Constraint-Based Workshops

5. Gene Deletions
January 24th, 2008

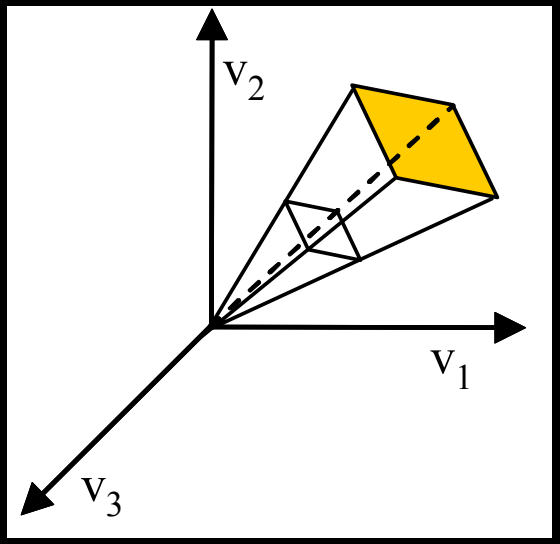
FBA Optimization Problem Statement

- Objective Function: A function that is maximized or minimized to identify optimal solutions
- Constraints: Place limits on the allowable values the solutions can take on.

Maximize: $c \cdot v$

Such that $S \cdot v = b = 0$

$LB \leq v \leq UB$



Review of Shadow Prices & Reduced Costs

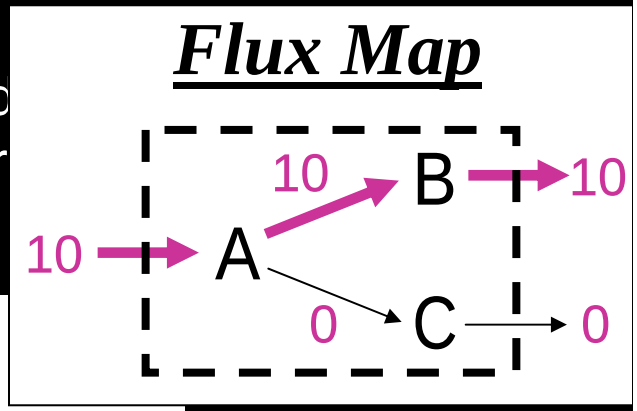
- Shadow Prices (SP):
  - One for each constraint or metabolite
  - dZ/db_i
  - SP<0 means adding metabolite (ie. change b=0 to b<0) would increase Z.
  - SP>0 means removing metabolite (ie. change b=0 to b>0) would increase Z.
- Reduced Costs (RC):
  - One for each variable or flux.
  - dZ/dv_j (for zero fluxes)
  - RC < 0 means increasing flux (v_j) would reduce Z.

Shadow Prices (1 per constraint)

SP_A = -1 →

- If we change b_A from zero to 1: we are saying the production of A has to be higher than the consumption of A by 1 unit (remember S·v=production-consumption).
- A lower consumption of A means that the flux through v1 will have to go down by 1 unit. Hence, dZ/db_A = -1.
- For example, if b_A = 1 then Z = 9.

SP_C=SP_B = EPS (~0) → This is because if you added B or C to the network they wouldn't allow for higher flux through v_1.




```
---- EQU massbalance

     LOWER      LEVEL      UPPER     MARGINAL

A      .          .          .        -1.000
B      .          .          .          EPS
C      .          .          .          EPS
```

Conclusions from Last Time

- You can use shadow prices and reduced costs to evaluate your results.
- For example: If you maximize growth rate and find zero growth, you can identify metabolites which are needed in order to grow (those with a negative shadow price).
  - This is useful if you are debugging a network.

Metabolic Network Example

Reaction List

v_1: A → B

v_2: A → C

b_1: → A

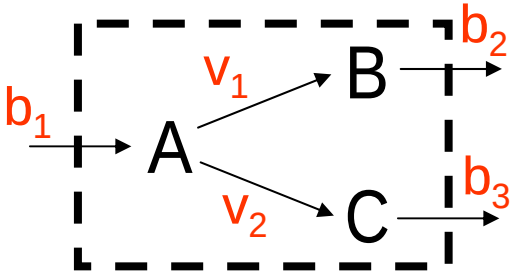
b_2: B →

b_3: C →

S Matrix

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & b_1 & b_2 & b_3 \end{array} \\ \begin{pmatrix} -1 & -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix} \end{array}$$

Metabolic Map



Maximize $Z = c \cdot v = v_1$

Such that $S \cdot v = 0$

$0 \leq v \leq 10$

Gene Deletion Algorithms:

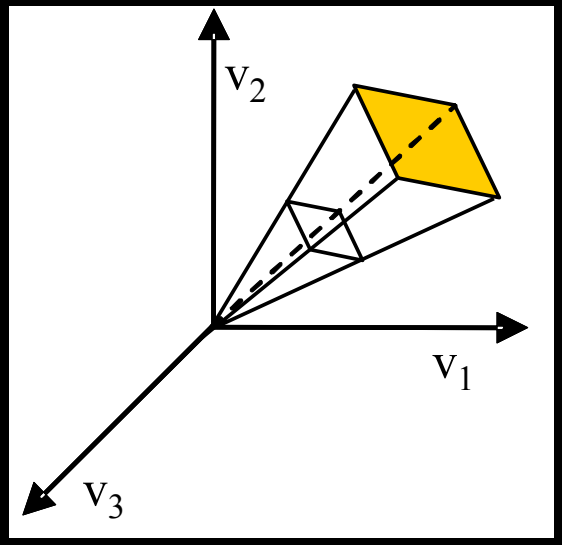
FBA
MOMA
ROOM

FBA: Wildtype vs. Knockout Mutant

Maximize: $c \cdot v$

Such that $S \cdot v = b = 0$

$LB \leq v \leq UB$

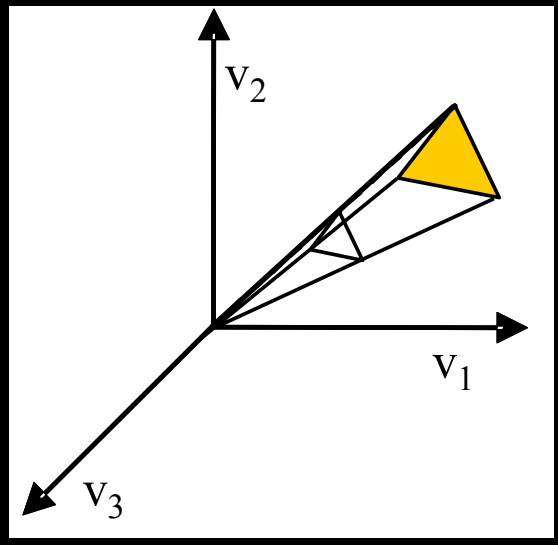


Maximize: $c \cdot v$

Such that $S \cdot v = b = 0$

$LB \leq v \leq UB$

$v_k = 0$

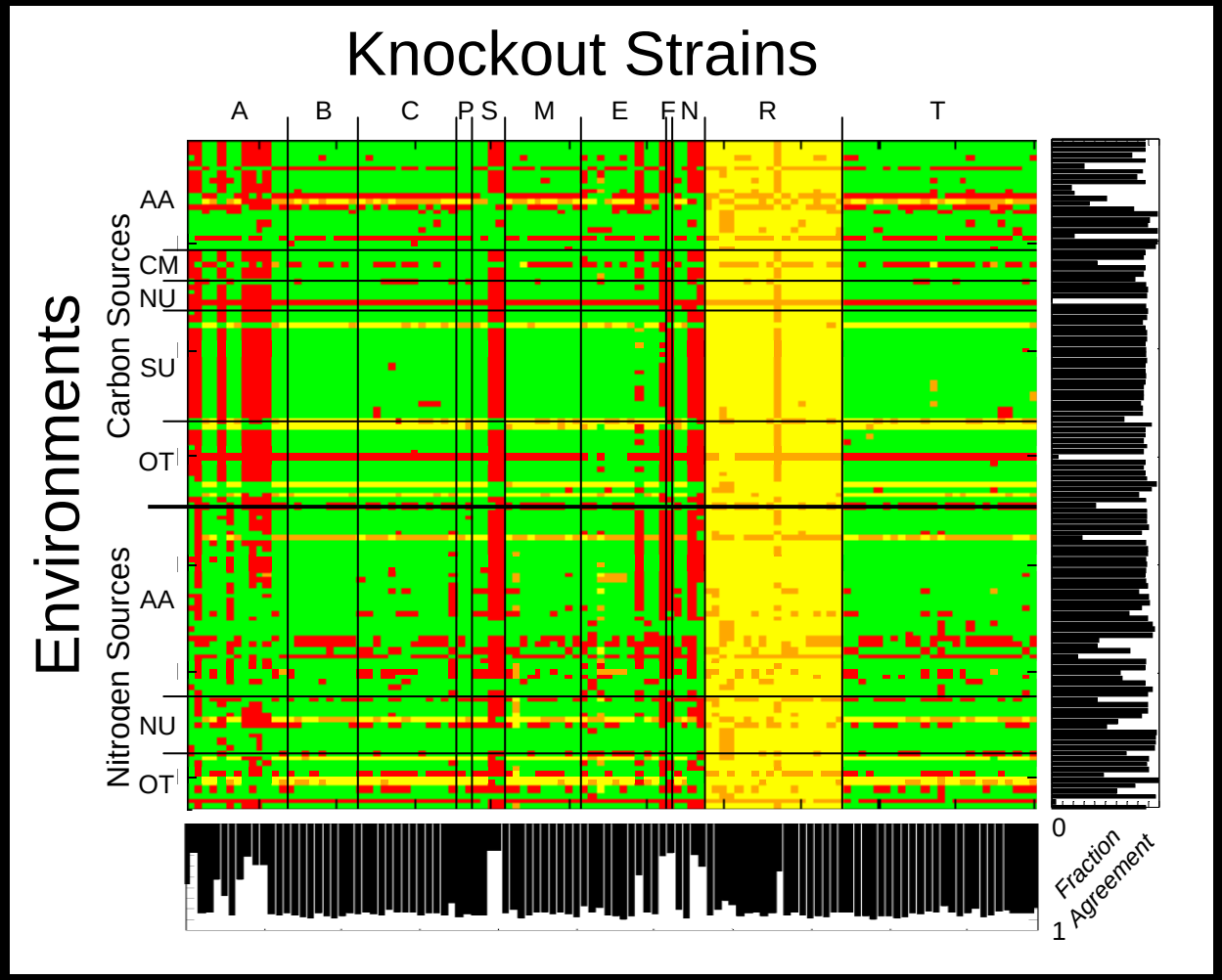


Model-Driven Discovery Via High-Throughput Testing

Comparison of model predictions and observed growth phenotypes

exp/met/reg		Percent
Reg and Met models predict correctly		
+/+/+	6222	45.3%
-/-/-	2094	15.2%
Reg model predicts correctly		
-/+/-	657	4.8%
+/-/+	0	0.0%
+/n/+	1350	9.8%
-/n/-	505	3.7%
Met model only predicts correctly		
+/n/-	242	1.8%
-/n/+	153	1.1%
-/-/+	0	0.0%
+/+/-	257	1.9%
Neither model predicts correctly		
-/+/+	702	5.1%
+/-/-	1568	11.4%
Met	8968	65.2%
Reg	10828	78.7%
Total	13750	100.0%

Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J. and Palsson, B.O. *Nature,* 429(6987):92-6 (2004).

