Understanding and predicting genetic interactions in yeast metabolism

Balázs Papp

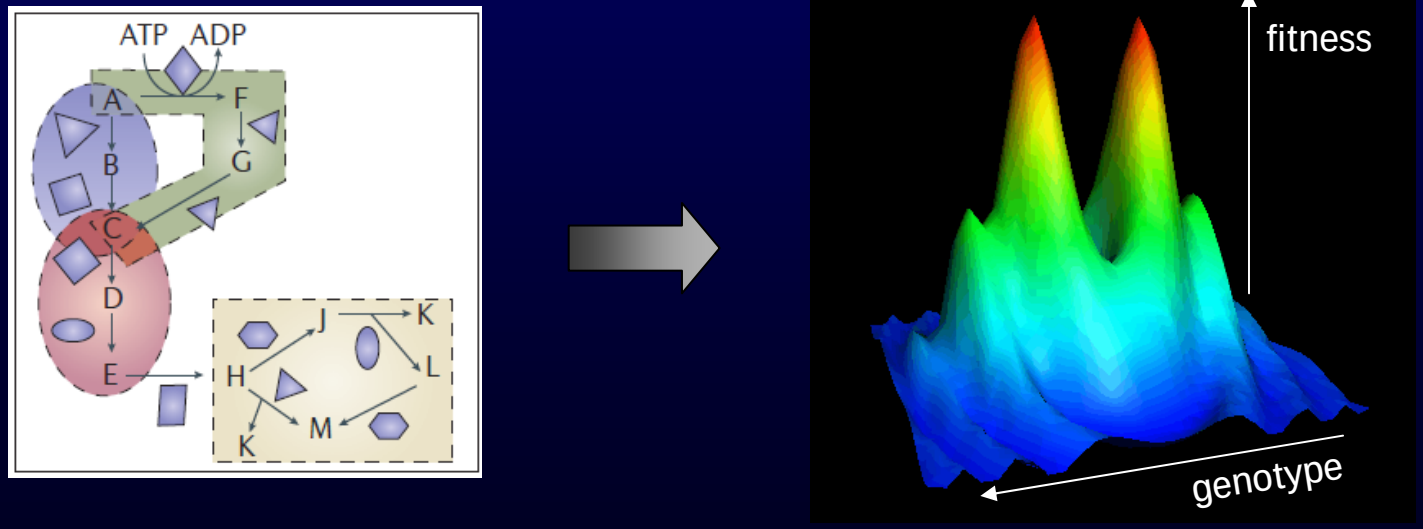
www.brc.hu/sysbiol

Biological Research Center
Szeged, Hungary

Cambridge Systems Biology Centre
Cambridge, UK

A major goal of evolutionary systems biology it to understand fitness landscapes

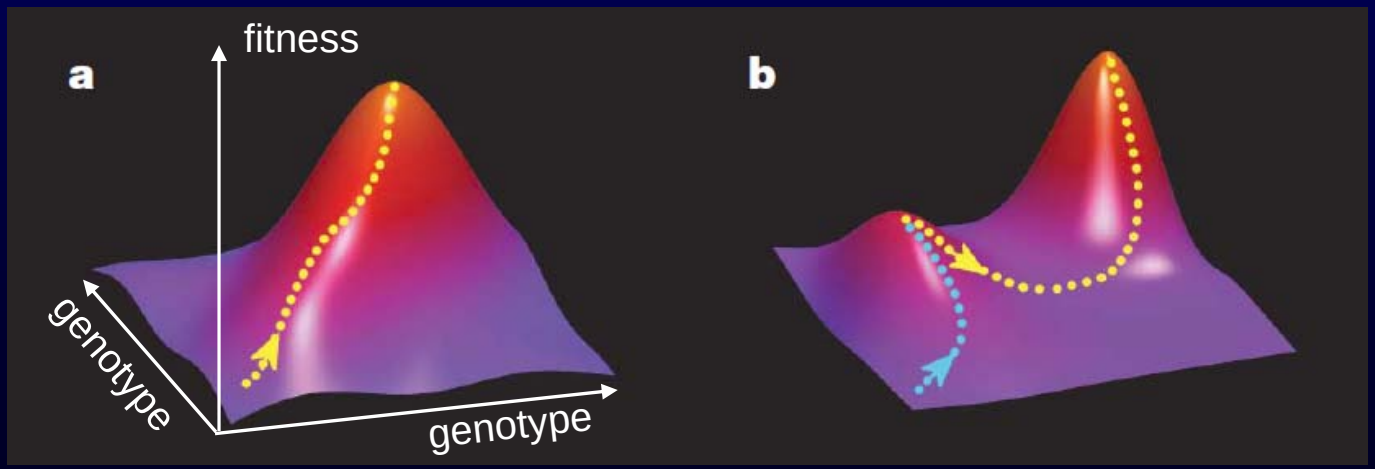
- Systems biology models provide a mapping between genotype and phenotype
- A framework to mechanistically understand mutational effects (fitness landscape)



fitness landscape

Why is it important to understand the genotype – phenotype map?

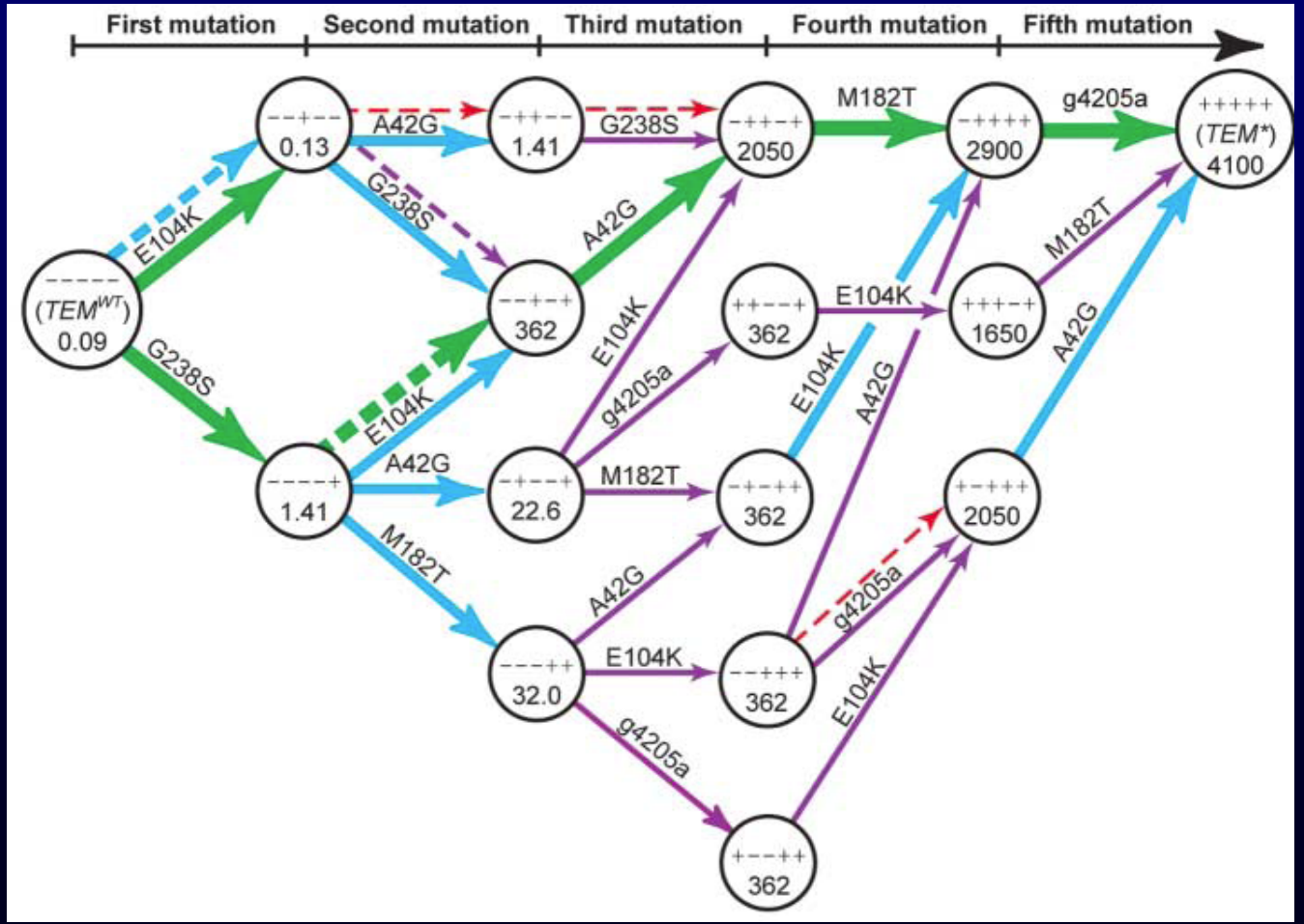
- Provides functional information
- Important for evolution: the shape of fitness landscape determines the accessible evolutionary trajectories



