# Mathematical Programming for Identification of Functional Differences in Genome-Scale Models

**Joshua J. Hamilton** and Jennifer L. Reed

Department of Chemical and Biological Engineering
University of Wisconsin-Madison

243rd ACS National Meeting
March 26, 2012

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

# Outline


1. Introduction and Background
2. CONGA Algorithm
3. Comparison of *E. coli* Metabolic Models

# Genome-Scale Models & Constraint-Based Modeling

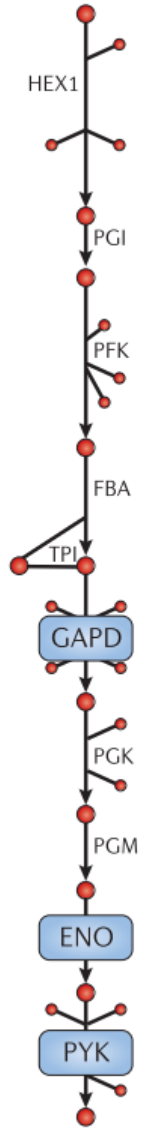
## Why models?

- Understanding cellular behavior
- Evaluate and design metabolic engineering strategies
- Contextualize high-throughput data

# Genome-Scale Models & Constraint-Based Modeling

## Why models?

- Understanding cellular behavior
- Evaluate and design metabolic engineering strategies
- Contextualize high-throughput data



# Genome-Scale Models & Constraint-Based Modeling

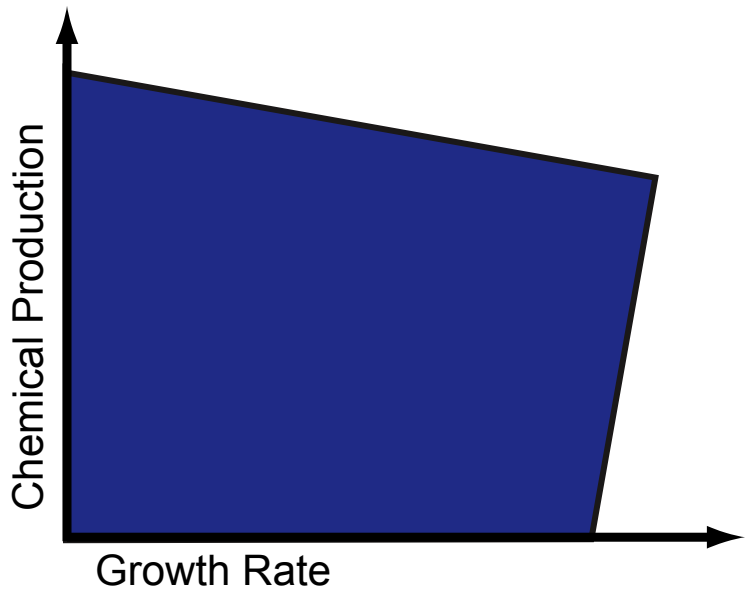
## Why models?

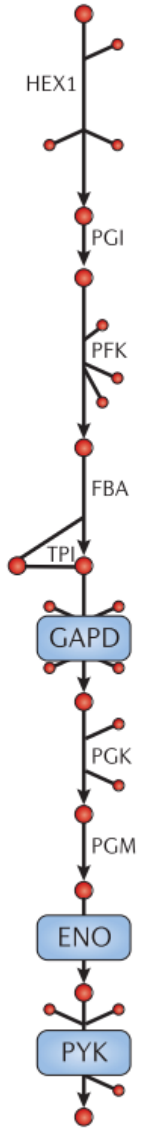
- Understanding cellular behavior
- Evaluate and design metabolic engineering strategies
- Contextualize high-throughput data

## Flux distribution subject to

- Steady-state mass balance constraints
- Limits on fluxes

## Solution Space



# Genome-Scale Models & Constraint-Based Modeling

## Why models?

- Understanding cellular behavior
- Evaluate and design metabolic engineering strategies
- Contextualize high-throughput data

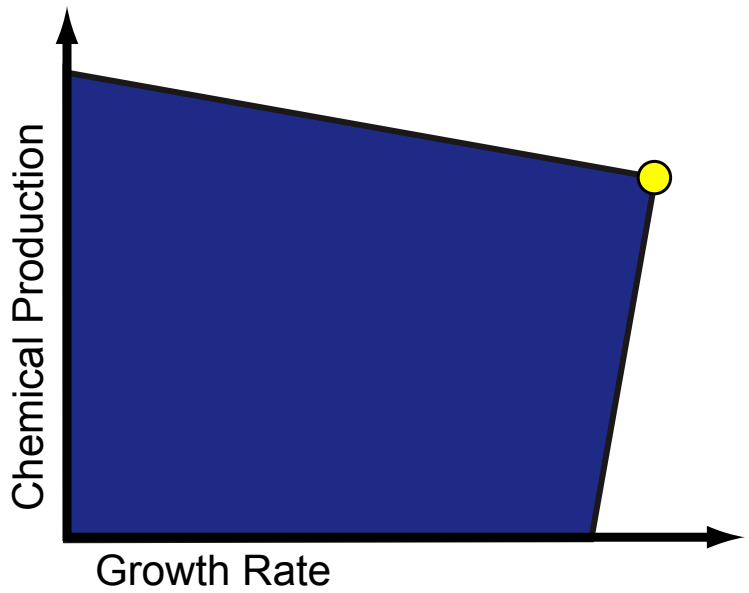
## Flux distribution subject to

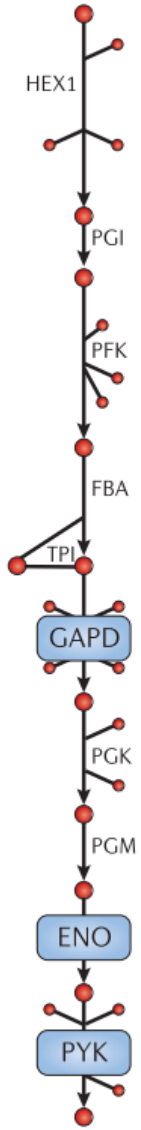
- Steady-state mass balance constraints
- Limits on fluxes

## And want to

- Maximize growth

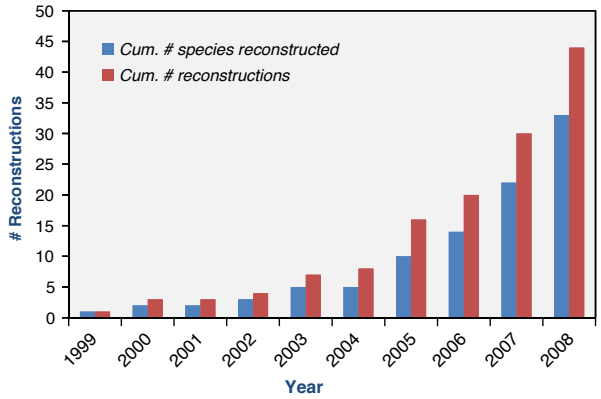
## Solution Space



# Advances in Model Development

- Number of available models growing exponentially [1]

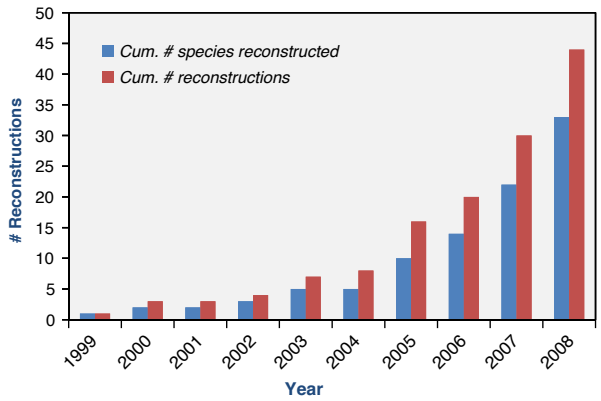



[1] Oberhardt et al, Mol Sys Bio 2009

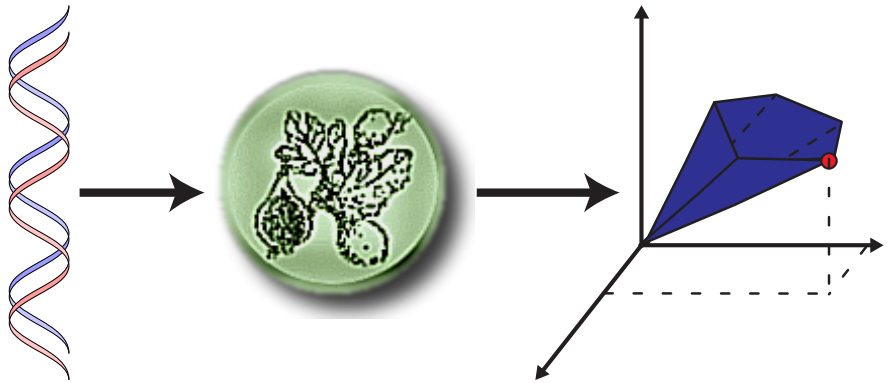
[2] Henry et al, Nature Biotech 2010

# Advances in Model Development

- Number of available models growing exponentially [1]




