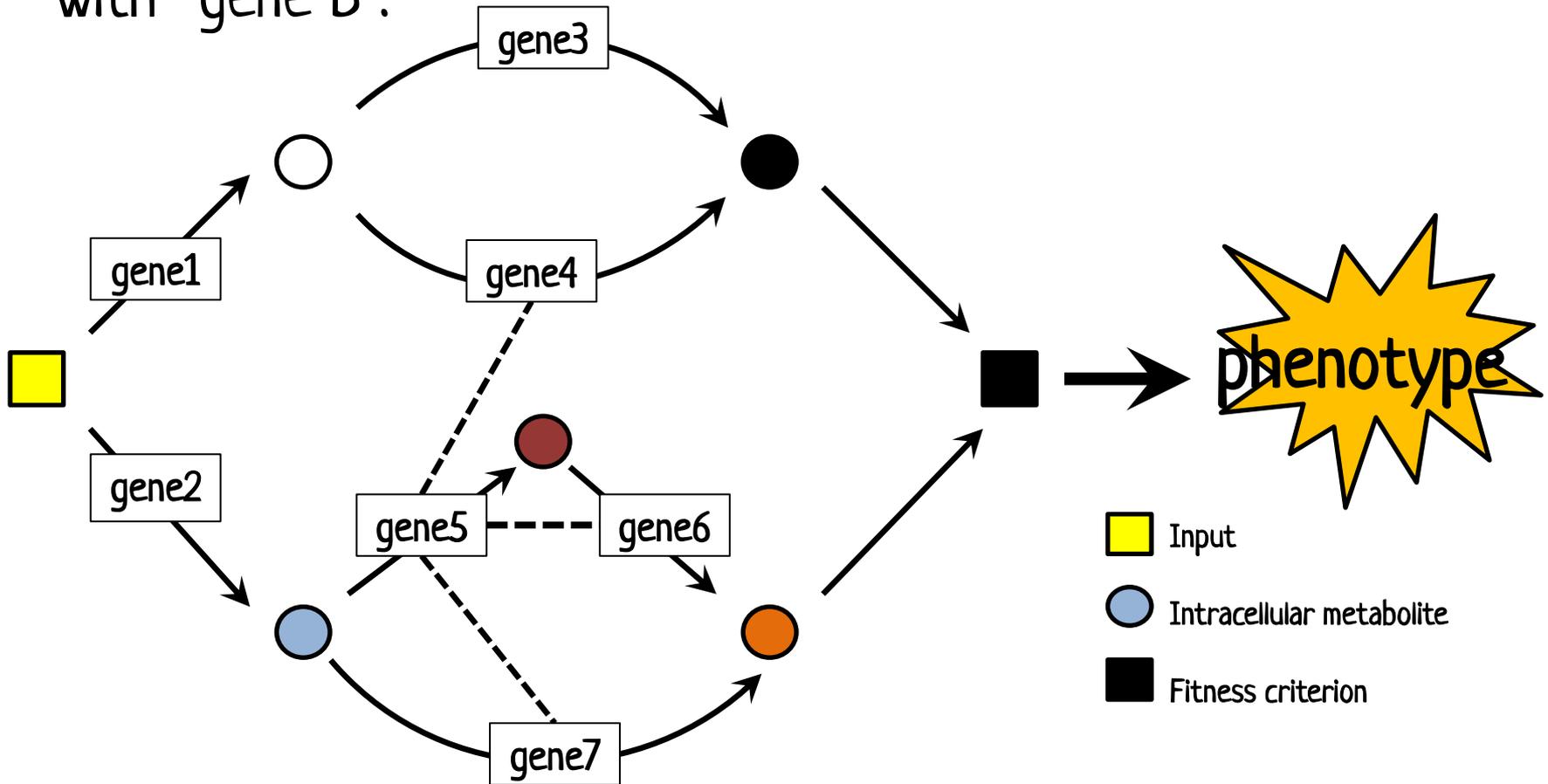


# EPISTASIS & METABOLISM

Presented by  
Chintan Joshi

# What is EPISTASIS - Definition

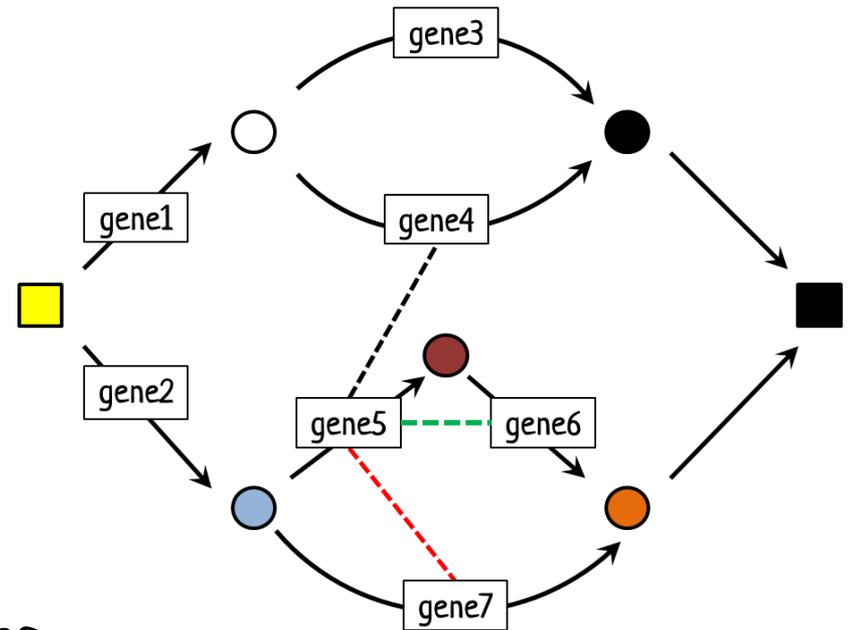
- When the mutant phenotype of “gene A” masks the mutant phenotype of “gene B”, then “gene A” is said to be EPISTATIC with “gene B”.



# What is EPISTASIS - Causes: Metabolism

- Epistasis is the phenomenon of interaction between two genetic elements by causing the growth to be affected synergistically.

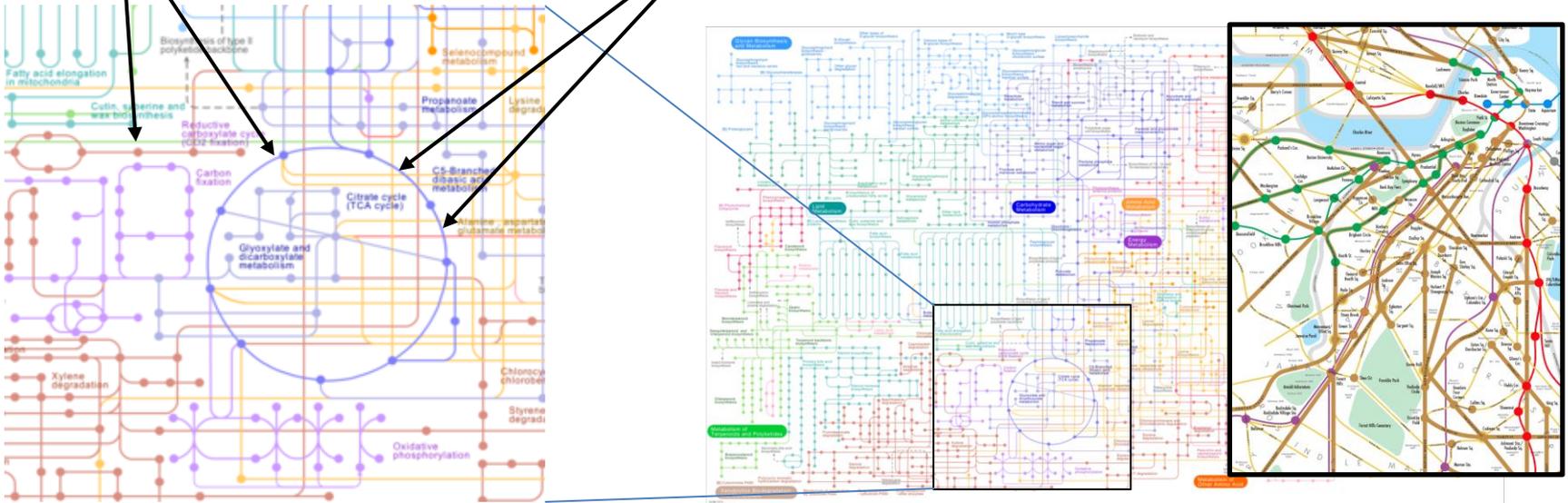
- In context of growth fitness:
  - **Synergistic(-):**
    - sharing same metabolite
    - amongst beneficial mutations
  - **Diminishing returns(+):**
    - linear pathways
    - Amongst deleterious mutations



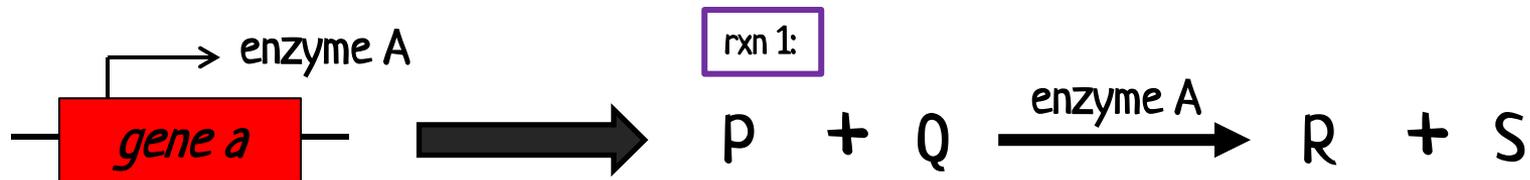
- Epistasis result from alleviating affects of physiological homeostasis (i.e. internal environment - concentration, pH etc.)?

