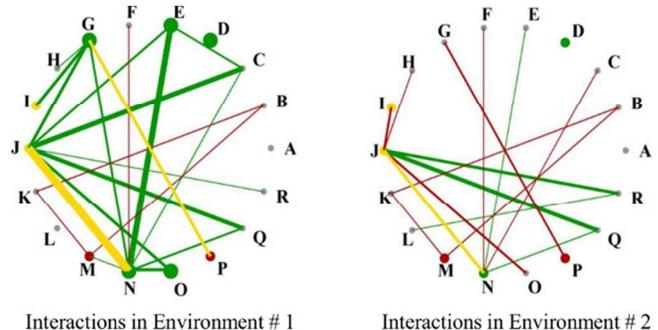




**Epistatic interactions among metabolic genes depend upon
environmental conditions**

Journal:	<i>Molecular BioSystems</i>
Manuscript ID:	MB-ART-03-2014-000181.R2
Article Type:	Paper
Date Submitted by the Author:	17-Jun-2014
Complete List of Authors:	Prasad, Ashok; Colorado State University, Department of Chemical and Biological Engineering Joshi, Chintan; Colorado State University, Chemical and Biological Engineering

Different environmental conditions lead to different metabolic flow, leading to changes in epistasis interactions between genes. Epistasis in turn affects the adaptive fitness landscape.



Epistasis interactions in two different growth conditions
80x39mm (300 x 300 DPI)

ARTICLE

Cite this: DOI: 10.1039/x0xx00000x
Received 00th January 2014,
Accepted 00th January 2014
DOI: 10.1039/x0xx00000x
www.rsc.org/

Epistatic interactions among metabolic genes depend upon environmental conditions

Chintan Jagdishchandra Joshi¹ and Ashok Prasad^{1,2,*}

¹Department of Chemical and Biological Engineering and ²School of Biomedical Engineering, Colorado State University, Fort Collins, Colorado 80523, United States of America

When the effect of the state of one gene is dependent on the state of another gene in more than an additive or neutral way, the phenomenon is termed epistasis. In particular positive epistasis signifies that the impact of the double deletion is less severe than the neutral combination, while negative epistasis signifies that the double deletion is more severe. Epistatic interactions between genes affect the fitness landscape of an organism in its environment and are believed to be important for the evolution of sex and the evolution of recombination. Here we use large-scale computational metabolic models of microorganisms to study epistasis computationally using Flux Balance Analysis (FBA). We ask what the effects of the environment are on epistatic interactions between metabolic genes in three different microorganisms: the model bacterium *E. coli*, the cyanobacteria *Synechocystis* PCC6803 and the model green algae, *C. reinhardtii*. Prior studies had shown that in standard laboratory conditions epistatic interactions between metabolic genes are dominated by positive epistasis. We show here that epistatic interactions depend strongly upon environmental conditions, i.e. the source of carbon, the Carbon/Oxygen ratio, and for photosynthetic organisms, the intensity of light. By a comparative analysis of flux distributions under different conditions, we show that whether epistatic interactions are positive or negative depends upon the topology of the carbon flow between the reactions affected by the pair of genes being considered. Thus complex metabolic networks can show epistasis even without explicit interactions between genes, and the direction and scale of epistasis are dependent on network flows. Our results suggest that the path of evolutionary adaptation in fluctuating environments is likely to be very history dependent because of the strong effect of the environment on epistasis.

Introduction

One of the central problems in biology is that of understanding the mapping between genotype and phenotype. It is now clear that a simple list of active genes do not sufficiently explain phenotype since genes interact in myriad intricate ways. The word “epistasis” has come to suggest the multiple deviations from mere additive effects displayed by genes in an organism. It was first coined by Bateson in 1909 as one genetic variant masking the effect of another [1]. Broadly speaking, when the effect of the state of one gene is dependent on the state of another gene in more than an additive or neutral way, the phenomenon is termed epistasis [2,3]. Epistatic interactions have been classified in multiple ways. For example directional or mean epistasis, also called magnitude epistasis, occurs when both mutations are either deleterious or beneficial, and may be further classified as either aggravating (negative) or buffering (positive) [3]. Aggravating, or negative, interactions between two genes lead to a reduction in fitness of the double mutation that is

greater than that expected by the two single mutations acting independently. Buffering, or positive, interactions occur when one mutation masks the effect of the other mutation [4]. Sign epistasis on the other hand occurs when the effect of one of the mutations changes sign in the background of the other mutation. Finally, the situation when both the mutations are separately deleterious but beneficial when they happen together has been named reciprocal sign epistasis [5-7].

Epistasis is evolutionarily important since epistatic effects can affect the shape of the evolutionary fitness landscape, or the adaptive landscape, that maps gene mutations to fitness. Arguments about the importance and role of epistatic effects played a major role in the debate between Sewall Wright and R. A. Fisher in the 1930s [8]. Epistasis is believed to be necessary for the evolution of sex and recombination [9]. Adaptive landscapes are typically thought of as rugged, with multiple fitness peaks and valleys, and it has been shown that a rugged fitness landscape requires the existence of reciprocal sign epistasis [5,6].

However epistatic effects are hard to uncover experimentally. Cellular metabolism is one arena of research that lends itself easily to the analysis of some kinds of epistatic interactions since it is relatively well understood, and genome-scale constraint based models using Flux Balance Analysis (FBA) do a reasonably good job in predicting intracellular fluxes [10,11] as well as the effect of perturbations [12,13]. A small but significant body of literature has emerged that uses these computational methods to search for epistatic interactions via gene deletions [4,14-16]. An advantage of computational methods is their ability to analyze all putative epistatic interactions; however the framework of FBA limits the analysis to mean or magnitude epistasis, and sign epistasis cannot be studied. Using these methods, it has been shown that metabolic networks of yeast and *E. coli* are characterized by the dominance of small positive epistatic interactions [4,14]. Epistatic interactions were shown to be largely either positive or negative between metabolic subsystems, allowing a redefinition of modularity between functional modules of cellular metabolism [4]. A key insight in this work was that epistasis in the context of FBA-based computational approaches is a consequence of network structure, with linearly connected pathways likely to show positive epistasis with each other, and branched pathways likely to show negative epistasis [4]. In the same work, it was also shown that positive epistasis is higher amongst functionally unrelated genes while negative epistasis was higher among functionally related genes. It has also been shown in later work that epistatic interactions are not absolute, but depend upon the effect being considered. For metabolic models this is most often a function representing “fitness”, and thus epistatic interactions depend upon the particular definition of fitness being used [15], in other words, different fitness functions capture different aspects of functional relationships between genes.

However, a phenotype constitutes the observable characteristic of a genotype in a particular environment. Relatively few experimental studies have analyzed the effect of changing environments on epistatic interactions. One paper analyzed a small set of 18 mutations showed that about a third of mutations analyzed exhibited the joint effect of both the environment and the genetic background [17]. Much more recently, a set of 5 beneficial mutations in *E. coli* were analyzed by constructing 32 double mutations and studying them in 1920 different environments. The effect of both the single mutations as well as epistatic interactions were found to be environmentally dependent [18]. Another equally recent experiment studied three variations of the well-studied Lac operon in *E. coli*, each of which contained three to six point mutations, in the presence or absence of IPTG, and again found strong dependence of epistatic effects on the environment [7].

These results suggest that despite their limitations, computational studies of epistasis in different environments could yield significant insight into the impact of fluctuating environments on the evolutionary process. However to date no such studies using metabolic models have been carried out. This paper seeks to fill that gap by using constraint-based models of metabolism to delineate the effects of the environment on epistatic interactions between metabolic genes in three different microorganisms, the model bacterium *E. coli*, the cyanobacteria *Synechocystis* PCC6803 and the model green algae, *C. reinhardtii*. Computational studies of epistasis have concentrated on yeast and *E. coli*. We therefore also present here the first computational analysis of epistatic interactions in photosynthetic organisms. We also perform a comparative analysis of epistasis in the central carbon metabolism between *E. coli* and *Synechocystis*.

