). doi:10.1074/mcp.M700311-MCP200 Medline

49. O. Khersonsky, D. S. Tawfik, Enzyme promiscuity: A Mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471 (2010).

50. M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38** (Database issue), D355 (2010). doi:10.1093/nar/gkp896 Medline

51. J. D. Orth *et al*., A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* **7**, 535 (2011). doi:10.1038/msb.2011.65 Medline

52. R. A. Majewski, M. M. Domach, Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnol. Bioeng.* **35**, 732 (1990). doi:10.1002/bit.260350711 Medline

53. A. Varma, B. O. Palsson, Metabolic capabilities of *Escherichia coli*: I. synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.* **165**, 477 (1993). doi:10.1006/jtbi.1993.1202 Medline

54. J. Pramanik, J. D. Keasling, Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* **56**, 398 (1997). doi:10.1002/(SICI)1097-0290(19971120)56:4<398::AID-BIT6>3.0.CO;2-J Medline

55. J. S. Edwards, B. O. Palsson, The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5528 (2000). doi:10.1073/pnas.97.10.5528 Medline

56. J. L. Reed, B. O. Palsson, Thirteen years of building constraint-based in silico models of *Escherichia coli*. *J. Bacteriol.* **185**, 2692 (2003). doi:10.1128/JB.185.9.2692-2699.2003 Medline

57. A. M. Feist, B. O. Palsson, The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* **26**, 659 (2008). doi:10.1038/nbt1401 Medline

58. N. E. Lewis *et al*., Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390 (2010). doi:10.1038/msb.2010.47 Medline

59. C. Shou *et al*., Measuring the evolutionary rewiring of biological networks. *PLOS Comput. Biol.* **7**, e1001050 (2011). doi:10.1371/journal.pcbi.1001050 Medline

60. R. U. Ibarra, J. S. Edwards, B. O. Palsson, *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**, 186 (2002). doi:10.1038/nature01149 Medline

61. G. T. Cocks, T. Aguilar, E. C. Lin, Evolution of L-1, 2-propanediol catabolism in *Escherichia coli* by recruitment of enzymes for L-fucose and L-lactate metabolism. *J. Bacteriol.* **118**, 83 (1974). Medline

62. I. Thiele, N. Jamshidi, R. M. Fleming, B. O. Palsson, Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: A knowledge base, its mathematical formulation, and its functional characterization. *PLOS Comput. Biol.* **5**, e1000312 (2009). doi:10.1371/journal.pcbi.1000312 Medline

63. Y. Zhang *et al*., Three-dimensional structural view of the central metabolic network of Thermotoga maritima. *Science* **325**, 1544 (2009). doi:10.1126/science.1174671 Medline

64. R. L. Chang, L. Xie, L. Xie, P. E. Bourne, B. Ø. Palsson, Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLOS Comput. Biol.* **6**, e1000938 (2010). doi:10.1371/journal.pcbi.1000938 Medline

65. B. D. Bennett *et al*., Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593 (2009). doi:10.1038/nchembio.186 Medline

66. A. Mitchell *et al*., Adaptive prediction of environmental changes by microorganisms. *Nature* **460**, 220 (2009). doi:10.1038/nature08112 Medline

67. P. Stover, V. Schirch, Serine hydroxymethyltransferase catalyzes the hydrolysis of 5,10-methenyltetrahydrofolate to 5-formyltetrahydrofolate. *J. Biol. Chem.* **265**, 14227 (1990). Medline

68. V. Schirch, S. Hopkins, E. Villar, S. Angelaccio, Serine hydroxymethyltransferase from *Escherichia coli*: Purification and properties. *J. Bacteriol.* **163**, 1 (1985). Medline

69. S. Hopkins, V. Schirch, Properties of a serine hydroxymethyltransferase in which an active site histidine has been changed to an asparagine by site-directed mutagenesis. *J. Biol. Chem.* **261**, 3363 (1986). Medline

70. R. Contestabile *et al*., L-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase. A subgroup of strictly related enzymes specialized for different functions. *Eur. J. Biochem.* **268**, 6508 (2001). doi:10.1046/j.0014-2956.2001.02606.x Medline

71. Y. Taniguchi *et al*., Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533 (2010). doi:10.1126/science.1188308 Medline

72. R. J. Heath, C. O. Rock, Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. *J. Biol. Chem.* **271**, 27795 (1996). doi:10.1074/jbc.271.44.27795 Medline

73. G. Sawers, A. Böck, Anaerobic regulation of pyruvate formate-lyase from *Escherichia coli* K-12. *J. Bacteriol.* **170**, 5330 (1988). Medline

74. J. Knappe, G. Sawers, A radical-chemical route to acetyl-CoA: The anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. *FEMS Microbiol. Rev.* **6**, 383 (1990). doi:10.1016/S0168-6445(05)80006-3 Medline

75. C. Hesslinger, S. A. Fairhurst, G. Sawers, Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in *Escherichia coli* that degrades L-threonine to propionate. *Mol. Microbiol.* **27**, 477 (1998). doi:10.1046/j.1365-2958.1998.00696.x Medline

76. R. Mahadevan, C. H. Schilling, The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264 (2003). doi:10.1016/j.ymben.2003.09.002 Medline

77. S. Ohno, *Evolution by gene duplication*. (Springer-Verlag, Berlin, New York, 1970), pp. xv, 160 p.

78. M. Soskine, D. S. Tawfik, Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572 (2010). doi:10.1038/nrg2808 Medline