# Systems View of Metabolic Systems

- Nodes (N) - Metabolites & Edges (E) - Reactions  Stoichiometry



- Basis for all forms of metabolic network modeling



# Flux Balance Analysis

Applying mass-balance:

$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$

$$\frac{dB}{dt} = v_1 + v_4 - b_2$$

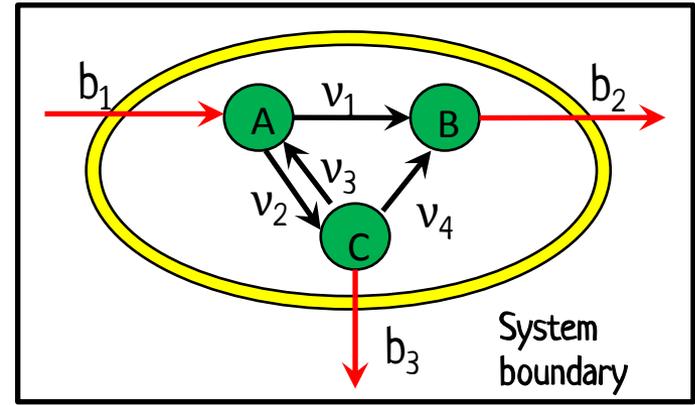
$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix}$$

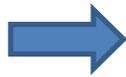
$$= \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$\leftarrow S \rightarrow$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \mathbf{v}$$



$$\frac{dz}{dt} = S \cdot v(z, t)$$



Infinite solutions

# Flux Balance Analysis

Applying mass-balance:

$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$

$$\frac{dB}{dt} = v_1 + v_4 - b_2$$

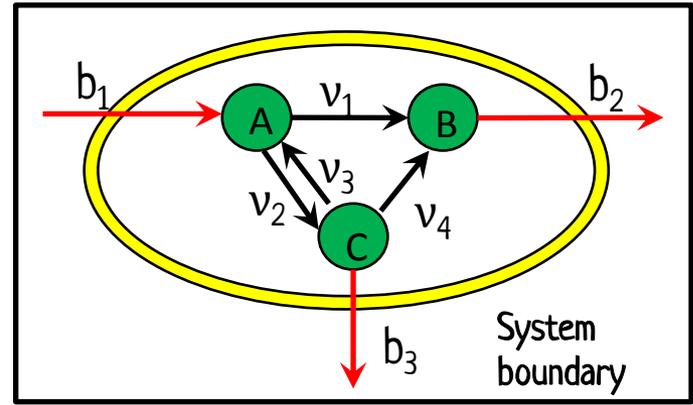
$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

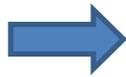
$S$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \mathbf{v}$$



At steady state:

$$\frac{dz}{dt} = S \cdot v(z, t)$$



Infinite solutions

# Flux Balance Analysis

Applying mass-balance:

$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$

$$\frac{dB}{dt} = v_1 + v_4 - b_2$$

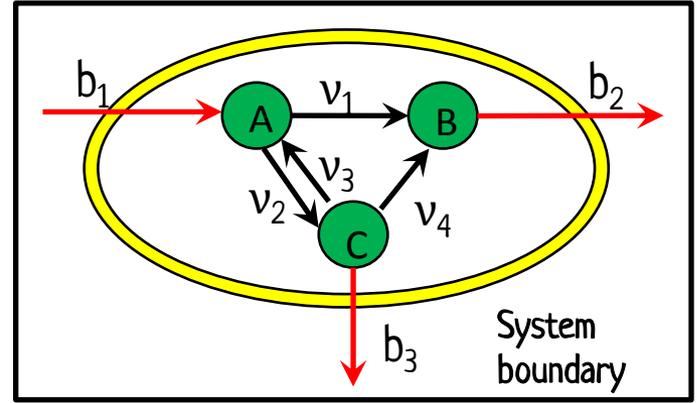
$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

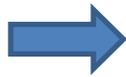
$$\leftarrow S \rightarrow$$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \mathbf{v}$$



At steady state:

$$\frac{dz}{dt} = S \cdot v(z, t)$$



Infinite solutions

$$\max(\alpha \cdot v_1 + \beta \cdot v_4)$$

$$0 \leq v_1, v_2, v_3, v_4 \leq v_{\max}$$

$$0 \leq b_2, b_3 \leq b_{\max}$$

Gene deletion can be simulated by adding more constraints.

# Flux Balance Analysis

Applying mass-balance:

$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$

$$\frac{dB}{dt} = v_1 + v_4 - b_2$$

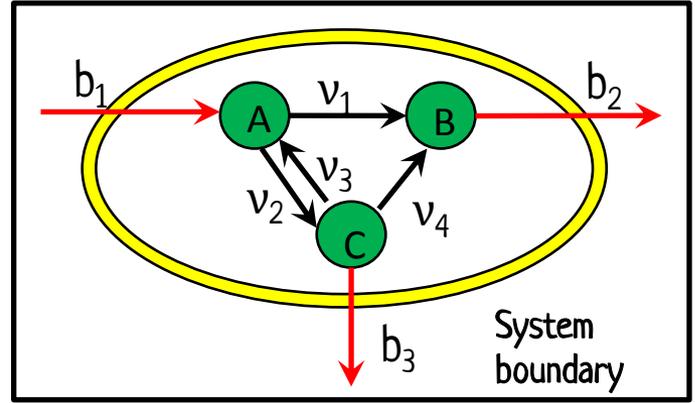
$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$S$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \mathbf{v}$$



At steady state:

$$\frac{dz}{dt} = S \cdot v(z, t)$$



Infinite solutions

$$\begin{aligned} &\max(\alpha \cdot v_1 + \beta \cdot v_4) \\ &0 \leq v_1, v_2, v_3, v_4 \leq v_{\max} \\ &0 \leq b_2, b_3 \leq b_{\max} \end{aligned}$$

Generalized form:

$$f = \max_{v(t)} \sum \alpha_i \cdot v_i(t)$$

$$\frac{dz}{dt} = 0 \Rightarrow S \cdot v = 0$$

$$v_{j, \min} \leq v_j \leq v_{j, \max}$$



**FBA (def):**  
Constraint based modeling to simulate flow of metabolite

Gene deletion can be simulated by adding more constraints.

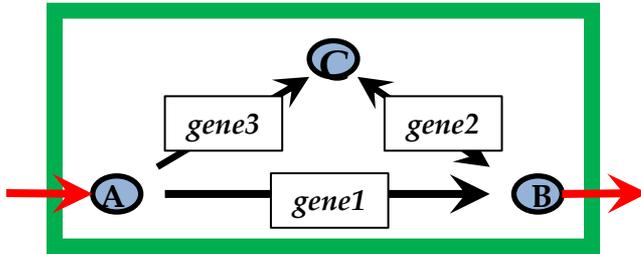


$$v_k = 0$$

# EPISTASIS & METABOLISM

- When the mutant phenotype of “gene A” masks the mutant phenotype of “gene B”, then “gene A” is said to be EPISTATIC with “gene B”.
- The dependence of mutant phenotype of “gene B” on mutant phenotype “gene A”, metabolically, is a result of functional association between “gene A” and “gene B”.
- In metabolism, functionality of a gene is characterized by three important components of metabolic network:
  - Enzyme expressed by these genes, and
  - Reactions catalyzed by these enzymes
  - Phenotype – fluxes carried through these reactions

# EPISTASIS (basic understanding)



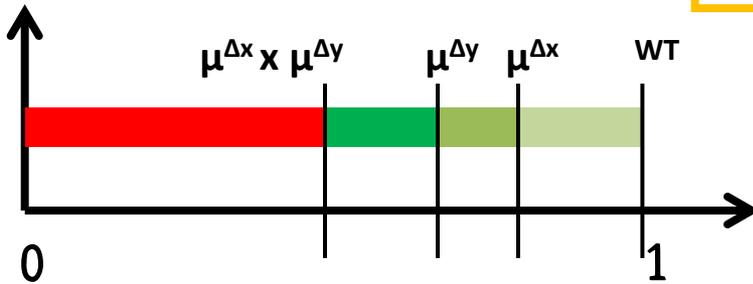
$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$f = \max_{\mathcal{G}(t)} \alpha_i \cdot \mathcal{G}_i(t)$$

$$\frac{dz}{dt} = 0 \Rightarrow S \cdot \mathcal{G} = 0$$

$$\mathcal{G}_{j, \min} \leq \mathcal{G}_j \leq \mathcal{G}_{j, \max}$$

$$\frac{dz}{dt} = S \cdot \mathcal{G}(z, t)$$



→ Negative Epistasis  
  
  
 } → Positive Epistasis

FBA provides a qualitative framework to visualize the functional relationship between various subsystems relative to growth rate.