Epistatic analysis is performed using double gene deletions on these three organisms under various different growth conditions. Our analysis throws up a number of novel conclusions. Prior work had

indicated that magnitude epistasis in metabolism is dominated by positive interactions in both yeast and *E. coli*. We show that while this remains true in an aerobic environment, epistasis in anaerobic conditions is dominated by negative epistasis. More generally we show that the increase in the C/O ratio leads to disappearance of large number of positive interactions. We find both differences and similarities in the epistatic interactions of similar genes between *E. coli* and *Synechocystis* under heterotrophic conditions, and show that these arise out of differences in network flows. We show therefore that epistatic interactions are not so much determined by network structure as they are by network flows, and *E. coli* under different carbon sources has different epistatic interactions. We find that under photoautotrophic conditions, the C/photon ratio affects epistatic interactions in the same way as the C/O ratio did in *E. coli*, and under conditions of unlimited light both *Synechocystis* and *Chlamydomonas* are characterized by the relative disappearance of positive interactions between metabolic genes. We thus show that the epistatic interactions uncovered by the computational analysis are not only dependent on the organization of the metabolic network, but also on the environmental conditions.

Results and discussion

Different carbon sources lead to different patterns of fluxes and epistatic interactions

To analyze the effect of environmental conditions on fluxes, we calculated flux distributions in the 174 different carbon sources (substrates) under which *E. coli* could grow, according to the previous model predictions of the model iAF1260 [19]. The reactions were assigned ranks in each growth conditions based on the absolute value of flux [Figure 1a]. We find that fluxes indeed change drastically across different growth conditions as shown by the change in the rankings of reactions [Figure 1a]. The magnitude of the change can be seen by the wide variations in the coefficient of variation of the ranks [Figure 1b and Supplementary Figure 1]. A number of reactions also showed flux reversal under different environment conditions [Figure 1c]. Taken together the data suggest that the topology of carbon flows can change significantly under different carbon substrates.

We chose 9 different carbon substrates to analyze the effect of environmental conditions on epistasis in greater depth in *E. coli*. After removing isozymes we generated a list of 93 genes [Supplementary Table 1], which were non-lethal and generated a flux perturbation in presence of at least one carbon source, and constructed 4278 (93 times 92 times half) double deletion mutations. We find that substrates with more carbon atoms generally result in a greater number of non-zero interactions [Figure 2a]. The only exception is Maltotriose which has fewer total interactions than Glucose. Note that a core set of negative interactions remain conserved under every examined carbon source, thus what is changing are positive interactions. The smallest total number of positive interactions was observed when the organism was grown in presence of formate. During its metabolism formate is used for formylation of tetrahydrofolate (THF) and converted to methylene-tetrahydrofolate (MLTHF). Methylene in MLTHF enriches glycine to serine. Serine is then sequentially converted to phosphoenolpyruvate (PEP), which feeds gluconeogenic and citric acid pathway. Metabolically speaking therefore, formate is quite different from glucose. It requires that a larger number of anabolic reactions be turned on, compared with glucose which requires decomposition to smaller molecules like pyruvate to form higher carbon derivatives. To quantify the metabolic distance between glucose and formate, we introduce the idea of a “metabolic path length”, which is defined as

the minimum number of steps required to form one carbon source from another. We, therefore, hypothesized that average path length between two substrates is proportional to the difference between numbers of positive interactions under different substrates. Consider the case of glucose and sucrose. These two metabolites differ by four reactions [Figure 2b], SUCRtex (sucrose transporter), SUCptspp (sucrose phosphate), FFSD (β -fructofuranosidase), and (XYLI2: a hexose isomerase).

We manually calculated the shortest path length from glucose to the other substrates like formate, formaldehyde, acetate, fumarate, ribose, sucrose, trehalose, and maltotriose. To quantify the notion of difference between number of interactions, taking both positive and negative interactions into account, we calculated the Root Mean Square difference (or the Euclidian distance) between the two dimensional vector representing the number of positive and the number of negative interactions respectively for pairs of growth conditions. In agreement with our hypothesis we find that increase in path length leads to an increase in difference of number of interactions [Figure 2c]. We find that interactions observed under glucose did not change drastically from interactions observed under sucrose, trehalose and maltotriose [Supplementary Figure 2]. Thus the short RMS distance between interactions in glucose and sucrose is due to the metabolic path length of just four reactions mentioned above that separate the two metabolites.

Amongst these 4278 pairs, 150 pairs interacted positively and 22 interacted negatively, in at least one growth condition. Out of these 172 pairs with non-zero interactions, interestingly, only one gene pair changed sign in different environmental growth conditions, b2779 (Enolase, ENO) – b3956 (Phosphoenolpyruvate (PEP) carboxylase, PPC). This pair interacted positively in presence of sugars (trehalose, ribose, glucose, sucrose, and maltotriose), interacted negatively in presence of aldehyde (formaldehyde) and did not interact in presence of carboxylate (formate, acetate, and fumarate). Analysis of flux distributions revealed that the positive interactions are the result of forward flux through ENO (catalyzing dehydration of 2-phosphoglycerate (2PG) to PEP). The product of ENO is PEP, which is a substrate for PPC (catalyzing carboxylation of PEP to oxaloacetate (OAA)). Thus, this linear chain of reaction results in positive interactions. However, in presence of formaldehyde, ENO has a backward flux (hydrolysis of PEP to 2PG). The burden of PEP utilization to make important cellular biomass components results in a synthetic lethal. Thus, under these conditions, the pathway bifurcation occurring due to flow of carbon results in negative interaction. In presence of carboxylates (formate, fumarate, and acetate), PPC carries no flux because OAA is made by TCA cycle. Thus, no interaction occurs between these two genes with carboxylates as substrates.

Interestingly, we find that none of the gene pairs interacted positively in all 9 growth conditions, but 5 gene pairs interacted negatively in all 9 growth conditions. Further, we find that out of these 150 positively interacting pairs, 46% of the interactions occur in either one of the growth condition, while only 18% of the negatively interacting gene pairs occur in either one of the growth condition. Our results indicate that negative interactions, in general, are more likely to persist than positive interactions. The list of 93 genes, and list of positive and negative interactions included in the analysis is presented in the Supplementary information.

Positive epistasis dominates Aerobic growth of *E. coli* and *Synechocystis*

It has previously been shown that metabolic epistatic interactions uncovered by flux balance analysis in *E. coli* and in yeast are dominated by positive or buffering interactions [14]. Photosynthetic

organisms have not been previously analyzed for epistatic interactions. We therefore performed a similar analysis on *Synechocystis* under heterotrophic aerobic growth on glucose, to compare genetic interactions between metabolic genes of *Synechocystis* with *E. coli* under similar environmental conditions (epistasis in autotrophic conditions is discussed later). For completeness, and to validate our method, we repeated the exercise for *E. coli*. A single gene deletion was performed to find all essential genes, and a double gene deletion was performed on the remaining set of non-essential genes. An essential gene, in our simulations, is defined as the gene that leads to a growth rate of less or equal to 10% of the wild-type growth rate. Epistasis values were calculated as shown in *Methods section*. The histogram of scaled epistasis values showed that *E. coli* and *Synechocystis*, under aerobic growth on glucose, is dominated by positive interactions, which were about 5-fold [Figure 3a] and 2.5-fold [Figure 3c] more than the negative interactions, respectively. When these deletions were categorized as a deletion in a particular subsystem of the organism, oxidative phosphorylation and glycolysis had the highest number of interactions with other subsystems in both the organisms. [Figure 3b & 3d]. Note that a previous study on epistasis in metabolic genes [14] reported a much larger number of positive interactions since they were reporting epistasis due to partial deletion of reactions rather than total deletion of genes.

