Applications of genome-scale metabolic reconstructions

Matthew A Oberhardt, Bernard Palsson and Jasson A Papin

Adi Shabi
Keren Yizhak

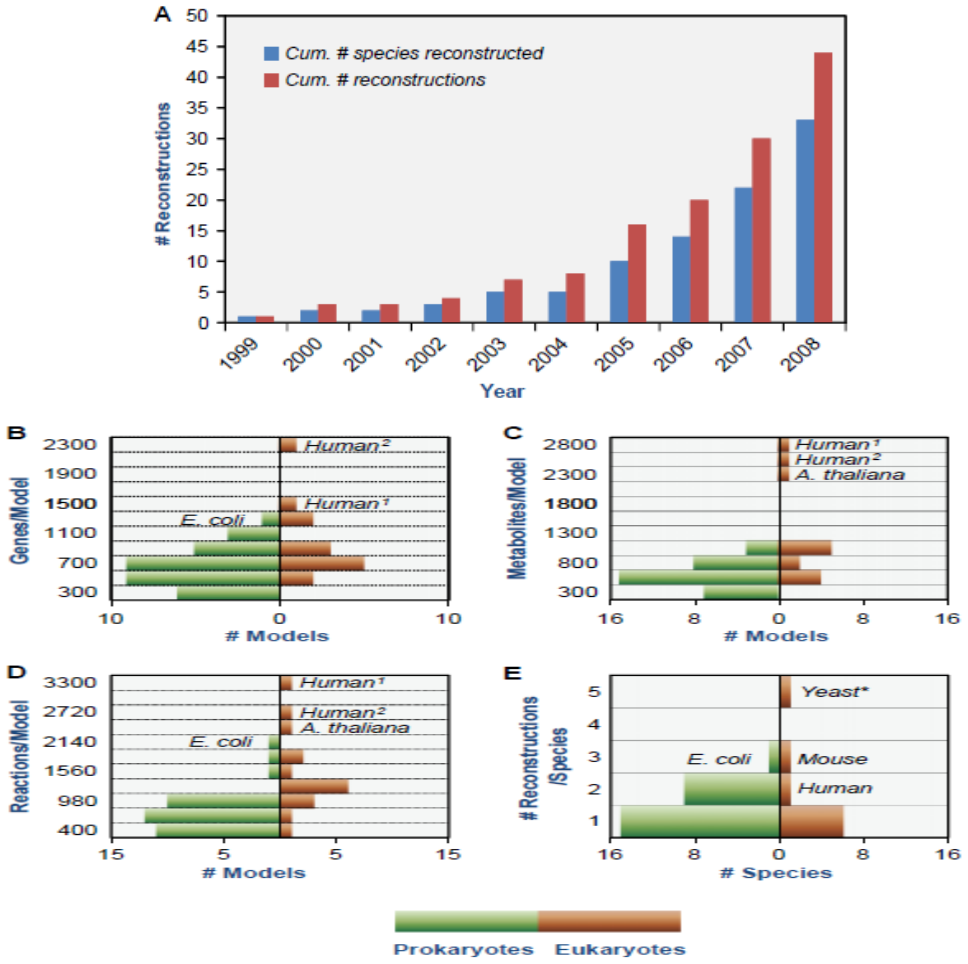
Outline

- Reconstructions statistics
- Applications
  - Contextualization of high-throughput data
  - Guidance of metabolic engineering
  - Directing hypothesis-driven discovery
  - Interrogation of multi-species relationships
  - Network property discovery

