Comparison of multiple *E. coli* models reveals unique metabolic phenotypes

March 5, 2013

Jonathan Monk

UCSD Bioengineering

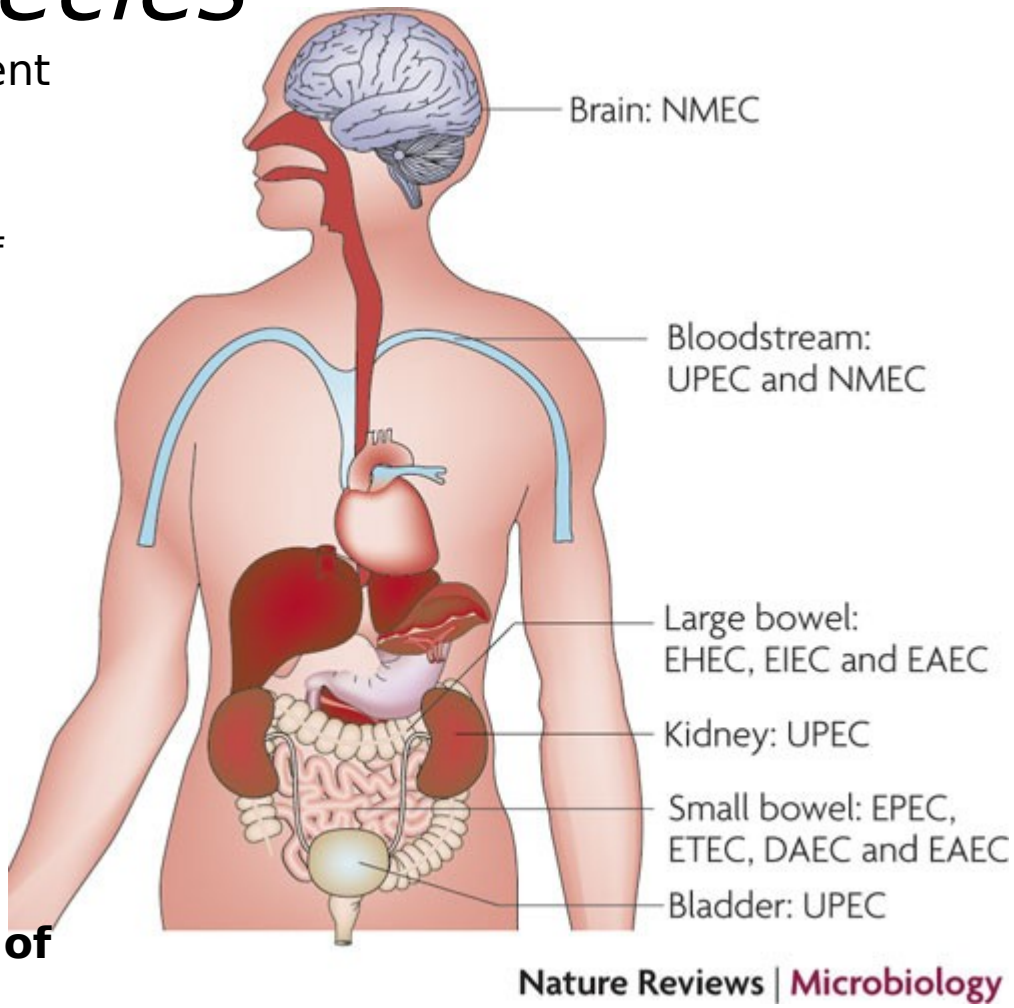
systems biology research group

Outline

- *E. coli*: from strain to species
- Core and Pan genomes/reactomes
- Metabolic Network Reconstruction Procedure
- Phenotypic Predictions
- Experimental Validation

Escherichia coli: from Strain to Species

- Predominant faculative anaerobe resident in the human gut
  - Most live as harmless commensals
  - Colonizes infant human gut within hours of birth
- Species also has many pathogenic members
  - Extraintestinal Pathogens (ExPEC)
    - Urinary infections, septicaemia and meningitis
  - Intestinal (InPEC):
    - 6 categories of intestinal infection:
      - EAEC, EIEC, EPEC, ETEC, EHEC and DAEC
- Additionally, has to survive outside gut for extended periods