Maximum number of positive interactions corresponds to maximum respiratory capacity in *E. coli*

We next varied the ratio of glucose to oxygen uptake ratio (C/O ratio) in *E. coli* and repeated the epistasis analysis [Figure 4a]. We varied the glucose to oxygen uptake ratios by changing the glucose uptake rate from 8 mmole/gDW/h to 64 mmole/gDW/h. Using experimentally determined specific glucose uptake rate of 8 mmoles/gDW/h [20], we calculated the maximum specific oxygen uptake rate (18.2 mmoles/gDW/h) required by the wild-type cell. We call this rate the maximum respiratory capacity of the wild-type cell. The simulated C/O ratio (0.4395) under the nominal conditions fell within the range of experimentally determined C/O ratios from various different experiments, 0.35 – 0.49 [21]. We find that as the C/O ratio is increased total number of buffering or positive interactions dramatically decrease, while the number of negative interactions remain approximately constant [Figure 4a]. We also noticed that most of the negative interactions remained robust throughout different C/O ratios.

The increase in the C/O ratio is analogous to the organism shifting from aerobic growth to anaerobic growth. This inspired us to evaluate the anaerobic condition, which corresponds to a scenario where C/O ratio goes to infinity. As *E. coli* is a facultative anaerobe, it is also able to grow under anaerobic conditions. We find that under anaerobic condition, the number of positive interactions almost vanish, while the negative interactions are unaffected [Figure 4b]. Thus environmental conditions resulting in excess of carbon substrate (in this case, glucose) help mutations that would otherwise be deleterious under maximum respiratory conditions to grow at optimal growth rates. In the presence of excess carbon, many positively interacting (under nominal conditions) double gene deletions do not interact with each other, leading to their disappearance.

Dominance of negative epistasis under high light conditions in *Synechocystis*

In order to study the effect of varying light conditions on epistatic interactions, we simulated autotrophic growth of *Synechocystis*

under very low to high light conditions. As before non-lethal genes (174 in number) which constrained at least one reaction were included in the analysis. We find that as the photons absorbed increase from 50 mmoles/gDW/h [Supplementary Figure 3e] to 60 mmoles/gDW/h [Supplementary Figure 3f], the number of positive interactions decrease, and under high or unconstrained light conditions, the positive interactions disappear entirely [Figure 5b]. Further, our analysis showed that except for one weakly interacting pair in low light, negative interactions remained unchanged, irrespective of the amount of light available to the organism [Supplementary Figure 3a-i].

We find that fluxes through reactions belonging to the following subsystems increase under high light conditions: Oxidative phosphorylation, photosynthesis, nitrogen metabolism, glyoxylate metabolism, and pyrimidine metabolism.

To test whether the disappearance of positive interactions is a more general property of photoautotrophic metabolism, a similar analysis was performed on another single cell photosynthetic organism, *C. reinhardtii* (iRC1080). We find results to be similar for *C. reinhardtii* as for *Synechocystis*. Under limited light conditions, number of positive interactions and negative interactions were comparable [Figure 5c]; while under high light, the number of positive interactions reduced considerably (~90%) and number of negative interactions remained same [Figure 5d].

Under high light conditions, autotrophic organisms suffer from reduced growth rate [22-25]. Three main changes that occur during such an environmental condition are: (i) reduction in growth rate owing to increase in damage and *de novo* synthesis of photosynthetic proteins [23,24], (ii) increase in the photo-respiratory flux [25] and (iii) decrease in carbon fixation [25]. This reduction in growth rate is not captured in our model due to absence of pathways for damage of photosynthetic proteins. However, it must be noted that the model does correctly predict an increase in the photo-respiratory flux [26]. Since the negative effects of high light cannot be properly accounted for under the current model framework, we cannot comment on how realistic the results of the FBA optimization under high light are. However they do correspond with the case of a high C/O ratio in *E. coli*. Thus similar to excess nutrients, excess light too, leads to a reduction in the number of buffering interactions. Note that both *E. coli* growing under Formate, essentially a 1-carbon source, and *Synechocystis* growing under CO₂; show the dominance of negative interactions.

Chlamydomonas and *Synechocystis* had 2 gene pairs and 1 gene pair respectively which were weakly negatively interacting under low light and non-interacting in high light. In *Chlamydomonas*, the 2 pairs belonged to acetyl-CoA transport across various compartments and the other gene pair belonged to energy production via ATPase in thylakoid membranes. In *Synechocystis*, the weakly interacting genes belonged to ferredoxin oxidoreductase and Glutamate dehydrogenase.

There are three main types of molecules absolutely required for a mutant to grow even at sub-optimal growth rates: (i) ATP, (ii) electron carriers and (iii) carbon. We hypothesize that these mutants were limited by electron carriers and/or ATP when under limited light. However, under high light, there would be a relative excess of these electron carriers and/or ATP. This enrichment of electron carriers and/or ATP under high light helps the organism to grow at optimal growth rate. In presence of high light, the carbon fixation efficiency ($V_{\text{CO}_2 \text{ fixed}}/V_{\text{hv utilized}}$) reduces, as a consequence of which mutations tend to be less deleterious and are able to achieve optimal growth rate, resulting in no interaction between genes.

Epistatic Interactions are dependent on carbon flow in the network

If epistasis in metabolic genes depends on carbon flows in the network, identical genes in two organisms should display mostly similar epistatic interactions, while the differences should be attributable to differences in carbon flow patterns. We compared scaled epistasis amongst gene pairs that constrained identical reactions in both organisms. Of 74 such gene pairs, we found that 54 have identical types of epistasis (positive or negative). Out of these 15 are negatively interacting, while 39 are positively interacting in both the organisms. Some interactions which are common to both are as follows: positive interactions among Glycolysis and TCA cycle, positive and negative interactions within Glycolysis, positive interactions among Pentose phosphate pathway and oxidative phosphorylation, positive interactions among TCA cycle and oxidative phosphorylation [Figure 3b and 3d]. However out of the remaining 20 gene pairs, 18 were mismatches. The mismatches (positive to negative) occur between succinate dehydrogenase (SUCDi), genes belonging to lower glycolysis, and NADH dehydrogenase (NADH); while mismatches (negative to positive) also occur amongst genes belonging to lower and middle glycolysis (Enolase (ENO), phosphoglycerate kinase (PGK), and triose phosphate isomerase (TPI)).

It was not possible to discern the reason for the mismatches from the flux distribution of the entire network, due to its complexity. We therefore decided to perform a reaction-wise epistasis analysis of a subnetwork consisting of reactions involved in glycolysis/gluconeogenesis, TCA cycle, and pentose phosphate pathway. Here by reaction-wise epistasis we mean the non-additive effects of deleting two reactions from the metabolic network. This is equivalent to assuming that each reaction is constrained by a different gene. In reality gene deletions may constrain more than one reaction, making their effect harder to interpret. This sub-network was made up of 33 reactions. It can be seen that large numbers of interactions remain the same in both the organisms [Figure 6]. In *Synechocystis*, 25 and 39 reaction pairs interacted positively and negatively, respectively. Comparing the positively interacting reaction pairs in *Synechocystis* to *E. coli*, we found that 18 reaction pairs interacted positively, and 7 pairs did not interact in *E. coli*. However on comparing the negatively interacting reaction pairs in *Synechocystis* to *E. coli*, we found that 25 reaction pairs interacted negatively but 14 pairs interacted positively in *E. coli*.