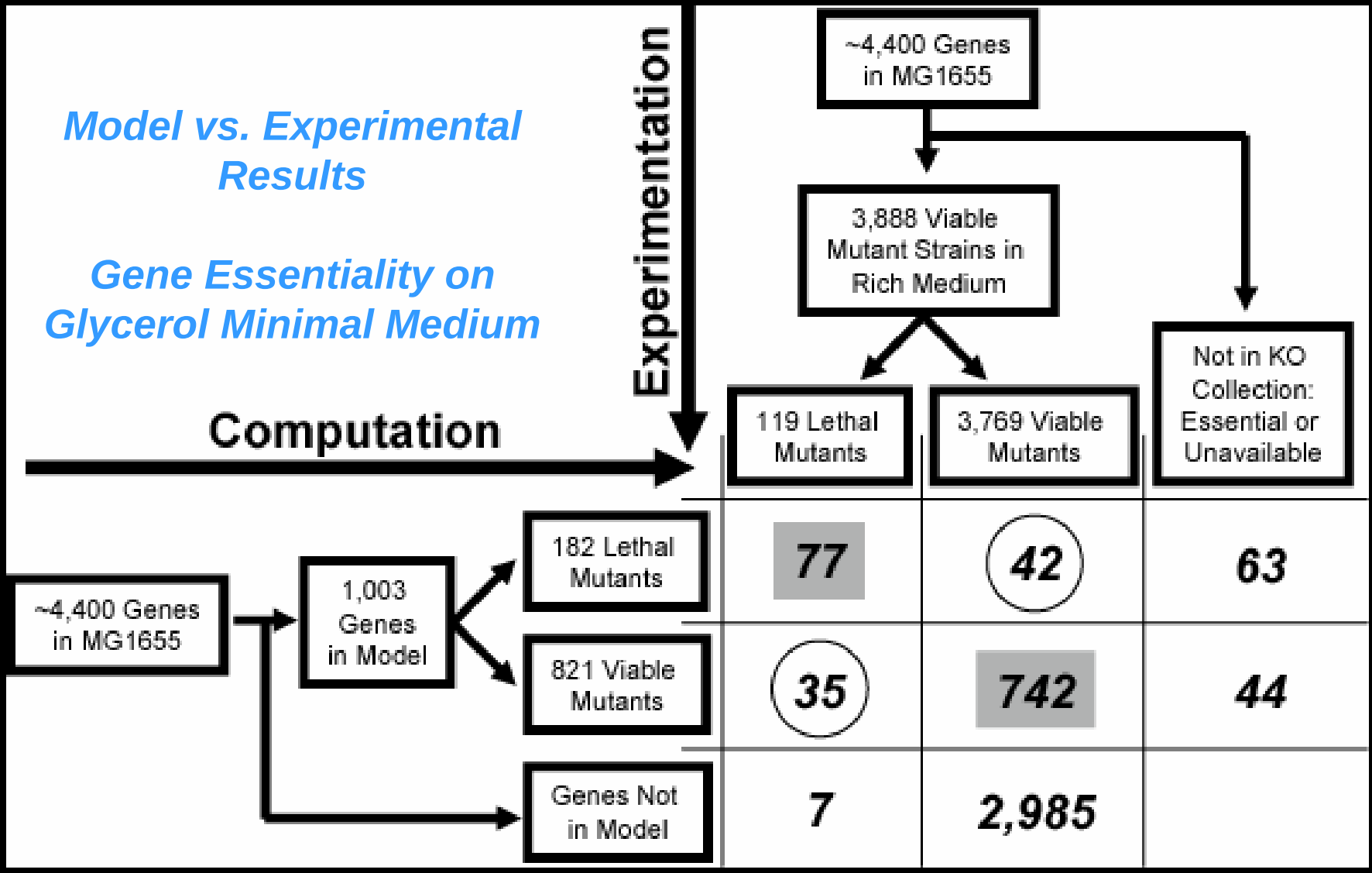


125 Environmental Conditions x 110 KO Mutants

Joyce, et al. J. Bacteriology. 188(23): 8259-8271 (2006)

Model vs. Experimental Results

Gene Essentiality on Glycerol Minimal Medium



Overall Model is 91 % Accurate

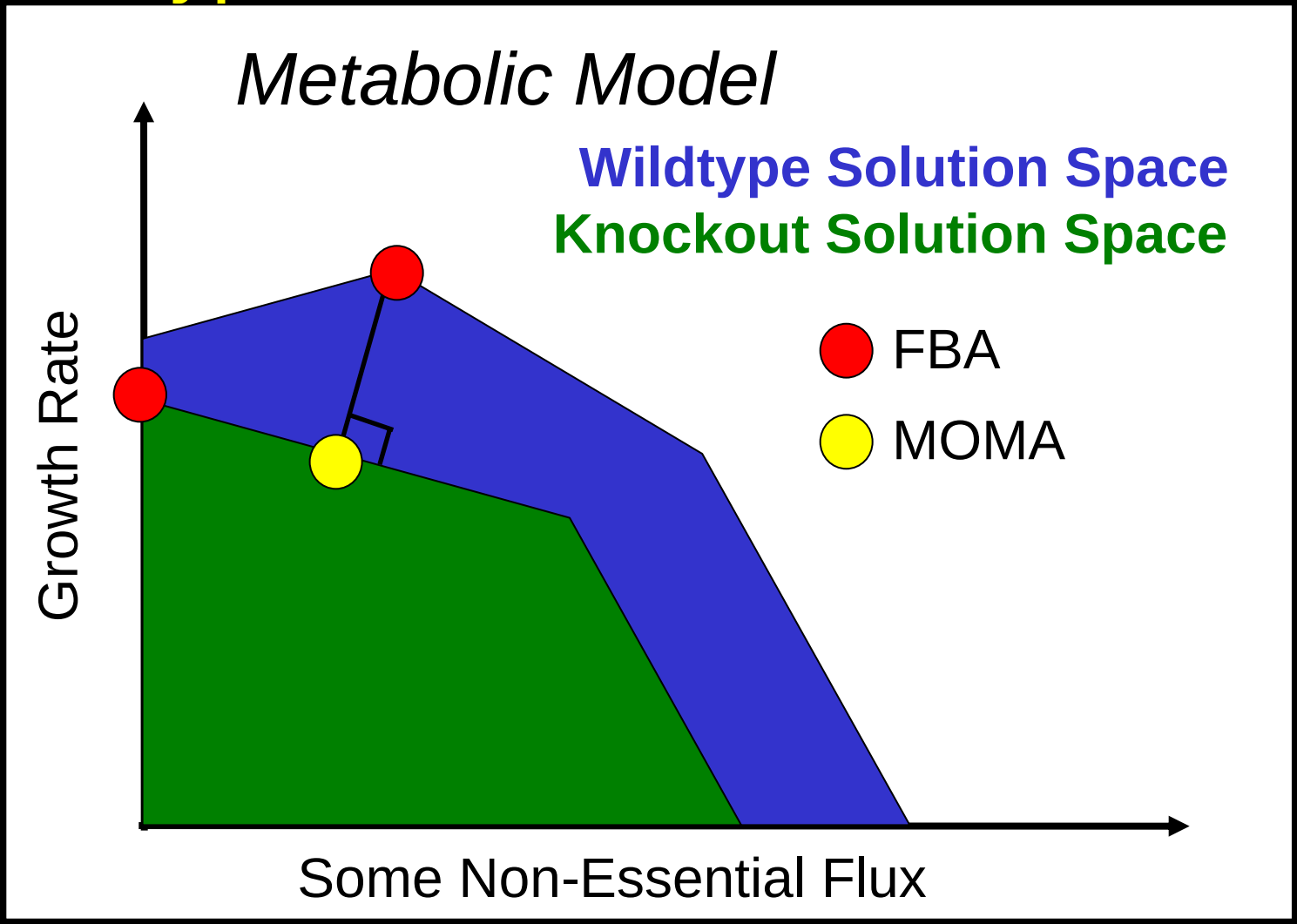
Discrepancies Generate Hypotheses

TABLE 2. Discrepancies between experimental identification and model prediction for essential and nonessential genes[a]

