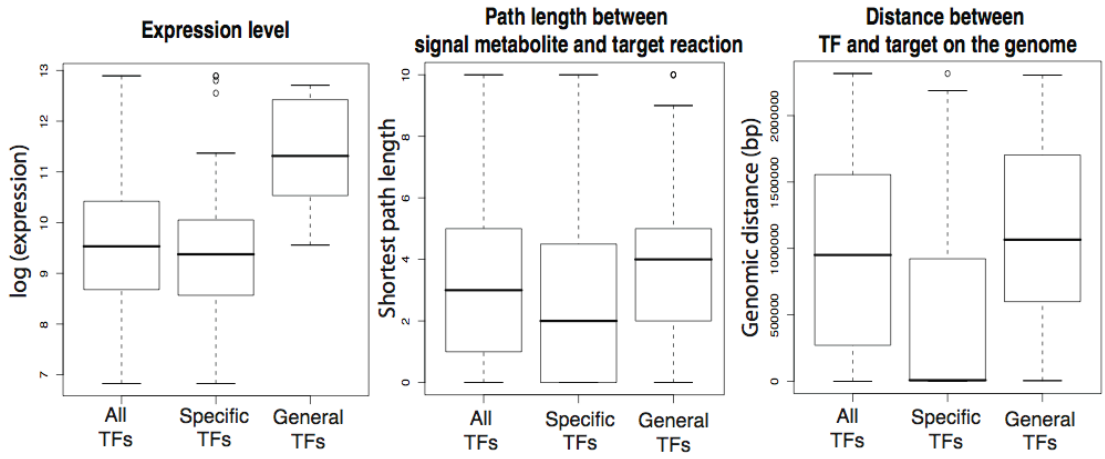
Supplementary Figure 1

Differences in various genomic properties of specific and general TFs



This figure shows three genomic / functional properties in which specific and general TFs differ significantly.

(Left panel) General TFs have higher expression levels than specific TFs; a TFs expression level was defined using a cut-off value across 225 Affymetrix arrays. This cutoff value was defined as follows:

For any given gene, if M is the highest expression value across all arrays and C is the sum of the third quartile and 1.5 x IQR of the expression vector, then the expression measure is C if M>C or M otherwise.

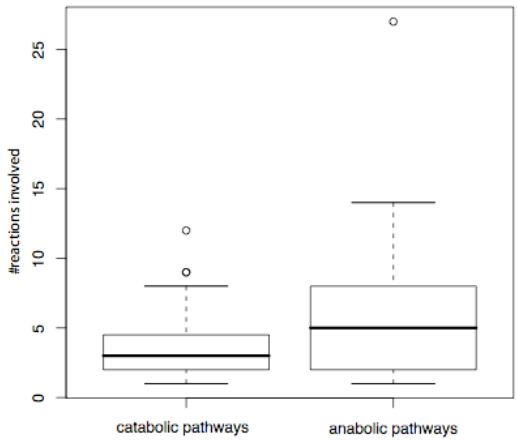
The use of C ensures that we do not assign an outlier, that is more an artifact than real, as the expression measure. Note that these observations are generally robust to the use of M as the expression measure, irrespective of whether it is much larger than C or not. A high expression value is more reasonable than the average or median of expression values as a TF may be expressed under only one or two conditions and under such circumstances, the median or the average would underestimate the expression value.

(Centre panel) Metabolites that feed back to general TFs are separated from target enzymes in the metabolic network by distances greater than specific TFs.

(Right panel) Finally, specific TFs are generally located very close to their target genes on the chromosome. The position of a TF on the chromosome was defined by its end coordinate where coupled transcription and translation would end. The chromosomal location of the target was taken as its start point since the TF binding site would be located closer to this position.

Supplementary figure 2

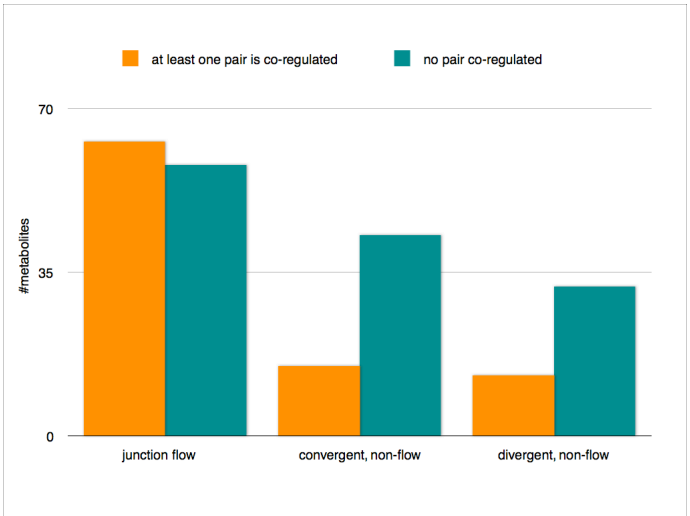
Anabolic pathways are longer than catabolic pathways