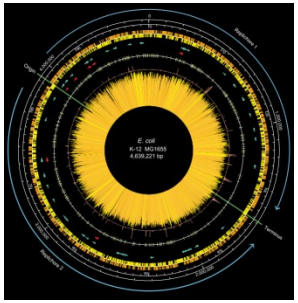
***E. coli* exhibits a remarkable variety of lifestyles!**

UCSD Bioengineering

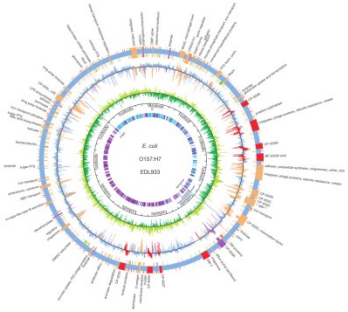
systems biology research group

K-12 is not representative of *E. coli species*

- History of K-12:
  - Isolated from feces of a convalescent diptheria patient in 1922
  - Adopted as a model organism in the 1940s
  - Likely underwent repeated subculture and/or storage in stab culture during interim 20 years
  - Later, underwent rounds of mutagenesis
  - UV light treatment to remove phage lambda
  - Acridine orange to remove F plasmid
  - Genome sequenced in 1997
- E. coli O157:H7 genome sequenced in 2001: remarkably has **1 million more base pairs than K-12**



E. Coli K-12
MG1655
4.6 Mbp



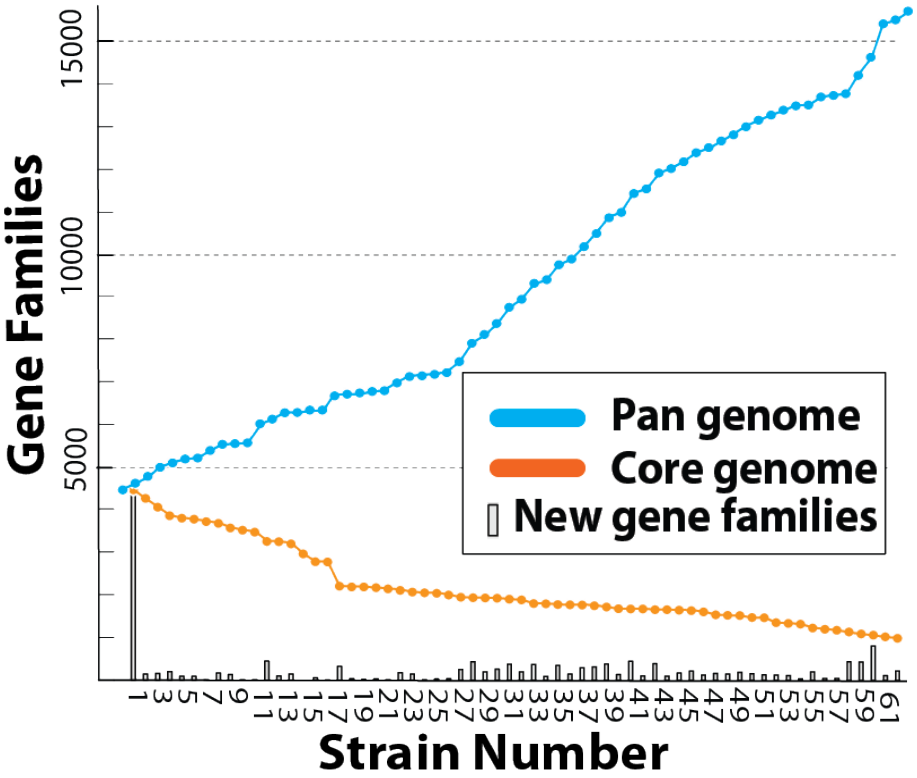
E. Coli O157:H7
EDL933
5.5 Mbp

Blattner, F. R. (1997). The Complete Genome Sequence of Escherichia coli K-12. *Science*, *277*(5331), 1453-1462. doi:10.1126/science.277.5331.1453

Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., et al. (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature*, *409*(6819), 529-33. doi:10.1038/35054089