- Current status of genome-scale metabolic reconstructions

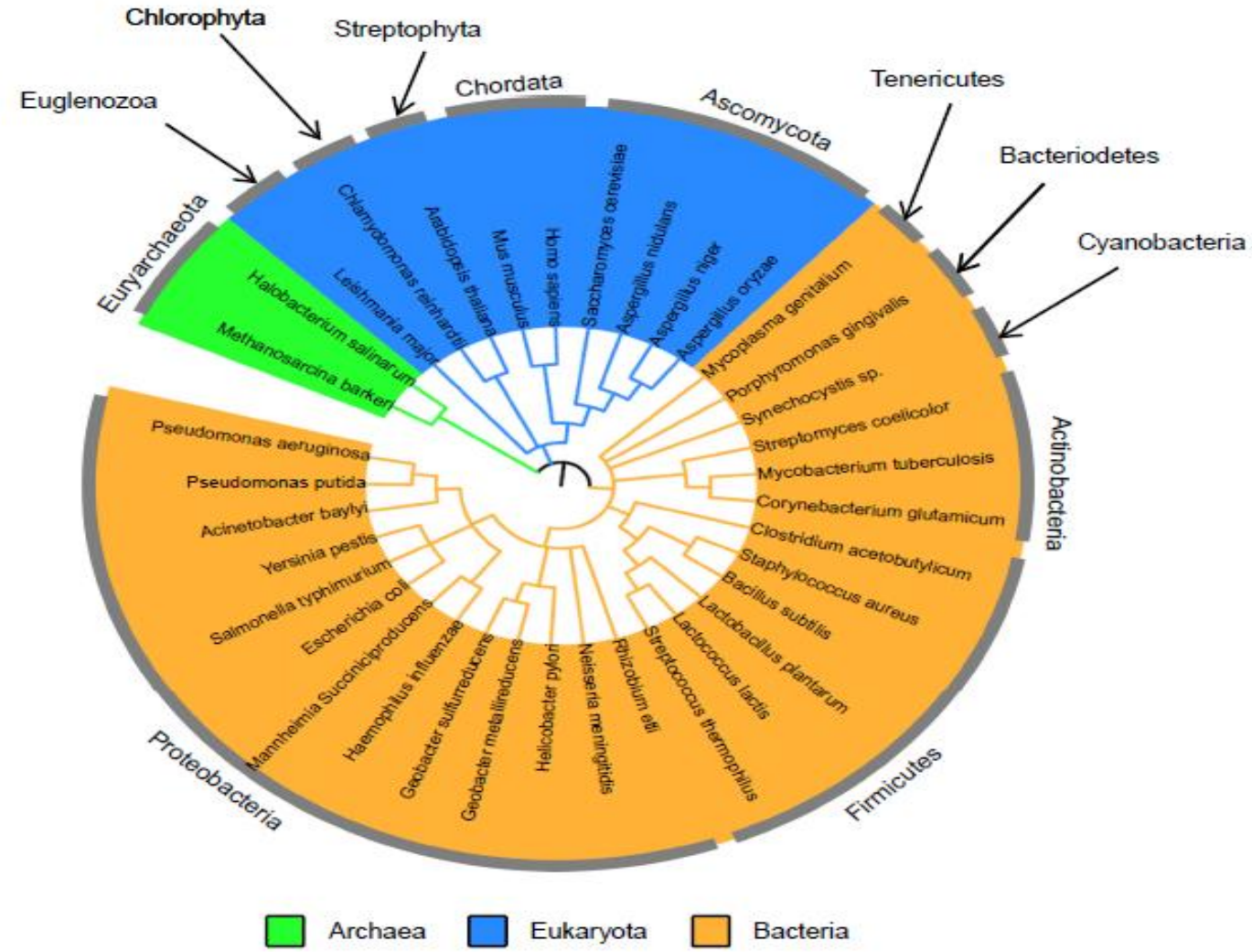
Reconstructions statistics

- Over 50 genome-scale metabolic reconstructions have been published



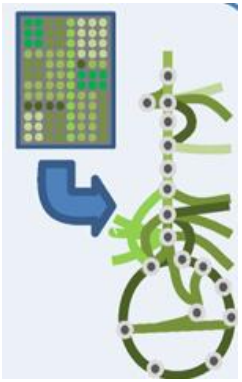
Currently available reconstructions

Currently available reconstructions



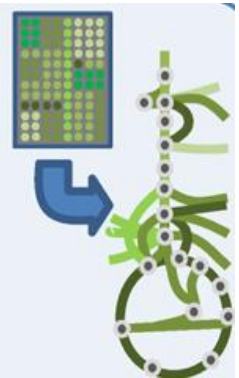
Outline

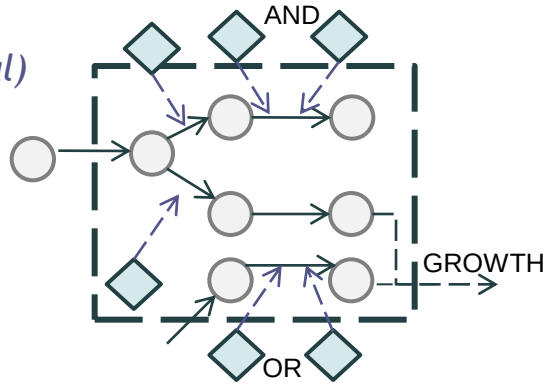
- Reconstructions statistics
- Applications
  - Contextualization of high-throughput data



Contextualization of high-throughput data

- Metabolic reconstruction, as a framework on which other data types can be overlaid, serves as powerful tool for contextualizing high-throughput data
  - imposing constraints based on experimental dataset of: gene & protein expression data, C13 flux data, high performance liquid chromatography
  - physiological states can directly be compared with *in silico* phenotypes: growth on a given media, gene essentiality data
  - multiple high-throughput data types can be analyzed in concert – giving integrated picture of cell function
    - Tissue-specific metabolic activities in *H. sapiens (Shlomi et al)*



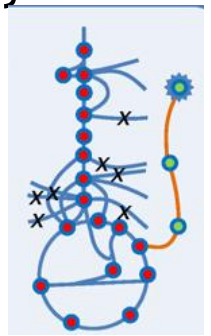


Outline

- Reconstructions statistics
- Applications
  - Contextualization of high-throughput data
  - Guidance of metabolic engineering

Guidance of metabolic engineering

- Selectively alter cell metabolism to improve a targeted cellular function
  - Increasing production of value-added chemicals
  - Increasing respiration rate of *G. sulfurreducens* predominant metal-reducing bacteria – bioremedation capabilities
  - Scale-up for bulk production of a vaccine against the pathogen, *Neisseria meningitides*
- Metabolic GENREs are uniquely capable of predicting secondary effects of a given metabolic perturbation



Ensemble modeling for strain development of L-lysine-producing *Escherichia coli* (Contador et al.)

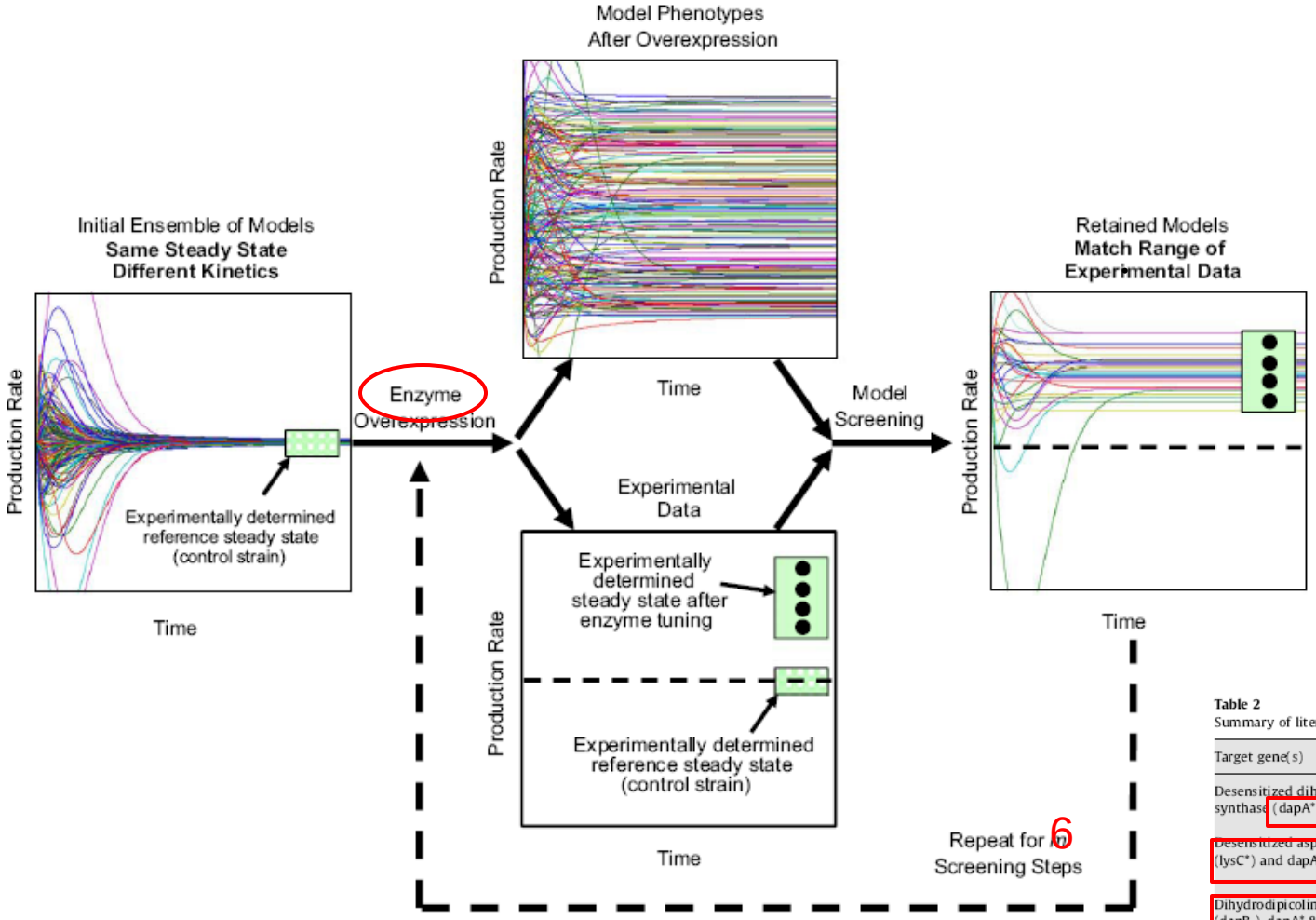
- In order to produce strains with improved yields of specific metabolites, increase the pathway flux through manipulation of single or multiple genes

- Selection of these genes is not trivial
  - uncharacterized enzyme kinetics
  - complicated network interaction
  - unexpected regulation

- Generation of a set of kinetic models that describe a set of enzyme over-expression phenotypes that produces increased levels of L-lysine (known from literature Kazima et al.)
  - allows for the generation of further targets for testing