# QUANTIFYING EPISTASIS

**NOTE:** We classify gene interactions as **AGGRAVATING** (negative) or **BUFFERING** (positive).

## Quantifying Epistasis:

(Segre et al., Nature 2005)

$$W_X = \frac{\mu^{X-}}{\mu^{wt}}; \quad W_{XY} = \frac{\mu^{XY-}}{\mu^{wt}}$$

$$\mathcal{E} = \frac{W_{XY} - W_X W_Y}{\tilde{W}_{XY} - W_X W_Y},$$

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

$\mu^{wt}$ ,  $\mu^{X-}$ , and  $\mu^{XY-}$  are growth rates of wild-type, mutant in gene X, and mutant in genes X and Y; **found using FBA.**

$W_X$  is fitness value of a single mutant in gene X

$W_{XY}$  is fitness value of double mutant in gene X and gene Y

$\mathcal{E}$  is the epistasis for the interaction between gene X and gene Y.

For subsequent figures of scaled epistasis and interactome:

**Red:** negative interaction

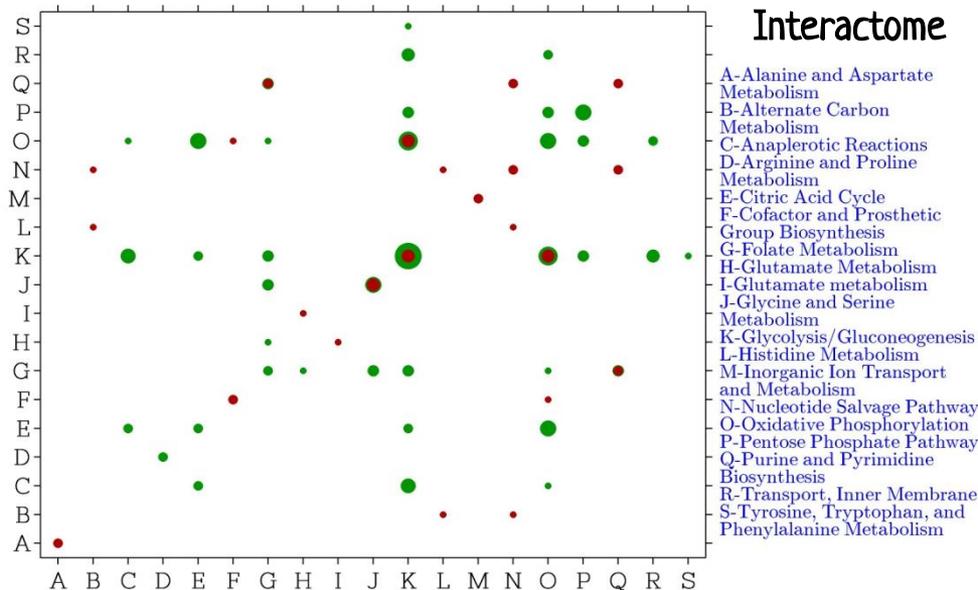
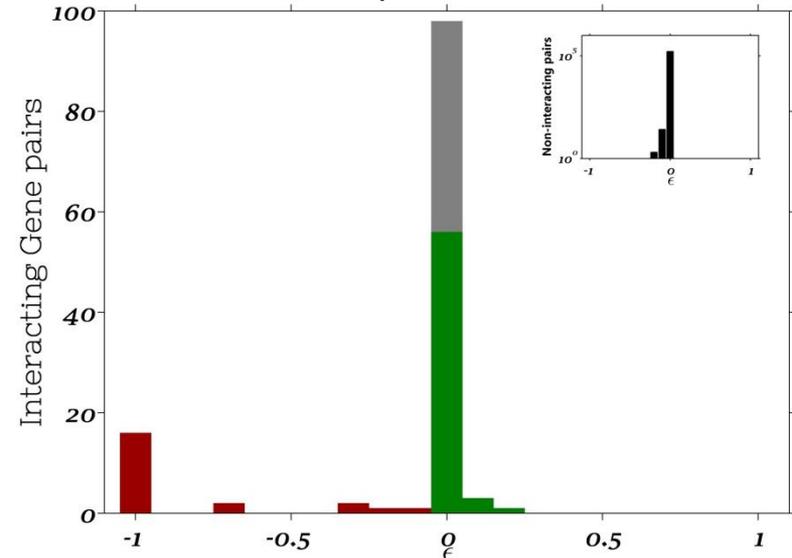
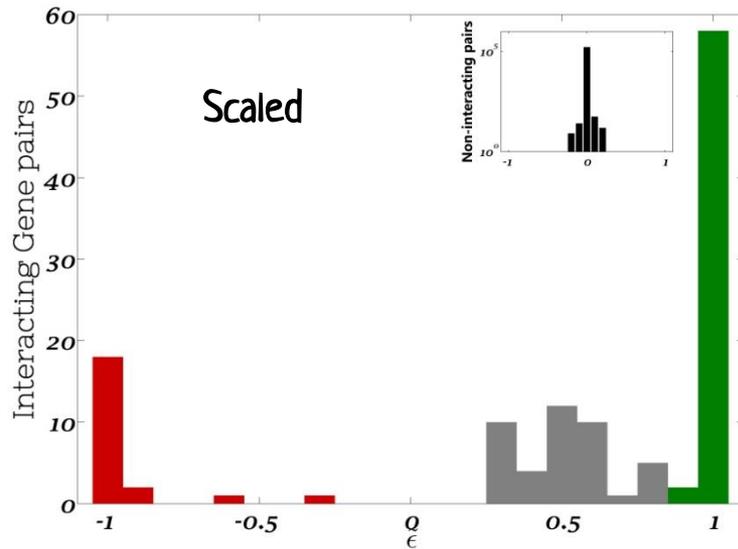
**Green:** positive interaction

We, first performed a single gene deletion and removed all the genes which result in a lethal phenotype. We also removed genes which were unable to constrain any reaction..

## Known so far and our direction...

- Positive epistasis is higher amongst functionally unrelated genes and deleterious mutations
- Dominance of negative epistasis amongst functionally related genes.
- Epistatic interactions are not absolute, but characteristic of gene-gene-phenotype triad
- How does environment picture into epistasis?

# *Escherichia coli* (Aerobic w/ Glucose – Scaled and Nonscaled Epistasis)



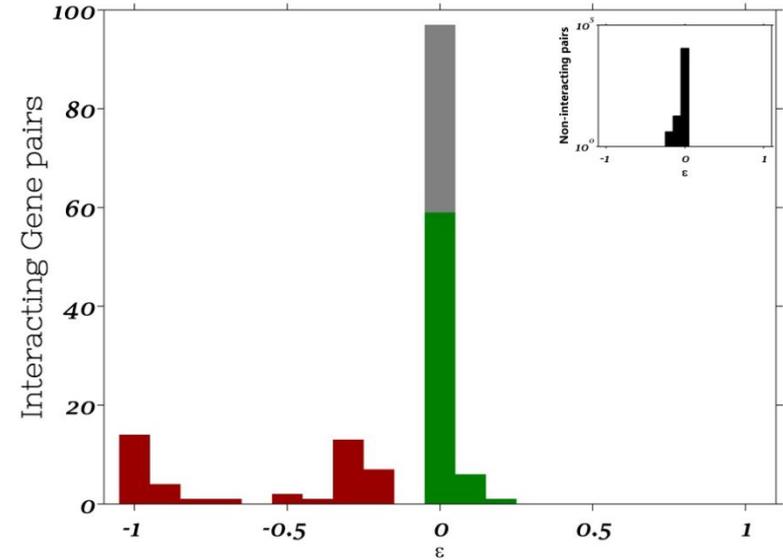
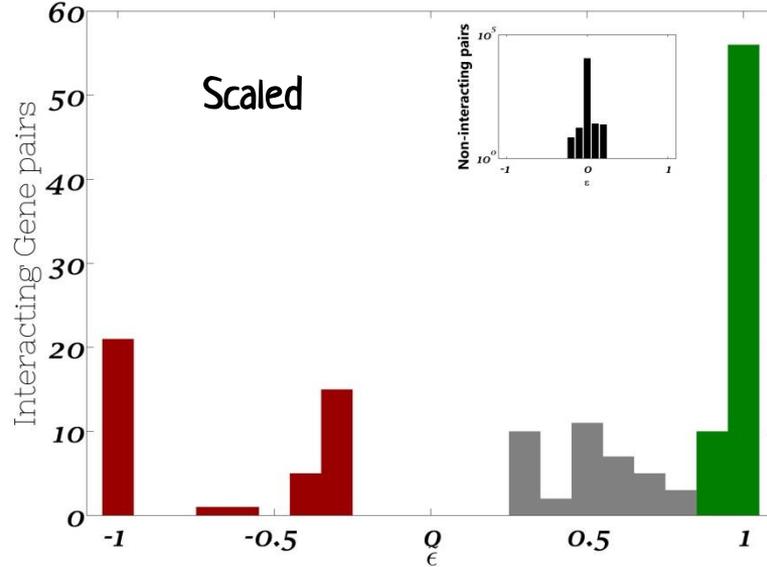
Number of negative interactions = 22  
 Number of positive interactions (blue + green) = 102

~5 fold more positive interactions than negative interactions. Interactome map is dominated by positive interactions.