We analyzed these 14 mismatches manually and determined that they arise due to differences in carbon flow. In *E. coli*, reactions catalyzed by glucose 6-phosphate dehydrogenase (G6PDH2), 6-phosphogluconolactonase (PGL), and phosphogluconate dehydrogenase (GND) interact positively with each of the reactions catalyzed by ribose 5-phosphate isomerase (RPI), ribose 5-phosphate epimerase (RPE), Transketolase 1 (TKT1), and Transaldolase (TALA). In *E. coli*, any deletion in oxidative pentose phosphate pathway (G6PDH2r and PGL) results in the same metabolic flux redistribution. In the absence of oxidative pentose phosphate pathway, operation of TALA, TKT1 and TKT2 is reversed such that ribose 5-phosphate, xylulose 5-phosphate, and ribose 1-phosphate (R1P) is produced in both the organism. However, in *Synechocystis*, other than phosphopentomutase (PPM), R1P can only be produced by the decomposition of adenosine. There are many other reactions that can produce R1P, in *E. coli*. This is why interaction amongst oxidative pentose phosphate reactions and TKT1, RPI, RPE, and TALA is positive in *E. coli*. Similar reasons can be attributed to other 4 interactions occurring amongst reactions in glycolysis [Figure 3]. The mismatches (epistasis sign-change) account for about 21% (14/64) of the total interactions which were positive or negative. Thus epistatic interactions are affected by metabolic flows,

which are in turn affected by the environmental condition of an organism.

Experimental

Flux Balance Analysis (FBA)

Flux Balance Analysis (FBA) is a mathematical framework used to calculate the flow of the metabolites through the metabolic network at steady state [27]. FBA was performed using the COBRA Toolbox [28]. In brief, FBA involves writing down an M by N stoichiometric matrix, S corresponding to the metabolic reactions for each organism. Here M is the number of metabolites and N is the number of reactions. Under steady state conditions the system of differential equations representing the chemical reaction system become a system of linear equations in the fluxes,

$$\sum_{j=1}^N S_{ij} v_j = 0 \quad (1)$$

Here, v is a vector of reaction flux and S_{ij} represents the stoichiometric coefficient for i^{th} metabolite in j^{th} reaction. To find the fluxes an objective function is chosen that is believed to be optimized by the organism, such as its growth rate. This makes it a linear programming problem (LPP) that can be solved by standard techniques by imposing additional constraints, discussed below, in addition to Eq. 1 [27,29]. The objective function most commonly used for such models is an equation describing the growth rate of the organism. Growth rate reactions are described as:

$$\sum_{j=1}^N c_j v_j \rightarrow \mu \quad (2)$$

In the above equation, c_j and v_j refer to the weight in final biomass and the flux of the product of the j^{th} reaction respectively, and μ refers to the growth rate of the organism. Maximization of growth rate was used as the objective function for all the simulations conducted in this study. Additional constraints are constructed in the following way:

1. Incorporating measured or experimentally estimated uptake and secretion rates of metabolites.
2. Incorporating a global limit on the upper and lower bounds of each reaction flux.

$$\alpha_j \leq v_j \leq \beta_j \quad (3)$$

α_j and β_j are the lower and upper limits placed on each reaction flux, v_j , respectively. Reversible reactions can take either negative or positive values of fluxes, while irreversible values were constrained to take only positive values. Further, if any reactions were turned off, inactivated or deleted, the flux through the reaction was set to zero: $v_j = 0$. The linear programming problem was implemented using COBRA Toolbox with Gurobi 4.6.1 on MATLAB R2011b [28].

Simulation of growth conditions in various organisms

For our analyses of different cellular metabolism, we chose genome scale models of *Escherichia coli* K12 MG1655 (iAF1260) [19], *Synechocystis sp.* 6803 (iJN678) [26], and *Chlamydomonas reinhardtii* (iRC1080) [30]. Growth conditions included in the analyses were: *E. coli* (**Aerobic**: Formate, Formaldehyde, Acetate, Fumarate, Ribose, Glucose, Sucrose, Trehalose, Maltotriose, and **Anaerobic**), *Synechocystis sp.* (**Autotrophic**: high light, limited light, and **Aerobic**: Glucose), and *C. reinhardtii* (**Autotrophic**: high light, limited light). For simulation of different carbon sources in *E. coli*, we normalized carbon uptake to 8 mmoles for 6 carbon atoms

in the molecule. For example, if 8 mmoles/gDW/h of glucose was used; then 12 mmoles/gDW/h of fumarate, a four carbon molecule, was used. Limited light conditions were simulated by setting the maximum light uptake to the optimal value calculated for wild-type cells. However, high light conditions were simulated by leaving light uptake unconstrained. Non-lethality criterion for a mutant was set to more than 10% (or 0.1 times) of wild-type growth rate, correct to first order of magnitude.

For aerobic growth of *E. coli* (model name: iAF1260) under various carbon sources, simulations were performed by applying the following constraints: (i) maximum uptake rate of the desired carbon substrate (EX_glc(e), EX_sucr(e), EX_for(e), EX_fum(e), EX_rib-D(e), EX_malttr(e), EX_tre(e), EX_fald(e), or EX_ac(e)) was set to 8 mmoles/gDW/h per 6 molecules of carbon in the substrate [19], while uptake rates of all other carbon sources were set to zero; maximum oxygen uptake rate (EX_o2(e)) was left unconstrained [19]; and all the other constraints were same as reported in the original article of the published model. Heterotrophic growth of *Synechocystis sp.* PCC6803 (model name: iJN678) was simulated by setting the maximum glucose uptake rate (EX_glc(e)) to 0.85 mmoles/gDW/h [31]; leaving maximum oxygen uptake rate unconstrained; and setting the uptake rates of other sources of carbon and light to zero. Autotrophic growth of *Synechocystis sp.* PCC6803 was simulated by setting the maximum carbon dioxide uptake rate to 3.7 mmoles/gDW/h [26] and uptake rates of other carbon sources to zero; while minimum photon uptake rate corresponding to maximum growth rate was calculated and subsequently set to 54.0948 mmoles/gDW/h; and all the other constraints were used from the original article where the model was published. Autotrophic growth of *Chlamydomonas reinhardtii* (model name: iRC1080) was simulated by utilizing the constraints from the original article where the model was published [30]. The default constraints on flux of a reversible reaction was [-1000, 1000], and of an irreversible reaction was [0, 1000], unless specified here or in the original article where these models are published. These constraints are a norm in the field and have been used in numerous FBA studies [19,26,30].

Ranking of fluxes

Fluxes for each of the 174 conditions leading to growth as reported in the original publication of the *E. coli* (iAF1260) model [19] were simulated and ranked according to flux magnitudes. The directionality of reaction was ignored, in case of reversible reaction, because enzyme catalyzing the activity will be observed whether the reaction was operating in forward or reverse direction.

Calculation of Epistasis

Firstly, a single gene deletion was performed to remove any essential genes. Then, a double gene deletion was performed on the remaining set of genes. Epistasis values were calculated as shown below [4]. The epistasis value for the interaction between gene X and gene Y is represented by ε . This value can be calculated by:

$$\varepsilon = W_{XY} - W_X W_Y \quad (4)$$

Here $W_X = \mu^X / \mu^{wt}$, $W_{XY} = \mu^{XY} / \mu^{wt}$ are the fitness values for the single mutant and the double mutant, and μ^{wt} , μ^X , and μ^{XY} are growth rates of wild-type, the mutant in gene X, and mutant in genes X and Y. While this is the absolute level of epistasis we need to establish a standard to compare it with. We follow Ref. [4] and scale the epistasis value given by Equation (4) as follows:

$$\tilde{\varepsilon} = \frac{W_{XY} - W_X W_Y}{|\tilde{W}_{XY} - W_X W_Y|}; \quad (5)$$

$$\tilde{W}_{XY} = \begin{cases} \min(W_X, W_Y) & \text{for } W_{XY} > W_X W_Y \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The unscaled (ε) and scaled ($\tilde{\varepsilon}$) epistasis values can be then classified as shown in Table 1 below.

Table 1 Classification of different ranges of unscaled and scaled epistasis.