Rationale	Subsystem	Gene(s) (Blattner no.)
False negatives	**Essential Experimentally, Model Predicts Growth**	
Molecules not included in biomass	Ubiquinone biosynthesis	*ubiG* (b2232); *ubiH* (b2907)
	Pyridoxine biosynthesis	*pdxA* (b0052); *pdxB* (b2320); *pdxH* (b1638); *pdxJ* (b2564)
	Thiamine biosynthesis	*iscC* (b2530)
Model includes alternative pathways/ isozymes	Amino acid biosynthesis	*carA* (b0032); *carB* (b0033); *glpD* (b3426); *glyA* (b2551); *proA* (b0243); *proB* (b0242); *thrB* (b0003); *thrC* (b0004)
Model predicts impaired but not lethal phenotype	ATP synthase	*atpA* (b3734); *atpB* (b3738); *atpC* (b3731); *atpF* (b3736); *atpG* (b3733); *atpH* (b3735)
Regulatory effect on *glpK*	PTS/PEP metabolism	*crr* (b2417); *glpK* (b3926); *ppc* (b3956); *ptsI* (b2416); *fruR* (b0080)
False positives	**Non-Essential Experimentally, Model Predicts No Growth**	
Model biomass components which might not be essential components	Fatty acid and lipid biosynthesis	*cls* (b1249); *fabF* (b1095)
	Glycogen	*glgA* (b3429); *glgC* (b3430)
	LPS synthesis	*dgkA* (b4042); *gmhA* (b0222); *gmhB* (b0200); *lpxL* (b1054); *msbB* (b1855); *rfaC* (b3621); *rfaD* (b3619); *rfaE* (b3052); *rfaF* (b3620); *rfaG* (b3631); *rfaI* (b3627); *rfaJ* (b3626); *rfaL* (b3622)
	Spermidine synthesis	*pfs* (b0159); *speD* (b0120); *speE* (b0121)
Unaccounted-for transport mechanisms	Ammonium transport	*amtB* (b0451)
	Glycerol transport	*glpF* (b3927)
	Sulfate transport	*cysW* (b2423)
Unaccounted-for metabolic enzymes	Arginine biosynthesis	*argB* (b3959); *argC* (b3958); *argD* (b3359); *argG* (b3172)
	Aspartate biosynthesis	*aspC* (b0928)
	Branced amino acid biosynthesis	*ilvY* (b3773); *ilvE* (b3770); *lrp* (b0889)
	Central metabolic	*aldA* (b1415)
	Cofactor biosynthesis	*coaA* (b3974); *coaE* (b0103); *pabC* (b1096)
	Glycolytic	*pgi* (b4025)
	Lysine biosynthesis	*dapF* (b3809); *ushA* (b0480); *lysR* (b2839)
	Nucleotide biosynthesis and salvage	*pyrI* (b4244); *trxB* (b0888); *ndk* (b2518)

[a] Twenty-six false-negative cases in which the model incorrectly predicted growth of the gene deletion strain were identified, in addition to 42 false-positive cases in which the model incorrectly predicted that genes were essential. Each case is grouped based on the likely rationale for the discrepancy and the gene functional annotation.