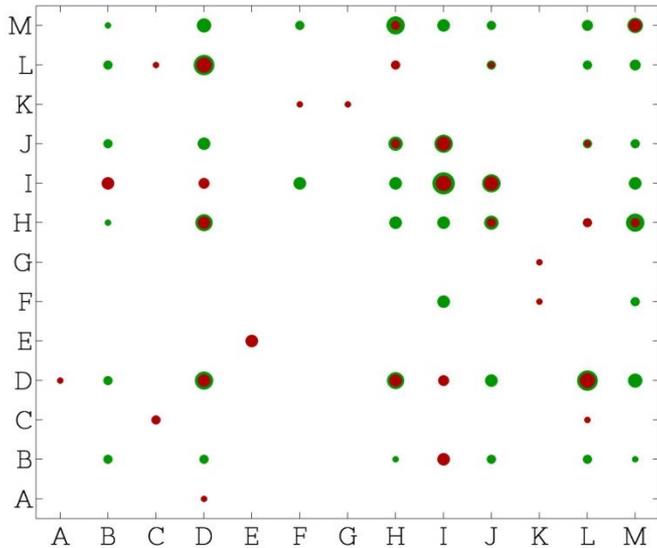
Largest number of interactions occur in Glycolysis/Gluconeogenesis and Oxidative Phosphorylation.

Interestingly, it has fairly large number of monochromatic interactions, compared to mixed interactions (+ and -).

# *Synechocystis* sp. PCC6803 (Aerobic w/ Glucose – Scaled and Nonscaled Epistasis)



## Interactome



- A-Carbon fixation
- B-Citrate cycle (TCA cycle)
- C-Glycerolipid metabolism
- D-Glycolysis/Gluconeogenesis
- E-Glyoxylate and dicarboxylate metabolism
- F-Nitrogen metabolism
- G-Others
- H-Oxidative phosphorylation
- I-Pentose phosphate pathway
- J-Photosynthesis
- K-Pyrimidine metabolism
- L-Pyruvate metabolism
- M-Transport

Number of negative interactions = 43

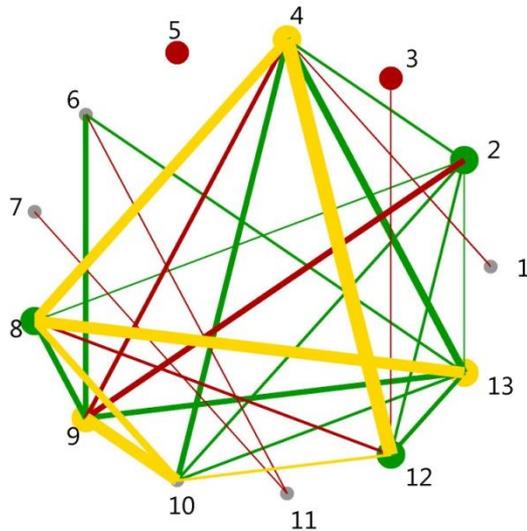
Number of positive interactions (blue + green) = 104

~2.5 fold more positive interactions than negative interactions. Interactome map is dominated by positive interactions.

Largest number of interactions (40) and interactions with largest number of subsystems (8) occur in Glycolysis/Gluconeogenesis. This is similar to *E. coli*.

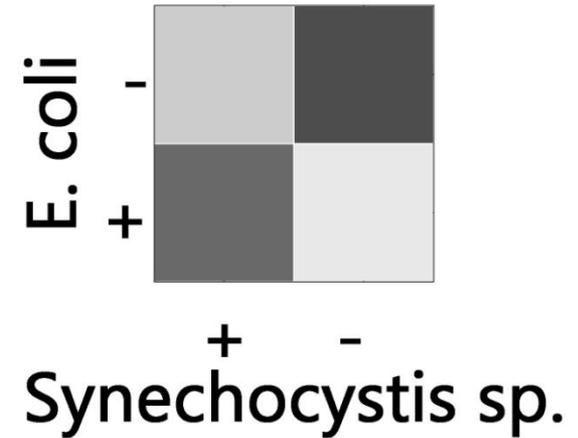
Interestingly, it has fairly large number of monochromatic interactions, compared to mixed interactions (+ and -).

# Differences in Aerobic growth (*Synechocystis* and *E. coli*)



## Synechocystis sp. PCC6803

- 1- Carbon fixation
- 2- Citrate cycle (TCA cycle)
- 3- Glycerolipid metabolism
- 4- Glycolysis/Gluconeogenesis
- 5- Glyoxylate and dicarboxylate metabolism
- 6- Nitrogen Metabolism
- 7- Others
- 8- Oxidative phosphorylation
- 9- Pentose phosphate pathway
- 10- Photosynthesis
- 11- Pyrimidine metabolism
- 12- Pyruvate metabolism
- 13- Transport

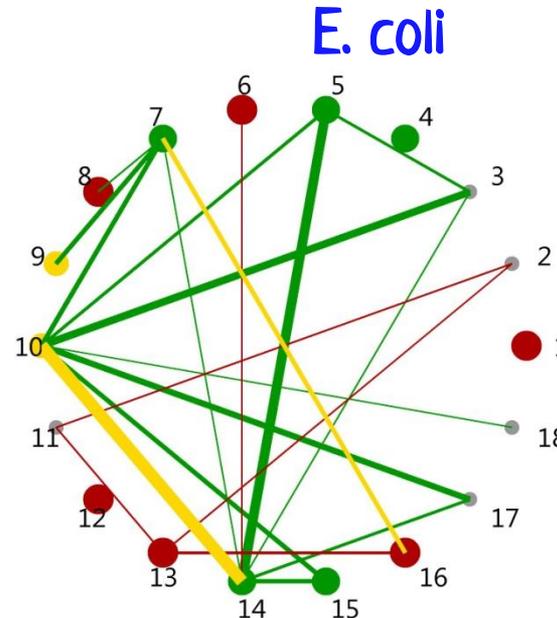


### From scaled:

Interaction amongst 49 out of 93 genes both common to *E. coli* and *Synechocystis* had identical interactions.

### From non-scaled:

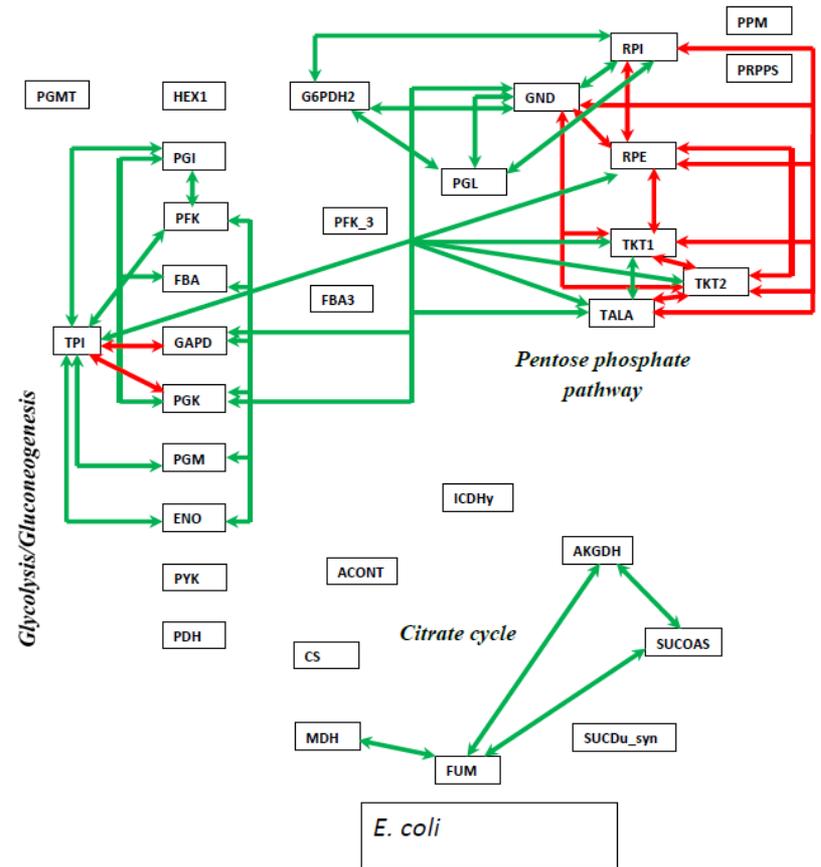
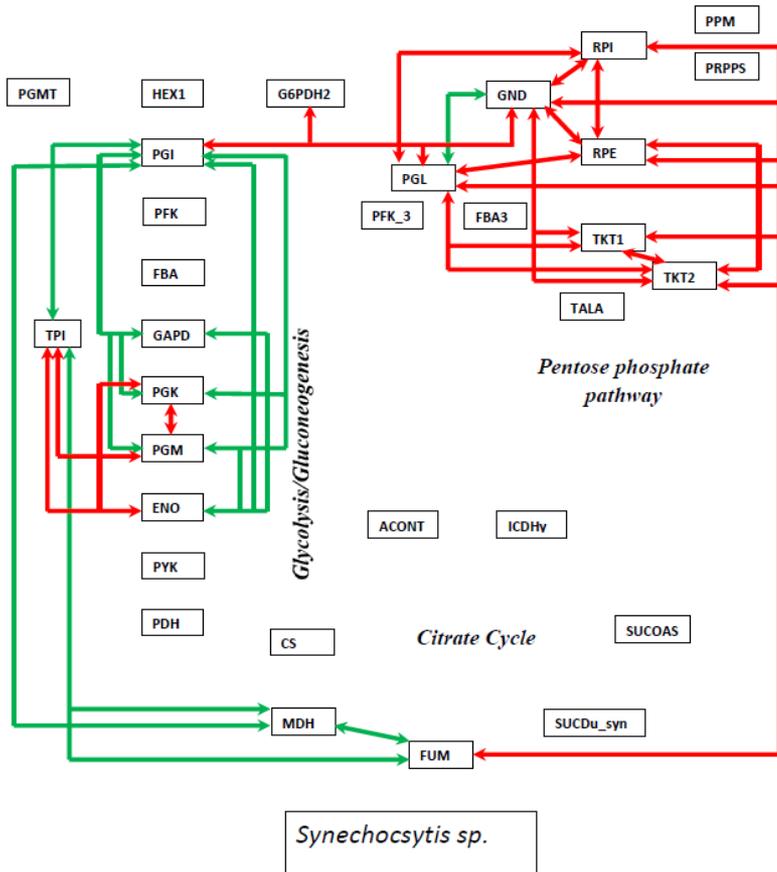
~80% of the interactions matched in both organisms.