	Unscaled Epistasis	Scaled Epistasis
No epistasis	$\varepsilon = 0$	$\tilde{\varepsilon} \approx 0$
Aggravating	$\varepsilon < 0$	$\tilde{\varepsilon} \approx -1$
Buffering	$\varepsilon > 0$	$\tilde{\varepsilon} \approx 1$

The scaled epistasis ($\tilde{\varepsilon}$) was used to classify the interactions into buffering (green) at $\tilde{\varepsilon} > \theta_+$; aggravating (red), including synthetic lethal at $\tilde{\varepsilon} = -1$ and strong synthetic sick at $\tilde{\varepsilon} < \theta_-$; and no epistasis otherwise. Here we used $(\theta_-, \theta_+) = (-0.25, 0.85)$. It must be noted here that denominator in Eq. 5 is an absolute value and will not change the sign of the epistasis.

Note that here we characterize the phenotype by the growth rate of the organism. Growth rate makes a good choice of phenotype because of the role of epistasis in selection dynamics [4], and it can be measured accurately using high throughput methods [4,32,33]. However, the mathematical framework of Flux Balance Analysis (FBA) used here to calculate growth rate requires maximization of growth rate. As a result, one is never able to calculate instances when the fitness of the mutants is higher than the fitness of the wild-type organism [15]. This is why sign epistasis cannot be studied using FBA. Therefore, we specify that our results are only relevant for epistatic interactions relative to growth rate.

Mapping gene pairs from one organism to another

For each of the genes involved in the pair, the E. C. numbers of the reactions constrained by the gene were found. These E. C. numbers were then searched for in the other organism. All the genes associated with the reactions with those E. C. numbers were found and pairs were created for the new organism, based on the pairs found in the source model. This mapping technique has been previously used to investigate the structure of enzyme-reaction association in microbial metabolism [34].

Calculation of RMS difference between interactions

The formula for calculating the root mean square (RMS) distance is:

$$D = \sqrt{(N_p - N'_p)^2 + (N_n - N'_n)^2} \quad (7)$$

The meaning of the symbols is as below:

D = RMS difference between interactions;

N_p, N_n = Number of positive and negative interactions in nominal case, respectively

N'_p, N'_n = Number of positive and negative interactions in growth condition, respectively.

Conclusions

Flux Balance Analysis of large-scale metabolic models is an attractive tool for studying epistatic interaction between genes computationally. It has been argued earlier that the sign of epistatic interactions between two genes gives us information about how the genes interact in the metabolic network. If the two genes belong to the same subsystem, a positive interaction suggests that they form a linear or sequential chain with respect to each other, while a negative interaction suggests that they are part of related pathways that form the same product [3, 4]. However, previous system-wide computational studies of epistasis have not considered the impact of environment conditions on predictions of epistatic interactions. Here, we systematically generated epistatic interaction network maps relative to growth rate for *E. coli*, *Synechocystis sp.*, and *C. reinhardtii* under various different environmental conditions, by which here is meant different substrates on which the organism grows. Analysis of these networks revealed that different environmental conditions yield different sets of epistatic interactions. Epistatic interactions therefore change with time as environmental conditions change.

We show that epistasis in anaerobic conditions is dominated by negative epistasis. More generally we show that the increase in the C/O ratio leads to disappearance of large number of positive interactions. We find both differences and similarities in the epistatic interactions of similar genes between *E. coli* and *Synechocystis* under heterotrophic conditions, and show that these arise out of differences in network flows. We find that under photoautotrophic conditions, the (CO₂/photon) ratio affects epistatic interactions in the similar way as the C/O ratio did in *E. coli*, and under conditions of somewhat high light *Synechocystis* tends to have lower positive interactions, and in unlimited light both *Synechocystis* and *C. reinhardtii* are characterized by a sharp decline in positive interactions between metabolic genes.

We also analyse *E. coli* under different carbon sources and show that it has different set of epistatic interactions, governed primarily by the flow of the carbon within the metabolic network. We thus show that the epistatic interactions uncovered by the computational analysis are not only dependent on the organization of the metabolic network, but also on the environmental conditions.

Our findings suggest that during adaptation in dynamically changing environment, the shape of the fitness landscape may be governed by the environmental history and the pattern of carbon flow in the current state of the metabolic network. Flux flows within similar parts of the metabolic network between two organisms under the same growth conditions gives rise to generally similar interactions. For example, the carbon flow through glycolysis in *E. coli* and *Synechocystis sp.*, under aerobic growth with glucose, will be similar (but not identical), and as a result the interactions occurring within the glycolysis pathway remain mostly similar as well. In both these organisms, under heterotrophic growth one molecule of glucose is catabolized to form two molecules of pyruvate, and is converted to acetyl-CoA, a precursor to the TCA cycle. However, *Synechocystis* grown photo-autotrophically will yield a different set of gene-gene interactions within the glycolysis pathway because these conditions require the formation of glucose and pyruvate from 3-phosphoglycerate.

A previous study has stressed the significance of the finding, on the basis of FBA, that positive epistasis is highly abundant between functionally unrelated genes in both *E. coli* and *S. cerevisiae* [14]. This study explained this phenomenon as occurring due to a second

mutation having a relatively smaller effect than the first. However we show that while positive epistasis is highly abundant compared with negative epistasis in many environmental conditions, in many other conditions it is no longer abundant, and in some cases, disappears entirely. Negative interactions however, in particular synthetic lethals, tend to remain conserved under different conditions. As previously noted, Ref. [14] calculate epistasis differently from us, i.e. they perform reaction deletions rather than gene deletions and they constrain flux through each reaction to 50% of its wild-type value rather than setting it to zero. In this paper we consider only epistasis due to loss-of-function mutations in genes.

Previous work has shown that selection pressures exerted due to changing environmental background resulted in different fitness landscapes. Complementary to these findings, our analysis with different growth conditions for *E. coli* show that positive interactions are more likely to change/disappear, while negative interactions are likely to stay conserved.

Our analysis shows that epistasis among metabolic genes that is predictable by FBA methods depends upon network flows. Therefore positive epistasis is not simply the result of the network topology connecting two genes being linear, as suggested in previous work, but network flows between two genes forming a linear topology. Similarly if network flows between two genes constitute a branched topology with the two genes on separate branches, we get negative epistasis between their deletions. Since FBA models do not have any transcriptional regulation, or nonlinear interactions between proteins, it is noteworthy that they show that epistatic effects can arise as a consequence of network structure alone.

We also show that excess nutrient uptake conditions result in a decrease in the number of positive interactions. Presumably, the excess of nutrient conditions result in enrichment of metabolites that under nominal conditions were limiting to the growth. This enrichment allows the organism to sustain the carbon, energy or electron flow in mutants, thereby changing deleterious mutants (under nominal conditions) to fit mutants (under excess of nutrient conditions). The behaviour of *E. coli* under excess carbon mirrored the behaviour of *Synechocystis* and *C. reinhardtii* under excess of light. In the latter case too, we found that formerly deleterious mutations become non-deleterious mutations as a result of which the most positive epistatic interactions between gene pairs vanish. Negative interactions that lead to synthetic lethality remain.

What is the importance of these epistatic predictions? An organism evolved in a specific niche should be, metabolically speaking, optimized to live in the niche. It should be expected therefore that loss-of-function mutations in metabolic genes are always accompanied by a decline in fitness. Given a single gene deletion that marginally decreases fitness; a second deletion with positive epistasis with the first is more likely to be selected in the population than one that further decreases fitness in a neutral way. This suggests that mutations during adaptation in varying environments are selectively directed by positive interactions occurring amongst deleterious mutations. This is in agreement with experimental studies that show that mutations that get fixed in populations undergoing environmental change, such as during the evolution of antibiotic resistance in bacteria, are deleterious in the background [35].

Since network flows can change if different substrates are being metabolized, epistatic interactions also change with change in substrate metabolized. This prediction agrees with previous experimental studies on effect of environment on epistasis and fitness landscape [7, 18]. Environmental conditions that change the flows can dramatically change the set of epistatic interactions, and thus the adaptive fitness landscape of the population. The environmental dependence of epistasis makes the task of piecing

together evolutionary history, and the role of epistasis in it, all the more difficult, since the specific evolutionary path followed by an organism, during adaptation in a variable environment, would be therefore highly dependent upon the specific environmental fluctuations that it encountered in its evolutionary history.