MOMA: Minimize Distance Between Wildtype & Mutant Flux Distributions



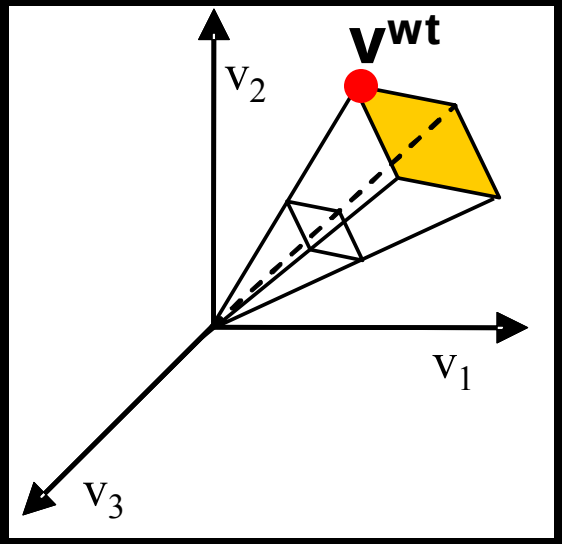
MOMA Prediction Algorithm

Maximize: $c \cdot v$

Such that $S \cdot v = b = 0$

$LB \leq v \leq UB$

SOLUTION = v^{wt}

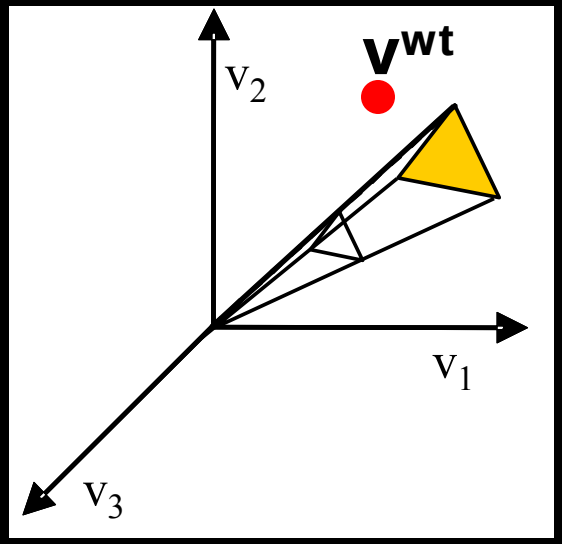


Minimize: $\sum (v_j^{wt} - v_j)^2$

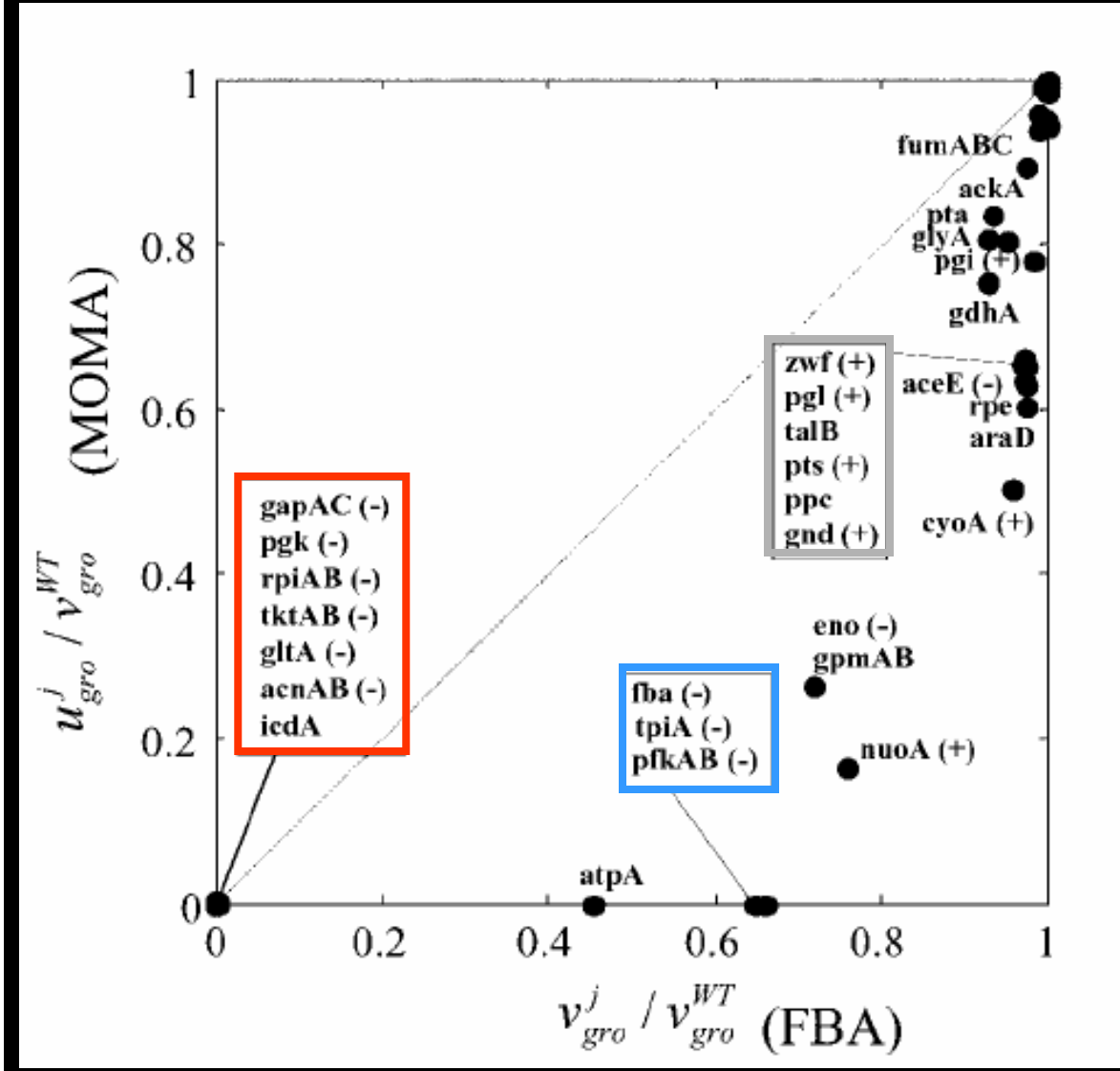
Such that $S \cdot v = b = 0$

$LB \leq v \leq UB$

$v_k = 0$



FBA vs. MOMA Mutant Growth Rate Predictions



The FBA mutant predictions for growth rate are always higher than the MOMA predictions.

Both FBA and MOMA predict lethal phenotypes, agreeing with experimental data

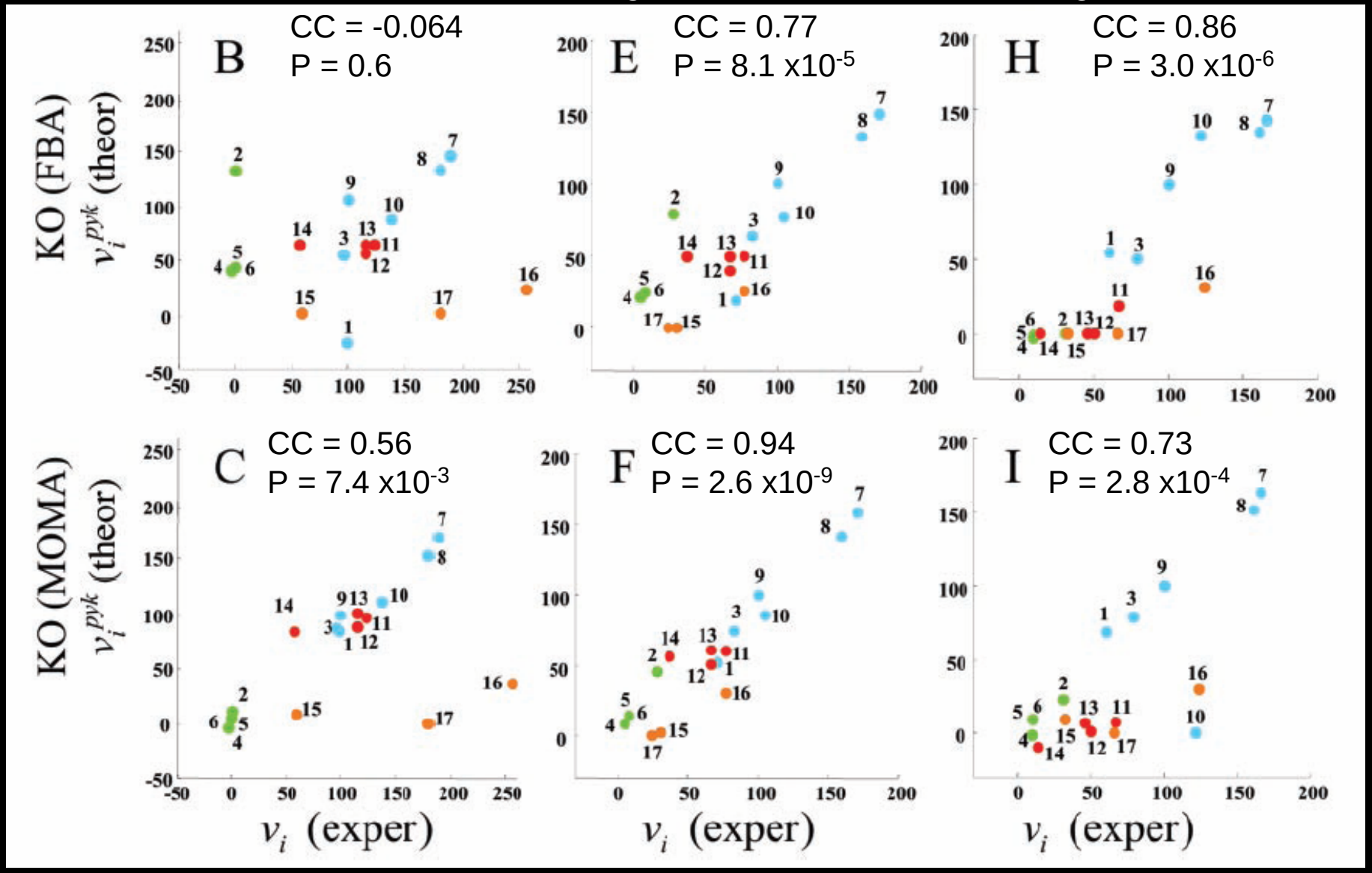
Both FBA and MOMA predict non-lethal phenotypes, agreeing with experimental data

Only MOMA predicts a lethal phenotype, agreeing with experimental data

Segre, et al. PNAS. 99(23): 15112-15117 (2002)