Ensemble Modeling

- Construction of dynamic models' ensemble (Monte Carlo sampling algorithm)
  - Span the space of kinetics allowable by thermodynamics
    - overcomes the difficult task of obtaining kinetic parameters
  - Anchored to the same steady state
- Screening process allows for learning from the behavior of the true system:
  - comparing models' predicted steady-state fluxes following perturbations with experimental data
  - Keeping only models with right predictions
- The retained models converge to an increasingly realistic and predictive subset



Increased production rate

- Eventually get a set of models that properly describes the known enzyme over-expression phenotypes

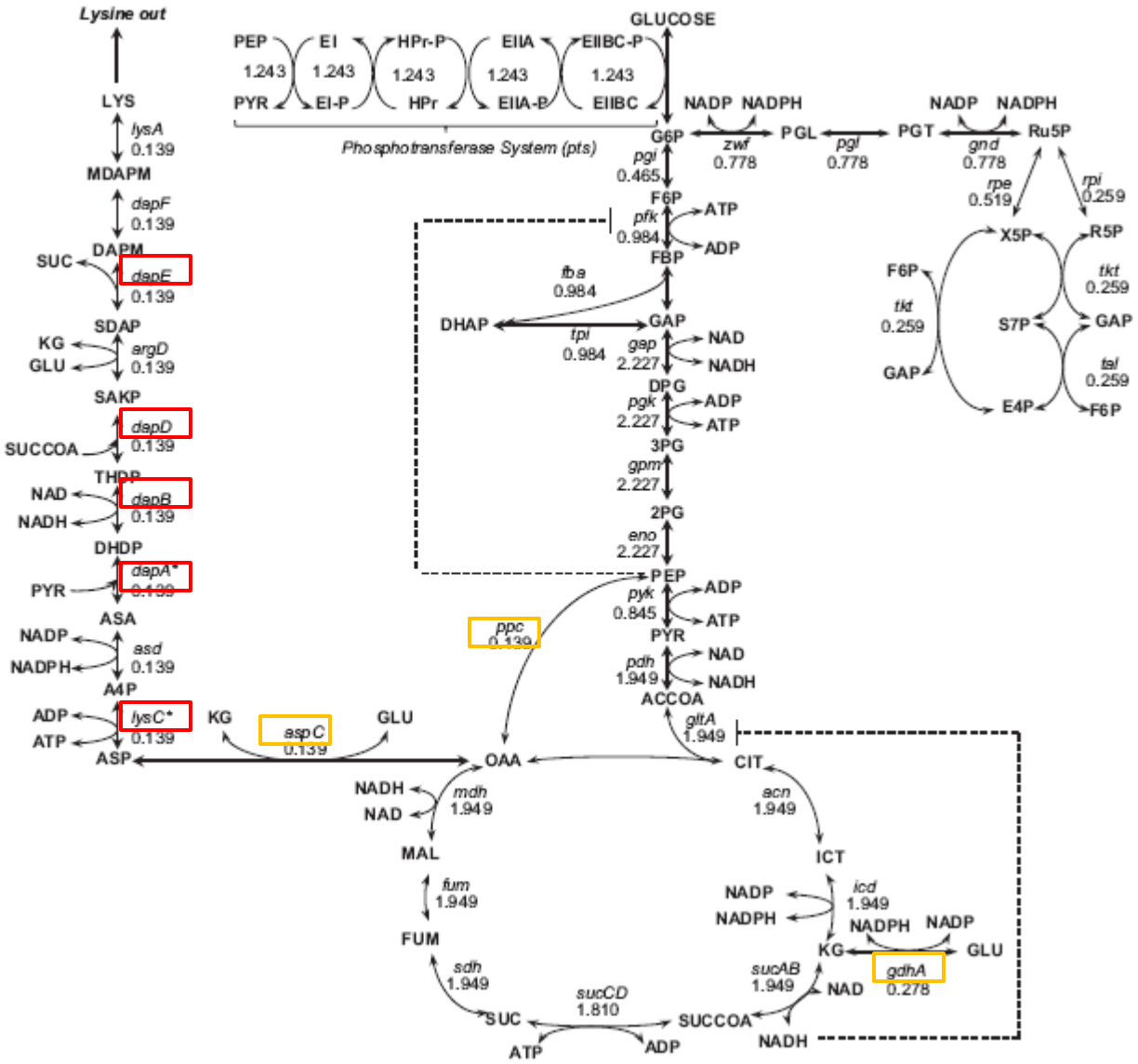
- This subset is more predictive as the additional data are used to refine the set of models

Table 2
Summary of literature used for screening phenotypes.

Target gene(s)	Phenotype	Reference
Desensitized dihydrodipicolinate synthase (dapA*)	*dapA** overexpression increases L-lysine production rate and yield	Kojima et al. (1996)
Desensitized aspartate kinase (lysC*) and dapA*	*dapA** & *lysC** simultaneous overexpression increases L-lysine production and yield	Kojima et al. (1996)
Dihydrodipicolinate reductase (dapB), dapA* & lysC*	*dapA**, *lysC**, & *dapB* simultaneous overexpression increases L-lysine production and yield	Kojima et al. (1996)
Tetrahydrodipicolinate succinylase(dapD), dapA*, lysC*, & dapB	*dapA**, *lysC**, *dapB*, & *dapD* simultaneous overexpression increases L-lysine production and yield	Kojima et al. (1996)
N-succinyl-L-diaminopimelate desuccinylase (dapE), dapA*, lysC*, & dapB	*dapA**, *lysC**, *dapB*, & *dapE* simultaneous overexpression increases L-lysine production and yield	Kojima et al. (1996)
*dapA**, *lysC**, *dapB*, *dapD*, & *dapE*	*dapA**, *lysC**, *dapB*, *dapD*, & dapE simultaneous overexpression increases L-lysine production and yield	Kojima et al. (1996)

Results

- Identified the next rate-limiting step in L-lysine production
- The final ensemble of models (six genes are over-expressed) were used to predict three new candidates for over-expression in order to further improve L-lysine production
  - Anabolic L-lysine pathway
  - In central metabolism


Lysine out
LYS
lysA 0.139
MDAPM
dapF 0.139
DAPM
SUC
dapE 0.139
SDAP
KG
GLU
argD 0.139
SAKP
dapD 0.139
SUCCOA
THDP
NAD
NADH
dapB 0.139
DHDP
dapA* 0.139
PYR
ASA
NADP
NADPH
asd 0.139
A4P
ADP
ATP
lysC* 0.139
ASP
KG
aspC 0.139
GLU
PEP
1.243
PYR
EI
1.243
EI-P
HPr-P
1.243
HPr
EIIA
1.243
EIIA-P
EIIBC-P
1.243
EIIBC
Phosphotransferase System (pts)
GLUCOSE
G6P
pgi 0.465
F6P
pfk 0.984
ATP
ADP
FBP
fba 0.984
DHAP
tpi 0.984
GAP
gap 2.227
NAD
NADH
DPG
pgk 2.227
ADP
ATP
3PG
gpm 2.227
2PG
eno 2.227
PEP
pyk 0.845
ADP
ATP
PYR
ppc 0.139
pdh 1.949
NAD
NADH
ACCOA
gltA 1.949
CIT
OAA
NADH
NAD
mdh 1.949
MAL
fum 1.949
FUM
sdh 1.949
SUC
sucCD 1.810
ATP
ADP
SUCCOA
NADH
acn 1.949
ICT
NADP
NADPH
icd 1.949
NADPH
NADP
KG
GLU
sucAB 1.949
NAD
gdhA 0.278
NADP
NADPH
G6P
zwf 0.778
PGL
pgl 0.778
PGT
NADP
NADPH
gnd 0.778
Ru5P
rpe 0.519
rpi 0.259
X5P
R5P
F6P
tkt 0.259
S7P
GAP
tkt 0.259
GAP
tal 0.259
E4P
F6P