Acknowledgements

We are grateful to Pankaj Mehta (Boston University) for suggesting that we look at C/O ratios in *E. coli* and to Christie Peebles (Colorado State University) for suggesting that we analyse anaerobic growth. Support from NSF grant CBET-1336236 (C.J.J and A.P.) and NSF CAREER grant PHY-1151454 (A.P.) is gratefully acknowledged.

References

1. W. Bateson, *Mendel's Principles of Heredity*. 1909, Cambridge: Cambridge University Press.
2. M. Breen, C. Kemena, P. Vlasov, C. Notredame, and F. Kondrashov, *Nature*, 2012, **490**, 535–8.
3. J. de Visser, T. Cooper, and S. Elena, *Proceedings. Biological sciences / The Royal Society*, 2011, **278**, 3617–24.
4. D. Segrè, A. Deluna, G. Church, and R. Kishony, *Nature genetics*, 2004, **37**, 77–83.
5. A. Dawid, D. Kiviet, M. Kogenaru, M. de Vos, and S. Tans, *Chaos (Woodbury, N.Y.)*, 2010, **20**, 026105.
6. F. Poelwijk, S. Tănase-Nicola, D. Kiviet, and S. Tans, *Journal of theoretical biology*, 2011, **272**, 141–4.
7. M. de Vos, F. Poelwijk, N. Battich, J. Ndika, and S. Tans, *PLoS genetics*, 2013, **9**, e1003580.
8. E. D. Brodie III, *Epistasis and the evolutionary process*, 2000, 3–19.
9. J. de Visser and S. Elena, *Nature reviews. Genetics*, 2007, **8**, 139–49.
10. R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer, *Science (New York, N.Y.)*, 2012, **336**, 601–4.
11. R. Schuetz, L. Kuepfer, and U. Sauer, *Molecular systems biology*, 2006, **3**, 119.
12. D. Segrè, D. Vitkup, and G. Church, *Proceedings of the National Academy of Sciences of the United States of America*, 2002, **99**, 15112–7.
13. S. Fong and B. Palsson, *Nature genetics*, 2004, **36**, 1056–8.
14. X. He, W. Qian, Z. Wang, Y. Li, and J. Zhang, *Nature genetics*, 2010, **42**, 272–6.
15. E. Snitkin and D. Segrè, *PLoS genetics*, 2010, **7**, e1001294.
16. L. Xu, B. Barker, and Z. Gu, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 10420–5.
17. S. Remold and R. Lenski, *Nature genetics*, 2004, **36**, 423–6.
18. K. Flynn, T. Cooper, F. Moore, and V. Cooper, *PLoS genetics*, 2013, **9**, e1003426.
19. A. Feist, C. Henry, J. Reed, M. Krummenacker, A. Joyce, P. Karp, L. Broadbelt, V. Hatzimanikatis, and B. Palsson, *Molecular systems biology*, 2006, **3**, 121.
20. E. Fischer, N. Zamboni, and U. Sauer, *Analytical biochemistry*, 2004, **325**, 308–16.
21. A. Kayser, J. Weber, V. Hecht, and U. Rinas, *Microbiology (Reading, England)*, 2005, **151**, 693–706.
22. J. Kopecna, J. Komenda, L. Bucinska, and R. Sobotka, *Plant Physiology*, 2012, **160**: 2239–2250.
23. B. Demmig-Adams, W. W. Adams III, *Annual Review of Plant Physiology and Plant Molecular Biology*, 1992, **43**: 599–626.
24. Y. Allahverdiyeva, M. Ermakova, M. Eisenhut, P. Zhang, P. Richaud P, et al., *Journal of Biological Chemistry*, 2011, **286**: 24007–24014.

ARTICLE

25. C. Hackenberg, A. Engelhardt, H. C. Matthijs, F. Wittink, H. Bauwe H, et al., *Planta*, 2009, **230**: 625-637.
26. J. Nogales, S. Gudmundsson, E. Knight, B. Palsson, and I. Thiele, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 2678–83.
27. J. Orth, I. Thiele, and B. Palsson, *Nature biotechnology*, 2010, **28**, 245–8.
28. J. Schellenberger, R. Que, R. Fleming, I. Thiele, J. Orth, A. Feist, D. Zielinski, A. Bordbar, N. Lewis, S. Rahmanian, J. Kang, D. Hyduke, and B. Palsson, *Nature protocols*, 2011, **6**, 1290–307.
29. A. Feist and B. Palsson, *Current opinion in microbiology*, 2010, **13**, 344–9.
30. R. Chang, L. Ghamsari, A. Manichaikul, E. Hom, S. Balaji, W. Fu, Y. Shen, T. Hao, B. Palsson, K. Salehi-Ashtiani, and J. Papin, *Molecular systems biology*, 2010, **7**, 518.
31. C. Yang, Q. Hua, and K. Shimizu, *Metabolic engineering*, 2002, **4**, 202–16.
32. A. Jakubowska and R. Korona, *PLoS ONE*, 2011, **7**, e33132.
33. G. Martin, S. Elena, and T. Lenormand, *Nature genetics*, 2007, **39**, 555–60.
34. C. Joshi and A. Prasad, *PloS one*, 2013. **In Press**.
35. X. Wang, G. Minasov, and B. Shoichet, *Journal of molecular biology*, 2002, **320**, 85–95.

Figure legends

Figure 1: Fluxes change depending upon growth conditions. (A) Flux ranks associated with reactions corresponding to the 174 growth (environmental) conditions; the color axis represents the rank of the reaction in any particular environment. (B) Coefficient of variation (σ/μ) of the rank calculated across 174 growth conditions for each reaction. (C) Histogram of reactions in *E. coli*, X-axis represents the number of environmental conditions in which a given reaction has positive (blue), negative (magenta), or zero (green) flux; the letters (in parentheses) on the Y-axis correspond to subsystems as listed in Supplementary Table 4. **NOTE:** For (C), we had to use reversible model, while (A) and (B), reactions must be irreversible.

Figure 2: Epistasis under various different carbon sources. (A) Number of positive and negative interactions for *E. coli*, when grown under different carbon sources. The histograms above each bar represents the distribution of epistasis values in each of the growth conditions. Pairs with no interaction are not shown for proper visualization of data; these pairs do peak at epistasis value of 0 (giving rise to a trimodal distribution). Red represents negative interactions, green represents positive interactions; (B) four reactions defining the path from sucrose to glucose; (C) RMS distance between positive and negative interactions varies based on the shortest path length between carbon source mentioned in green and glucose.

Figure 3: Epistatic interactions maps relative to aerobic growth on glucose for *Synechocystis* sp. PCC6803 and *E. coli*. (A) histogram of epistasis for *E. coli* under aerobic growth with glucose; (B) Epistatic map for *E. coli*; (C) histogram of epistasis for *Synechocystis* under aerobic growth with glucose (heterotrophic); and (D) Epistatic map for *Synechocystis*. Red represents negative interactions, Green represents strong positive interactions, and grey represents weak positive interactions. The inset graph in (a) and (c) represents non-interacting pairs (black). The subsystems corresponding to letters is present in Supplementary Table 4 & 5. The size of the dots is proportional to the number of interactions (this convention is followed in all of these types of plots in the paper).