- New tools enable rapid conversion from genome to model [2]



[1] Oberhardt et al, Mol Sys Bio 2009

[2] Henry et al, Nature Biotech 2010

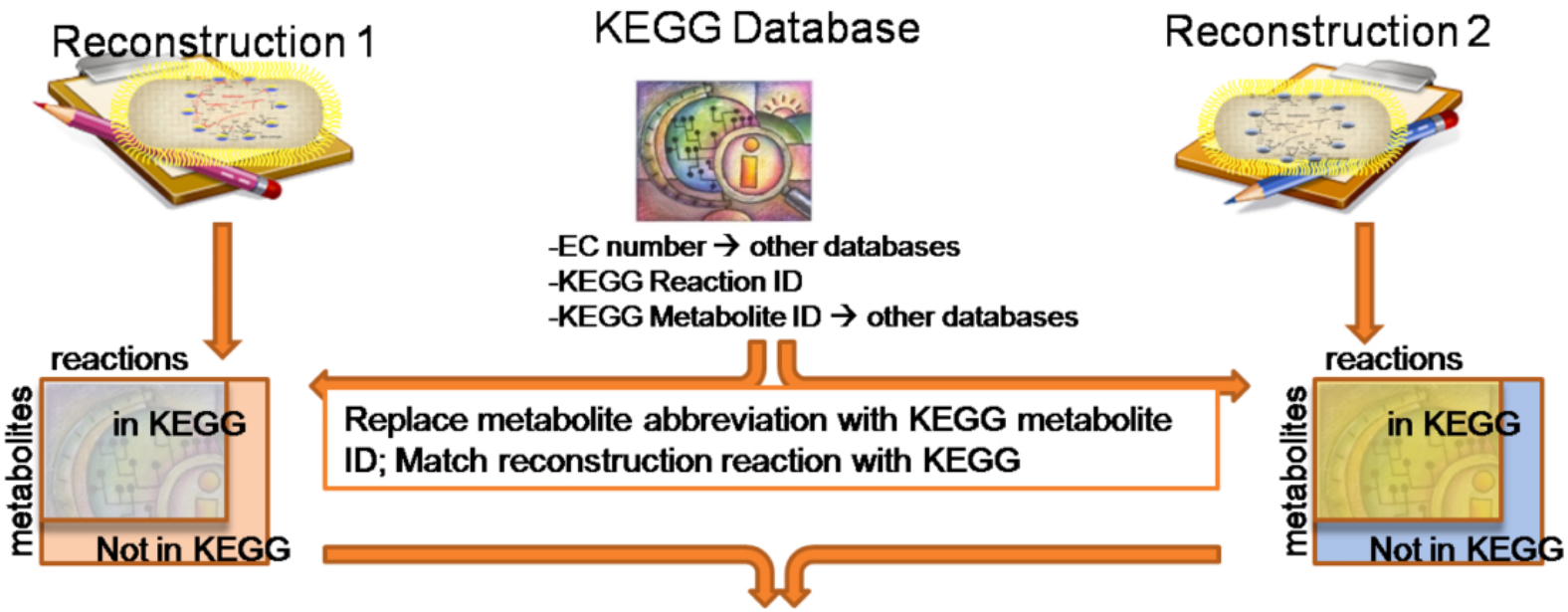
# Existing Approaches to Model Comparison

- Jamborees and network reconciliation align models at the reaction level and identify unique and shared network components

[3] Thiele et al, BMC Sys Bio 2011

# Existing Approaches to Model Comparison

- Jamborees and network reconciliation align models at the reaction level and identify unique and shared network components
- Require mapping of compounds and reactions across networks [3]




[3] Thiele et al, BMC Sys Bio 2011

# Existing Approaches to Model Comparison

- Jamborees and network reconciliation align models at the reaction level and identify unique and shared network components
- Require mapping of compounds and reactions across networks [3]
- Primarily descriptive, identifying only structural network differences [4]

[3] Thiele et al, BMC Sys Bio 2011

[4] Oberhardt et al, PLoS Comp Bio 2011

# CONGA: Comparison of Networks by Gene Alignment

## Advantages to Our Approach:

- Aligning models at the gene level can be done automatically
- Serves as a proxy for reaction-level alignments
- Mathematical programming can identify conditions under which structural network differences have a *functional* impact

# Outline

1 Introduction and Background

2 CONGA Algorithm

3 Comparison of *E. coli* Metabolic Models

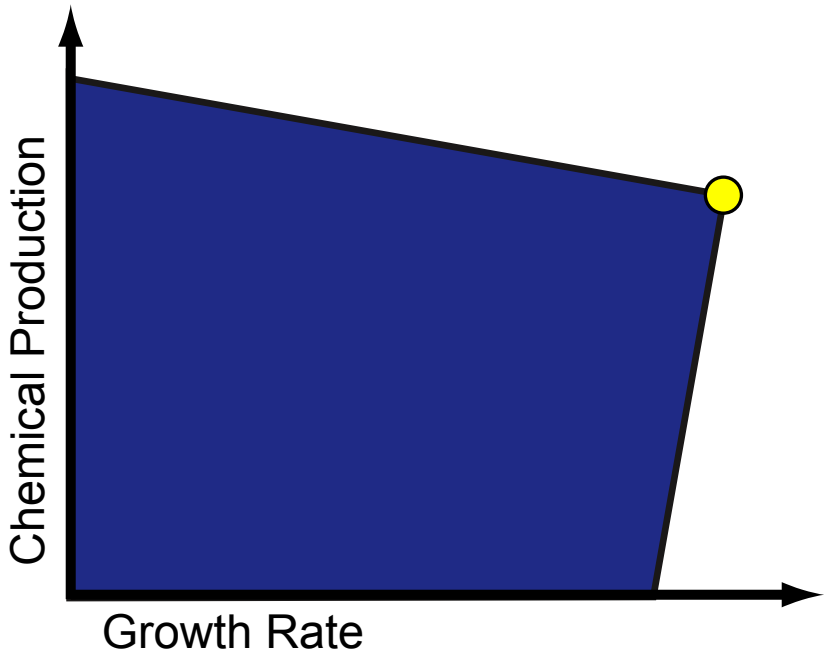
# CONGA Formulation

- Bilevel Program

## CONGA Formulation [5]

**max**
**s.t.** **max** cellular objective
**s.t.** cellular constraints

**max** cellular objective
**s.t.** cellular constraints



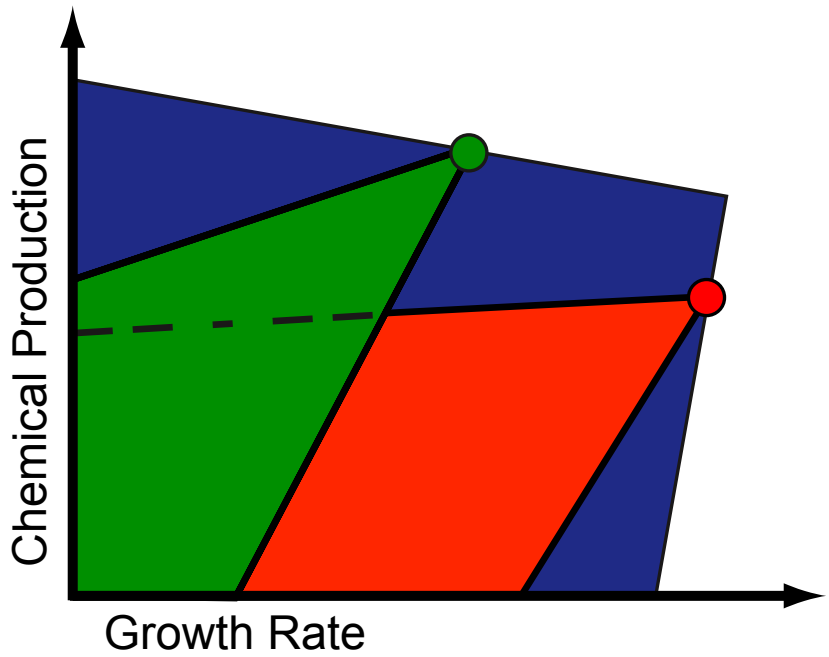

[5] Hamilton and Reed, PLoS ONE 2012

# CONGA Formulation

- Bilevel Program
  - Identify genetic deletions

## CONGA Formulation [5]

**max**
**s.t.** **max** cellular objective
**s.t.** cellular constraints
gene deletions
**max** cellular objective
**s.t.** cellular constraints
gene deletions