## E. coli

- 1- Alanine and Aspartate Metabolism
- 2- Alternate Carbon Metabolism
- 3- Anaplerotic Reactions
- 4- Arginine and Proline Metabolism
- 5- Citric Acid Cycle
- 6- Cofactor and Prosthetic Group Biosynthesis
- 7- Folate Metabolism
- 8- Glutamate Metabolism
- 9- Glycine and Serine Metabolism
- 10- Glycolysis/Gluconeogenesis
- 11- Histidine Metabolism
- 12- Inorganic Ion Transport and Metabolism
- 13- Nucleotide Salvage Pathway
- 14- Oxidative Phosphorylation
- 15- Pentose Phosphate Pathway
- 16- Purine and Pyrimidine Biosynthesis
- 17- Transport, Inner Membrane
- 18- Tyrosine, Tryptophan, and Phenylalanine Metabolism

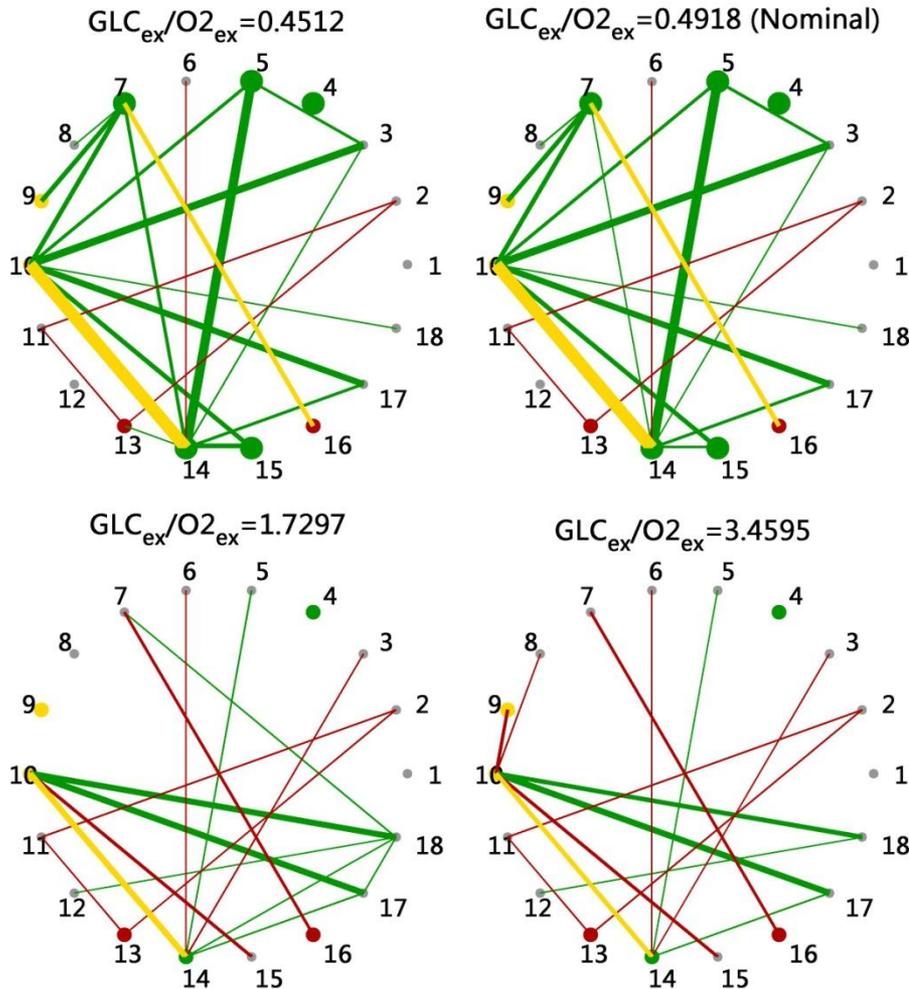
# Differences in Aerobic growth (*Synechocystis* and *E. coli*)



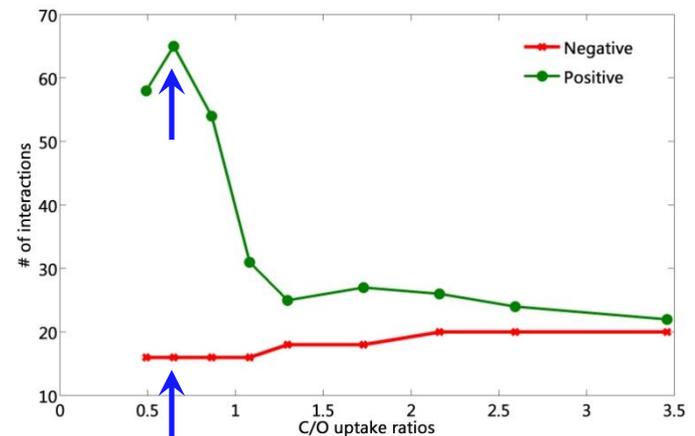
Matches are a result of network topology shared amongst Glycolysis, TCA cycle, Pentose phosphate pathway and Oxidative phosphorylation.

Scaling factors for the two organisms maybe different.

# Escherichia coli (Aerobic w/ Glucose – Disappearance of positive interactions)

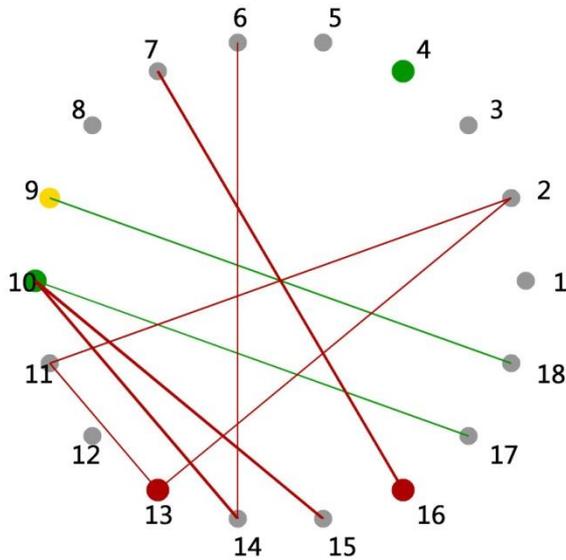


- 2. Alternate Carbon Metabolism (6)
- 3. Anaplerotic Reactions (1)
- 4. Arginine and Proline Metabolism (2)
- 5. Citric Acid Cycle (4)
- 6. Cofactor and Prosthetic Group Biosynthesis (6)
- 7. Folate Metabolism (2)
- 8. Glutamate Metabolism (1)
- 9. Glycine and Serine Metabolism (3)
- 10. Glycolysis/Gluconeogenesis (9)
- 11. Histidine Metabolism (1)
- 12. Membrane Lipid Metabolism (1)
- 13. Nucleotide Salvage Pathway (6)
- 14. Oxidative Phosphorylation (6)
- 15. Pentose Phosphate Pathway (3)
- 16. Purine and Pyrimidine Biosynthesis (3)
- 17. Transport (6)
- 18. Tyrosine, Tryptophan, and Phenylalanine Metabolism (1)

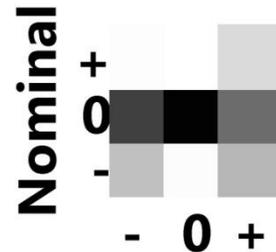


Disappearance of positive interactions – INTERACTIONS DO NOT REMAIN SAME!