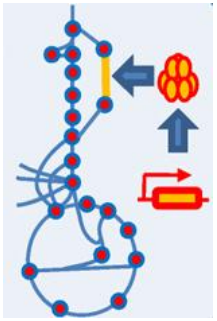
Outline

- Reconstructions statistics
- Applications
  - Contextualization of high-throughput data
  - Guidance of metabolic engineering
  - Directing hypothesis-driven discovery

Directing hypothesis-driven discovery

- Metabolic GENREs enable integration of large datasets for analysis of whole-cell phenotypes, and when wielded effectively, these analyses can be targeted to answer profound questions in biology

- Some biological questions investigated using metabolic GENREs involve cellular-level phenomena difficult to approach without a whole-cell model of metabolism

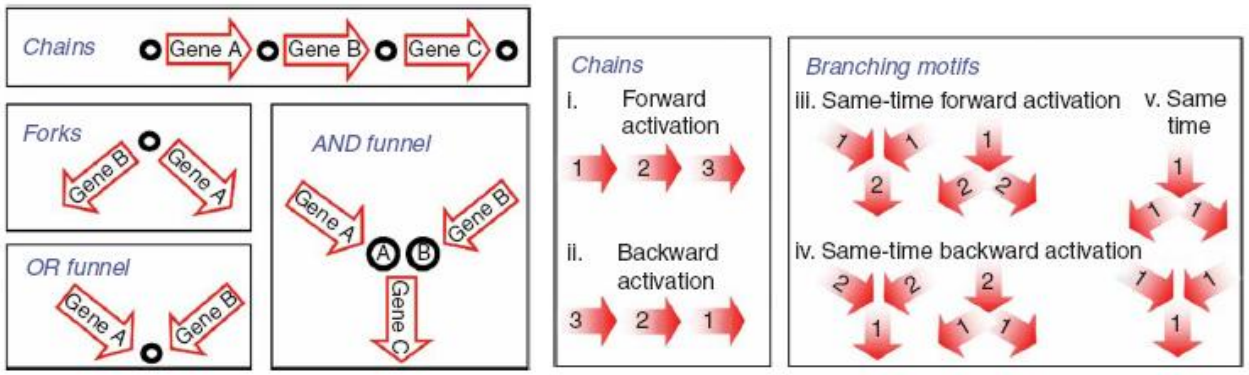


Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network (Chechik et al. 2008)

- The transcriptional timing of metabolic genes was studied using time courses of transcriptomic and proteomic data, as well as protein binding affinity data from ChIP-chip assays
- This analysis suggested that under relatively static environmental conditions, metabolism is primarily controlled through protein-level regulation (‘hierarchical control’)
- While during times of environmental change, transcriptional control guides metabolic function

Activity motifs

- Activity motifs describes a specific pattern of functional data
  - ordered timing of activation of the corresponding genes (Onsets of transcriptional responses)
  - ordered binding affinity to a transcription factor



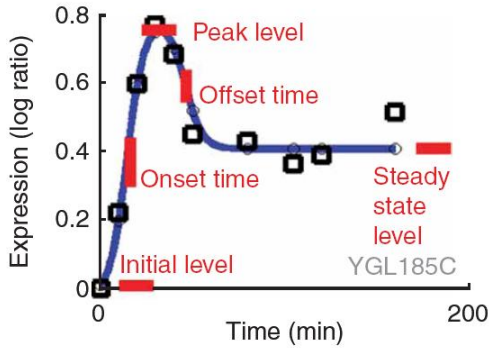
- Activity motifs can be identified by assessing the enrichment of activity patterns given the network wiring structure

Timing activity motifs

- Expression profiles of particular biological condition → transcription timing properties → wiring patterns → timing activity motif

- Enriched motifs uncover the principles of the organism's transcriptional response

- Cells use different timing activity motifs to optimize transcription timing in response to changing conditions:



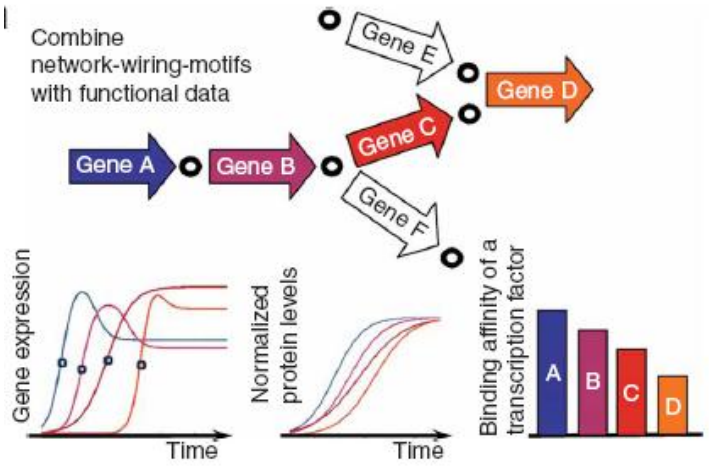
  - Forward activation to produce metabolic compounds efficiently
  - Backward shutoff to rapidly stop production of a detrimental product
  - Synchronized activation for co-production of metabolites required for the same reaction

Mechanism that can underlie the extensive ordered timing of transcriptional control

- Having differences in affinity of a common transcription factor for the promoters of various genes in the pathway can result in different transcription onsets

- ChIP-chip's continuous values can be interpreted as quantitative TF binding affinities
  - Multiple TF's
  - Across several conditions

- Binding activity motifs – linear chains of enzymes whose genes exhibited patterns of ordered affinity for a TF



Cont.

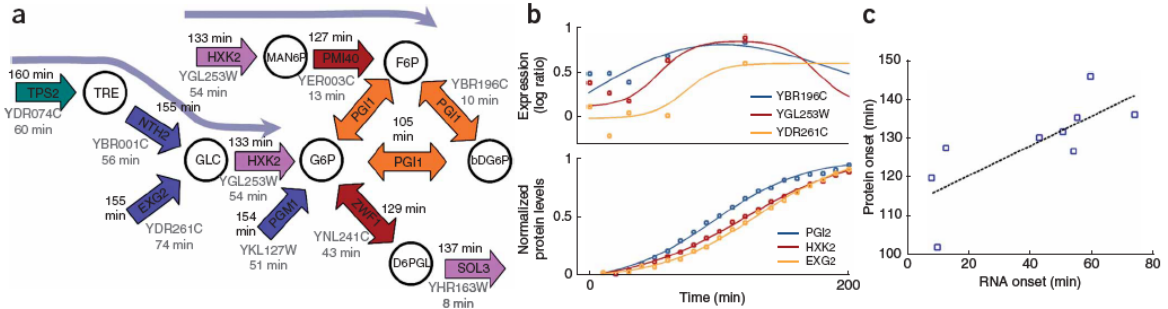
- Timing activity motifs significantly overlap with binding activity motifs
- Tuning of transcription factor binding affinities may play a significant role in the temporal regulation of metabolic transcription