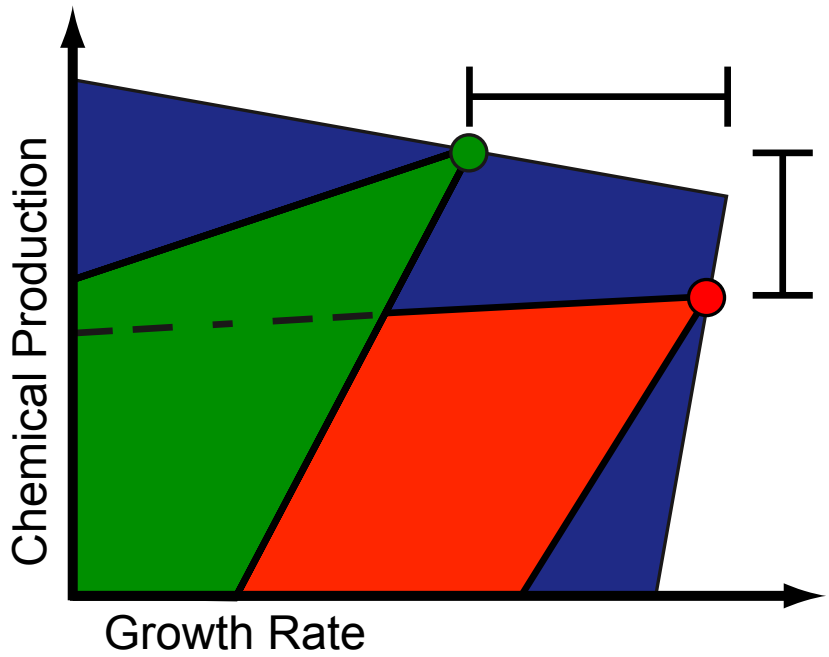

[5] Hamilton and Reed, PLoS ONE 2012

# CONGA Formulation

- Bilevel Program
  - Identify genetic deletions
  - Giving rise to a difference in phenotype

## CONGA Formulation [5]

**max** difference in cellular phenotype
**s.t.** **max** cellular objective
**s.t.** cellular constraints
gene deletions
**max** cellular objective
**s.t.** cellular constraints
gene deletions
number of deletions $\leq$ limit



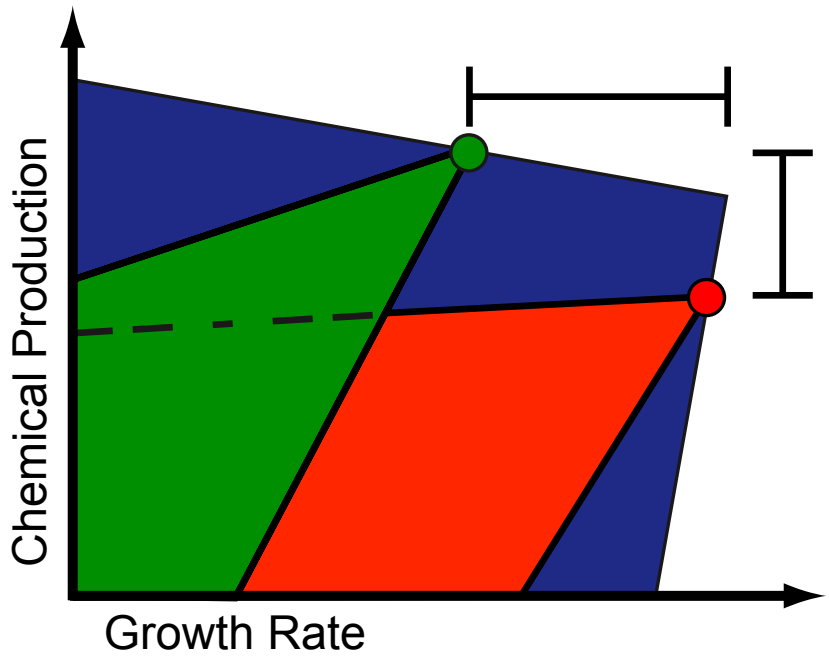

[5] Hamilton and Reed, PLoS ONE 2012

# CONGA Formulation

- Bilevel Program
  - Identify genetic deletions
  - Giving rise to a difference in phenotype
  - Which point to genetic differences w.r.t. that phenotype

## CONGA Formulation [5]

**max** difference in cellular phenotype
**s.t.** **max** cellular objective
**s.t.** cellular constraints
gene deletions
**max** cellular objective
**s.t.** cellular constraints
gene deletions
number of deletions $\leq$ limit
*orthologs deleted in common*




[5] Hamilton and Reed, PLoS ONE 2012

# Outline


1. Introduction and Background
2. CONGA Algorithm
3. Comparison of *E. coli* Metabolic Models

# Comparison of *E. coli* Metabolic Models

- Genome-scale models of *E. coli* are frequently used in metabolic engineering studies
- Updated version released every few years: *i*JR904 (2003) [6], *i*AF1260 (2007) [7]
- How does the updated model affect computational predictions?

[6] Reed et al, Genome Bio 2003
[7] Feist et al, Mol Sys Bio 2007

# Comparison of *E. coli* Metabolic Models

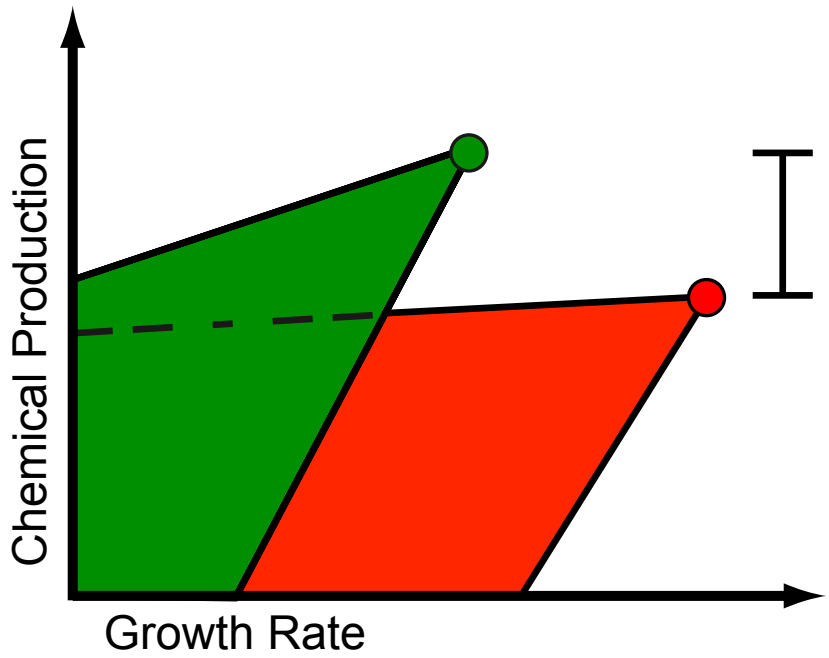
- Genome-scale models of *E. coli* are frequently used in metabolic engineering studies
- Updated version released every few years: *i*JR904 (2003) [6], *i*AF1260 (2007) [7]
- How does the updated model affect computational predictions?
  - Can we identify conditions where one model predicts higher yields than the other?
  - Are these strategies desirable metabolic engineering strategies?
  - What structural network differences are responsible?

