

Identification of Functional Differences in Cyanobacterial Metabolic Networks Using Constraint-Based Models

Joshua J. Hamilton and Jennifer L. Reed

Department of Chemical and Biological Engineering
University of Wisconsin-Madison

ASM 113th General Meeting
May 20, 2013



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON



Outline

- 1 Introduction and Background
- 2 CONGA Algorithm
- 3 Comparison of Cyanobacterial Metabolic Models



Genome-Scale Models

What are genome-scale models?

- Cells are microscopic chemical factories
- Collection of biochemical transformations based on genome
- Predict and modify cellular behaviors

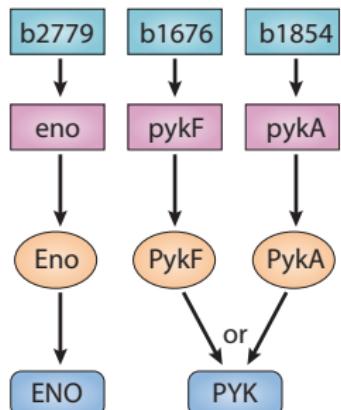
Reed et al, *Nat. Rev. Genet.* 2006

Genome-Scale Models

What are genome-scale models?

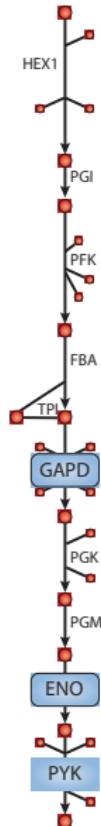
- Cells are microscopic chemical factories
- Collection of biochemical transformations based on genome
- Predict and modify cellular behaviors

Gene Associations



Stoichiometry

Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP → G6P + ADP + H	glk
PGI	[c]G6P ↔ F6P	pgi
PFK	[c]ATP + F6P → ADP + FDP + H	pfkA, pfkB
FBA	[c]FDP ↔ DHAP + G3P	fbaA, fbaB
TPI	[c]DHAP ↔ G3P	tpiA
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	gapA, gapC1, gapC2
PGK	[c]13DPG + ADP ↔ 3PG + ATP	pgk
PGM	[c]3PG ↔ 2PG	gpmA, gpmB
ENO	[c]2PG ↔ H ₂ O + PEP	eno
PYK	[c]ADP + H + PEP → ATP + PYR	pykA, pykF



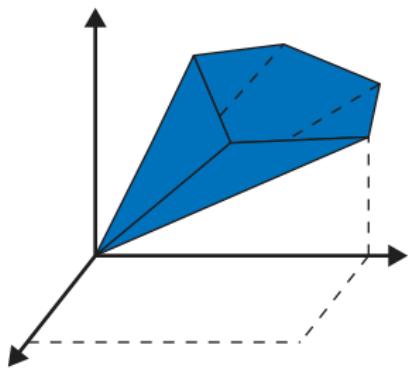
Reed et al, *Nat. Rev. Genet.* 2006

Constraint-Based Modeling

Constraints

Physiochemical constraints define feasible cellular behaviors

- Steady-state mass balance constraints
- Enzyme capacity constraints
- Reaction direction constraints



Price et al, *Nat. Rev. Microbiol.* 2004

Constraint-Based Modeling

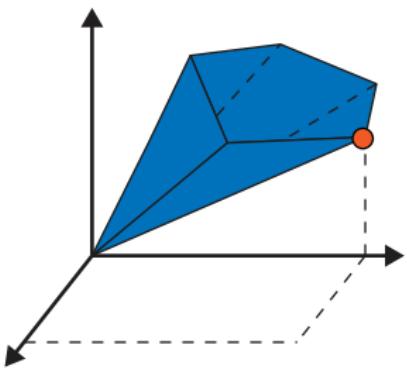
Constraints

Physiochemical constraints define feasible cellular behaviors

- Steady-state mass balance constraints
- Enzyme capacity constraints
- Reaction direction constraints

Objective Function

A function which specifies cellular behavior (i.e., maximize biomass)