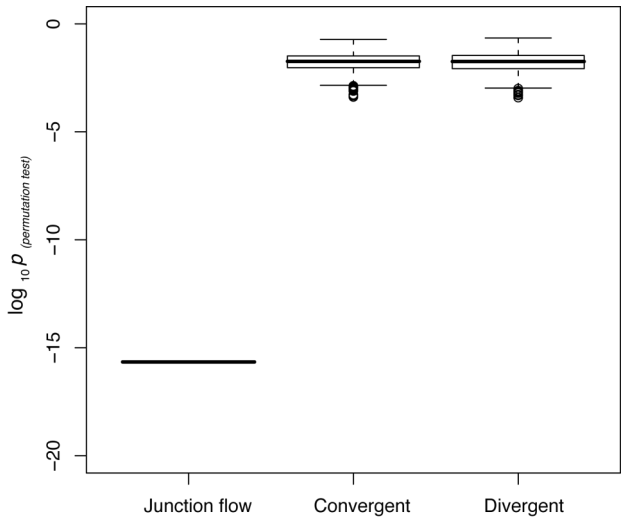
This figure compares the distributions of the numbers of reactions involved in anabolic and catabolic pathways.

Supplementary Figure 3

Co-regulation of at least one flow connected enzymes at junctions



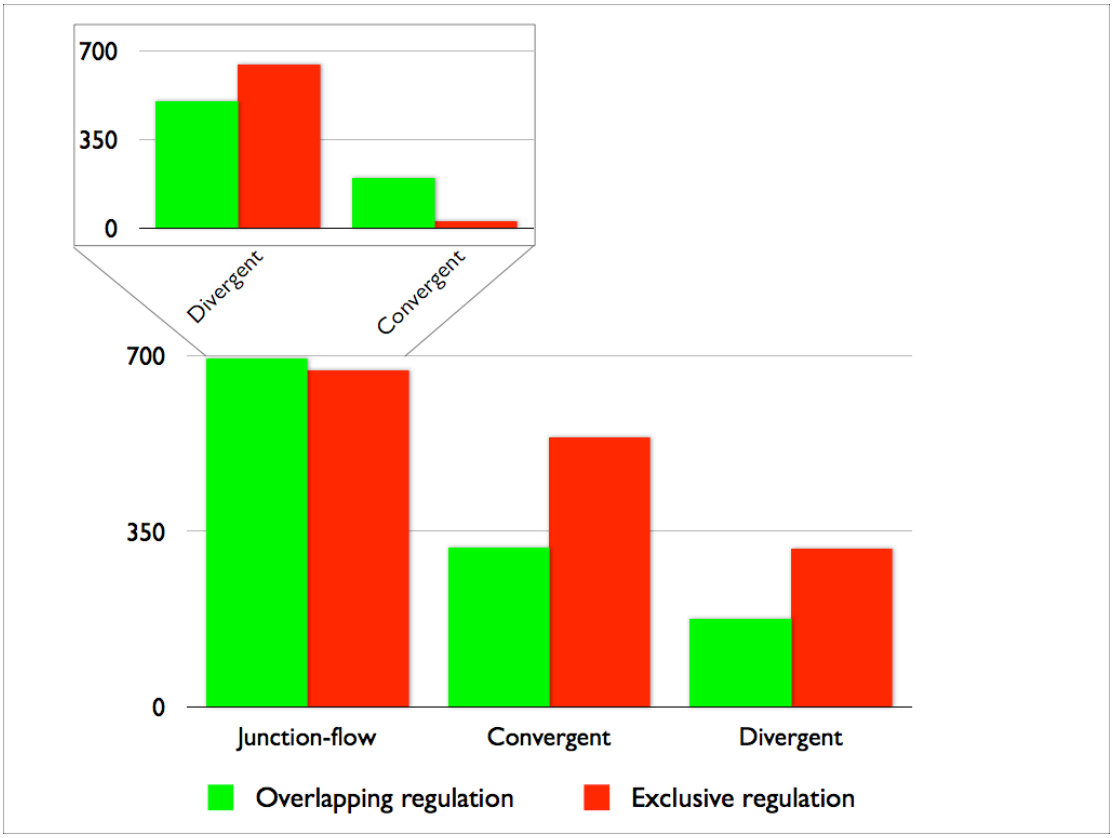
This figure gives the numbers of junction metabolites (with at least one pair of connected enzymes having known TFs) where at least one enzyme-pair connected in a flow (left) or non-flow (centre and right) is co-regulated by the same set of TFs. The difference between the flow and the non-flow configurations is significant (p = 1.2 x 10^{-4})