[6] Reed et al, Genome Bio 2003
[7] Feist et al, Mol Sys Bio 2007

# Identifying Model-Dominant Strategies

## Formulation

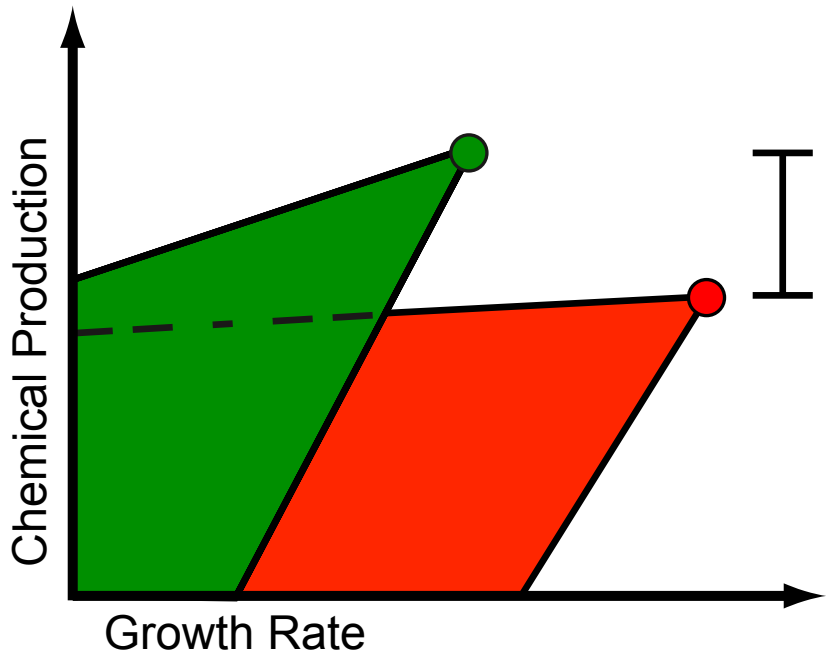
**max** difference in biochemical production
**s.t.** **max** cellular growth
**s.t.** cellular constraints
gene deletions
**max** cellular growth
**s.t.** cellular constraints
gene deletions
number of deletions $\leq$ limit
orthologs deleted in common



# Identifying Model-Dominant Strategies

## Formulation

**max** difference in biochemical production
**s.t.** **max** cellular growth
**s.t.** cellular constraints
gene deletions
**max** cellular growth
**s.t.** cellular constraints
gene deletions
number of deletions $\leq$ limit
orthologs deleted in common



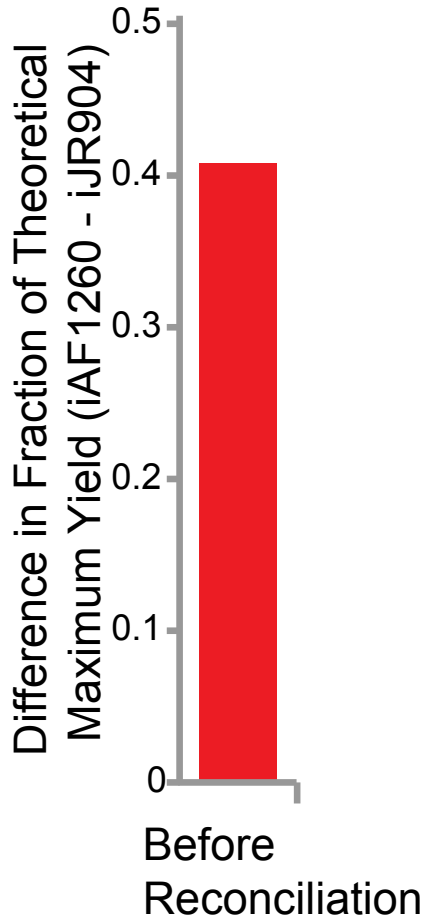

## CONGA identified model-dominant strategies for:

- Three fermentation products: ethanol, lactate, succinate
- Each *E. coli* model (*i*AF1260 and *i*JR904)
- Across multiple numbers of gene deletions

# Production of Ethanol: Deletion of *mhpF* and *adhC* [5]

- Deletion of *mhpF* and *adhC* (acetaldehyde dehydrogenase) is model-dominant in the *i*AF1260 model

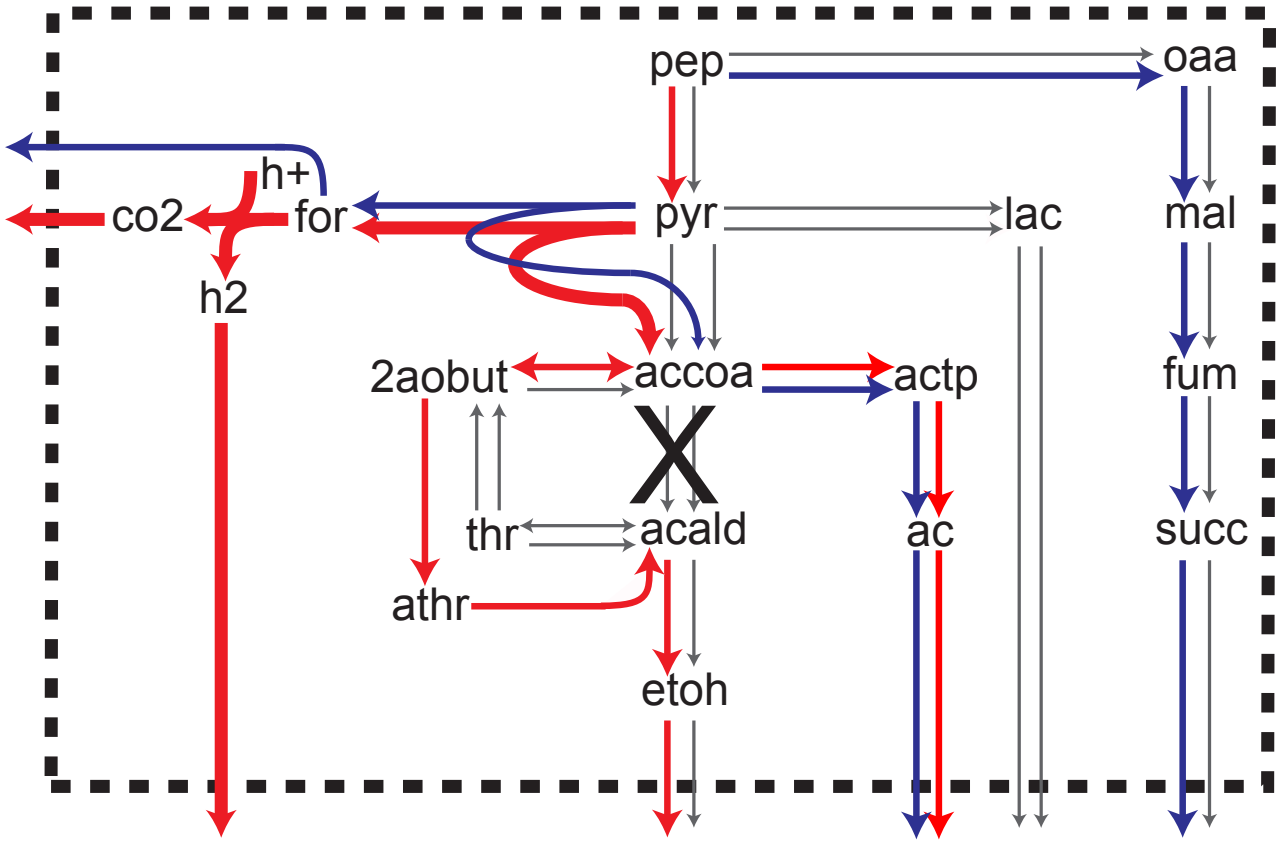
| | % of Theoretical Yield | |
|---|---|---|
| **Metabolite** | **iAF1260** | **iJR904** |
| Acetate | 30.7 | 40.5 |
| **Ethanol** | 40.8 | 0 |
| Succinate | 0.5 | 48.5 |

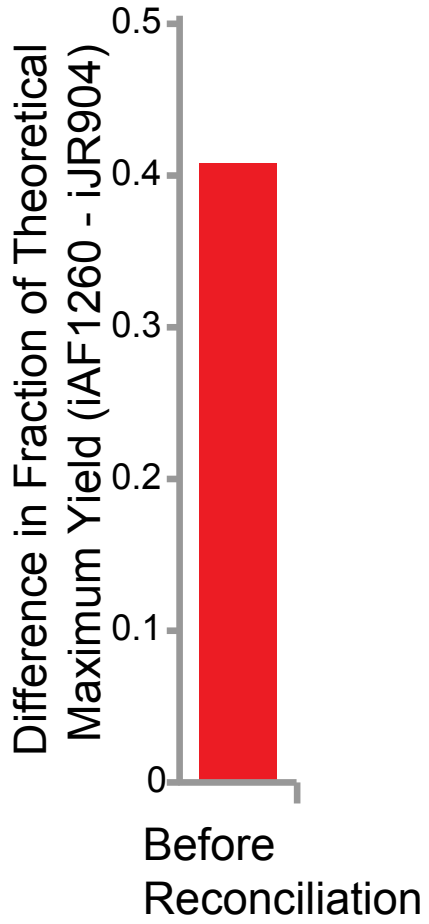



[5] Hamilton and Reed, PLoS ONE 2012

# Production of Ethanol: Deletion of *mhpF* and *adhC* [5]

- Deletion of *mhpF* and *adhC* (acetaldehyde dehydrogenase) is model-dominant in the *i*AF1260 model