Price et al, *Nat. Rev. Microbiol.* 2004



Existing Approaches to Model Comparison

- Growing interest in computational techniques to compare metabolic models

Thiele et al, *BMC Syst. Biol.* 2011

Oberhardt et al, *PLoS Comput. Biol.* 2011



Existing Approaches to Model Comparison

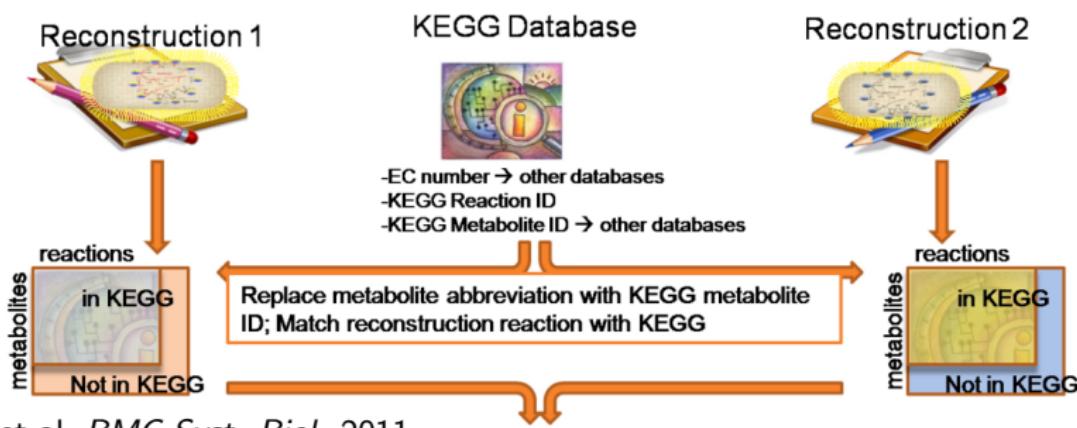
- Growing interest in computational techniques to compare metabolic models
- Jamborees and network reconciliation align models at the reaction level and identify unique and shared network components

Thiele et al, *BMC Syst. Biol.* 2011

Oberhardt et al, *PLoS Comput. Biol.* 2011

Existing Approaches to Model Comparison

- Growing interest in computational techniques to compare metabolic models
- 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



Thiele et al, *BMC Syst. Biol.* 2011

Oberhardt et al, *PLoS Comput. Biol.* 2011



Existing Approaches to Model Comparison

- Growing interest in computational techniques to compare metabolic models
- 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
- Primarily descriptive, identifying only structural network differences

Thiele et al, *BMC Syst. Biol.* 2011

Oberhardt et al, *PLoS Comput. Biol.* 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
- Relate genotype to phenotype

Hamilton and Reed, *PLoS ONE* 2012



Outline

- 1 Introduction and Background
- 2 CONGA Algorithm
- 3 Comparison of Cyanobacterial Metabolic Models