Example: a beta-lactamase allele conferring resistance *

- 5 point mutations increase resistance by 100,000-fold
- 102 trajectories are inaccessible out of the 120 possible mutational path linking wilde-type and resistant allele

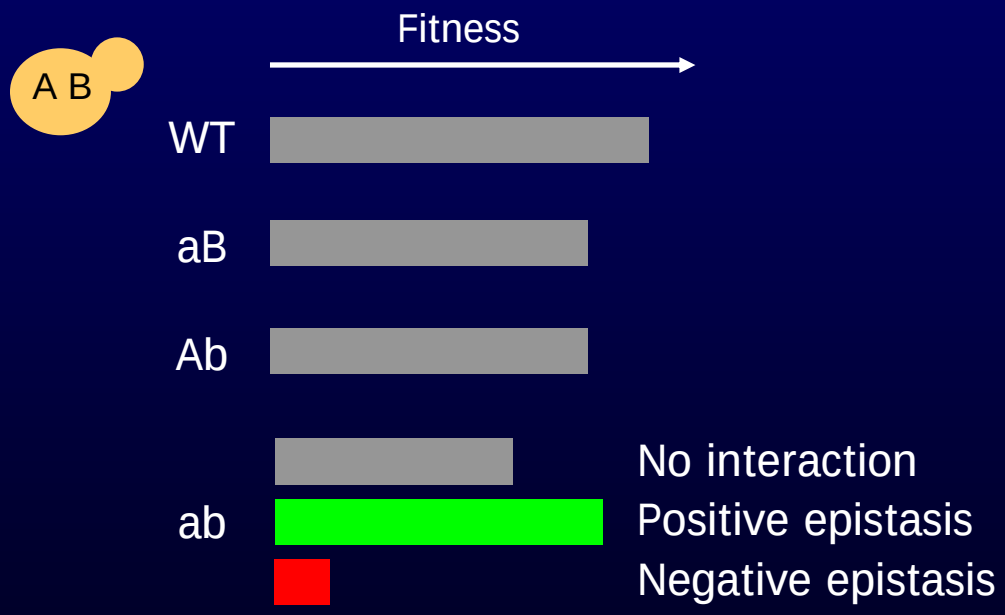
10 most probable trajectories

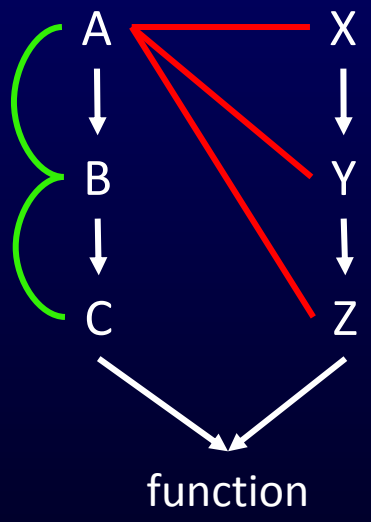


* Weinreich et al. 2006 Science 312: 111

Genetic interaction (epistasis) refers to the non-independence of mutational effects

Epistasis between two gene deletions:



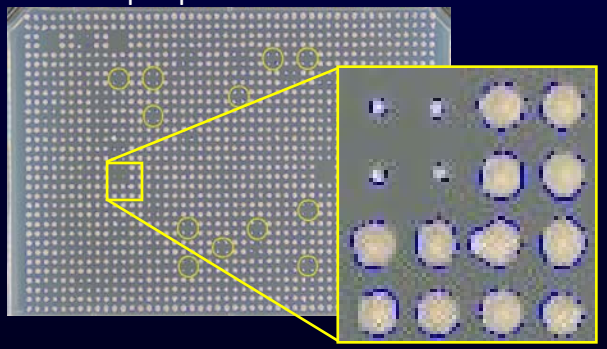


Generating epistasis data: metabolic SGA miniarray

Quantitative epistasis data generated by Charles Boone's lab in *S. cerevisiae*:

- ~ 378 000 gene pairs tested (~1000 genes), 1246 negative and 322 positive interactions identified
- Genes include enzymes, transporters and regulators

SGA output plate



Colony detection

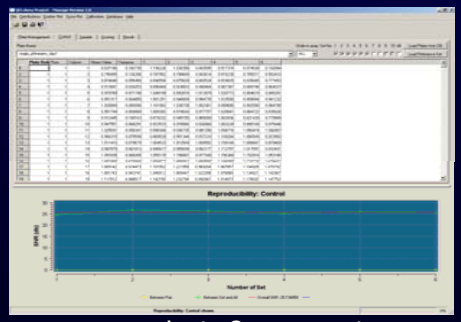


Image analysis & processing

Data Correction:

1) Systematic effects (plate/position)

2) Competition effects

3) Measure DM fitness
   Estimate SM fitness

What network properties correlate with epistasis?

We use the latest genome-scale metabolic network reconstruction of yeast*:

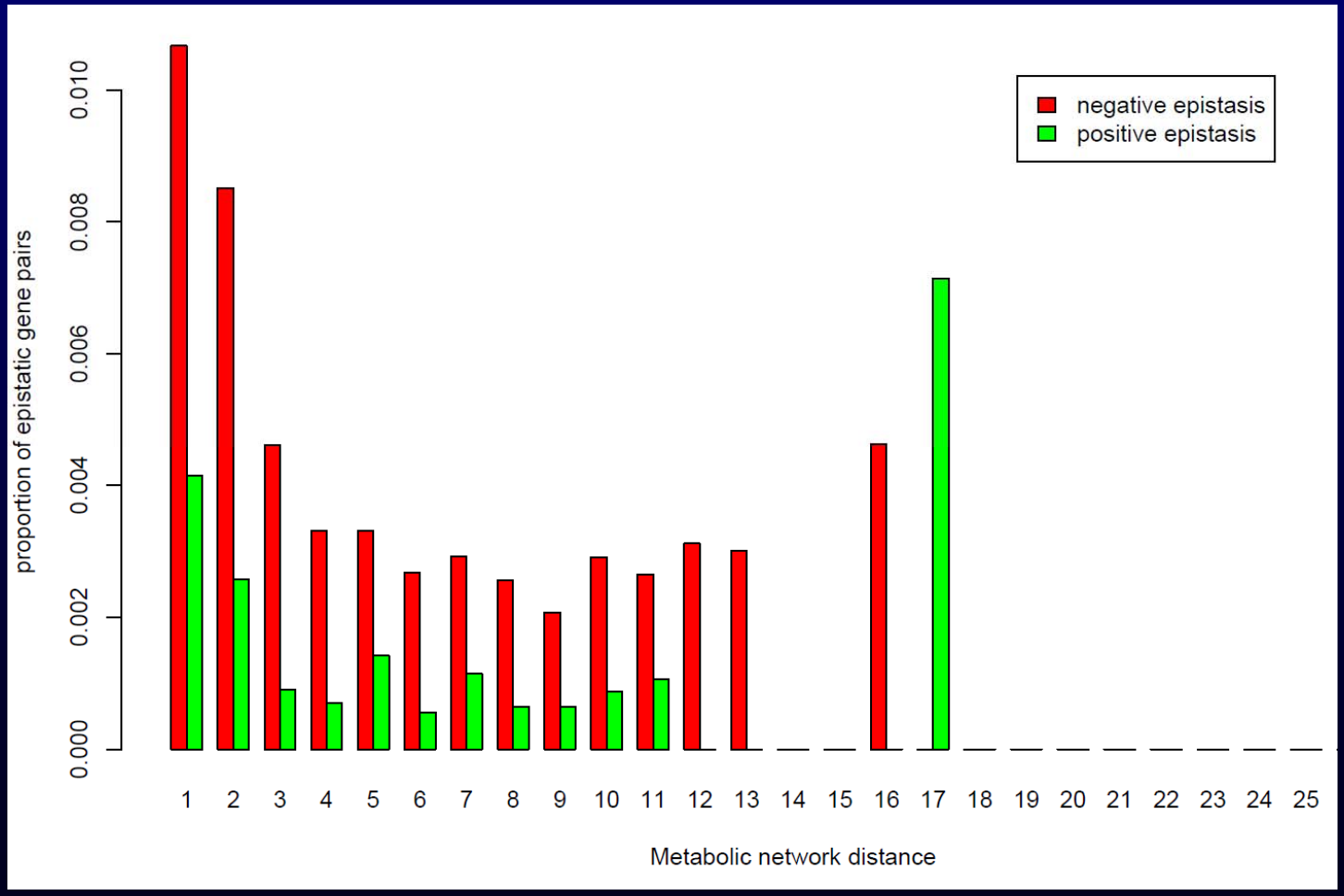
• 904 genes, 1412 biochemical reactions (395 transport) connecting 1228 metabolites

• Information on isoenzymes, enzyme complexes, reaction reversibility, cellular compartments, transport reactions.

* Mo et al. (2009) *BMC Syst Biol* **3**: 37

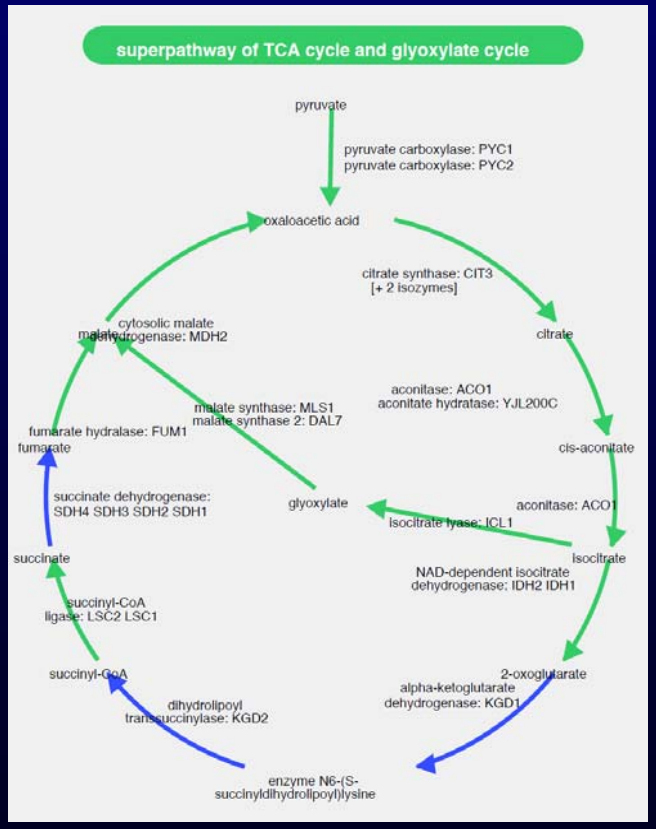
Interacting gene pairs are close in the network

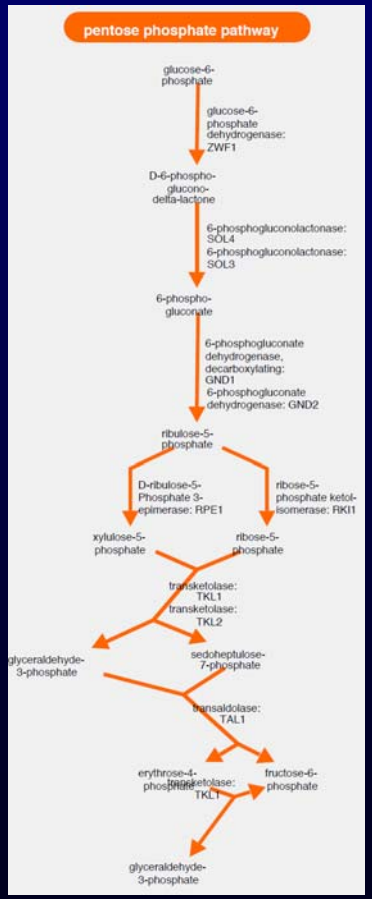
Both positive and negative genetic interactions are enriched among gene pairs separated by short network distances

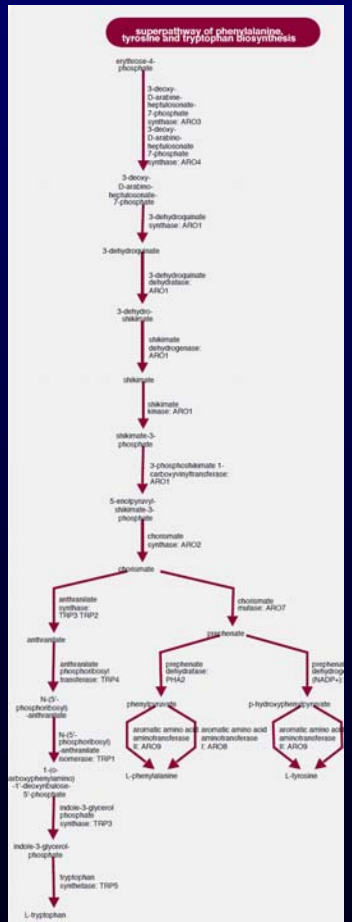


Interacting gene pairs are often in the same metabolic subsystem

Both positive (5x) and negative (3.9x) genetic interactions are enriched within ’traditionally defined’ metabolic subsystems

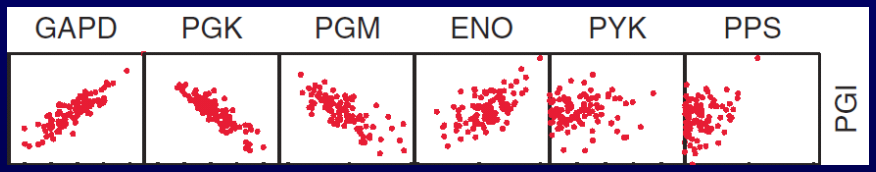




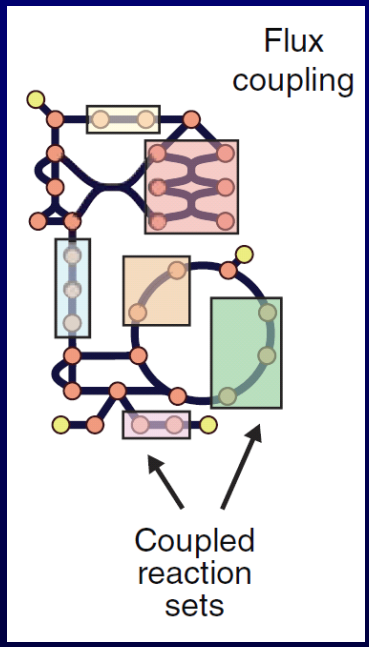


Strong overrepresentation of interactions within functional modules

Unbiased, systems-level module definition:
reaction sets with correlated (coupled) fluxes