This figure shows that the chance that at least one pair of enzymes connected at any junction metabolite would be co-regulated is higher in the flow than in the non-flow configuration. The distributions shown here are log 10 of the p-values obtained from a permutation test described in the main text. All p-values in the flow configuration are given as < 2.2 x 10^{-16}. See methods section in main text for statistical information.

Supplementary Figure 4

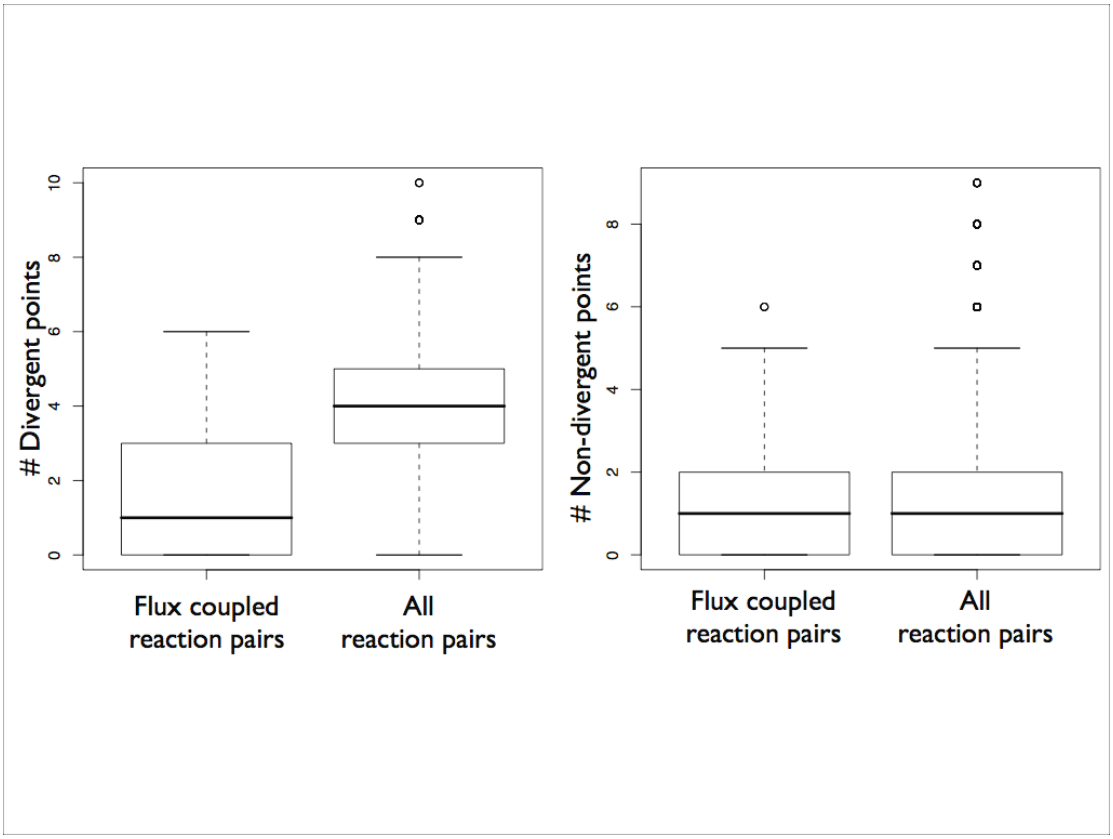
Overlapping and exclusive regulation of enzyme pairs connected at fork point metabolites



The distributions of overlapping (at least one common TF) and exclusive (no common TFs) regulation among junction-flow, convergent and divergent reaction pairs, all of which are connected at fork point metabolites are shown here. The inset shows the difference in this distribution between junction-flow reactions connected at divergent and convergent points.