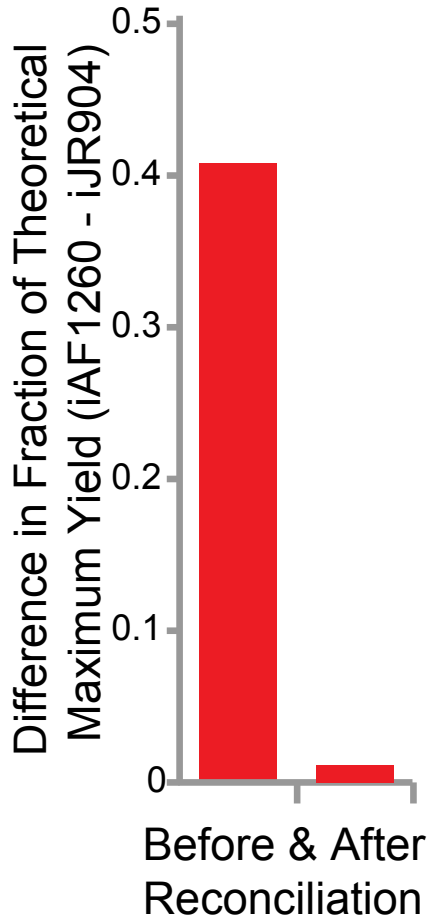




[5] Hamilton and Reed, PLoS ONE 2012

# Production of Ethanol: Deletion of *mhpF* and *adhC* [5]

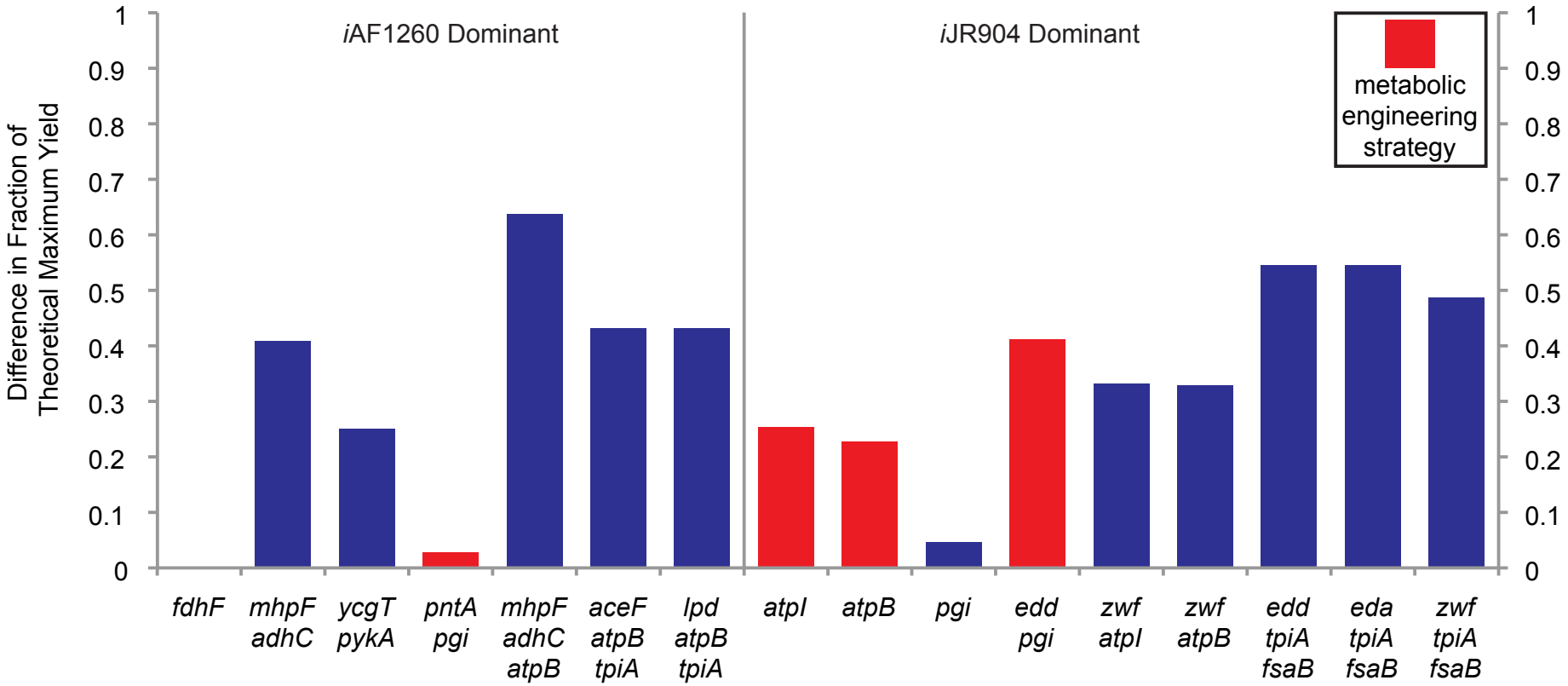
- Reconciliation of the *i*JR904 model eliminates the phenotypic difference

| | % of Theoretical Yield | |
|---|---|---|
| **Metabolite** | **iAF1260** | **iJR904** |
| Acetate | 30.7 | 30.8 |
| **Ethanol** | 40.8 | 39.7 |
| Succinate | 0.5 | 0.5 |