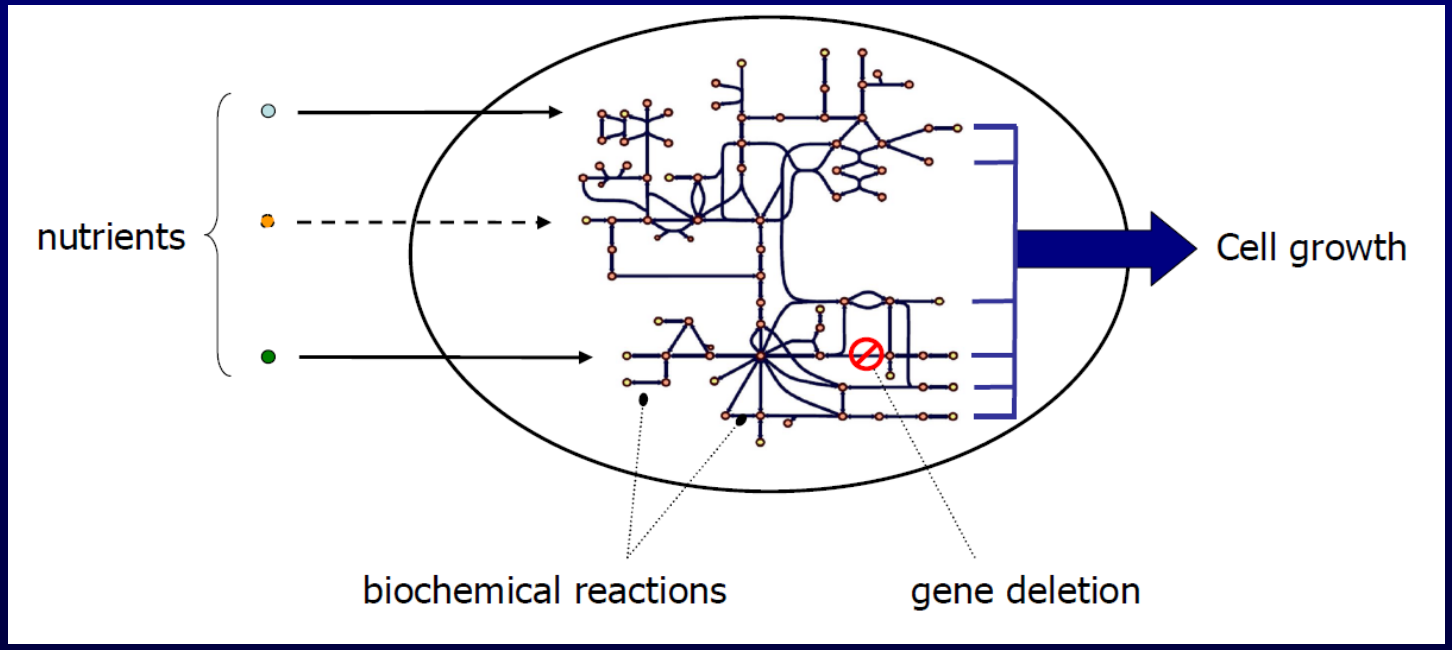
in silico flux correlations



→ Directionally coupled pairs are 35x enriched in positive epistasis and 10x enriched in negative epistasis

Can we predict individual genetic interactions using a systems biology model?

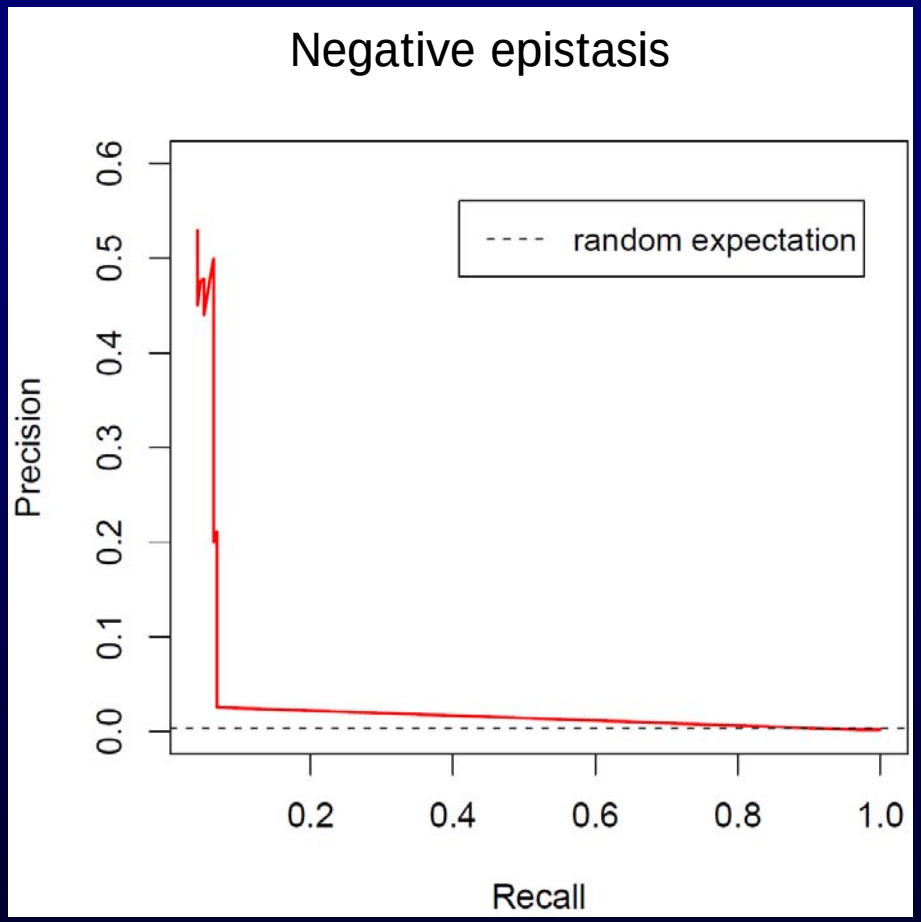
Predicting epistasis using flux balance analysis