Protein timing activity motifs

- The levels of active enzymes are only partially determined by mRNA levels – general high correlation, with significant intergene variation
  - Multiple subsequent regulatory control
  - Protein half-life

- Time course of protein levels were measured for genes participating in timed motifs after exposure to DTT (induces expression activation)
  - The level of protein product roughly resembles a scaled, time-delayed integral of its mRNA level
  - Onset time of protein activation in good correlation with the onset time of the mRNA activation

- Conservation of ordered timing relationships



Outline

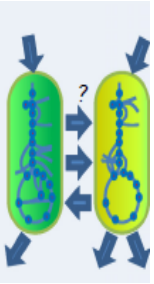
- Reconstructions statistics
- Applications
  - Contextualization of high-throughput data
  - Guidance of metabolic engineering
  - Directing hypothesis-driven discovery
  - Interrogation of multi-species relationships

Interrogation of multi-species relationships

"In many cases it is through the interactions of species that the most interesting phenotypes emerge"

- Interactions between species
- Interactions between different cell types

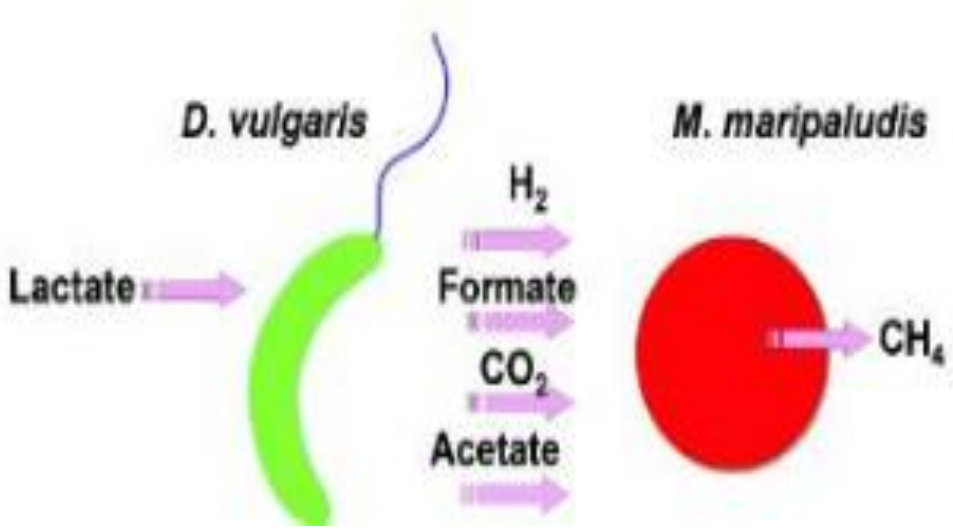
These understandings may help in bridging the phenotype-genotype gap

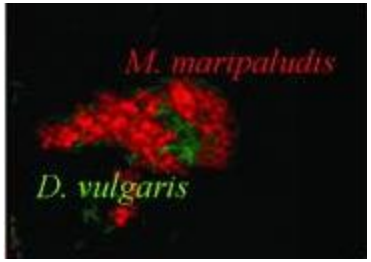


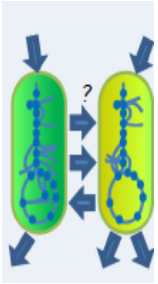
Metabolic modeling of a mutualistic microbial community

(Stolyar et a l, 2007)

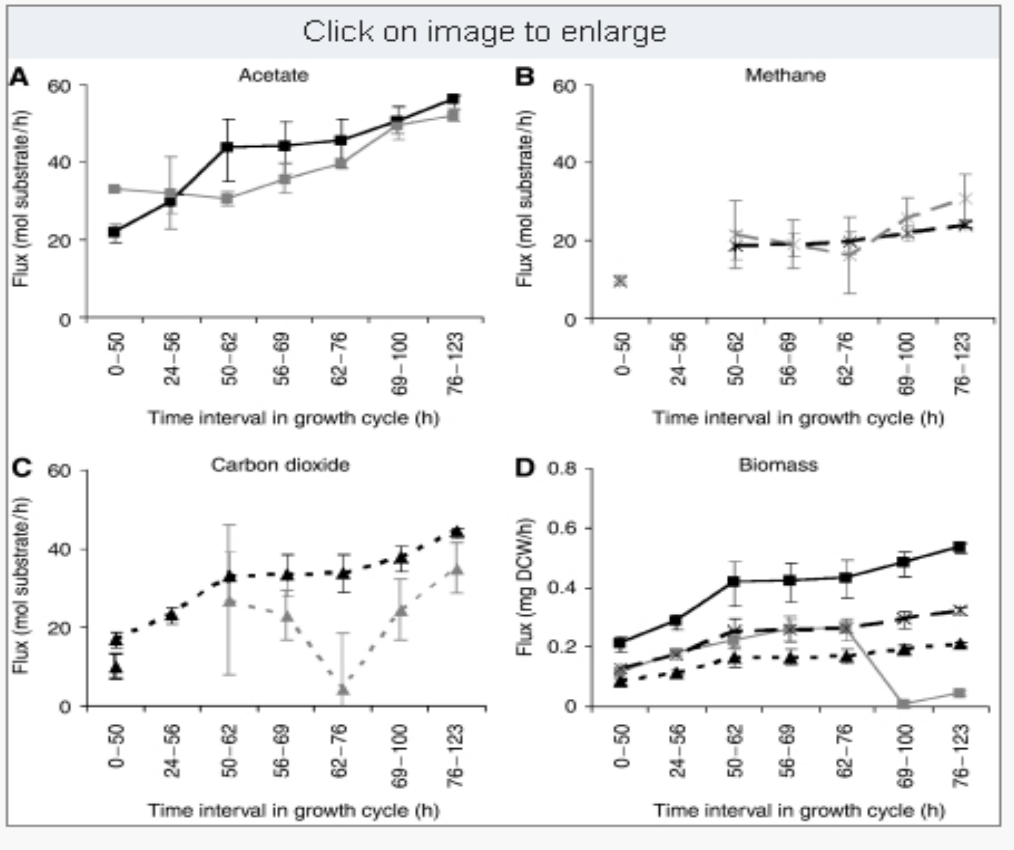
- Producing and analyzing the first multispecies stoichiometric metabolic model
- Prediction of several ecologically relevant characteristics

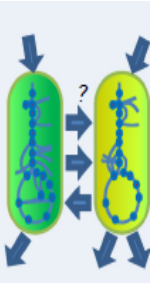






- A three compartment model:
  - D. vulgaris metabolic model
  - M. maripaludis metabolic model
  - culture medium





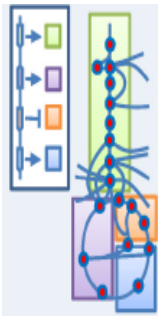
Outline

- Reconstructions statistics
- Applications
  - Contextualization of high-throughput data
  - Guidance of metabolic engineering
  - Directing hypothesis-driven discovery
  - Interrogation of multi-species relationships
  - Network property discovery

Network property discovery

"Complex cellular networks can spawn emergent phenomena that would be undetectable by reductionist approaches"

- Existence of loops
- Optimal pathway usage
- Pathway redundancy
- Metabolite connectivity

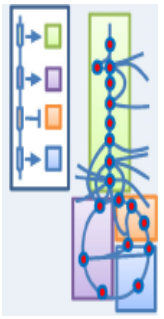


Investigating the metabolic capabilities of *Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing* alternative drug targets