Supplementary Figure 5

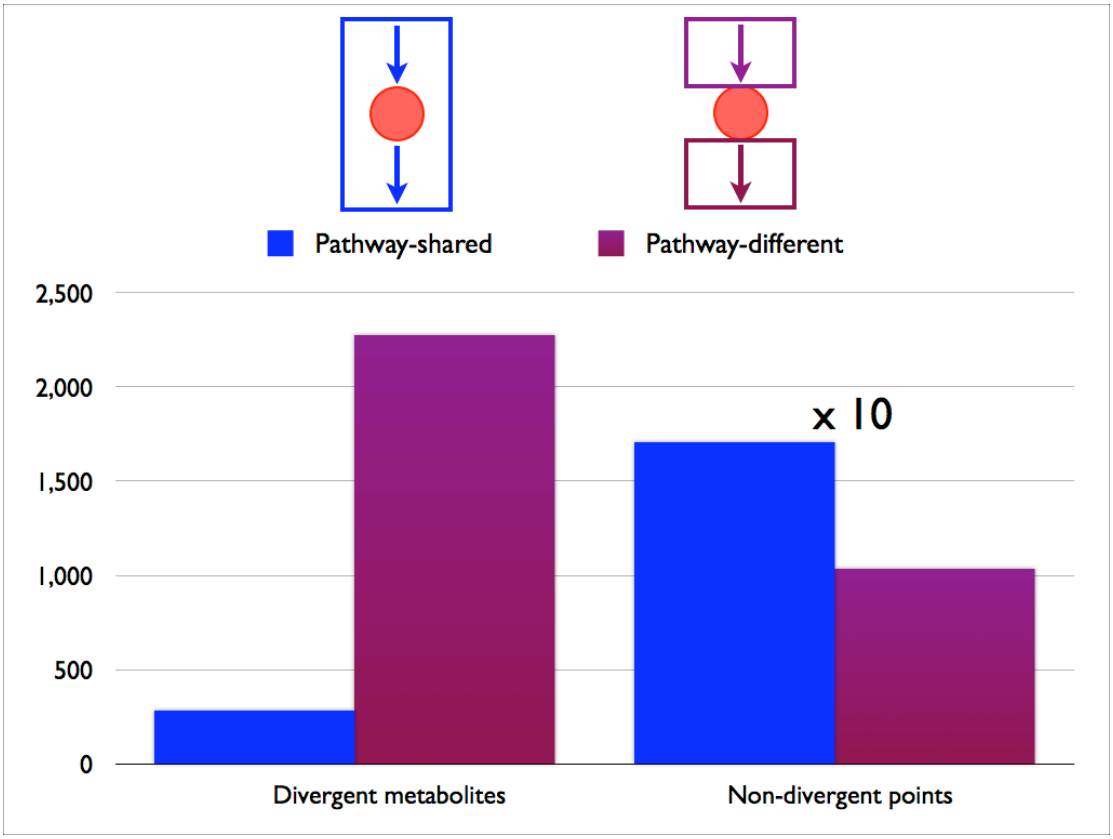
Flux coupling and divergent / non-divergent reactions



Box-plots of the distributions of (left) divergent and (right) non-divergent metabolites separating pairs of flux-coupled enzymes are shown here. These distributions are compared against those for all pairs of enzyme in the metabolic network.

Supplementary Figure 6

Divergent points and pathway boundaries



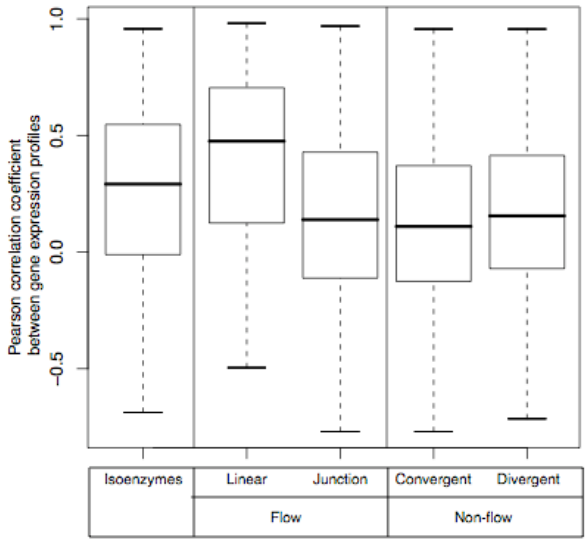
This figure shows that pairs of reactions connected at a divergent metabolite tend not to share a common Ecocyc pathway membership, while those connected at non-divergent metabolites show a greater tendency to belong to the same pathway. The numbers for non-divergent points is multiplied by 10 for display purposes.