CONGA Formulation

- Bilevel Program

CONGA Formulation

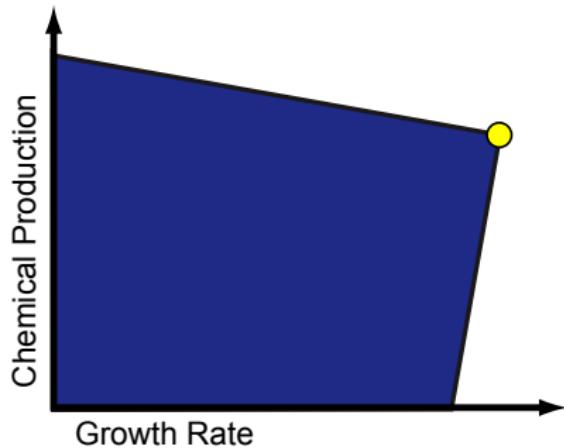
max

s.t. **max** cellular objective

s.t. cellular constraints

max cellular objective

s.t. cellular constraints



Hamilton and Reed, *PLoS ONE* 2012

CONGA Formulation

- Bilevel Program
 - Identify genetic deletions

CONGA Formulation

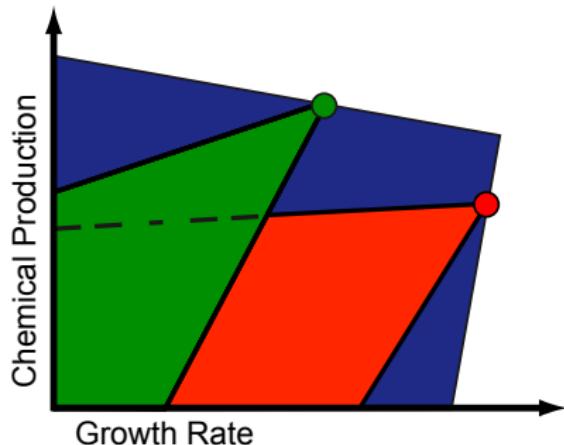
max

s.t. **max** cellular objective

s.t. cellular constraints
gene deletions

max cellular objective

s.t. cellular constraints
gene deletions



Hamilton and Reed, *PLoS ONE* 2012

CONGA Formulation

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

CONGA Formulation

max difference in cellular phenotype

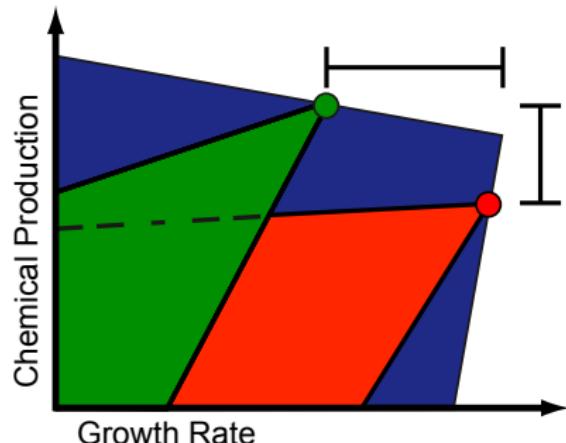
s.t. **max** cellular objective

s.t. cellular constraints
gene deletions

max cellular objective

s.t. cellular constraints
gene deletions

number of deletions \leq limit



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

max difference in cellular phenotype

s.t. **max** cellular objective

s.t. cellular constraints
gene deletions

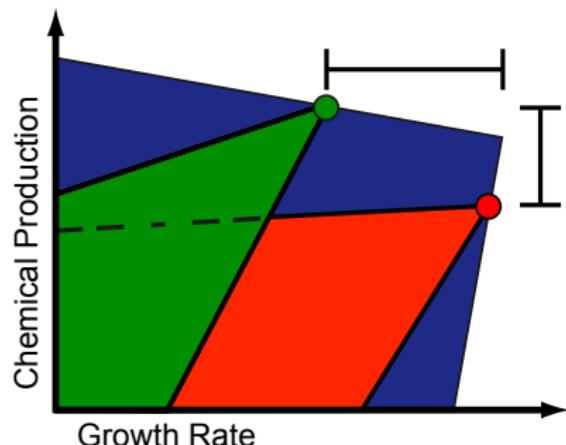
max cellular objective

s.t. cellular constraints
gene deletions

number of deletions \leq limit

orthologs deleted in common

Hamilton and Reed, PLoS ONE 2012





Outline

- 1 Introduction and Background
- 2 CONGA Algorithm
- 3 Comparison of Cyanobacterial Metabolic Models



Development of Cyanobacterial Metabolic Model

Synechococcus sp. PCC 7002 Model

- Constructed a draft model based on *Cyanothece* model
- CONGA used to identify genetic knockouts giving rise to different phenotypes

Vu et al, *PLoS Comput. Biol.* 2012
Hamilton and Reed, *PLoS ONE* 2012



Development of Cyanobacterial Metabolic Model

Synechococcus sp. PCC 7002 Model

- Constructed a draft model based on *Cyanothece* model
- CONGA used to identify genetic knockouts giving rise to different phenotypes

CONGA Identifies Structural Differences

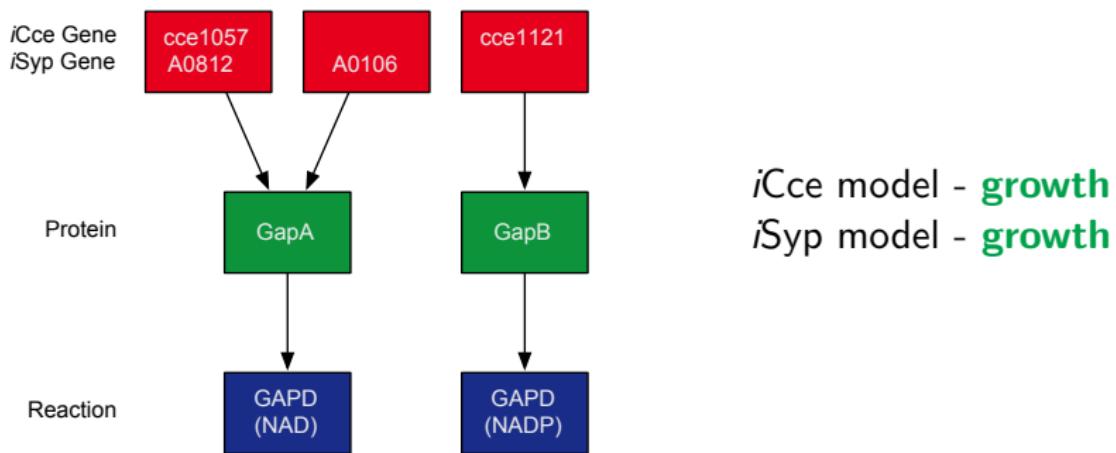
- Genetic - a gene-protein-reaction (GPR) association differs between models
- Ortholog - genes encoding enzymes with identical functions have different sequences
- Metabolic - one organism contains an enzyme which enables it to carry out a unique biochemical transformation