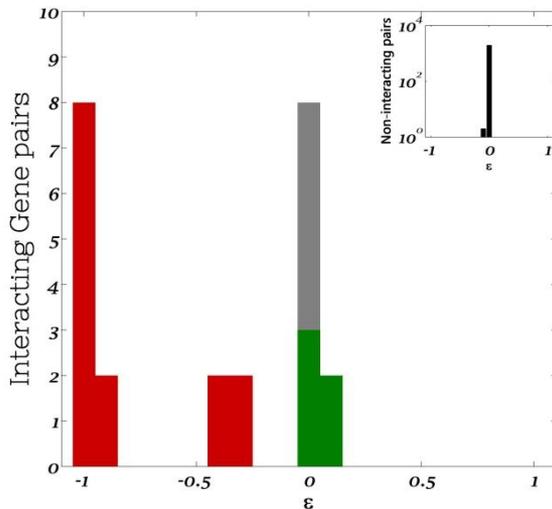
# *Escherichia coli* (Anaerobic w/ Glucose – Disappearance of positive interactions)



Anaerobic glucose=325.0062



- 2. Alternate Carbon Metabolism (6)
- 3. Anaplerotic Reactions (1)
- 4. Arginine and Proline Metabolism (2)
- 5. Citric Acid Cycle (4)
- 6. Cofactor and Prosthetic Group Biosynthesis (6)
- 7. Folate Metabolism (2)
- 8. Glutamate Metabolism (1)
- 9. Glycine and Serine Metabolism (3)
- 10. Glycolysis/Gluconeogenesis (9)
- 11. Histidine Metabolism (1)
- 12. Membrane Lipid Metabolism (1)
- 13. Nucleotide Salvage Pathway (6)
- 14. Oxidative Phosphorylation (6)
- 15. Pentose Phosphate Pathway (3)
- 16. Purine and Pyrimidine Biosynthesis (3)
- 18. Tyrosine, Tryptophan, and Phenylalanine Metabolism (1)

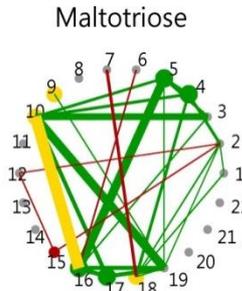
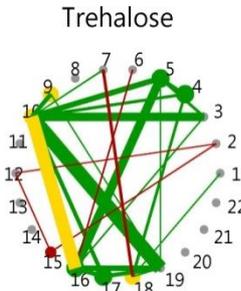
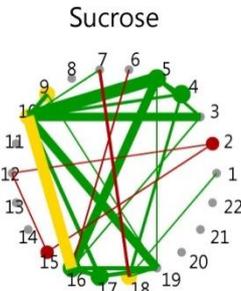
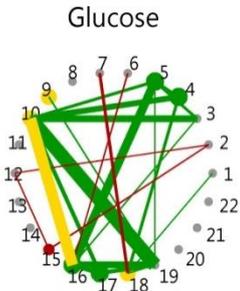
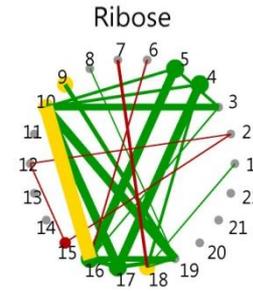
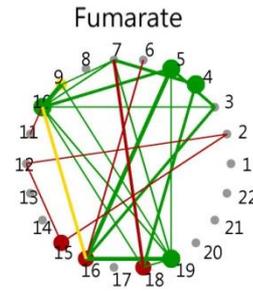
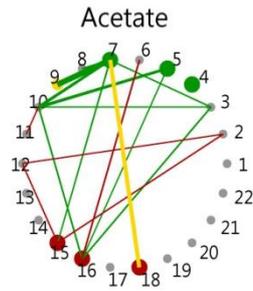
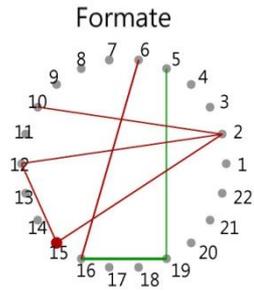


Negative interactions = 14

Positive Interactions = 5

This indicates that under non-optimal nutrient uptakes, negative interactions increase and positive interactions decrease.

# Different carbon sources result in different interactions



1. Alanine and Aspartate Metabolism
2. **Alternate Carbon Metabolism**
3. Anaplerotic Reactions
4. Arginine and Proline Metabolism
5. Citric Acid Cycle
6. **Cofactor and Prosthetic Group Biosynthesis**
7. Folate Metabolism
8. Glutamate Metabolism
9. Glycine and Serine Metabolism
10. Glycolysis/Gluconeogenesis
11. Glyoxylate Metabolism
12. **Histidine Metabolism**
13. Inorganic Ion Transport and Metabolism
14. Membrane Lipid Metabolism
15. **Nucleotide Salvage Pathway**
16. **Oxidative Phosphorylation**
17. Pentose Phosphate Pathway
18. Purine and Pyrimidine Biosynthesis
19. Transport, Inner Membrane

PRPPS: PRPP + AMP  $\rightarrow$  R5P + ATP

ADK1: ATP + AMP  $\rightarrow$  2 ADP

R15PK: ATP + R15BP  $\rightarrow$  ADP + PRPP

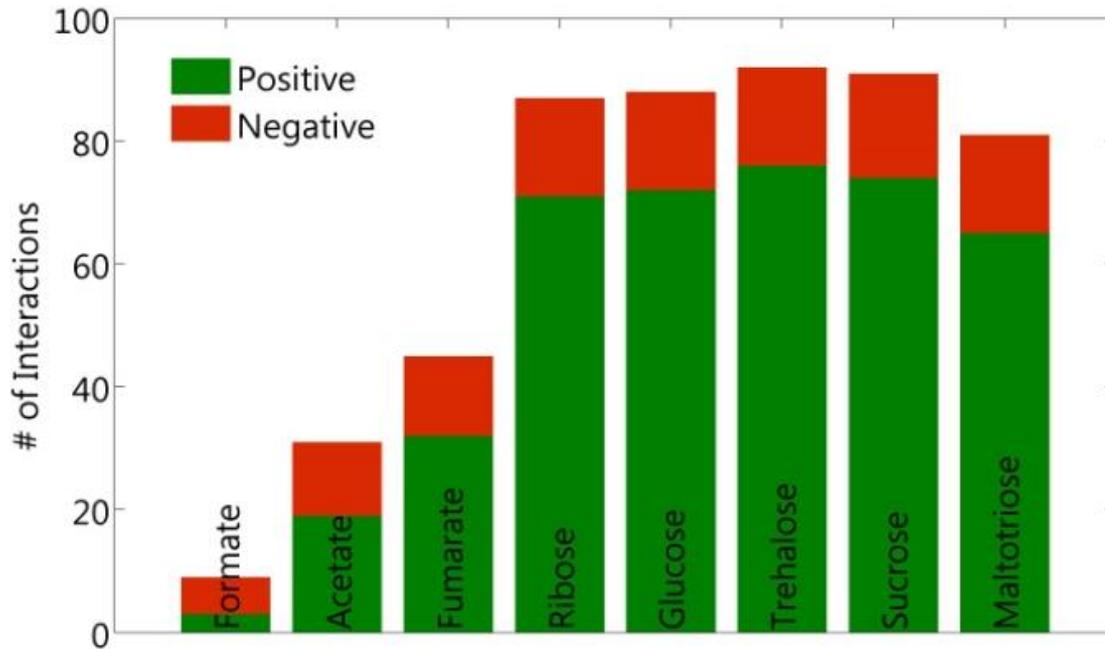
RPM: R5P  $\rightarrow$  R1P

R1PK: R1P  $\rightarrow$  R15BP

**URIDK2r : CYTK1/2 (15),  
ADK : PRPPS (15-12),  
ADK : R15BPK (15-2),  
GTHOr : TRDR (6-16),  
PRPPS : R15BPK (12-15)**

All of these are a result of synthetic lethals!

# *E. coli*: Interactions differ as the growth condition changes



A path is a sequence of reactions which leads to the conversion of one metabolite to another.

**Acetate – Formate: 8**  
**Acetate – Fumarate: 6**  
**Acetate – Ribose: 10**  
**Acetate – Glucose: 11**  
**Fumarate – Ribose: 11**