Supplementary Figure 7

Co-regulation of Isozymes

#reactions mapped to more than one enzymatic conversions ("enzrxn" ID in ecocyc)		97
#pairs of enzymatic conversions linked to the same reaction		196
#pairs of enzyme genes belonging to different enzymatic conversions and linked to the same reaction	total	267
	with known regulators	111
	with identical sets of regulators	8
	with overlapping sets of regulators	69
	with mutually exclusive sets of regulators	34
Median expression correlation		0.29

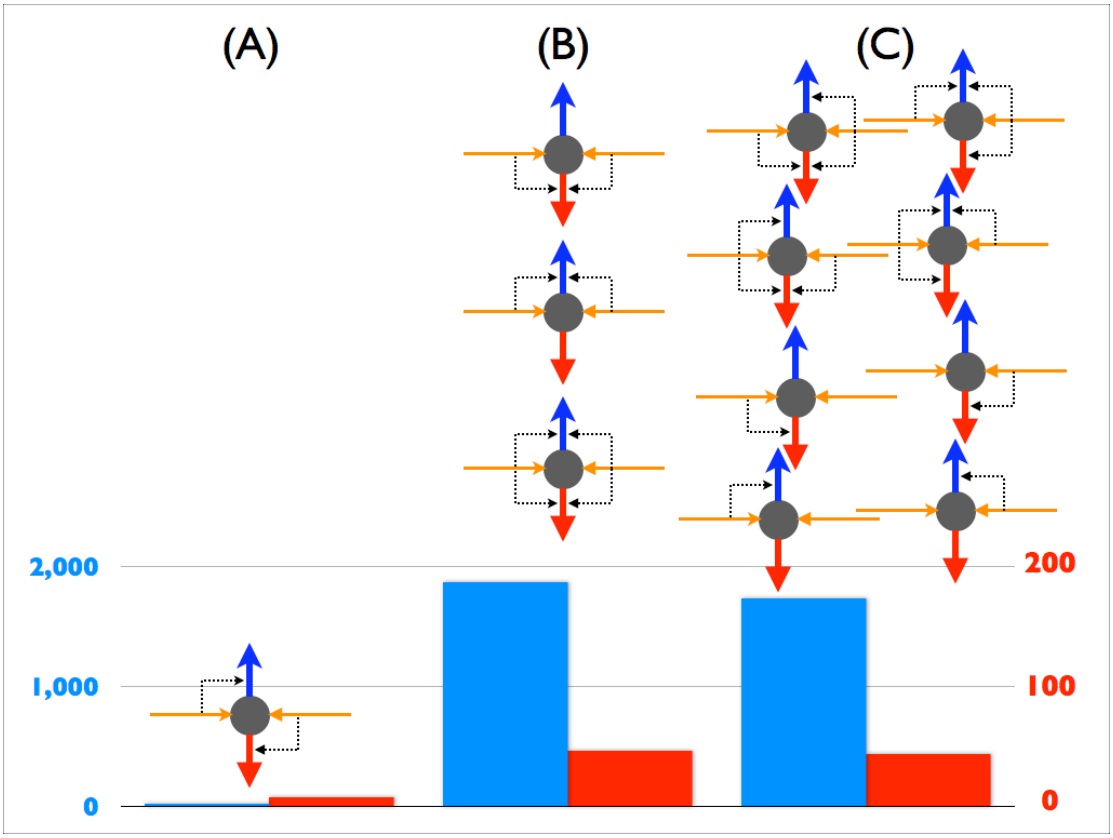
Distribution of co-expression correlations for isozymes compared with those for flow and non-flow enzyme-pairs



Isoenzyme pairs have significantly smaller expression correlation than linear-flow enzyme pairs (p=4.4 x 10^{-4}) but much higher correlation than the flow and non-flow enzyme pairs at junctions (p=8.1 x 10^{-7}).

Supplementary Figure 8A

Evaluation of linear-switch (Ihmels et al. 2004, schematic A) **pattern involving reactions catalysed by isozymes**