FBA vs. MOMA Flux Level Predictions

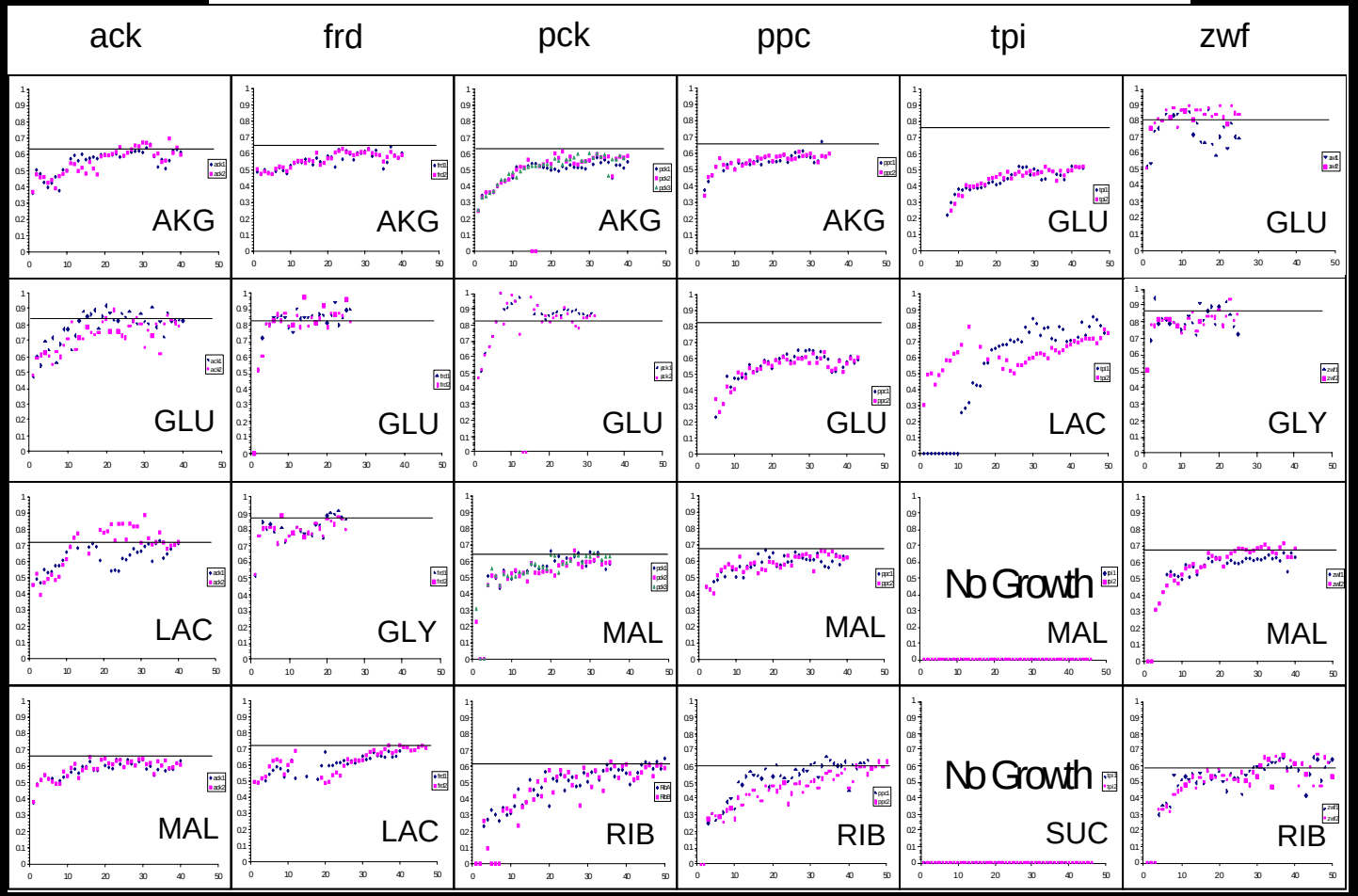
Carbon Limited Growth | High Carbon Growth | Nitrogen Limited Growth



CC = Correlation Coefficient and P = p-value

Segre, et al. PNAS. 99(23): 15112-15117 (2002)

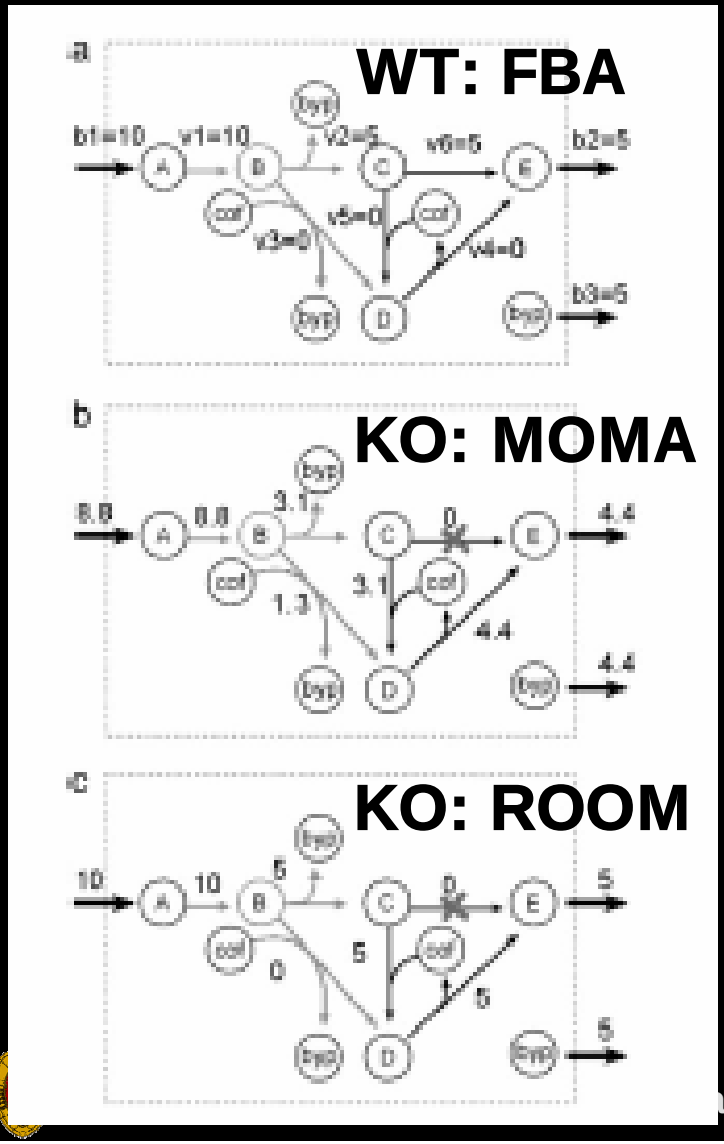
Deletion Strain Evolution



- *39 of 50 cases correctly predicted computationally*
- *Parallel cultures exhibit similar endpoint phenotypes*
- *Average GR increase of 87% observed*

Fong et al. Nature Genetics. 36(10): 1056-1058 (2004)