Figure 4: Epistasis under varying glucose-to-oxygen uptake ratios. (A) Number of positive (green) and negative (red) interactions corresponding to each $v_{\text{glu}}/v_{\text{O}_2}$ (C/O specific uptake) ratio. The clock diagrams shown in insets represent interactions amongst subsystem at (left to right) C/O uptake ratio = 0.4918, C/O uptake ratio = 1.7297, and C/O uptake ratio = 3.4595. The letters correspond to subsystems as listed in Supplementary Table 4. The size of the dots pertaining to each subsystem indicates the number of epistatic interactions, with Green = positive; Red = Negative and Yellow = mixed (both negative and positive). All clock diagrams shown in the paper follow these conventions. (B) Histogram of scaled epistasis of *E. coli* for anaerobic growth with glucose. Histogram is read as distribution of scaled epistasis based on the non-scaled epistasis value of the interacting pair. Red represents negative interactions, green represents strong positive interactions, and gray represents weak positive interactions. The inset graph represents non-interacting pairs (black). (C) Difference between aerobic growth of *E. coli* (nominal) and anaerobic growth of *E. coli*, represented by a grey map where darker the value higher is proportion of total gene interaction in that category. A perfect black corresponds to all pairs having same type of interaction; a perfect white corresponds to no pairs in the region. (D) Clock diagram representing interaction between genes belonging to various subsystems in *E. coli* under anaerobic growth conditions.

Figure 5: Epistasis interactions amongst reactions belonging to three compartments Glycolysis, Citrate cycle, and pentose phosphate pathway for cells grown aerobically with glucose. The overall picture represents the flow of carbon in *E. coli*. The black arrows indicate the direction of flow in *E. coli* and *Synechocystis*. The yellow arrow indicates reaction operating in reverse direction in *Synechocystis*. The grey arrow indicates significantly less (<10% of proportion of reaction flux through *E. coli*) flux through the reaction in *Synechocystis*. The orange arrows indicates significantly less (<10% of proportion of reaction flux through *Synechocystis*) flux through the reaction in *E. coli*. The green lines indicate the differences in epistasis which was negative for *Synechocystis* but positive for *E. coli*.

Figure 6: Histograms of scaled epistasis for photoautotrophic organisms under limited light and high light conditions. For *Synechocystis* sp. PCC6803, (A) limited light, (B) high light; for *C. reinhardtii* (C) limited light, (D) high light. Red represents negative interactions, green represents strong positive interactions, and grey represents weak positive interactions. The inset graph represents non-interacting pairs (black).