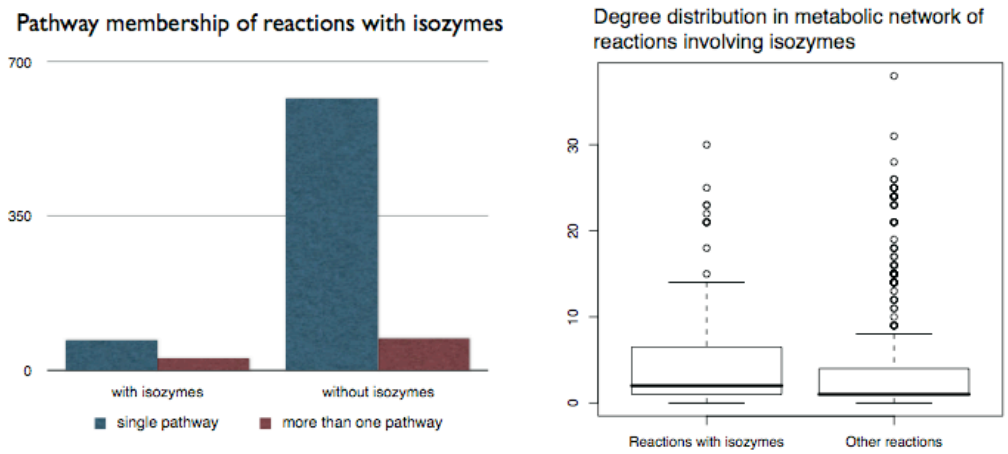


This figure (blue bars, left axis) shows the numbers of quadruplets of enzymes, involving one pair of isozymes and a pair of upstream / downstream enzyme pair, that fall under different regulatory scenarios. Co-regulation is defined by an expression correlation greater than 0.30. The red bars (right axis) give the numbers of reactions with isozymes at which at least one quadruplet follows the given pattern.

In this figure, the orange arrows represent two isozymes for the same reaction. The blue and the red arrows represent other reactions that are in a flow-configuration with the isozymes. For the sake of convenience, the reaction with isozymes is shown as producing the metabolite. The dotted lines connect pairs of co-regulated enzymes. Thus, (A) corresponds to the linear switch pattern illustrated by Ihmels and colleagues (2004) in yeast. In (B), both isozymes are co-regulated with the same reactions and (C) represents all other patterns where there is at least one co-regulatory relationship.

Supplementary Figure 8B

Connectivity of reactions catalysed by isozymes

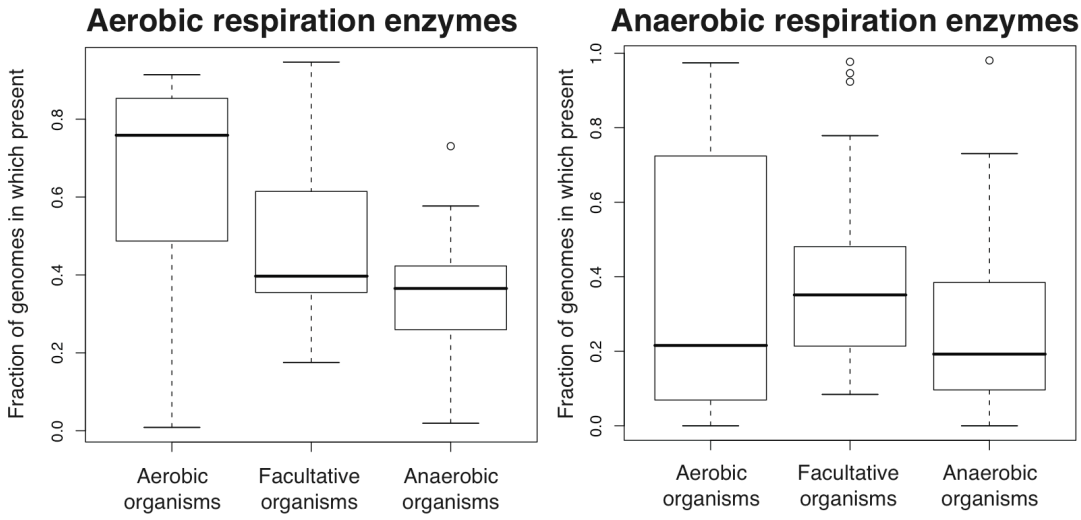


(Left panel) Shows the number of pathways assigned by Ecocyc to reactions that involve isozymes and compares this with those not involving isozymes. Reactions involving isozymes are enriched for being involved in more than one pathway.

(Right panel) Shows the degree distribution, in the metabolic network, of reactions involving isozymes and compares this with those not involving isozymes. Reactions involving isozymes are slightly more connected than those that do not.