Core vs Pan genome

- **Core genome:** Genes present in **every** member of a species "essence of the species"
- **Pan genome:** Variable genes present in **any** member of a species
- For *E. coli*: currently 15,000 gene families predicted in pangenome, ~ 2,000 in core genome
- Estimated to be nearly 45,000 in pan genome of *E. coli*



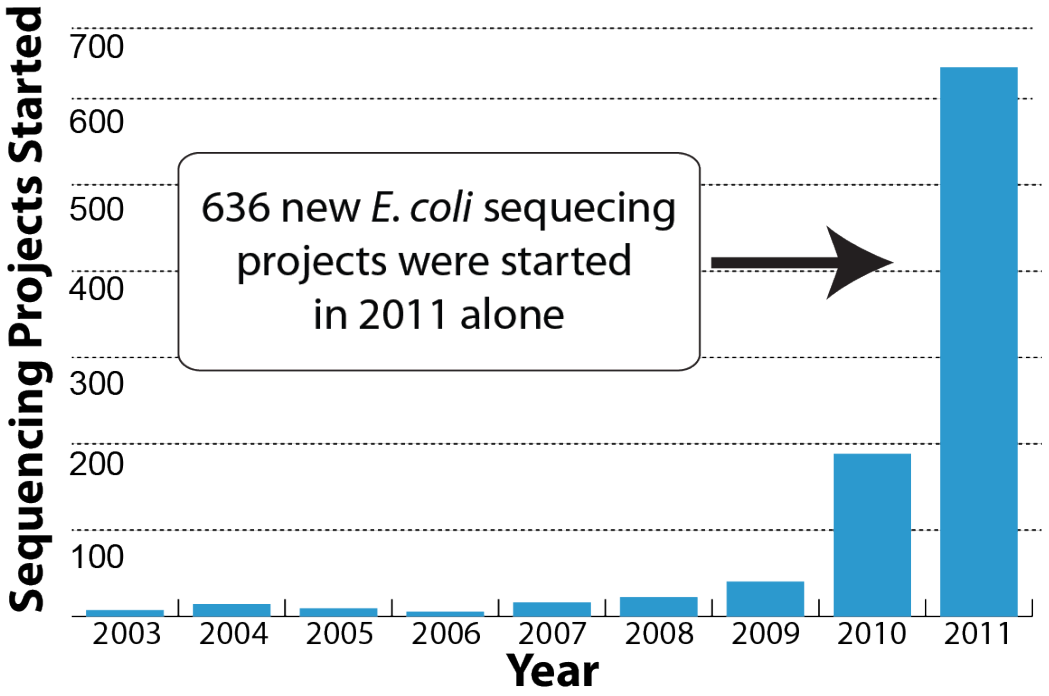
Snipen, L., Almøy, T., & Ussery, D. W. (2009). Microbial comparative pan-genomics using binomial mixture models. *BMC genomics*, *10*, 385. doi:10.1186/1471-2164-10-385

Lukjancenko, O., Wassenaar, T. M., & Ussery, D. W. (2010). Comparison of 61 sequenced Escherichia coli genomes. *Microbial ecology*, *60*(4), 708-20. doi:10.1007/s00248-010-9717-3

UCSD Bioengineering

systems biology research group

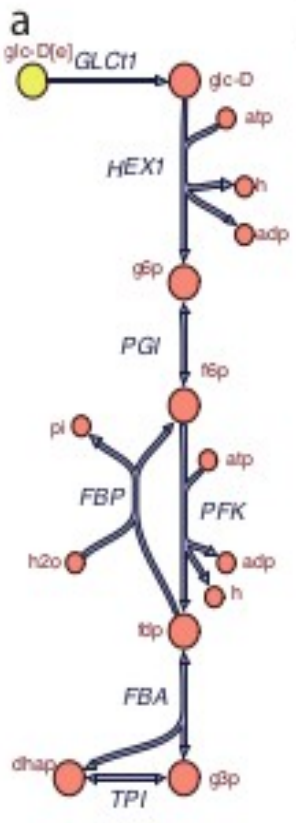
Increase in *E. coli* genome sequences



- Need tools to analyze these sequences
- Metabolic reconstructions are one solution
- Many core genes are metabolic