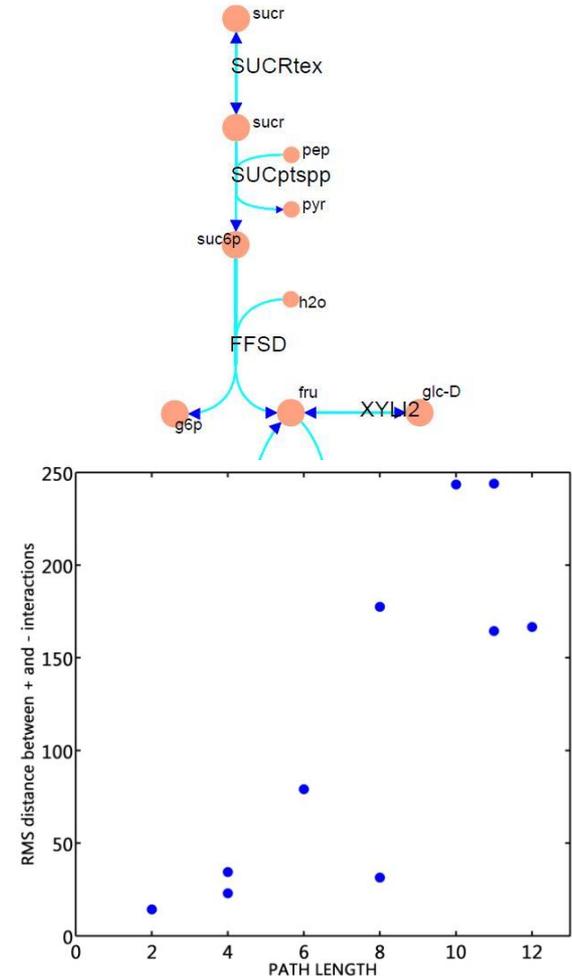
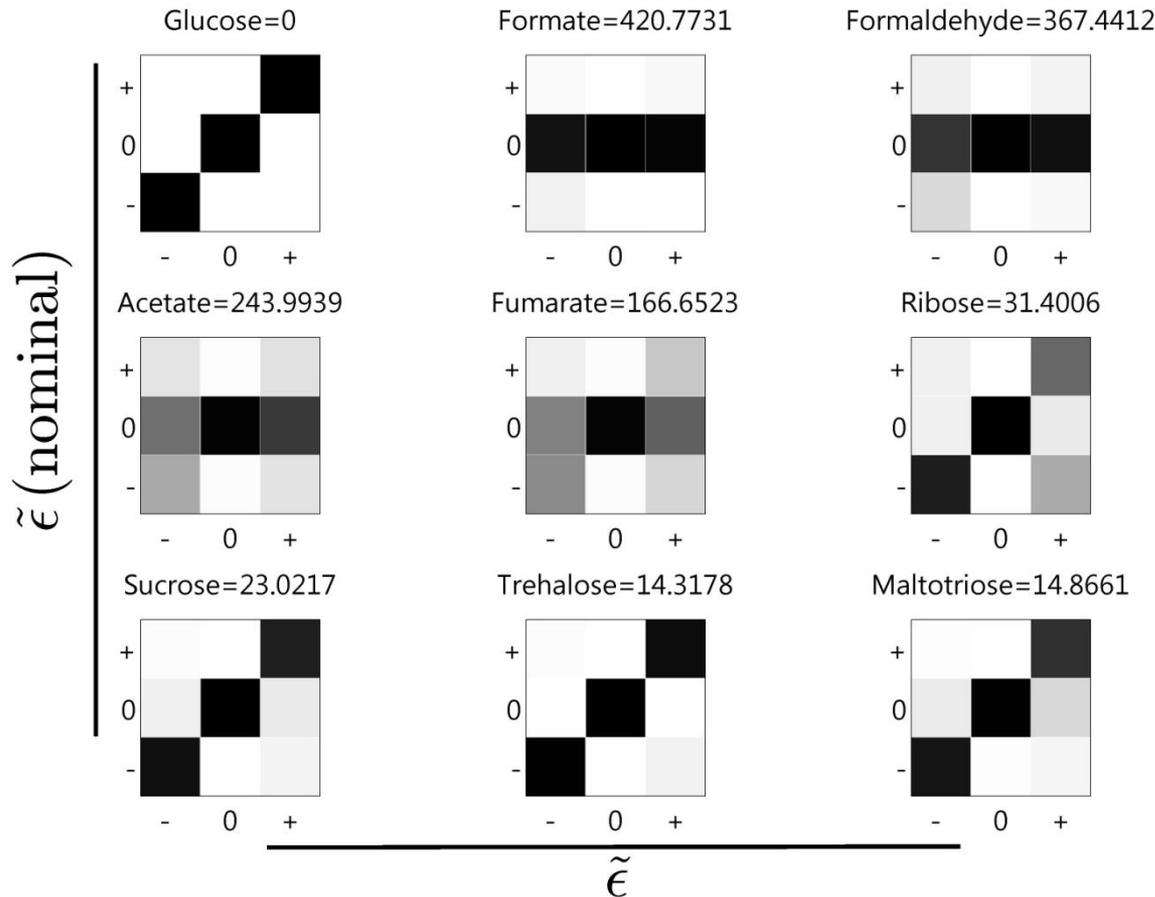
**Fumarate – Glucose: 12**  
**Ribose – Glucose: 8**  
**Sucrose – Glucose: 4**  
**Glucose – Trehalose: 2**  
**Trehalose – Sucrose: 4**

Path length  
(manually)

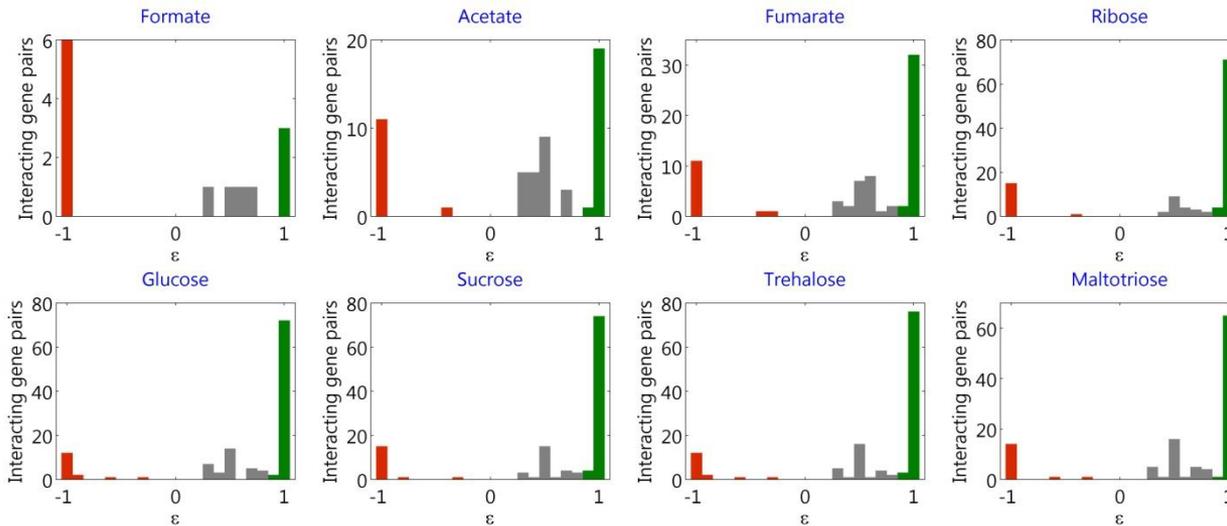
**Hypothesis: The difference between number of interactions (positive interactions) is directly proportional to the path length between two carbon sources.**

# *E. coli*: Interactions differ as the growth condition changes

**Hypothesis:** The difference between number of interactions (positive interactions) is directly proportional to the path length between two carbon sources.



# *E. coli*: Lower the carbon atoms higher the interaction



Formate is converted to CO<sub>2</sub> before carbon is assimilated into higher carbon compounds.

Formate and Formaldehyde both had higher negative interactions than positive interactions.

**Formate – 1**

**Formaldehyde – 1**

**Acetate – 2**

**Fumarate – 4**

**Ribose – 5**

**Glucose – 6**

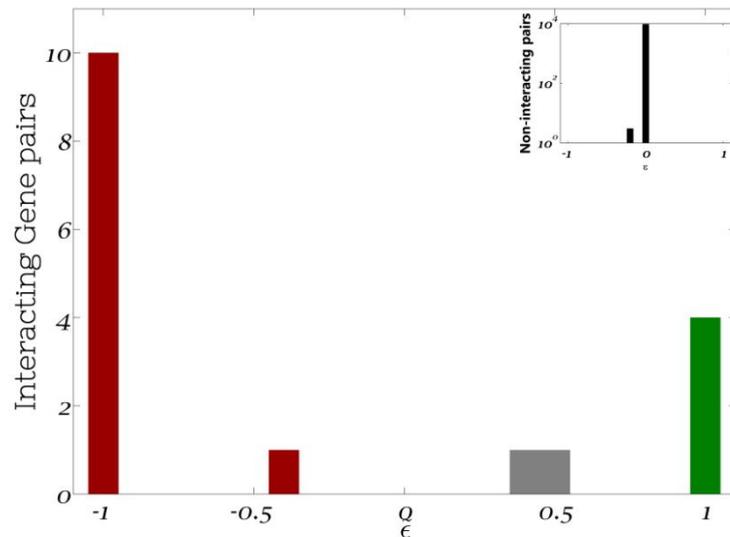
**Sucrose – 12**

**Trehalose – 12**

**Maltotriose – 18**

**Hypothesis: Smaller the uptake molecule, larger is the portion of the negative epistatic interactions.**

# *Synechocystis* sp. PCC6803 (Constrained Autotrophic)



Under CONSTRAINED light, set to WT photon flux:

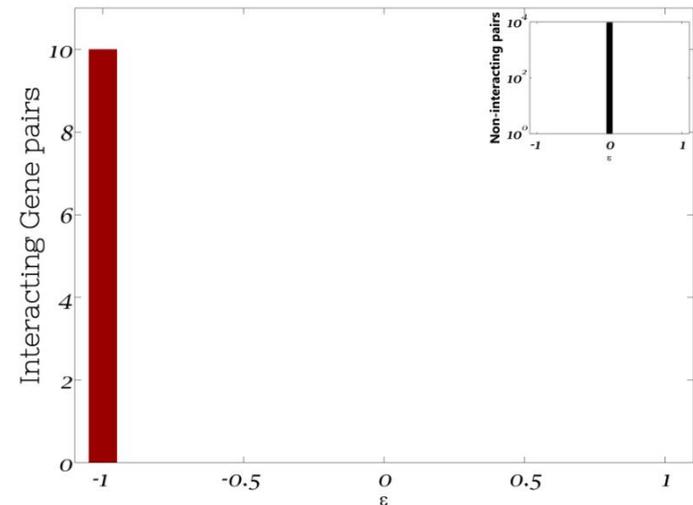
Number of negative interactions = 22

Number of positive interactions (blue + green) = 18

Comparable number of positive and negative interactions. Interactome map is more sparse in negative interactions than in positive interactions.