Supplementary Figure 9

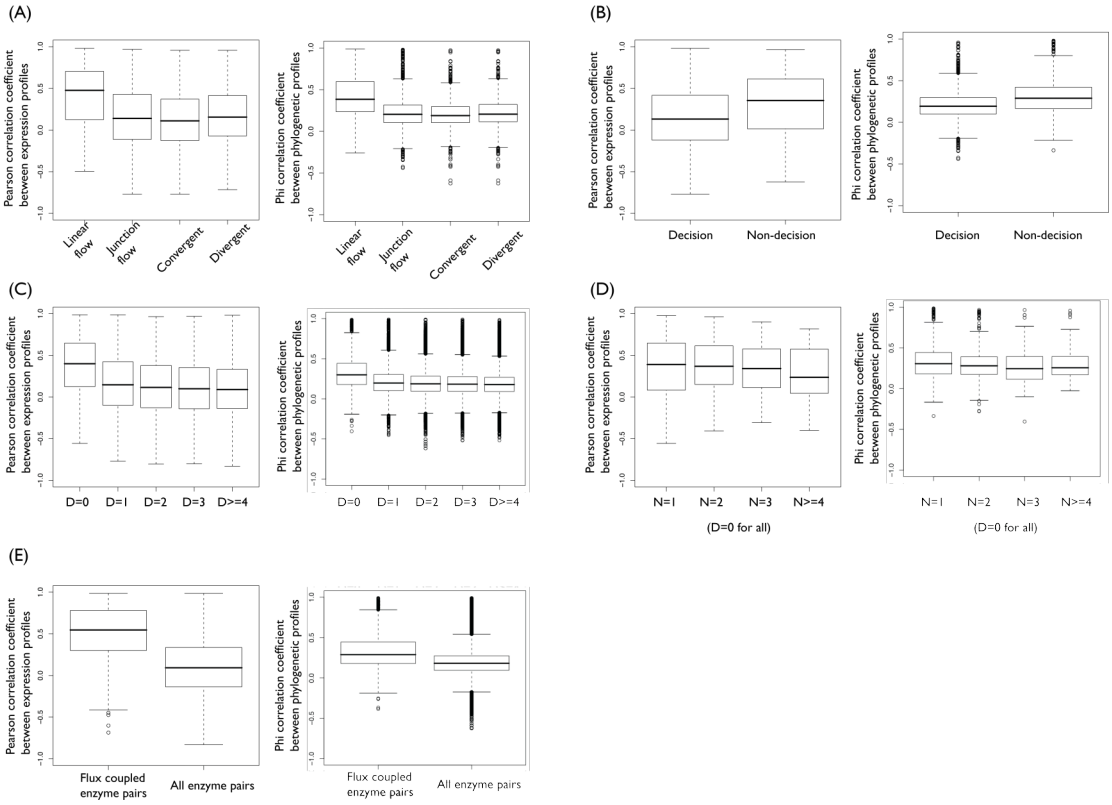
Differences between aerobic and anaerobic organisms in their content of aerobic and anaerobic respiration pathways



Genes belonging to aerobic respiration pathways (TCA and AERESPOND-PWY in Ecocyc) are more conserved in aerobic and facultative organisms than in anaerobes. However, genes from anaerobic energy generation pathways (ANARESP1-PWY, ANARESPACC-PWY, ANARESPDON-PWY and FERMENTATION-PWY in Ecocyc) are not differently conserved between anaerobic and aerobic organisms, though more conserved in facultative organisms. This could be due to the ability of different organisms to make use of different substrates for fermentation.

Supplementary Figure 10

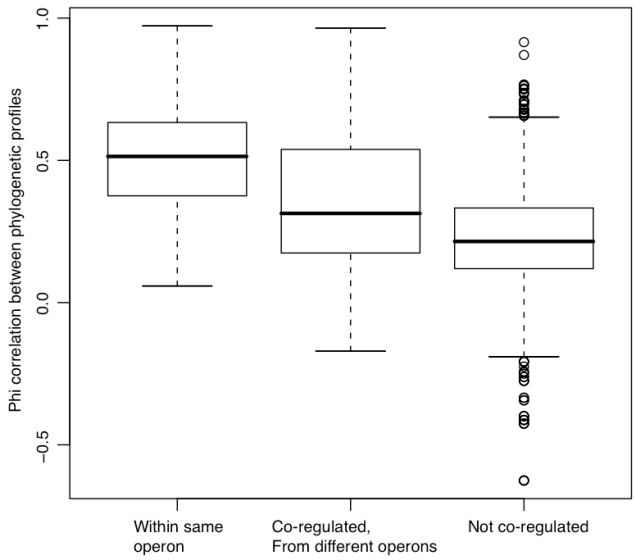
Correlation between co-expression and co-evolutionary patterns



This figure shows the correlation between co-expression and co-evolution in various instances (A) linear reactions compared to reaction pairs connected at fork points – junction-flow, convergent and divergent connections (b) junction-flow reaction pairs connected at divergent points against those connected at non-divergent points (c) impact of the number of divergent points separating two enzymes on their co-expression and co-evolution (d) as in c, but for non-divergent points (e) co-expression and co-evolution of flux-coupled enzymes relative to all pairs of enzymes in the dataset.