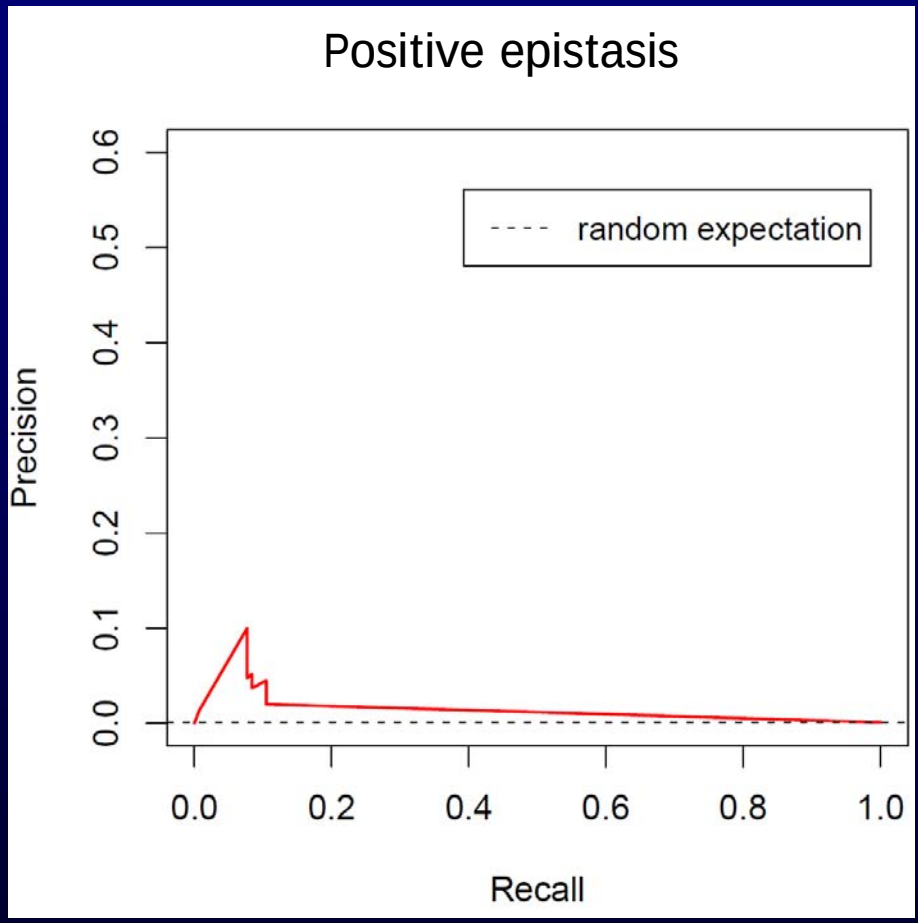


- FBA is good at predicting single mutant viability (~90% accuracy)*
- But it's unknown how well it predicts multiple mutations

* Kuepfer et al. (2005) Genome Res 15: 1421

Prediction accuracy

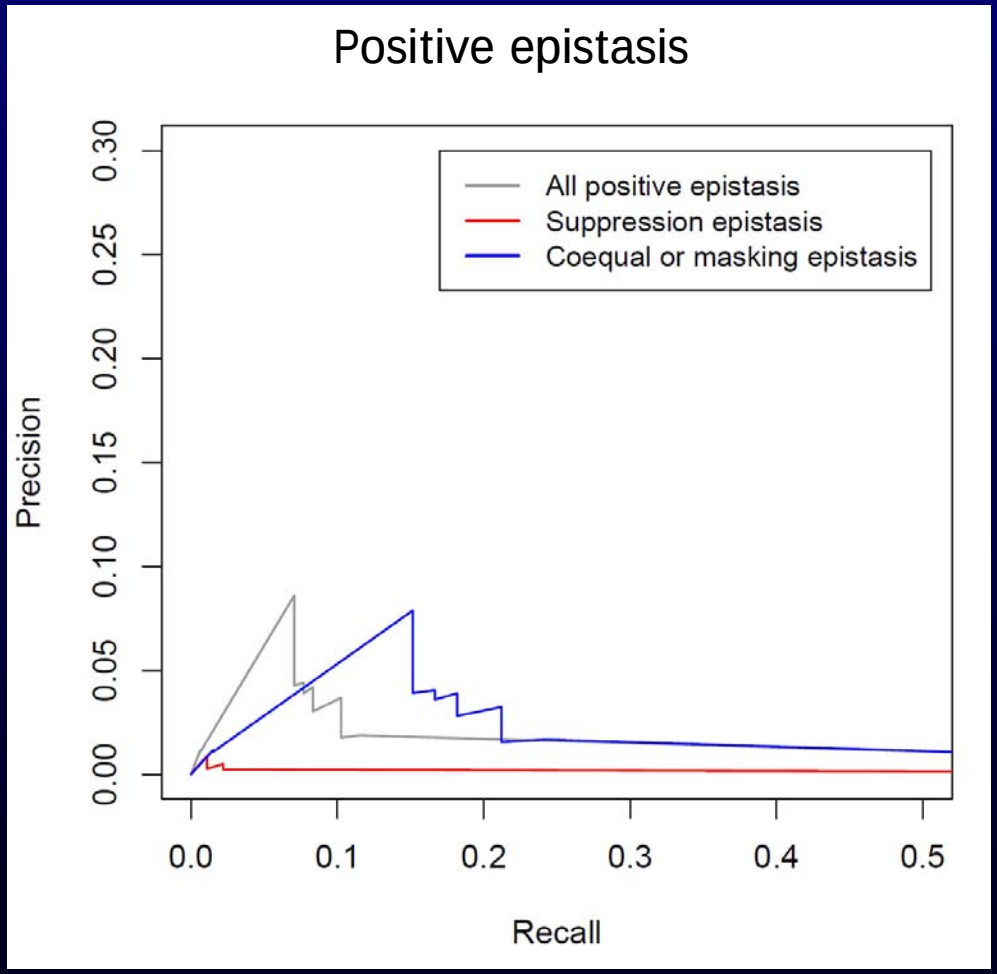




Recall: fraction of true interactions correctly predicted (true positive rate)
Precision: fraction of predicted interactions that are correct (positive predictive value)

The FBA approach cannot predict ‘suppression’ epistasis

- Exclusion of suppression interactions increases recall of positive epistasis predictions by 2-fold



Using the model to understand epistasis among duplicate pairs

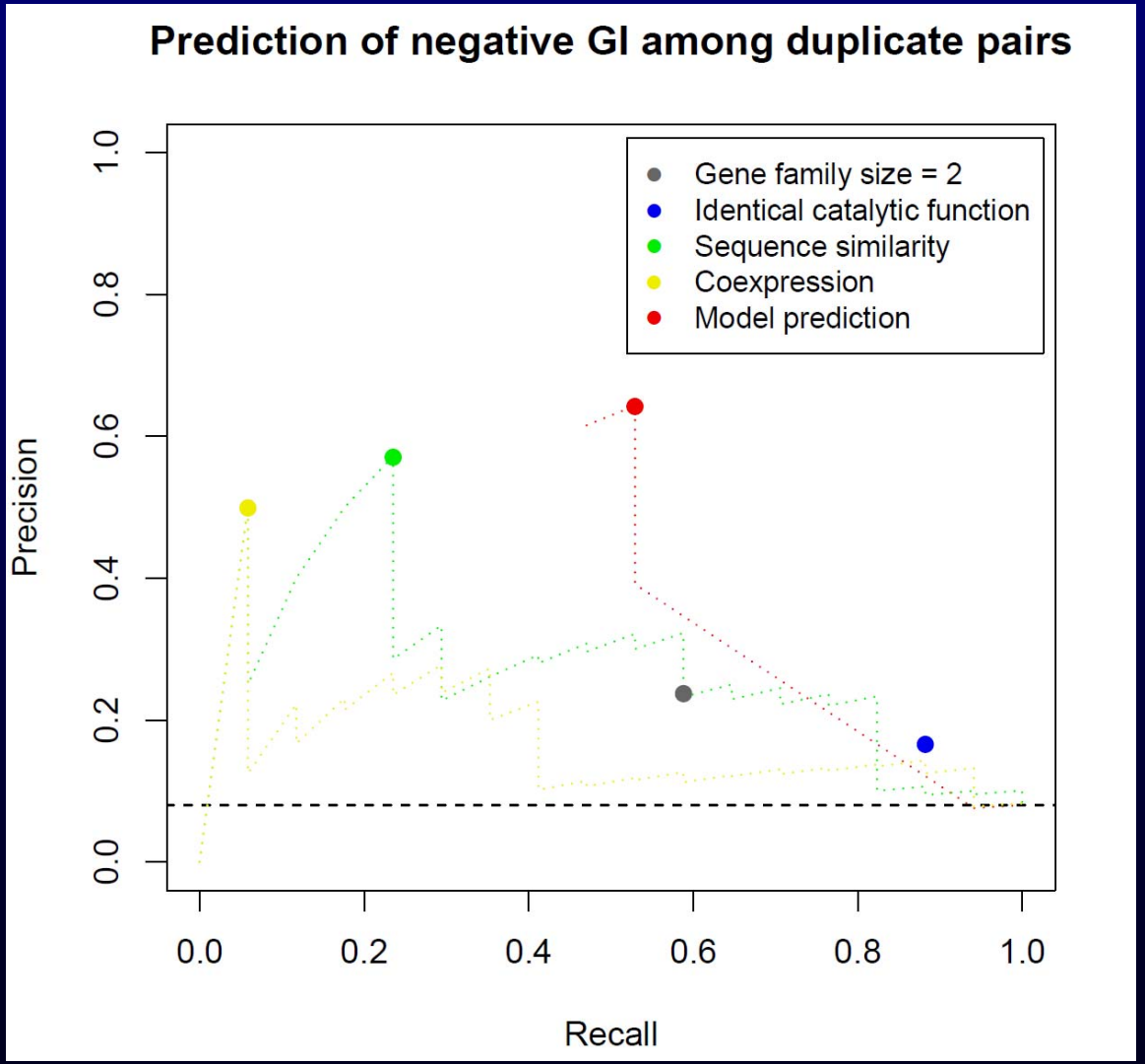
Previous reports: many duplicates do not show negative epistasis under a given condition (e.g., ~ 83%, Musso et al. 2008)