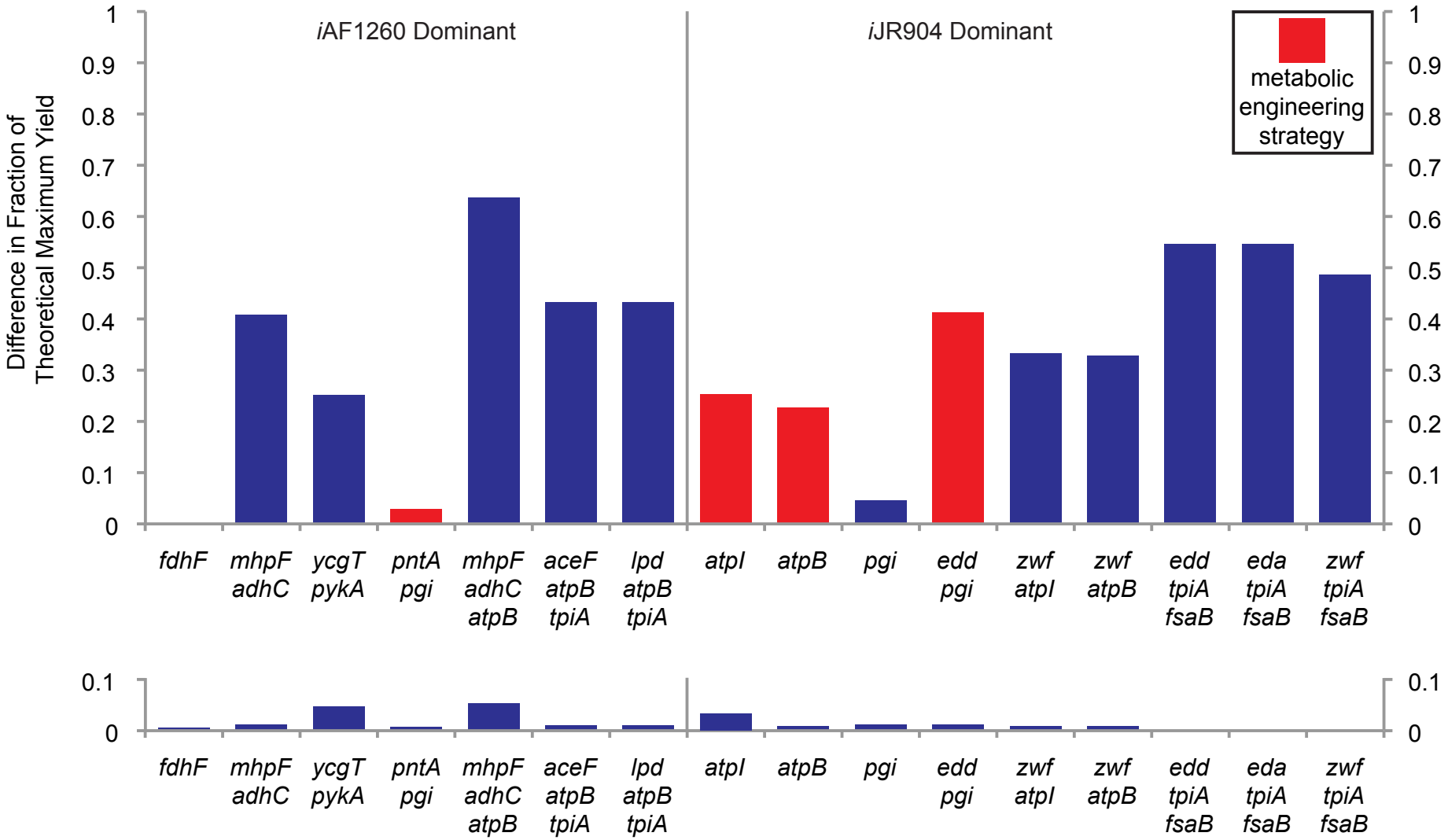

[5] Hamilton and Reed, PLoS ONE 2012

# Results for Ethanol [5]



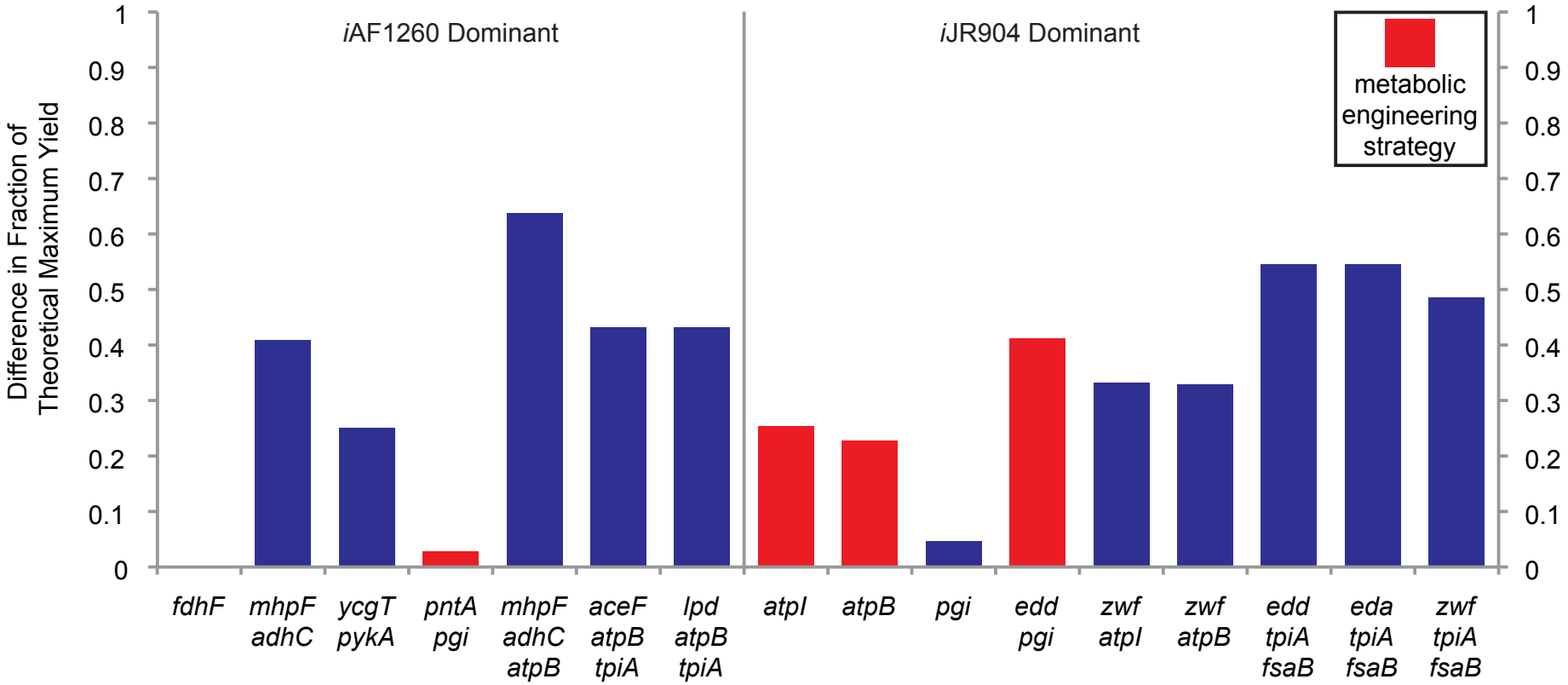

[5] Hamilton and Reed, PLoS ONE 2012

# Results for Ethanol [5]




[5] Hamilton and Reed, PLoS ONE 2012

# Results for Ethanol [5]




- Metabolic engineering strategies: among the top three strategies identified by OptORF [8]

[5] Hamilton and Reed, PLoS ONE 2012

[8] Kim and Reed, PLoS Comp Bio 2010

# Results Across All Fermentation Products [5]

- Five metabolic differences accounted for 75% of all strategies we identified!

| **Metabolic Difference** |
|---|
| 1,2-Propanediol Synthesis |
| Ethanol Synthesis |
| Hexokinase |
| Hydrogen Transport |
| Succinate Transport |

- Model-dominant strategies were also metabolic engineering strategies in 40% of cases

[5] Hamilton and Reed, PLoS ONE 2012

# Results Across All Fermentation Products [5]

- Five metabolic differences accounted for 75% of all strategies we identified!

| **Metabolic Difference** |
|---|
| 1,2-Propanediol Synthesis |
| Ethanol Synthesis |
| Hexokinase |
| Hydrogen Transport |
| Succinate Transport |

- Model-dominant strategies were also metabolic engineering strategies in 40% of cases
- Good models require an accurate representation of the pathway of interest

[5] Hamilton and Reed, PLoS ONE 2012

# Conclusions

- Developed a mathematical programming method for comparison of genome-scale network models

# Conclusions

- Developed a mathematical programming method for comparison of genome-scale network models
  - Identifies structural network differences relevant in a particular context
  - Facilitates network reconciliation - multiple models exist for *C. acetobutylicum*, *S. cerevisiae*, and more!
  - Selection of background strains for metabolic engineering
  - Can be applied to other perturbation strategies

# Conclusions

- Developed a mathematical programming method for comparison of genome-scale network models
  - Identifies structural network differences relevant in a particular context
  - Facilitates network reconciliation - multiple models exist for *C. acetobutylicum*, *S. cerevisiae*, and more!
  - Selection of background strains for metabolic engineering
  - Can be applied to other perturbation strategies
- Motivate a shift from *identifying* to *understanding* impact of differences between organisms

# Acknowledgements

**Prof. Jennifer Reed**

- Dr. Dave Baumler
- Camo Cotten
- Joonhoon Kim
- Wai Kit Ong
- Chris Tervo
- Trang Vu
- Xiaolin Zhang