Citrate cycle is the only subsystem with positive interactions.

Interestingly, there are no subsystem pairs with both types of interactions. This could be the sign of modularity, or a result of model detailedness.



Under UNCONSTRAINED light, set to maximum photon flux available:

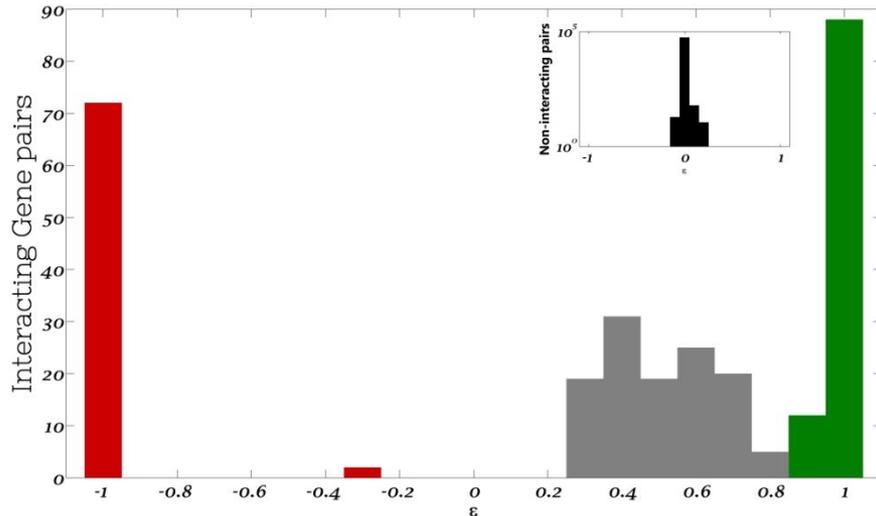
Number of negative interactions = 10

Number of positive interactions (blue + green) = 0

There are no positive interactions. We suspect this is the result of unlimited light, and ability of the network to maintain the ratio of light uptake to CO<sub>2</sub> uptake.

**In autotrophic systems, negative interactions are dominating!**

# *Chlamydomonas reinhardtii* (Negative interactions are dominating)



Under CONSTRAINED light, set to WT photon flux:

Number of negative interactions = 74

Number of positive interactions (blue + green) = 219

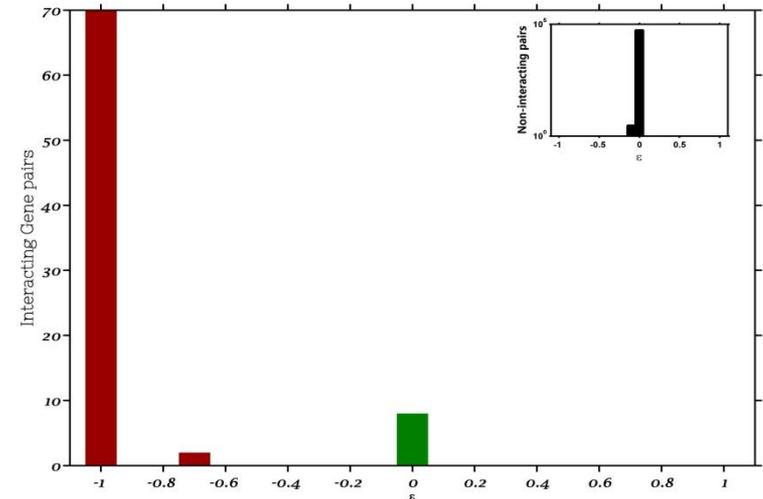
~3 fold more positive interactions than negative interactions.

Interactome map is dominated by positive interactions.

Removing the non-decisive interactions, we have 100 interactions, making it comparable.

Glycolysis/Gluconeogenesis interacts with highest number of subsystems (20), and produces largest number of interactions. (36). We found 20 subsystem pairs which contain mixed interactions spanned over 10 subsystems.

NOTE: Please refer excel file for subsystem codes.



Under UNCONSTRAINED light, set to maximum photon flux available:

Number of negative interactions = 72

Number of positive interactions (blue + green) = 8

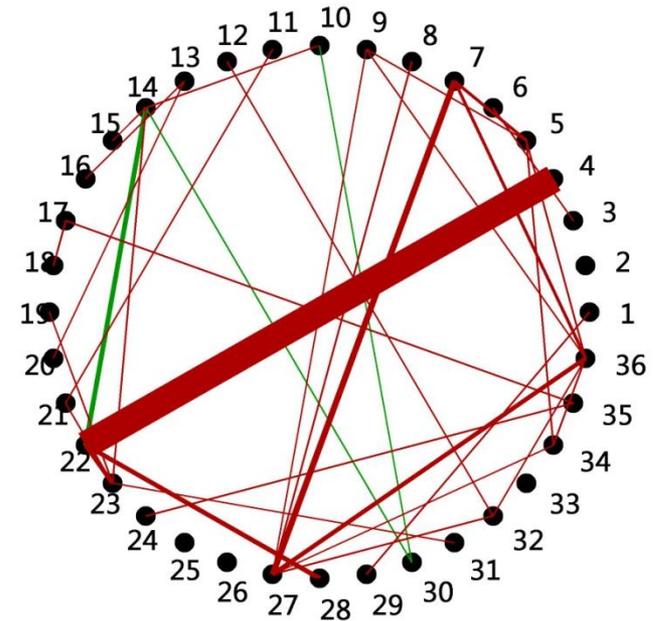
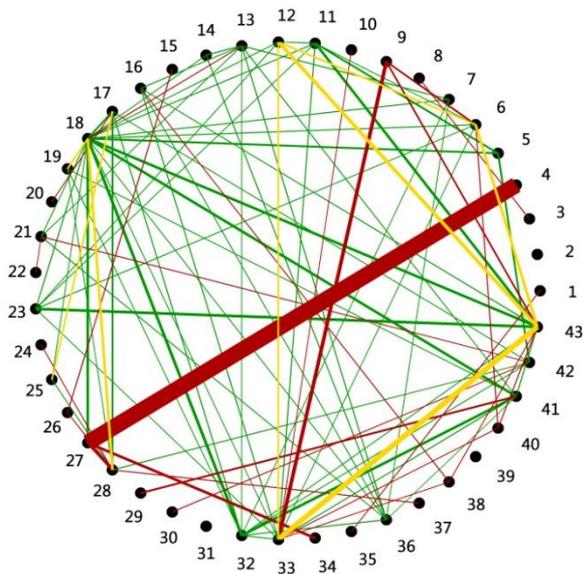
~9 fold more negative interactions than positive interactions.

Interactome map is dominated by negative interactions.

It would be interesting to find out if any of the single mutants can improve their growth rates, if more light was provided. We have used photon flux to be 100 mmoles/gDW/hr.

NOTE: Please refer excel file for subsystem codes.

# *Chlamydomonas reinhardtii* (Unconstrained/Constrained Autotrophic)



Hypothesis:

Positive epistasis: exists amongst deleterious mutations

Negative epistasis: exists amongst beneficial mutations

No epistasis: exists amongst non-essential mutations

Due to low levels of electron carriers mutations tend to be deleterious.

In presence of high nutrient conditions, mutations tend to be beneficial or non-essential resulting in loss of negative epistasis.

## Conclusions

- 1. Interactions observed for Synechocystis and E. coli under aerobic growth for glucose are similar.**
- 2. Interactions are different depending upon the substrate available:-**
  - a. Number of interactions is related to the path length between two substrates.**
  - b. Smaller the uptake molecule larger the proportion of negative interactions (yet to be understood!)**
- 3. Non-optimal substrate uptake ratios lead to decrease in positive interactions.**
- 4. Fitness landscape across various environmental conditions is rugged and maybe governed by network topology of the organism.**

Thank you! 😊

# ISSUES, STRENGTH & DISCUSSION

## ISSUES:

- Our method can not evaluate effects of intragenic mutations which affect the protein properties. This can be fixed by partially shutting off fluxes.
- Our method is not capturing the effect of mutations which might allow survivability even if the organisms growth requirements are not met exactly. Such a method might give some idea of stable mutations leading to branches on a phylogenetic tree, and capture very long-term effects.
- Our method is insensitive to sequence in which mutations occur. For example, if geneX and geneY are mutated sequentially, the result will be same as if geneY and geneX were mutated sequentially. However, only one case is taken in histogram, we showed both on interactome map.

## STRENGTHS & DISCUSSION:

- Our method evaluates effects of intergenic mutations which affect the network properties.
- Our method captures the effect of environmental factors on the network. We showed that difference in environmental conditions can change epistasis landscape. We are only capturing the long-term effects; i.e. after the organism has achieved a steady state.
- Our method captures the relationship between input bottlenecks and long term evolution. Our results suggests that input bottlenecks are necessary to generate positive epistasis, it may help an organism to continue to move on an epistasis landscape for a longer time. If there were no constraints on input fluxes, the organism might soon come to a dead-end on an evolutionary path due to negative interactions.