Supplementary figure 11

Role of operons in enzyme co-evolution

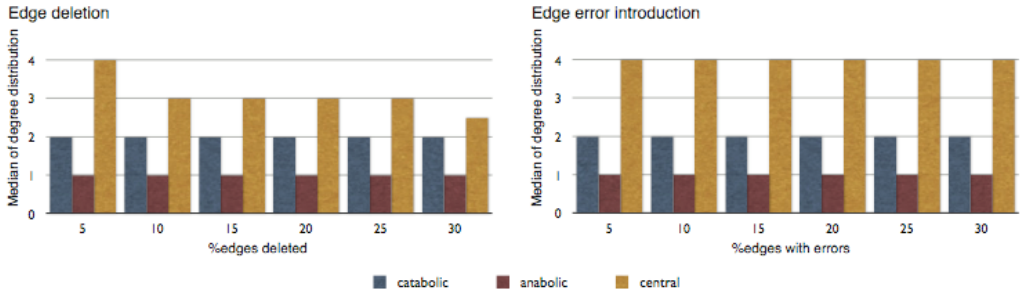


Co-regulated enzyme pairs from within the same operon tend to co-evolve more than those that are encoded in different operons, which in turn have a higher co-expression correlation than those that are not known to be co-regulated. Here, co-regulation is defined based on the transcriptional regulatory network.

Supplementary figure 12A

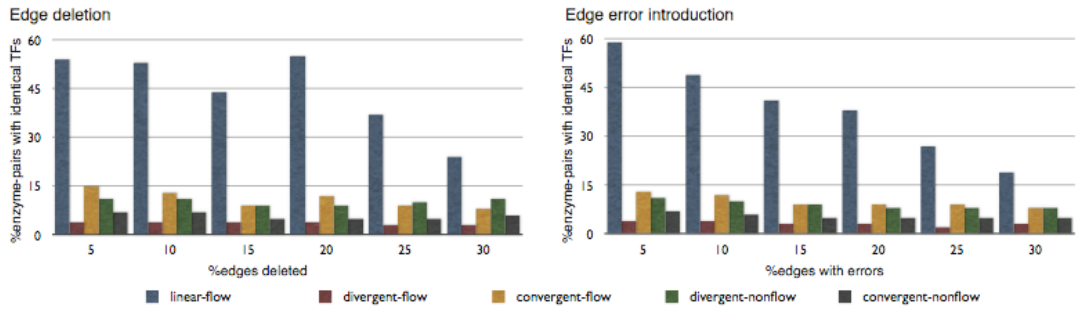
Robustness of results to completeness of- and errors in the transcriptional regulatory network

(A) Global regulation – catabolism, anabolism and central metabolism



This figure shows the medians of the in-degree distributions of catabolic (blue), anabolic (magenta) and central metabolic (orange) enzymes in a variety of transcriptional regulatory networks where deletion or errors are introduced in 5%, 10%, 15%, 20%, 25% and 30% of all regulatory interactions. Edge deletion has some impact only on central metabolism due to the high density of regulatory interactions involving these enzymes.

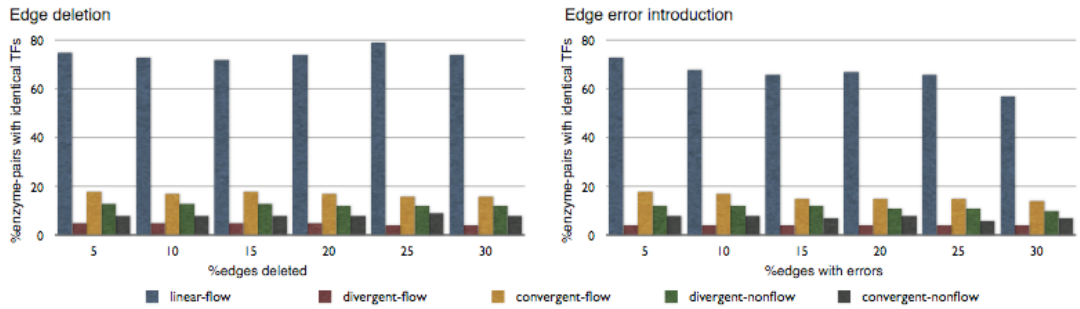
(B) Local regulation – flow and non-flow regulation



This figure shows the numbers of enzyme-pairs with identical TFs in linear-flow, divergent-flow, convergent-flow, divergent-nonflow and convergent-nonflow configurations when the transcriptional regulatory network is altered as described above. Greater variation is seen in the linear-flow system, probably because of the smaller numbers of such pairs that are present.

Supplementary figure 12B

Robustness of results to completeness of- and errors in the metabolic network



This figure shows the effect of deleting or introducing errors in 5%, 10%, 15%, 20%, 25% and 30% of edges in each of the representations of the metabolic network. The effect is measured by the percentage of enzyme pairs with identical TFs.