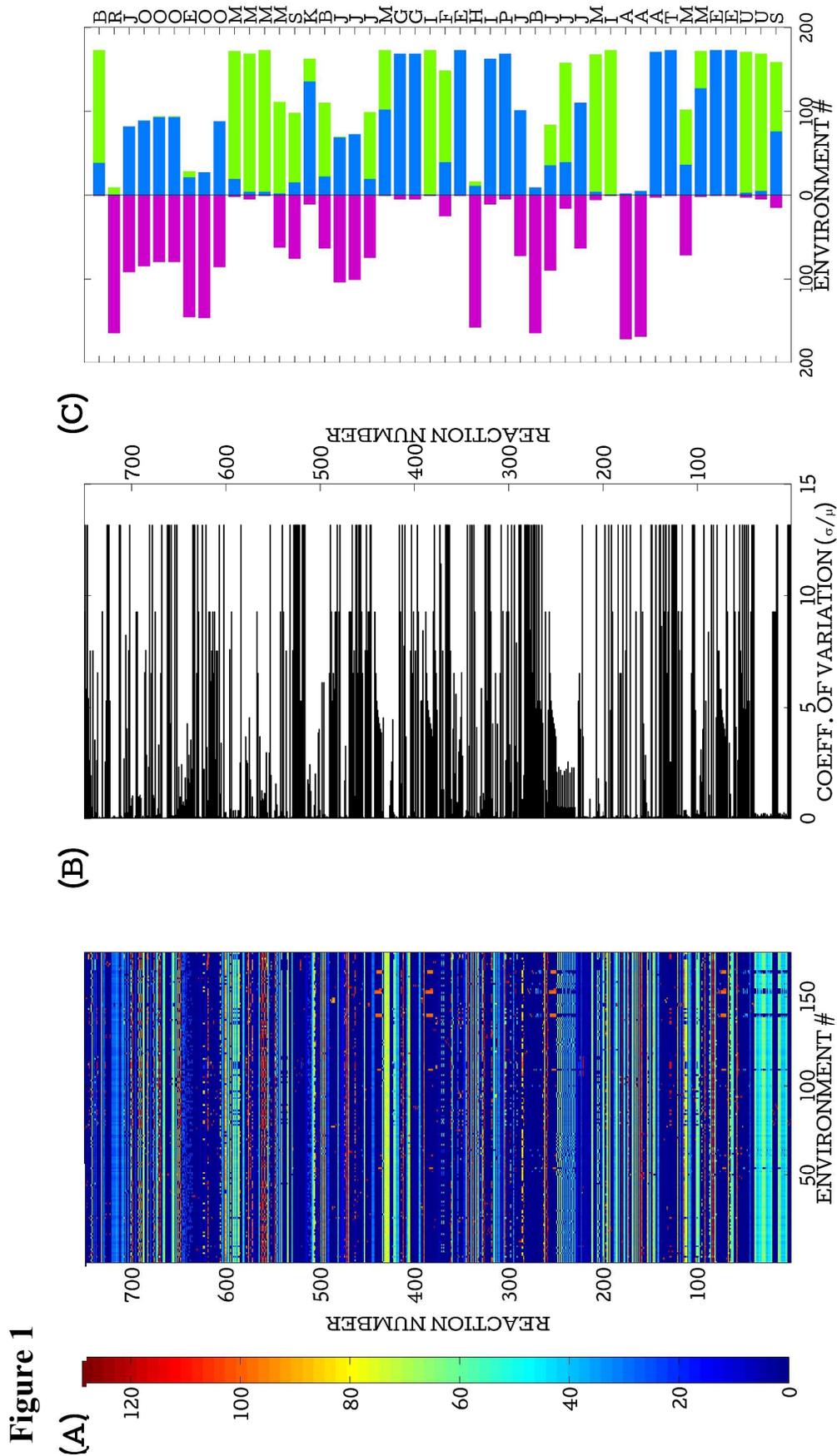


Figure 2

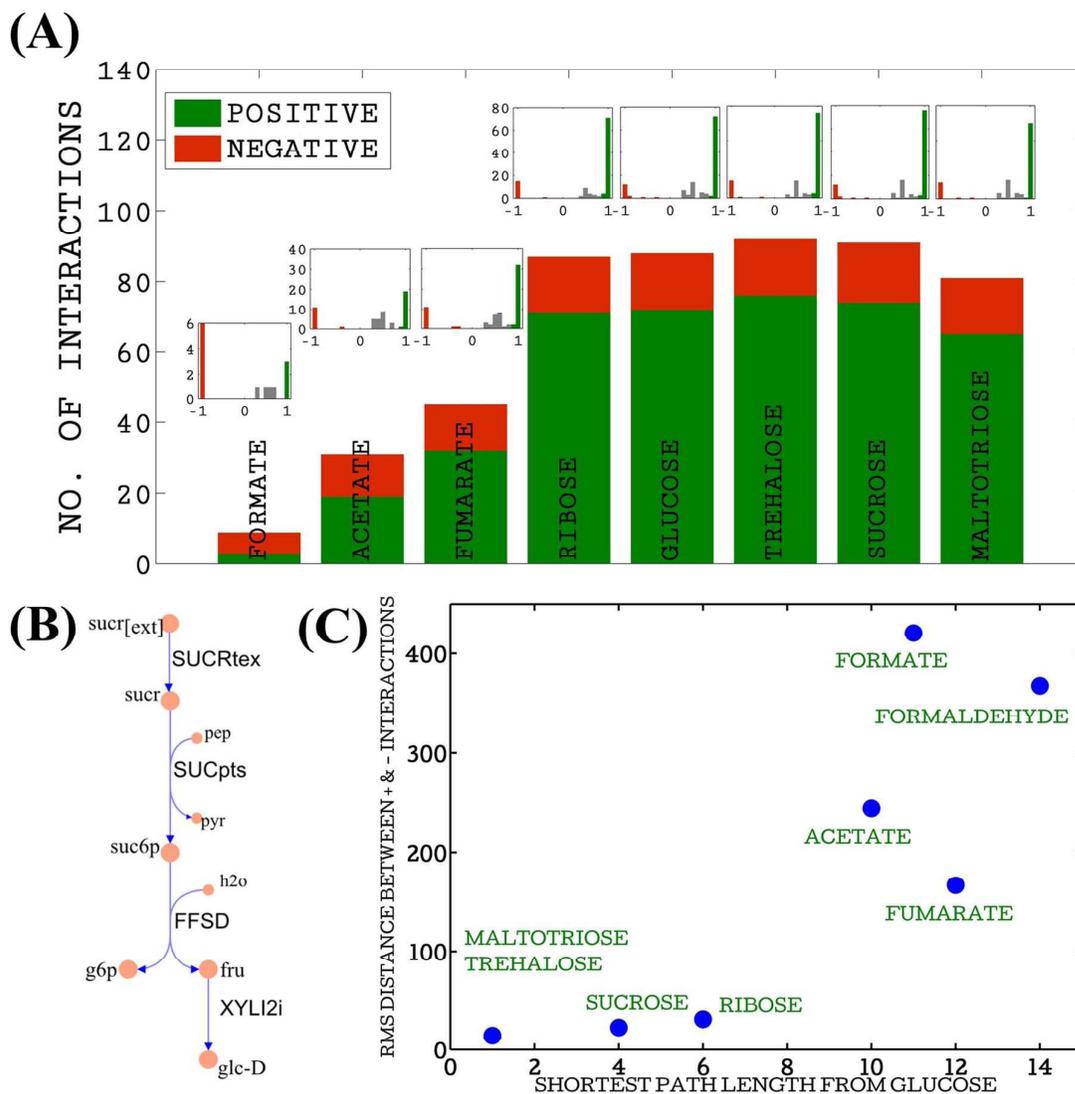


Figure 3

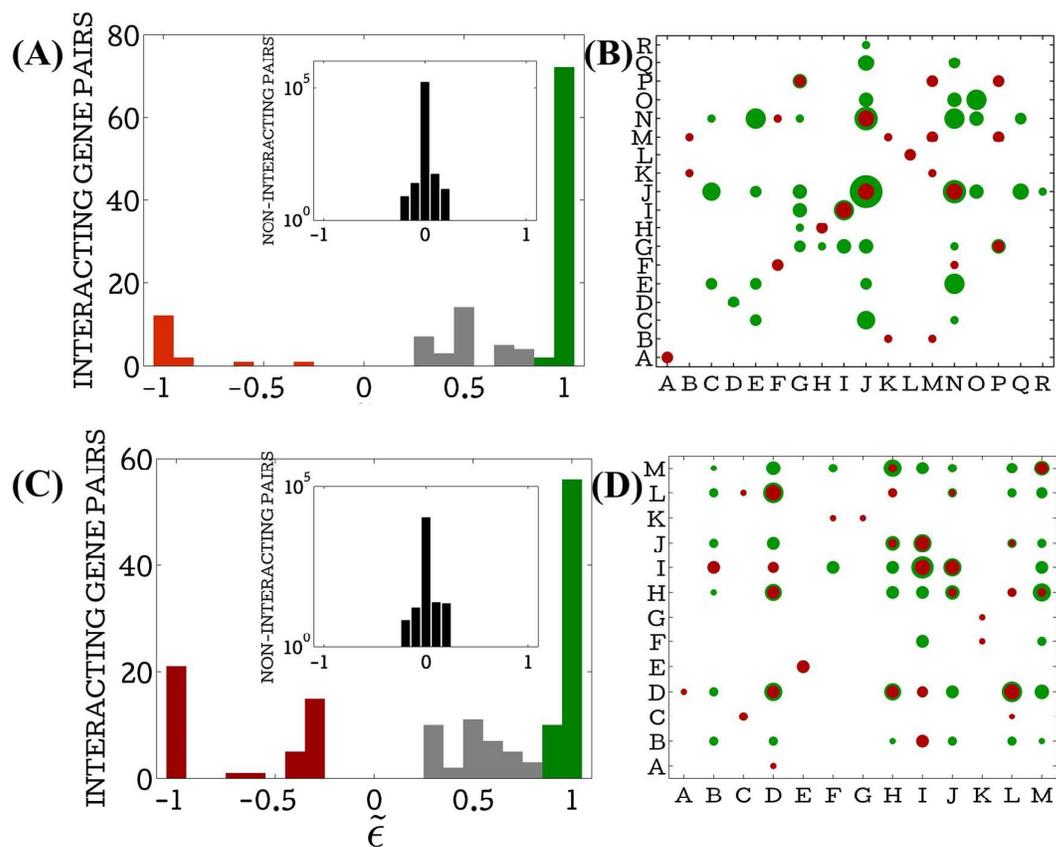


Figure 4

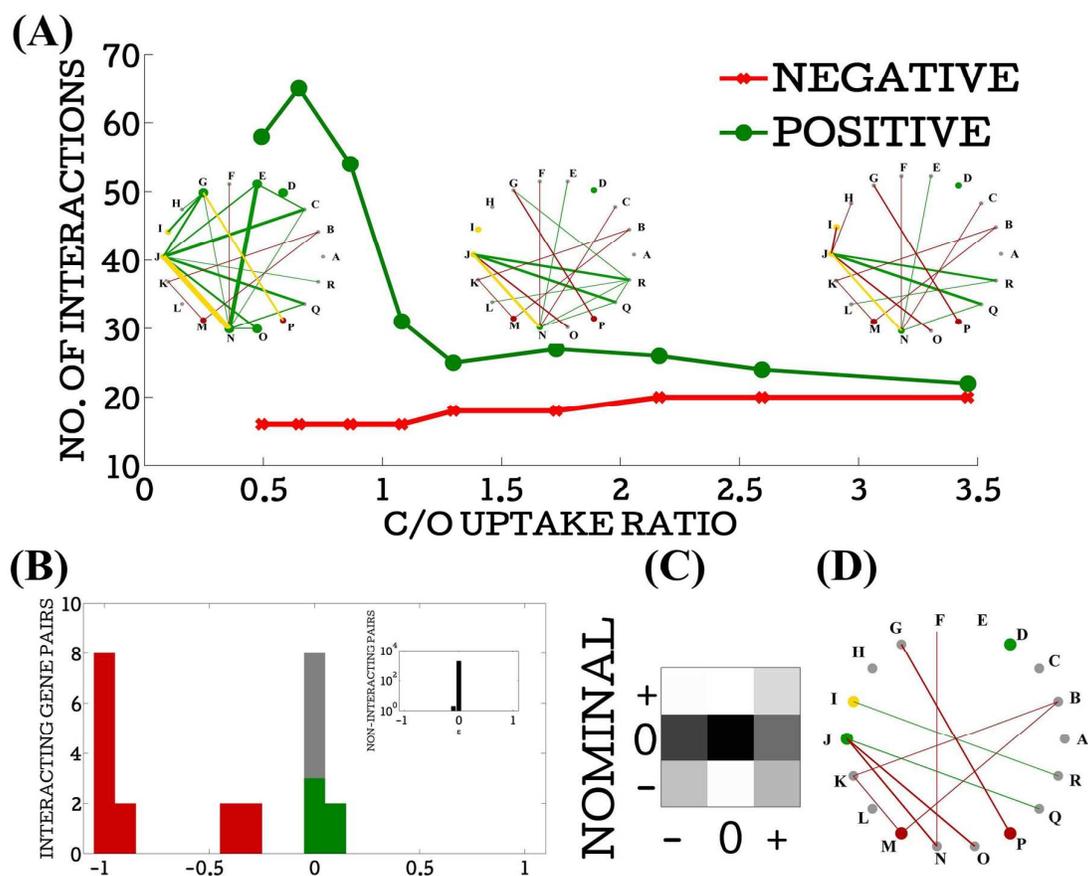


Figure 5