ROOM: Minimize the Number of Fluxes that Change



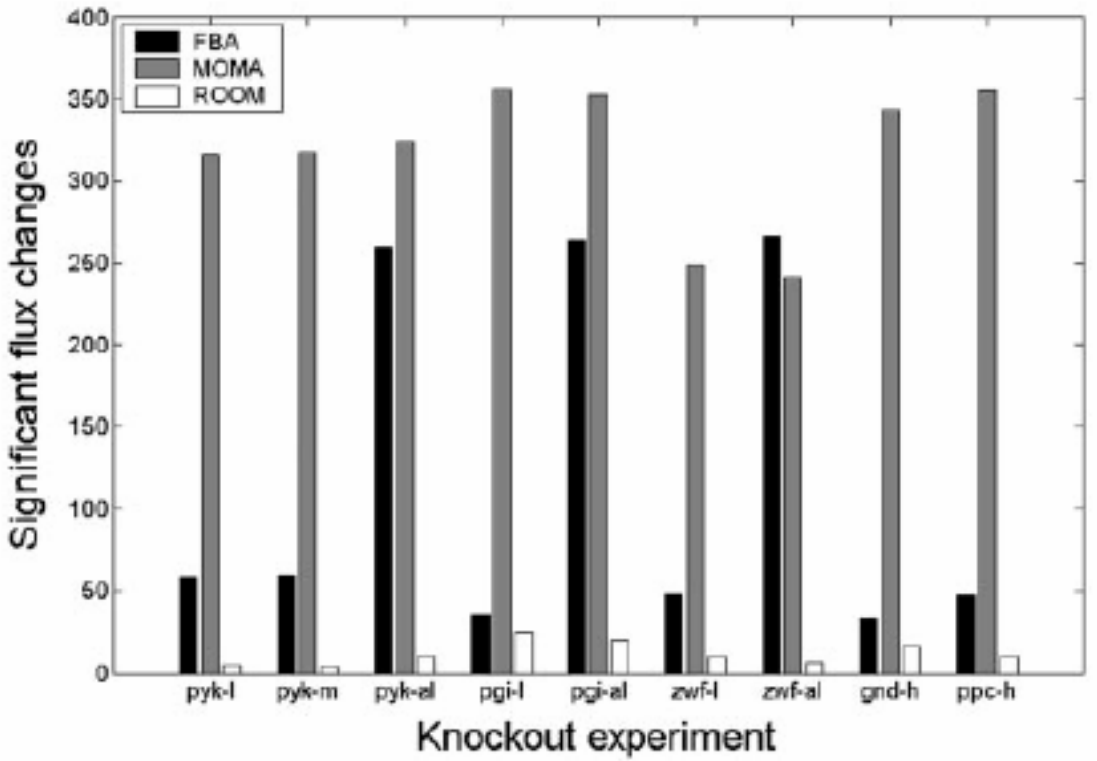


Fig. 4. Number of significant flux changes between the flux distribution of the wild-type strain and the flux distributions predicted by FBA, MOMA, and ROOM for five knocked-out organisms, under different growth conditions. The marking on the *x* axis is explained in the caption of Fig. 3.

Shlomi et al. PNAS (2005). 102(21):7695-700

Method Comparison to Experimental Data

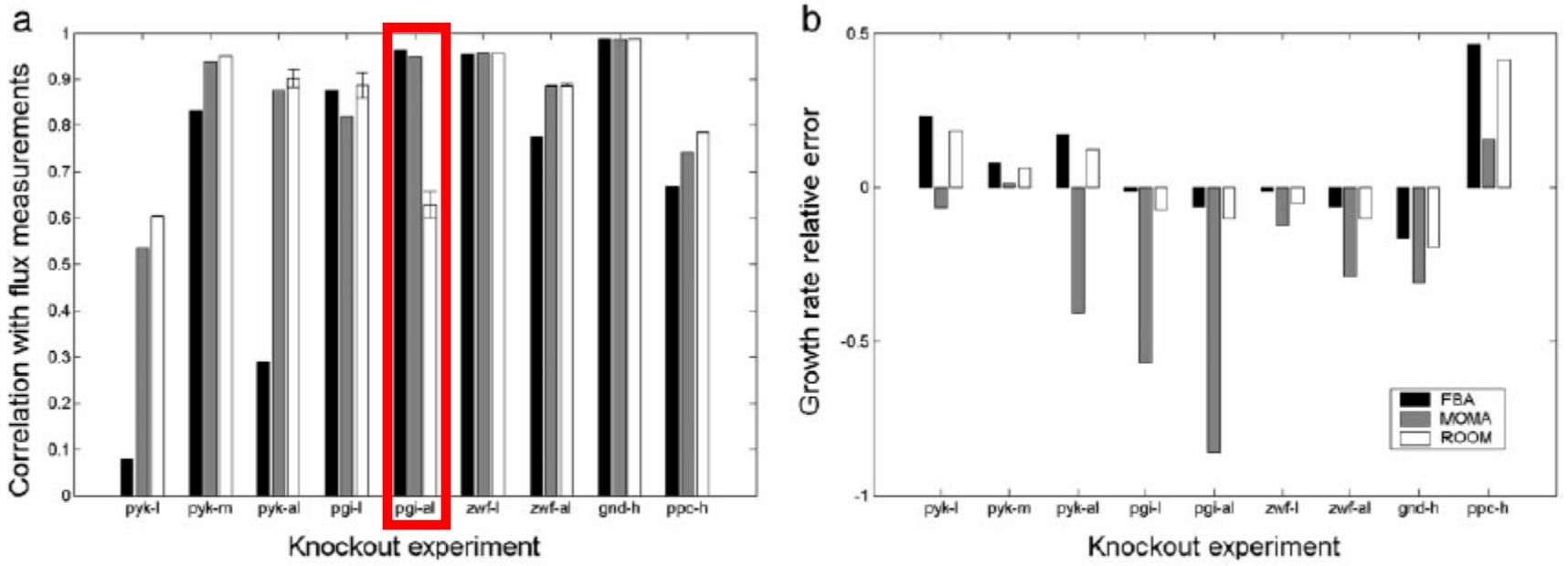


Fig. 3. Flux and growth-rate comparison among FBA, MOMA, and ROOM for five knocked-out organisms, under different growth conditions. The marking *x-y* on the *x* axis denotes knockout of gene *xxx-y* in a mutant strain grown on media *y*. l, m, h, and al stand for glucose-low, glucose-medium, glucose-high, and ammonia-low, respectively. (*a*) Pearson correlations between experimental fluxes and predictions. (*b*) Relative errors in growth rate predictions, calculated by subtracting the experimentally measured growth rate from the predicted growth rate and dividing by the experimentally measured growth rate.

- In 8 out of 9 cases ROOM has better or equal prediction capabilities with respect to flux over MOMA.
- MOMA tends to more significantly underpredict growth rate (quantitative comparison)

Shlomi et al. PNAS (2005). 102(21):7695-700

Final Points

- FBA will always predict higher (or equal) growth rates as compared to MOMA or ROOM.
- The MOMA solution is unique given a single wildtype flux distribution.
- The ROOM solution is not unique, there are often multiple flux distributions with the same number of altered fluxes.
- FBA better at predicting adaptive evolutionary outcomes.