Present study: only ~8% of duplicates in the metabolic network show negative epistasis in the experiments

Can we predict which duplicates show epistasis?

Can we predict which duplicates show negative epistasis?

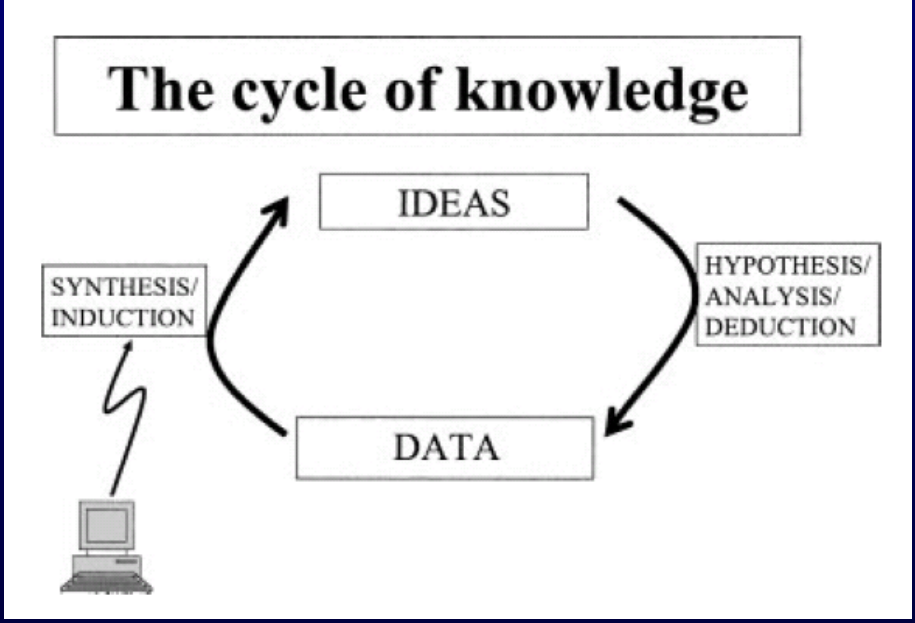
The metabolic model predicts epistatic duplicates with high accuracy, much better than sequence similarity, coexpression, functional identity or gene family size



Conclusions so far

- The genome-scale model can explain some global properties of epistasis networks
- Even the simple case of epistasis among duplicates is a systems-level property that cannot be well captured by measures of duplicate similarity alone
- Apparently, the model has high precision for negative epistasis, but very low recall: misses most observed interactions…

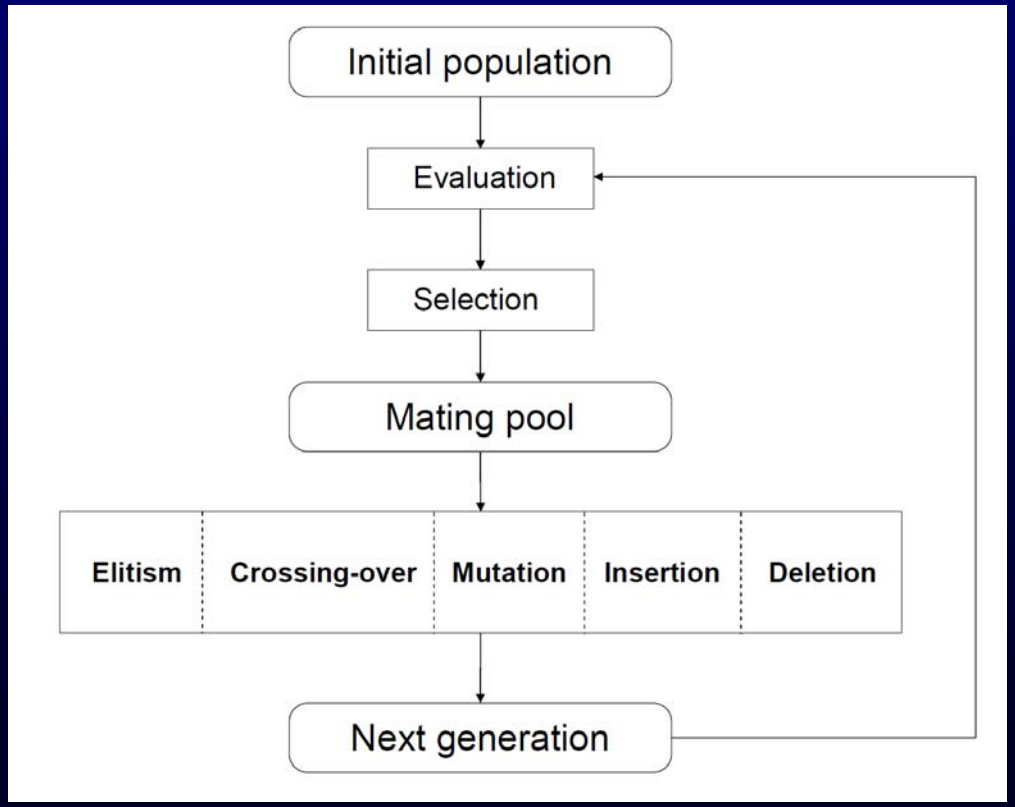
Hypothesis generation and testing is an iterative process



Can we use discrepancies between epistasis data and predictions to update the *in silico* model?