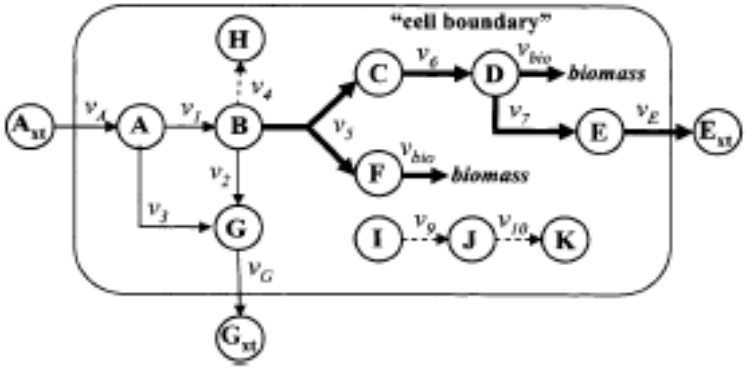
(Jamshidi and Palsson, 2007)

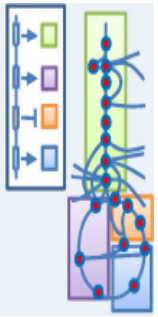
- the emergence of multi-drug resistant (MDR) strains of tuberculosis hails the need to develop additional medications for treatment
- Using flux coupling analysis in the context of know drug targets they proposed new alternative, but equivalent drug targets
- Single enzyme drug targets actually knock out complete pathways

Terminating the activity of any other enzyme in that pathway should have the same effect

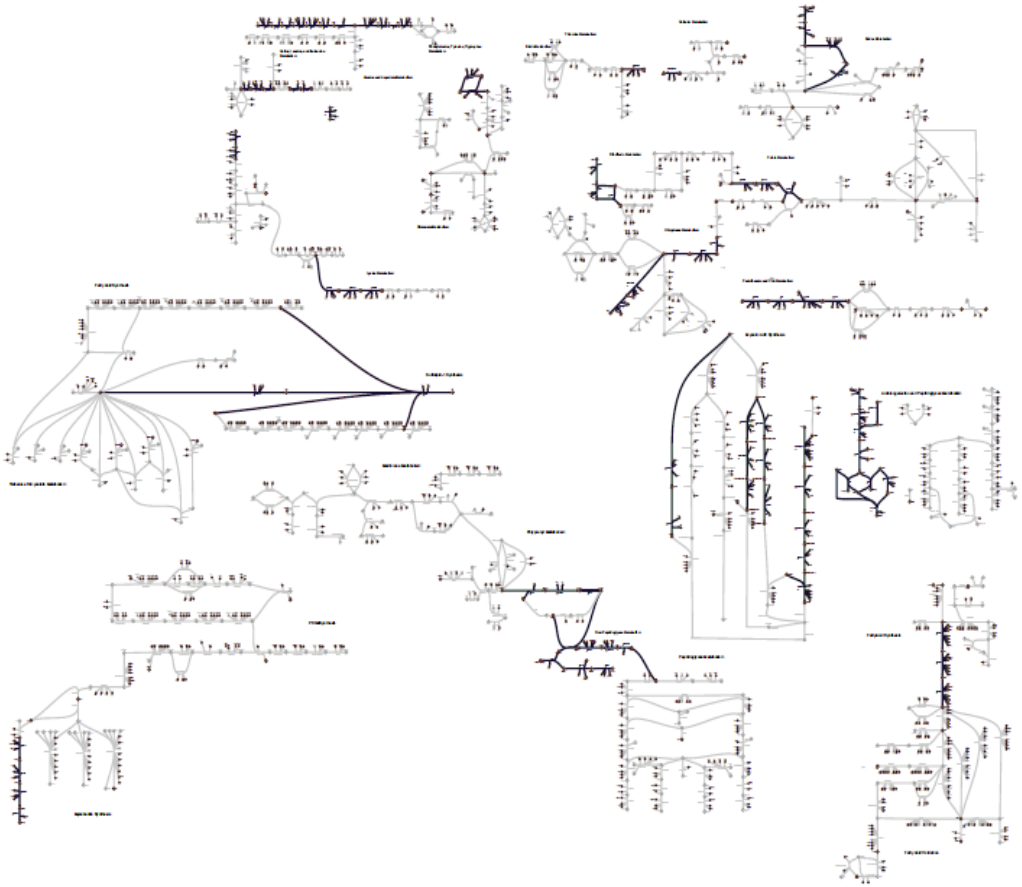


Coupling reaction example





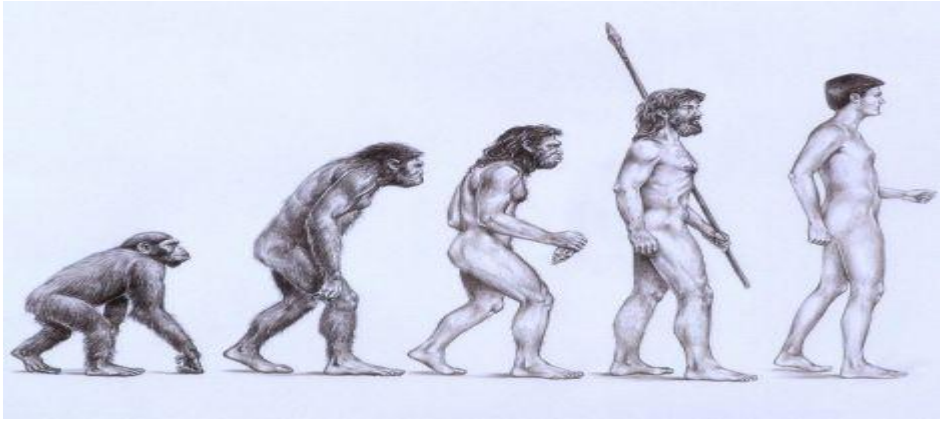
25 drug targets –
CR sets



Exploring evolutionary relationships

- Investigating functional evolution of metabolic and regulatory networks

- Deciphering the adaptive properties underlying the structure and function of metabolic networks



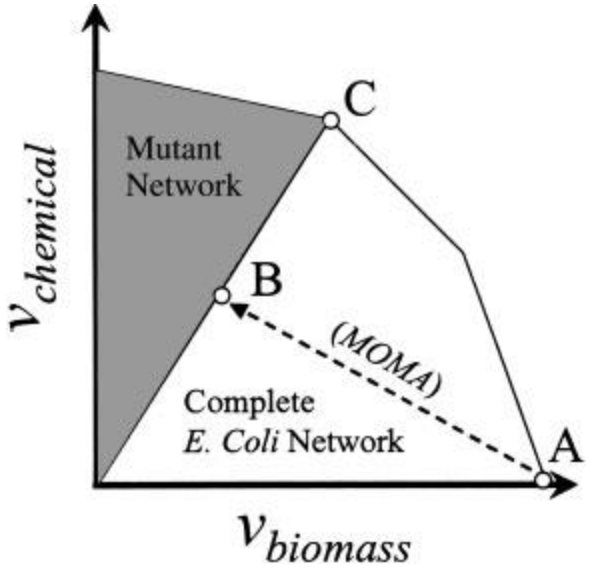
In Silico Design and Adaptive Evolution of Escherichia coli for Production of Lactic Acid

(Fong et al, 2005)