Funding

- National Science Foundation: Graduate Research Fellowship Program

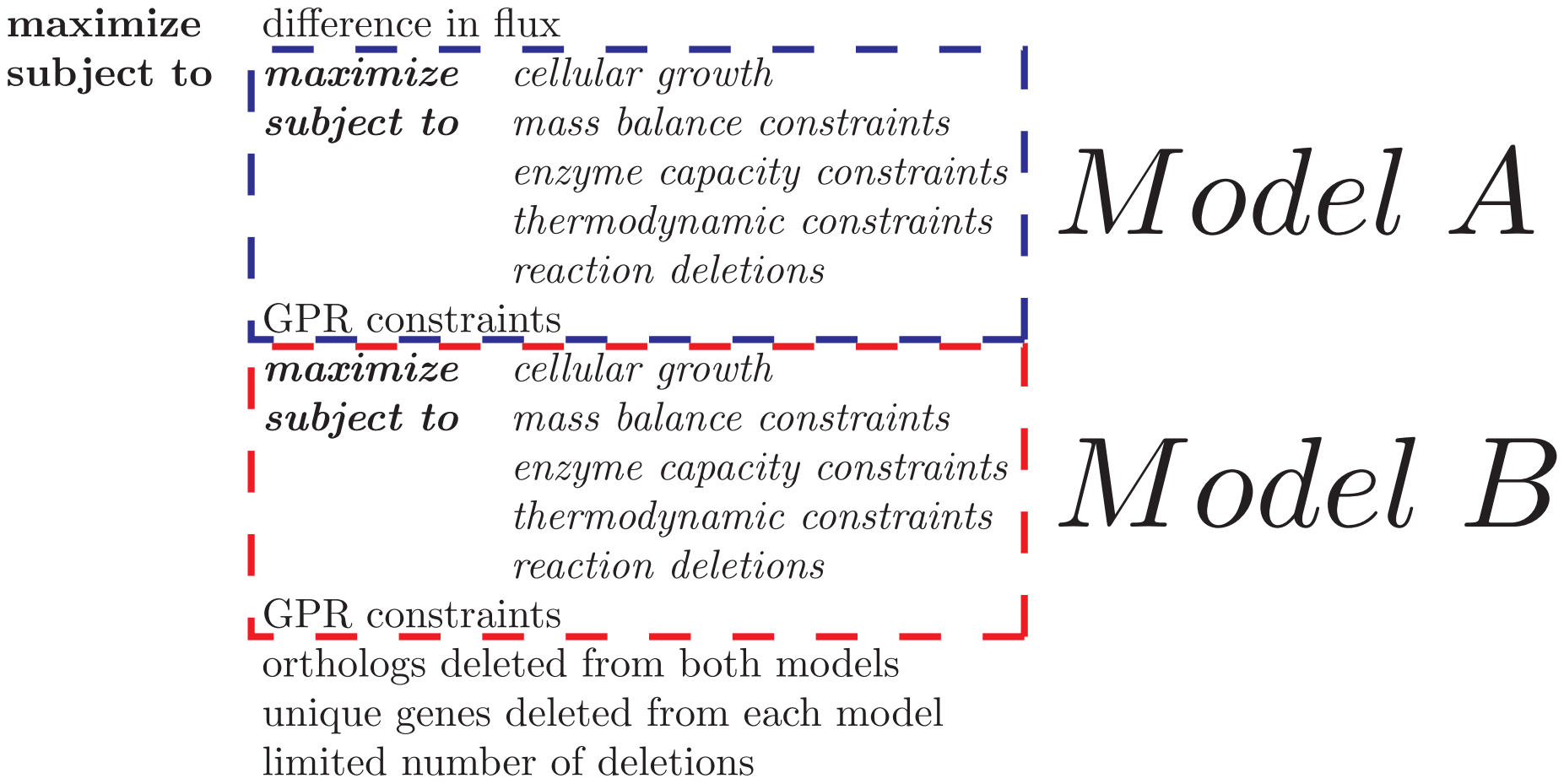

- Department of Energy: Genomic Science Program


U.S. DEPARTMENT OF ENERGY

# Questions?

- Developed a mathematical programming method for comparison of genome-scale network models
  - Identifies structural network differences relevant in a particular context
  - Facilitates network reconciliation and model development
  - Selection of background strains for metabolic engineering
  - Can be applied to other deletion strategies
- Motivate a shift from *identifying* to *understanding* impact of differences between organisms

# Formulation: Textual

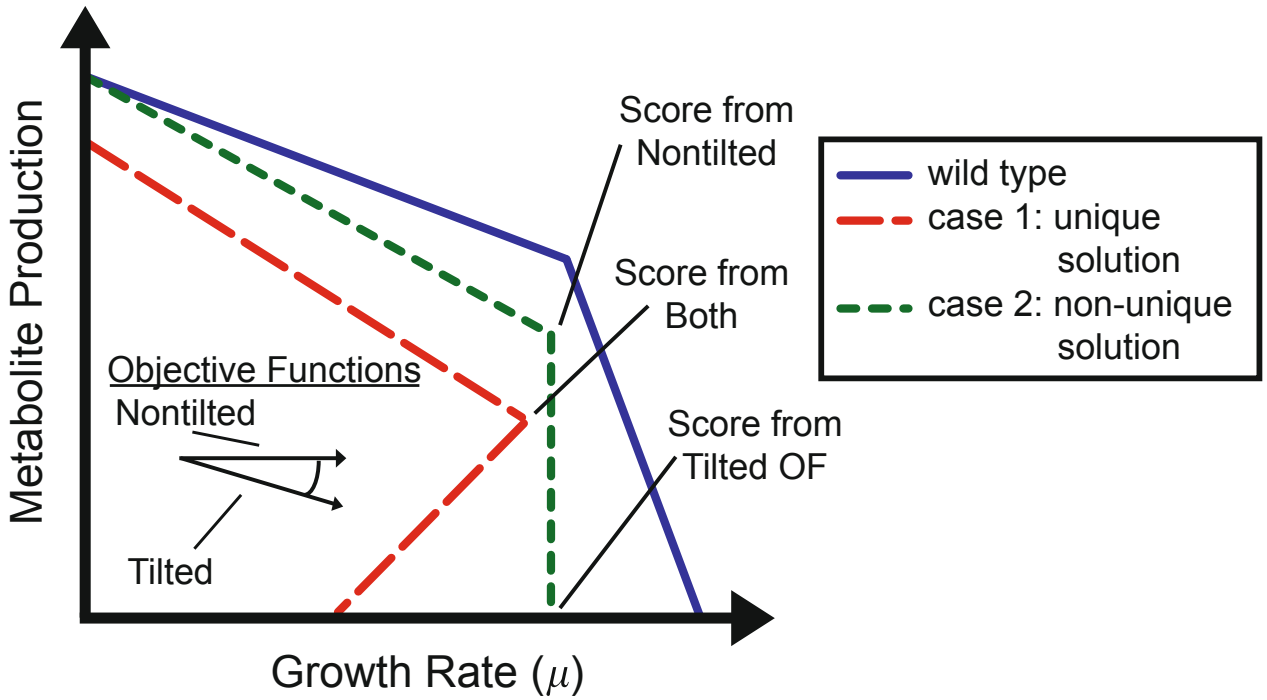


# Formulation: Mathematical

$$
\begin{aligned}
\max \quad & v_{chem_A} - v_{chem_B} \\
\text{s.t.} \quad & \max \quad v_{BM_A} - \gamma v_{chem_A} \\
& \text{s.t.} \quad \sum_j S_{ij} v_j = 0 && \forall i \in I \quad \forall \text{ Species A} \\
& \qquad \alpha_j \leq v_j \leq \beta_j && \forall j \in J \quad \forall \text{ Species A} \\
& \qquad v_j = 0 && \forall j \in J \mid y_j = 0 \quad \forall \text{ Species A} \\
\text{s.t.} \quad & \max \quad v_{BM_B} - \gamma v_{chem_B} \\
& \text{s.t.} \quad \sum_j S_{ij} v_j = 0 && \forall i \in I \quad \forall \text{ Species B} \\
& \qquad \alpha_j \leq v_j \leq \beta_j && \forall j \in J \quad \forall \text{ Species B} \\
& \qquad v_j = 0 && \forall j \in J \mid y_j = 0 \quad \forall \text{ Species B} \\
& y_j = f(z_{\hat{g}}, w_{\hat{p}}) \quad \forall \text{GPR}(j, \hat{p}, \hat{g}) \in J, P, G && \forall \text{ Species A and B} \\
& \sum_g (1 - z_g) \leq K && \forall \text{ Species A and B} \\
& z_{g_A} = z_{g_B} \quad \forall (z_{g_A}, z_{g_B}) \in O
\end{aligned}
$$

# Nonuniqueness: Tilting the Inner Objective

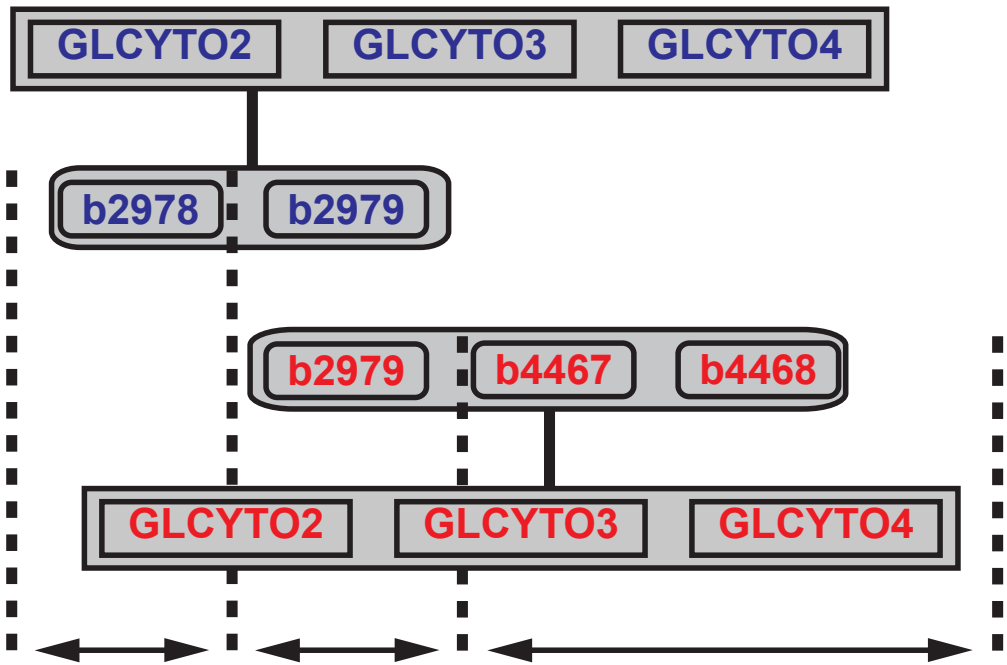
- In some instances a production flux may be nonunique
- May result in CONGA identifying incorrect solutions
- By imposing a penalty on the inner objective, we ensure the lowest production level is selected [9]