Core E. coli Metabolic Network

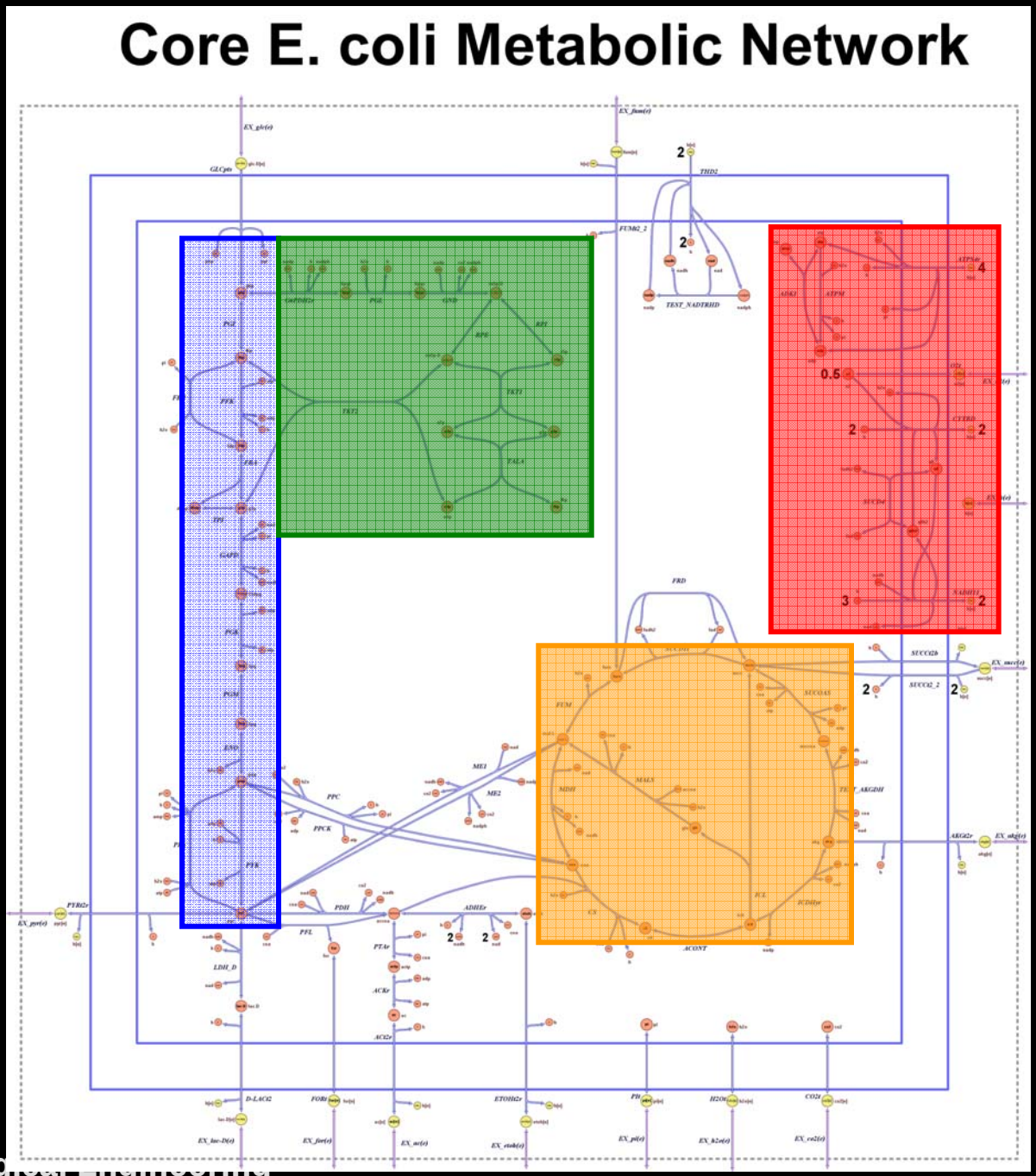
Glycolysis

Pentose Phosphate Pathway

TCA Cycle

Oxididative Phosphorylation

101 Genes
63 Metabolites
62 Reactions



ROOM, MOMA and FBA Predictions for Mutants

```
calcobj.. Obj=e=sum( j,c(j)*v(j) );
sum_y.. minnumber=e=sum(subj, y(subj));
u_restriction(subj).. v(subj)-y(subj)*(UpperLimits(subj)-wU(subj))=l=wU(subj);
l_restriction(subj).. v(subj)-y(subj)*(LowerLimits(subj)-wL(subj))=g=wL(subj);
eucl_distance.. distance=e=sum(j, sqr(wildtype_v(j)-v(j)));

Model FBA /massbalance, calcobj/;
Model ROOM /massbalance,sum_y,u_restriction,l_restriction/;
Model MOMA /massbalance,eucl_distance/;

*****************************************************************************************
*This section calculates the FBA solution for maximizing biomass
*for the wildtype strain and stores the fluxes in the wildtype_v parameter
v.lo(j)=LowerLimits(j);
v.up(j)=UpperLimits(j);
c('Biomass')=1;
solve FBA using lp maximizing Obj;
wildtype_v(j)=v.l(j);

*Defines allowable variation before becoming significant for ROOM calculations
wU(subj)=wildtype_v(subj)+delta*abs(wildtype_v(subj))+epsilon;
wL(subj)=wildtype_v(subj)-delta*abs(wildtype_v(subj))-epsilon;

v.fx('TKT1')=0;
*****************************************************************************************
*This section calculates the ROOM and MOMA solutions for the appropriate knockout
*indicated by the line v.fx('rxnname')=0;
*It also calculates the FBA solution for this same knockout
solve ROOM using mip minimizing minnumber;
mutant_room(j)=v.l(j);
solve MOMA using nlp minimizing distance;
mutant_moma(j)=v.l(j);
solve FBA using lp maximizing Obj;
mutant_fba(j)=v.l(j);
```

Define Which Reaction to Delete

Knockout Calculations

1. What are the maximum growth rates for the wildtype and mutant strains predicted using: MOMA, ROOM and FBA for the following cases:
   - tpi mutant (glucose aerobic)
   - pgi mutant (glucose aerobic)
   - acnA, acnB double mutant (glucose aerobic)

Knockout Calculations

2. If you remove ACONT, all methods predict a lethal phenotype. Looking at the shadow prices for the Mutant FBA prediction what metabolite can this mutant no longer produce that is needed for biomass production?

1.496 3pg + 3.7478 accoa + 1.0789 akg + 55.703 atp + 0.361 e4p + 0.0709 f6p + 0.129 g3p + 0.205 g6p + 55.703 h2o + 3.547 nad + 18.225 nadph + 1.7867 oaa + 0.5191 pep + 2.8328 pyr + 0.8977 r5p

→ 55.703 adp + 3.7478 coa + 41.025 h + 3.547 nadh + 18.225 nadp + 55.703 pi

Knockout Calculations (1. Ans)

- tpi mutant (delete TPI reaction)
  - 0.49(WT FBA)
  - 0.08(Mutant MOMA)
  - 0(Mutant ROOM)*
  - 0.35(Mutant FBA),
- pgi mutant (delete PGI reaction)
  - 0.49(WT FBA)
  - 0.47(Mutant MOMA)
  - 0.262 (Mutant ROOM)*
  - 0.49(Mutant FBA)
- acnA+acnB mutant (delete ACONT reaction)
  - 0.49(WT FBA)
  - All predict methods predict 0

** Alternative Optimal Solutions with different growth rates are possible*

Knockout Calculations (2. Ans)

2. Only a few compounds have negative shadow prices:

akg = a-ketoglutarate

icit = isocitrate

glx = glyoxylate

akg_e = α-ketoglutarate

→ Only akg is part of the biomass equation.