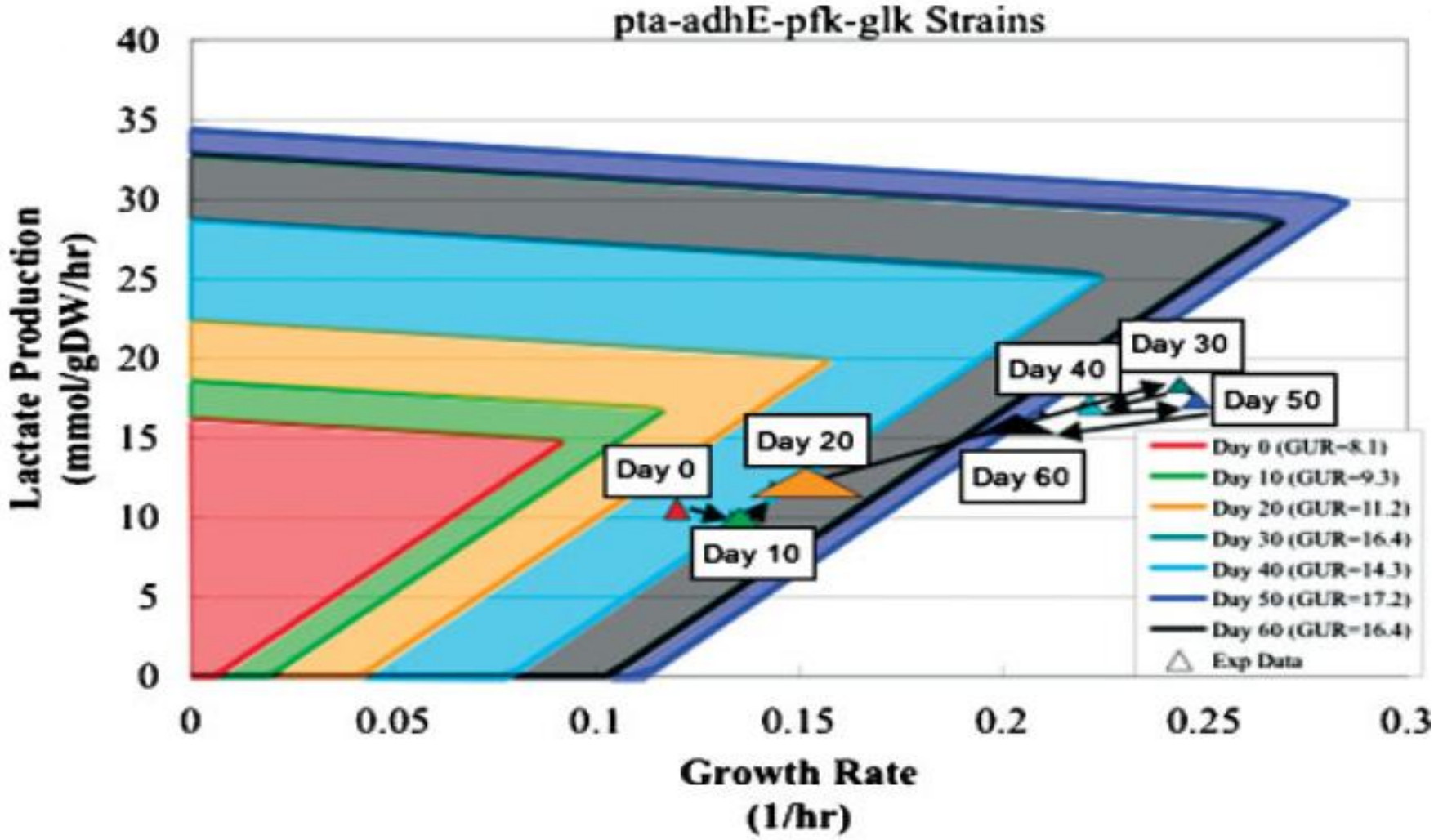
- Strain design for metabolite overproduction - OptKnock
- Adaptive evolution of the engineered strains can lead to improved production capabilities

maximize bioengineering objective
(through gene knockouts)
subject to
maximize cellular objective
(over fluxes)
subject to
- fixed substrate uptake
- network stoichiometry
- blocked reactions identified by outer problem

number of knockouts ≤ limit



Results


pta-adhE-pfk-glk Strains
Lactate Production (mmol/gDW/hr)
40
35
30
25
20
15
10
5
0
0
0.05
0.1
0.15
0.2
0.25
0.3
Growth Rate (1/hr)
Day 0
Day 10
Day 20
Day 30
Day 40
Day 50
Day 60
Day 0 (GUR=8.1)
Day 10 (GUR=9.3)
Day 20 (GUR=11.2)
Day 30 (GUR=16.4)
Day 40 (GUR=14.3)
Day 50 (GUR=17.2)
Day 60 (GUR=16.4)
Exp Data