An optimization method to automate network refinement

We used a genetic algorithm* to search for model modifications that improve overall epistasis predictions



* Rocha et al. (2008) *BMC Bioinf* **9**: 499

Evaluation of each modified model is based on how well they discriminate between epistatic and non-epistatic gene pairs

Gene pair	Predicted epistasis	Empirical epistasis
1.	None	None
2.	None	None
3.	None	SL
4.	None	None
5.	None	None
…	…	…

Modification to model

Gene pair	Predicted epistasis	Empirical epistasis
1.	None	None
2.	None	None
3.	SL	SL
4.	None	None
5.	None	None
…	…	…

Model parameters being optimized:

- biomass composition
- reaction presence / absence

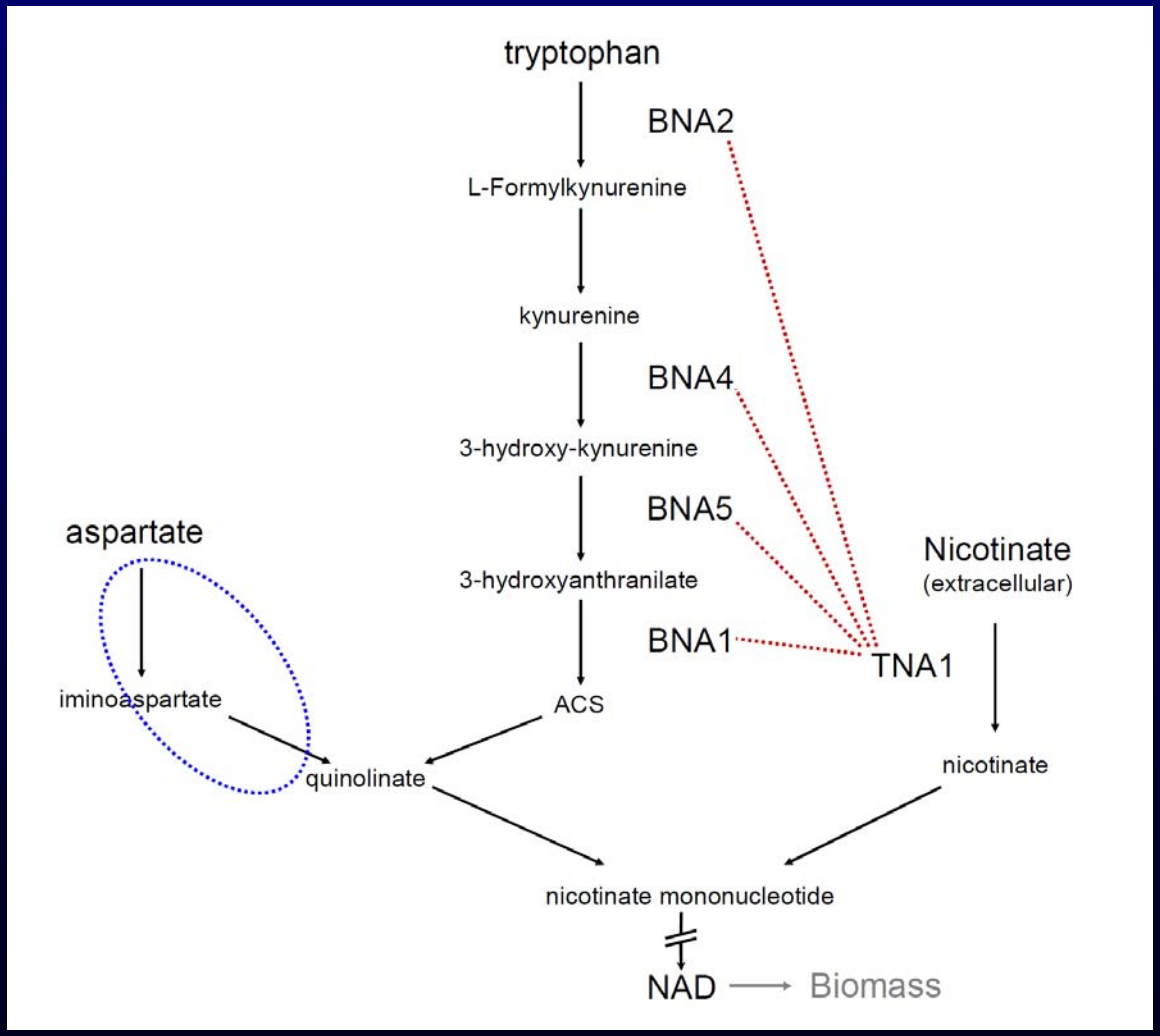
Few changes can double the true positive rate

A non-redundant set of suggested modifications from a single run:

Reaction to be inactivated	Increase in true positives when applied individually	Increase in true positives when applied together with other changes
ADK3m	1	6
PDHcm	2	2
QULNS	4	6
RNMK	1	3
SUCC2tr	0	5

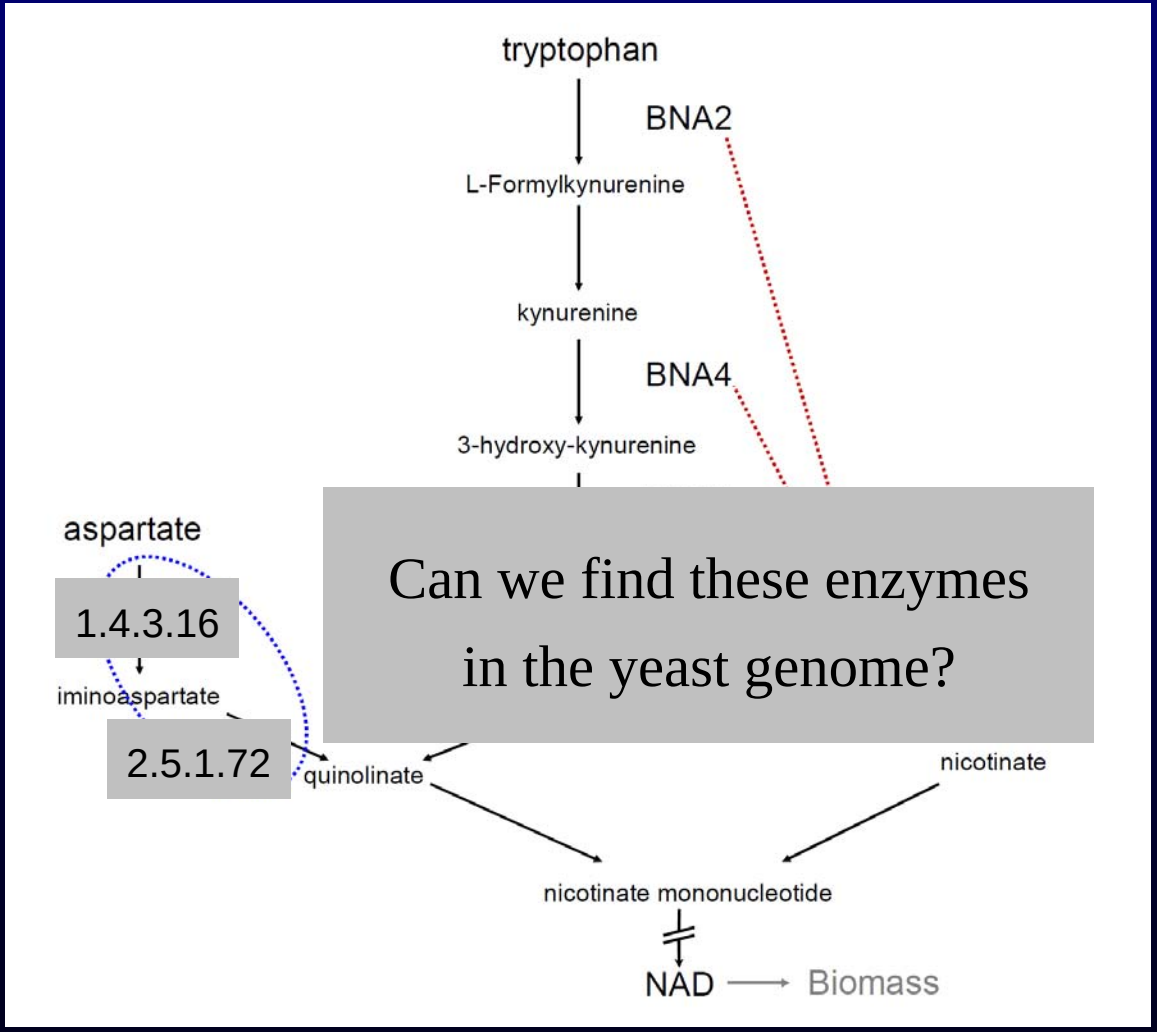
Example of suggested reaction removals

A significant improvement can be made by inactivating a 2-step pathway involved in NAD biosynthesis



Example of suggested reaction removals

A significant improvement can be made by inactivating a 2-step pathway involved in NAD biosynthesis



Support from bioinformatics analysis

EC 1.4.3.16

L-aspartate oxidase: there are some homologous sequences in the yeast genome, but these are other oxidoreductases (succinate dehydrogenase, fumarate reductase). This is supported by crystal structural data on *E. coli* L-aspartate oxidase (these share the same fold).