Vu et al, *PLoS Comput. Biol.* 2012
Hamilton and Reed, *PLoS ONE* 2012

Refinement of Draft Model

- Genetic and orthology differences may point to missing genetic content or other structural errors

Example: Deletion of GAPD gives different phenotypes

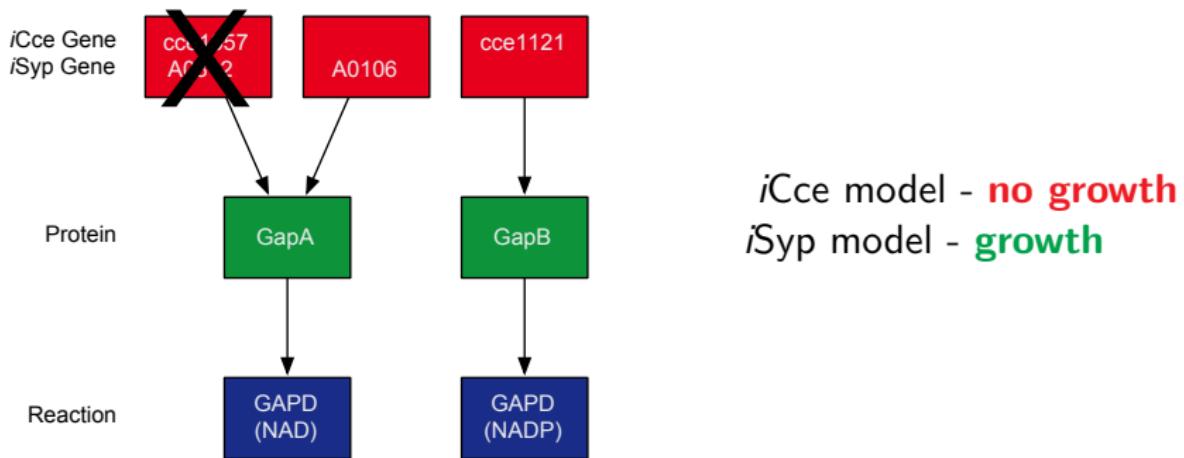


Hamilton and Reed, *PLoS ONE* 2012

Refinement of Draft Model

- Genetic and orthology differences may point to missing genetic content or other structural errors

Example: Deletion of GAPD gives different phenotypes

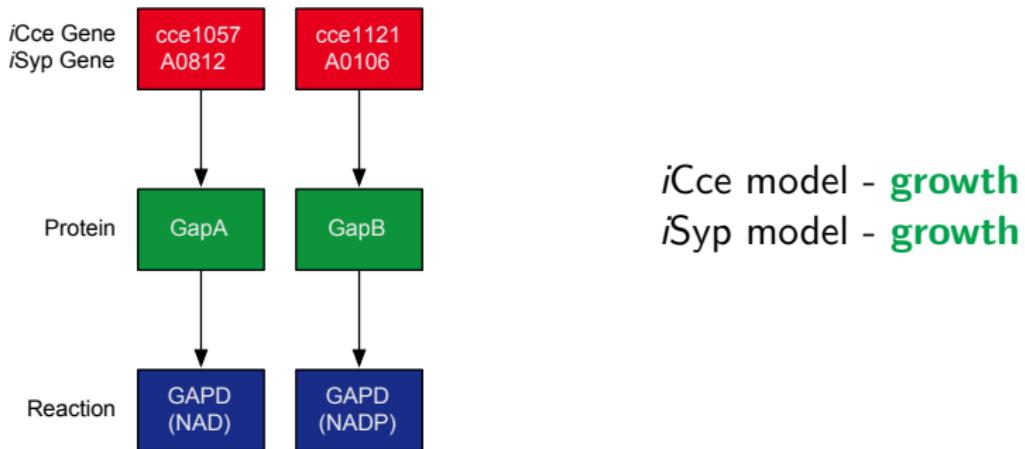


Hamilton and Reed, *PLoS ONE* 2012

Refinement of Draft Model

- Genetic and orthology differences may point to missing genetic content or other structural errors