Current status of genome-scale metabolic reconstructions

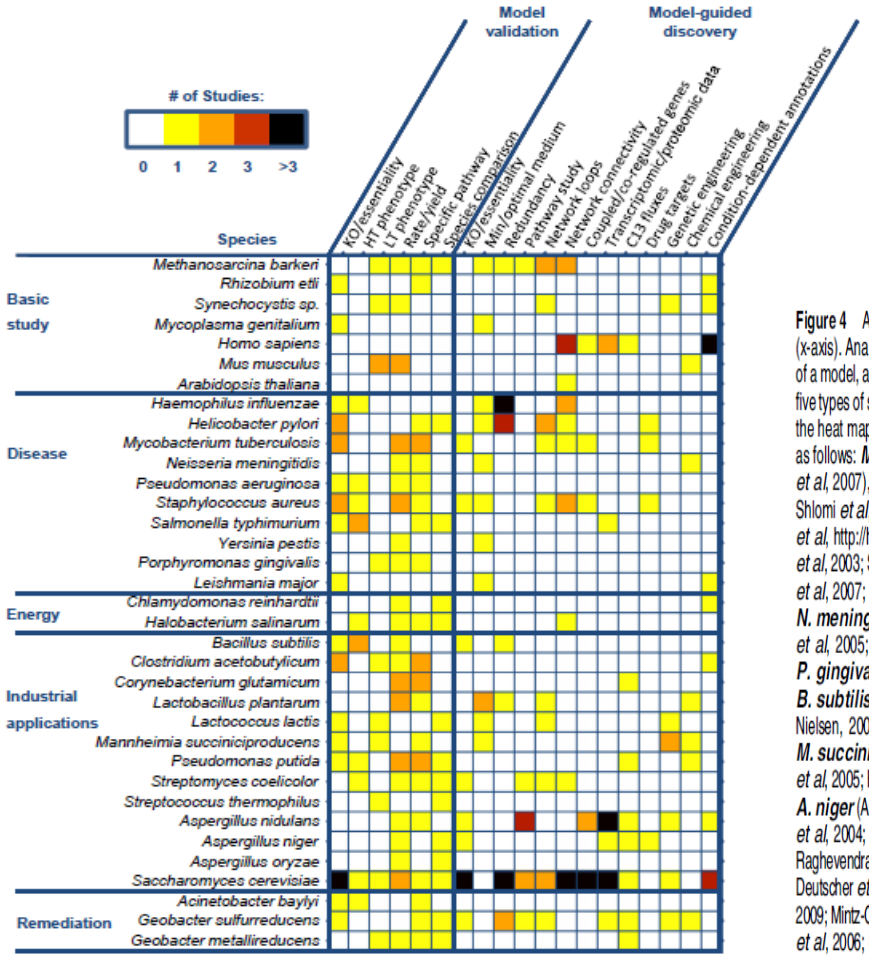


Figure 4 Analyses of metabolic GENREs. A heat map of studies that have been published for all reconstructed species (y-axis) using a variety of analysis techniques (x-axis). Analyses are broken into two categories: 'model validation,' indicating the use of a technique in a publication of a new metabolic GENRE to establish the validity of a model, and 'model-guided discovery,' indicating the use of a technique either in an original metabolic GENRE publication or in a follow-up study to perform one of the five types of studies outlined in this paper. Species are grouped according to the most relevant broad category to which the metabolic GENRE has been applied. Colors in the heat map indicate the number of publications performing the given analysis on a metabolic GENRE of the organism. References for each species in the figure are as follows: ***M. barkeri*** (Becker *et al*, 2006; Feist *et al*, 2006; Kun *et al*, 2008; Mahadevan and Lovley, 2008; Wright and Wagner, 2008), ***R. etli*** (Resendis-Antonio *et al*, 2007), ***Synechocystis sp.*** (Kun *et al*, 2008), ***M. genitalium*** (Suthers *et al*, 2009), ***H. sapiens*** (Duarte *et al*, 2007; Ma *et al*, 2007; Vo *et al*, 2007; Shlomi *et al*, 2008, 2009; Veeramani and Bader, 2009), ***M. musculus*** (Sheikh *et al*, 2005; Quek and Nielsen, 2008; Selvarasu *et al*, 2009), ***A. thaliana*** (Radrich *et al*, http://hdl.handle.net/10101/npre.2009.3309.1), ***H. influenza*** (Edwards and Palsson, 1999; Papin *et al*, 2002a; Papin *et al*, 2002b; Price *et al*, 2002; Price *et al*, 2003; Schilling and Palsson, 2000), ***H. pylori*** (Price *et al*, 2002, 2003; Schilling *et al*, 2002; Papin *et al*, 2002b; Thiele *et al*, 2005; Becker *et al*, 2006; Guimera *et al*, 2007; Kun *et al*, 2008; Wright and Wagner, 2008), ***M. tuberculosis*** (Beste *et al*, 2009; Beste *et al*, 2007; Jamshidi and Palsson, 2007; Kun *et al*, 2008), ***N. meningitides*** (Baart *et al*, 2007a, 2007b), ***P. aeruginosa*** (Oberhardt *et al*, 2008), ***S. aureus*** (Becker and Palsson, 2005; Becker *et al*, 2006; Heinemann *et al*, 2005; Kun *et al*, 2008; Samal *et al*, 2006), ***S. typhimurium*** (Abuoun *et al*, 2009; Raghunathan *et al*, 2009), ***Y. pestis*** (Navid and Almaas, 2009), ***P. gingivalis*** (Mazumdar *et al*, 2009), ***L. major*** (Chavali *et al*, 2008b), ***C. reinhardtii*** (Boyle and Morgan, 2009), ***H. salinarum*** (Gonzalez *et al*, 2008), ***B. subtilis*** (Oh *et al*, 2007; Henry *et al*, 2009a, b), ***C. acetobutylicum*** (Lee *et al*, 2008a; Senger and Papoutsakis, 2008a), ***C. glutamicum*** (Kjeldsen and Nielsen, 2009; Shinfuku *et al*, 2009), ***L. plantarum*** (Teusink *et al*, 2006, 2009; Stevens *et al*, 2008), ***L. lactis*** (Oliveira *et al*, 2005; Kun *et al*, 2008), ***M. succiniciproducens*** (Kim *et al*, 2007; Song *et al*, 2008; Lee *et al*, 2008c), ***P. putida*** (Nogales *et al*, 2008; Puchalka *et al*, 2008), ***S. coelicolor*** (Borodina *et al*, 2005; Hiratsuka *et al*, 2008; Kun *et al*, 2008), ***S. thermophilus*** (Pastink *et al*, 2009), ***A. nidulans*** (David *et al*, 2006, 2008; Panagiotou *et al*, 2008, 2009), ***A. niger*** (Andersen *et al*, 2008; Thykaer *et al*, 2009), ***A. oryzae*** (Vongsangnak *et al*, 2008), ***S. cerevisiae*** (Famili *et al*, 2003; Forster *et al*, 2003; Daran-Lapujade *et al*, 2004; Duarte *et al*, 2004; Prinz *et al*, 2004; Kuepfer *et al*, 2005; Patil and Nielsen, 2005; Becker *et al*, 2006; Cakir *et al*, 2006; Herrgard *et al*, 2006, 2008; Raghevendran *et al*, 2006; Samal *et al*, 2006; Usaite *et al*, 2006; Bundy *et al*, 2007; Harrison *et al*, 2007; Rokhlenko *et al*, 2007; Shlomi *et al*, 2007; Chechik *et al*, 2008; Deutscher *et al*, 2008; Kun *et al*, 2008; Mahadevan and Lovley, 2008; Nookaew *et al*, 2008; Notebaart *et al*, 2008; Wright and Wagner, 2008; Zelle *et al*, 2008; Cimini *et al*, 2009; Mintz-Oron *et al*, 2009; Mo *et al*, 2009), ***A. baylyi*** (Durot *et al*, 2008), ***G. sulfurreducens*** (Izallalen *et al*, 2008; Kun *et al*, 2008; Leang *et al*, 2009; Mahadevan *et al*, 2006; Mahadevan and Lovley, 2008; Risso *et al*, 2008; Scheibe *et al*, 2009; Segura *et al*, 2008), ***G. metallireducens*** (Sun *et al*, 2009).

THE DAILY NEWS

www.dailynews.com THE WORLD'S FAVOURITE NEWSPAPER - Since 1879

Ruppin's group takes over the field

H.Sapiens (Jerby et al, 2009; Folger et al, 2009 Benyamini et al, 2010) **E.coli** (Shabi et al, 2010; Yizhak et al, 2010; Zur et al, 2010) **M.tubercculosis** (Eilam et al, 2010; Zarecki et al ,2010; Stempler et al, 2010) **Y.pestis** (Jerby et al, 2011; Folger et al, 2011 Benyamini et al, 2011) **P.putida** (Shabi et al, 2011; Yizhak et al, 2011; Zur et al, 2011; Mintz et all, 2011 **S.cerevisiae** (Eilam et al, 2011; Zarecki et al ,2011; Stempler et al, 2011)

A.nidulans (Jerby et al, · 2012; Folger et al, 2012 Benyamini et al, 2012) **A.baylyi** (Shabi et al, 2012; Yizhak et al, 2012; Zur et al, 2012) **A.oryzae** (Eilam et al, 2012; Zarecki et al ,2012; Stempler et al, 2012) **S.coelicolor** (Jerby et al, 2013; Folger et al, 2013 Benyamini et al, 2013) **L.plantarum** (Shabi et al, 2013; Yizhak et al, 2013; Zur et al, 2013; Mintz et al. ,2013) **M.musculus** (Eilam et al, 2011; Zarecki et al ,2011; Stempler et al, 2011)

Towards a new era

High-Throughput Reconstruction and Optimization of 130 New Genome-Scale Metabolic Models