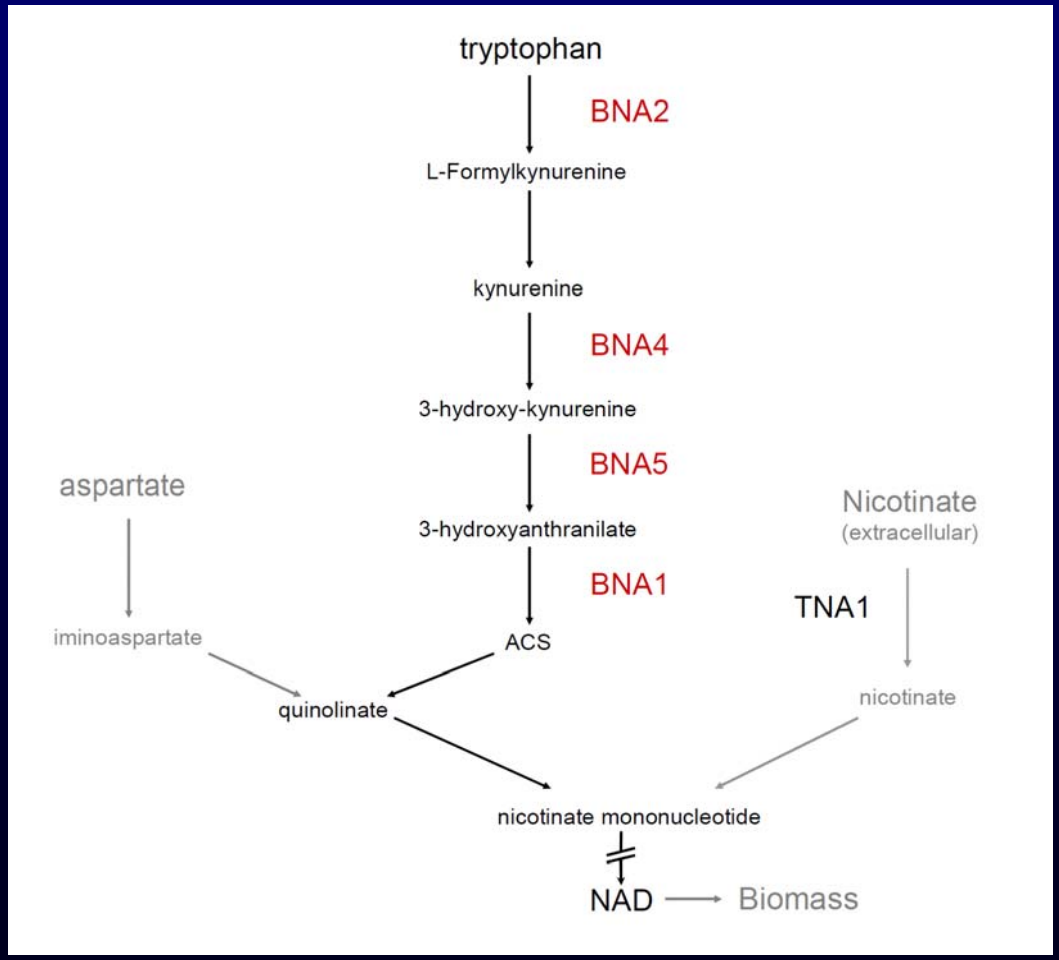
EC 2.5.1.72

Quinolinate synthase: an iterative PSI-BLAST search did not recover any hit from the yeast genome

These reactions might have been erroneously included in the yeast reconstruction

Testing specific predictions of the modified model

Prediction: BNA1, BNA2, BNA4 and BNA5 should be essential in the absence of nicotinate in the medium



Testing specific predictions of the modified model

Experimental support from the literature: Δ*bna1*, Δ*bna2*, Δ*bna4* and Δ*bna5* show growth reduction in the absence of nicotinate (Panozzo et al. 2002)

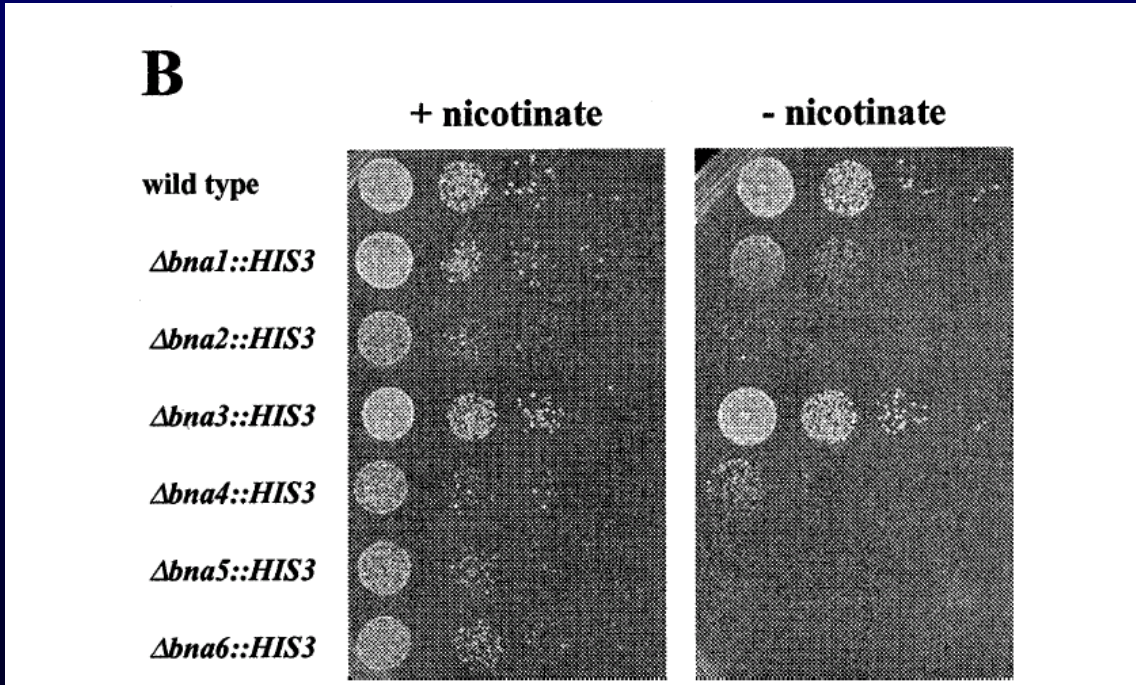


Fig. 3. The nicotinic acid auxotroph phenotype of wild-type *S. cerevisiae* and the Δ*bna* mutants. A: Wild-type *S. cerevisiae* (IW303)

Acknowledgements

Papp and Pál labs (Szeged):

Balázs Szappanos

Károly Kovács

Ferenc Honti

Csaba Pál

Toronto:

Michael Costanzo

Anastasia Baryshnikova

Brenda Andrews

Charles Boone

Cambridge:

Steve Oliver

Düsseldorf:

Gabriel Gelius-Dietrich

Martin Lercher

Minnesota:

Chad Myers

HFSP Human Frontier Science Program

Open bioinformatics postdoc position:

www.brc.hu/sysbiol