Example: Deletion of GAPD gives different phenotypes

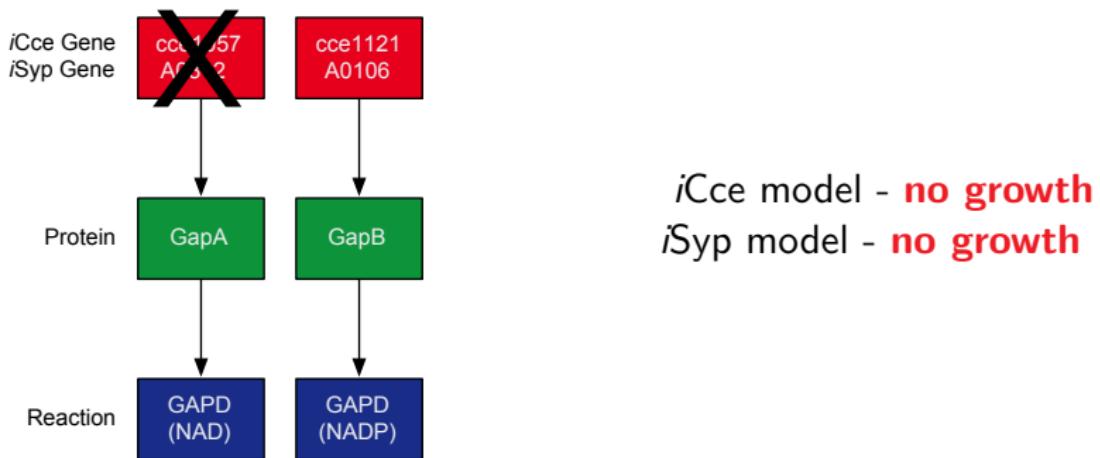


Hamilton and Reed, *PLoS ONE* 2012

Refinement of Draft Model

- Genetic and orthology differences may point to missing genetic content or other structural errors

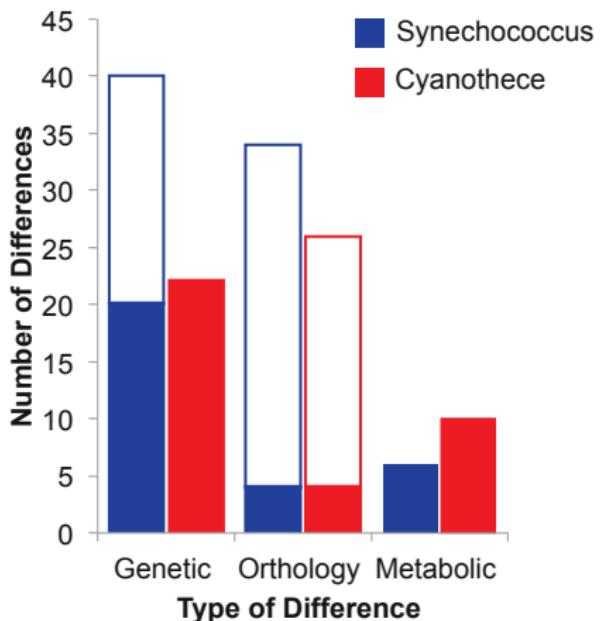
Example: Deletion of GAPD gives different phenotypes



Hamilton and Reed, *PLoS ONE* 2012

Initial vs Final Reconstruction

- We call the updated *Synechococcus* model *iSyp611*
- Model refinement reduced the number of genetic and orthology differences between models



	Before	After
Genes	542	611
Proteins	461	533
Reactions	491	552
Metabolites	529	542