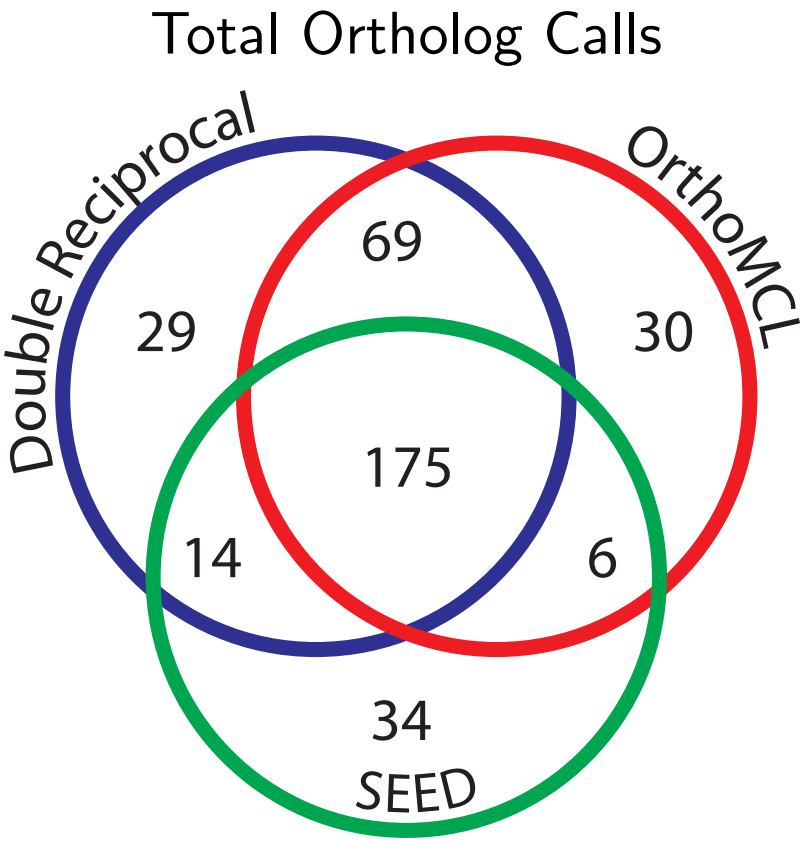

[9] Feist et al, Metab Eng 2010

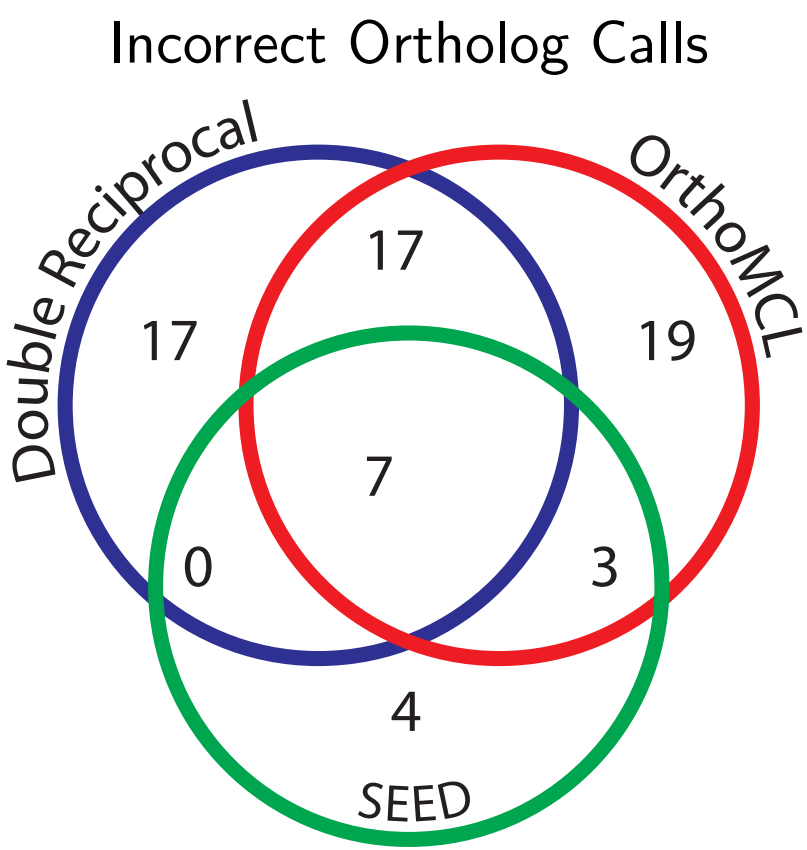
# Variable Reduction Procedure

- Computational difficulty of integer programs scales as the number of binary variables (genes)
- Consolidating groups of genes can ease the computational burden



# Comparison of Methods for Orthology Prediction

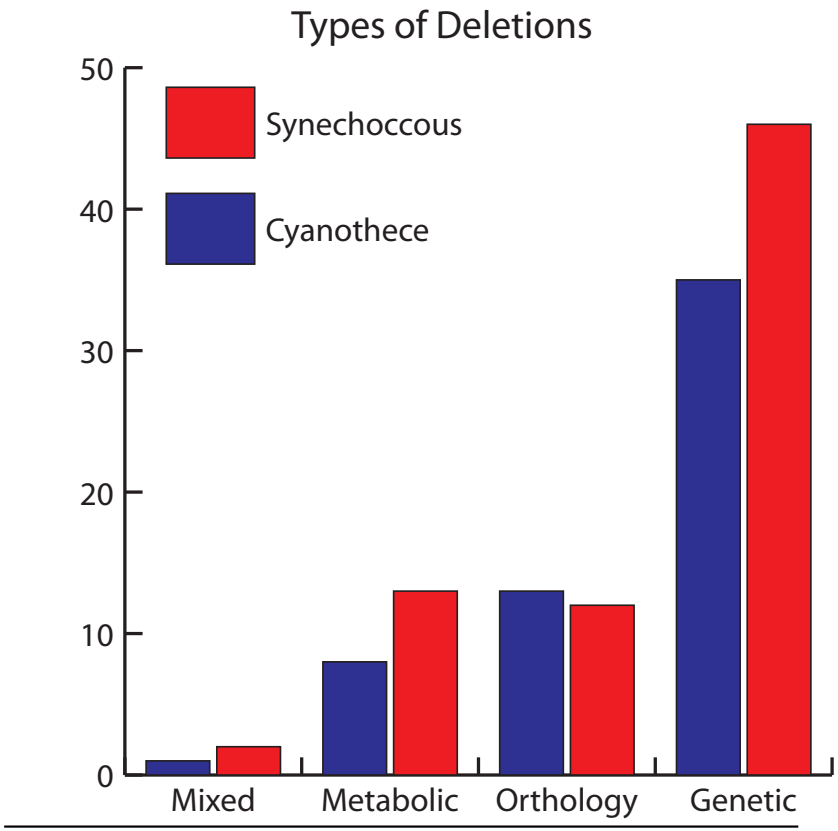



- BLAST and OrthoMCL call the most orthologs

# Model Development

- Developed a metabolic model for *Synechococcus* sp. PCC 7002 [5] from a model of *Cyanothece* sp. ATCC 51142 [10]




| | **Before** | **After** |
|---|---|---|
| Genes | 542 | 611 |
| Reactions | 491 | 552 |

[5] Hamilton and Reed, PLoS ONE 2012

[10] Vu et al, PLoS Comp Bio 2012