Hamilton and Reed, *PLoS ONE* 2012



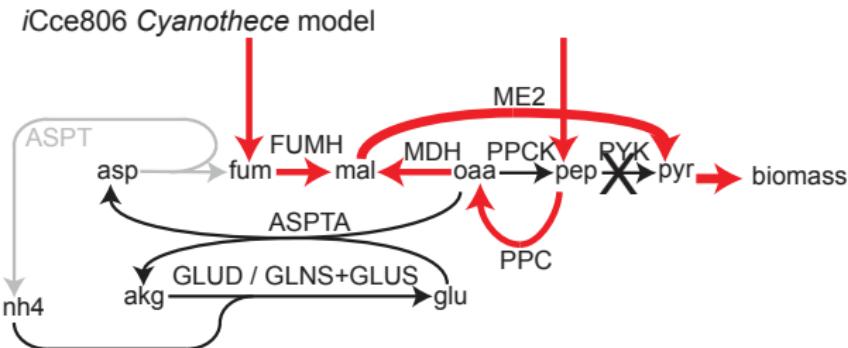
Metabolic Differences Remain

- Metabolic differences suggest unique biochemical capabilities

Hamilton and Reed, *PLoS ONE* 2012

Metabolic Differences Remain

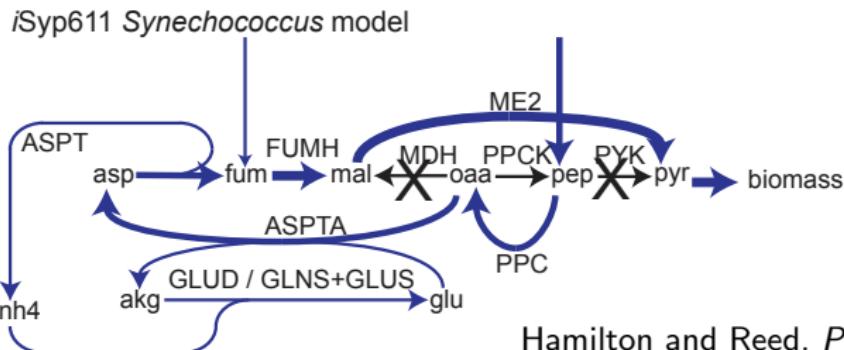
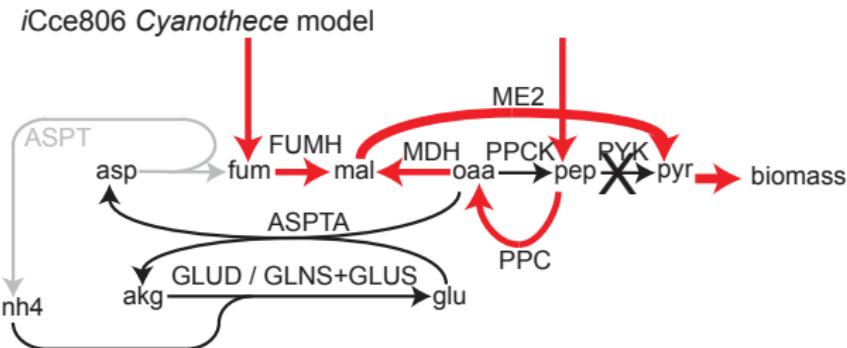
- The deletion $\Delta PYK \Delta MDH$ is lethal only in the *iCce806* model



Hamilton and Reed, *PLoS ONE* 2012

Metabolic Differences Remain

- The deletion $\Delta PYK \Delta MDH$ is lethal only in the *iCce806* model



Hamilton and Reed, *PLoS ONE* 2012



Metabolic Differences Remain

- Metabolic differences suggest unique biochemical capabilities
- Remain hypotheses until tested experimentally

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 development
 - Examine impact of functional differences on organismal fitness
 - Other applications - metabolic engineering, drug targeting



Conclusions

- Developed a mathematical programming method for comparison of genome-scale network models
 - Identifies structural network differences relevant in a particular context
 - Facilitates network development
 - Examine impact of functional differences on organismal fitness
 - Other applications - metabolic engineering, drug targeting
- Motivate a shift from *identifying* to *understanding* impact of differences between organisms

Acknowledgements

Prof. Jennifer Reed

- Dr. Joonhoon Kim
- Camo Cotten
- Matt Long
- Klaus Lovendahl
- Wai Kit Ong
- Chris Tervo
- Trang Vu
- Xiaolin Zhang



Formulation: Textual

maximize difference in flux
subject to

maximize cellular growth

subject to mass balance constraints

enzyme capacity constraints

thermodynamic constraints

reaction deletions

GPR constraints

maximize cellular growth

subject to mass balance constraints

enzyme capacity constraints

thermodynamic constraints

reaction deletions

GPR constraints

orthologs deleted from both models

unique genes deleted from each model

limited number of deletions

Model A

Model B



Formulation: Mathematical

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