Genome Scale Metabolic Reconstructions

- Take into account all known metabolic reactions in an organism
- Flux Balance Analysis can be used to examine these networks:
  - Allows calculation of phenotypic states
  - Bridges genotype with phenotype through GPRs (gene-protein-reaction relation)



b

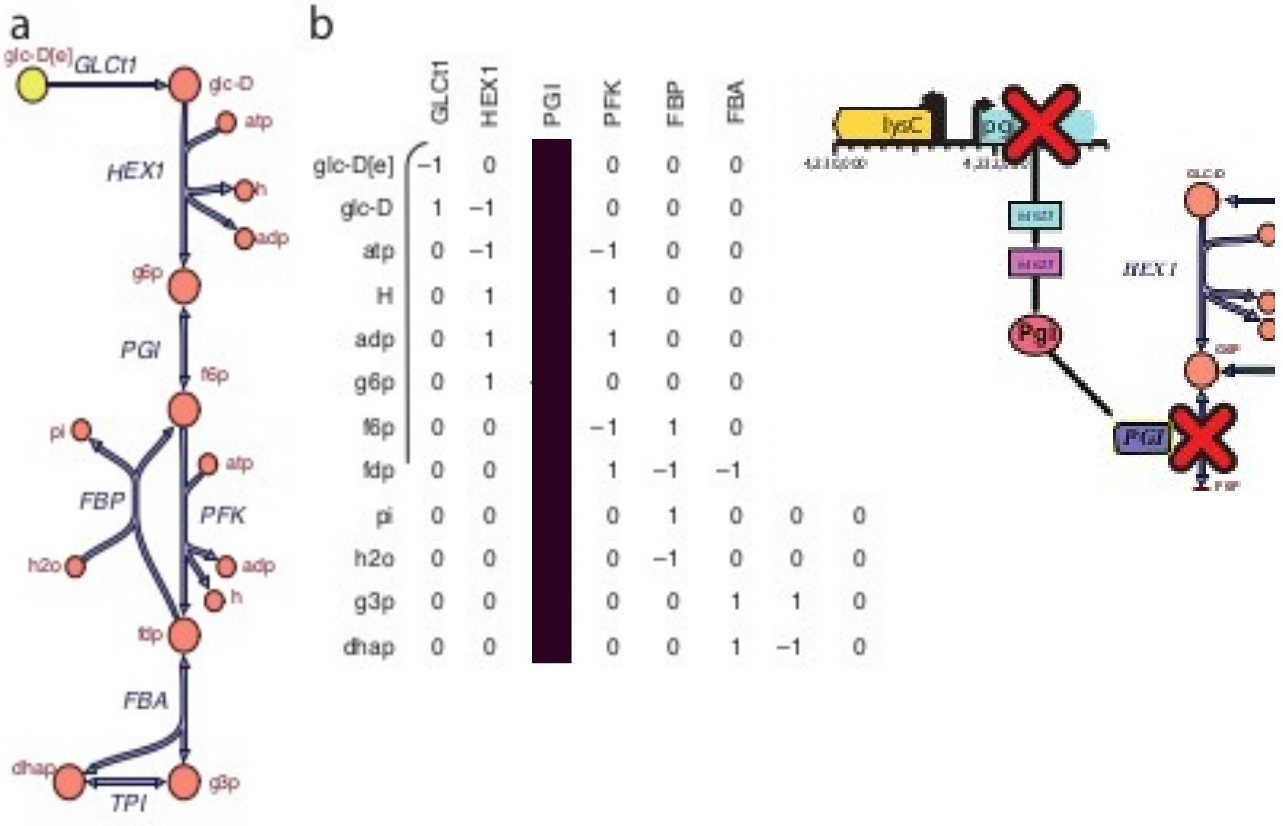
	GLCt1	HEX1	PGI	PFK	FBP	FBA	TPI	EX_glc	
glc-D[e]	−1	0	0	0	0	0	0	−1	
glc-D	1	−1	0	0	0	0	0	0	
atp	0	−1	0	−1	0	0	0	0	
H	0	1	0	1	0	0	0	0	
adp	0	1	0	1	0	0	0	0	
g6p	0	1	−1	0	0	0	0	0	
f6p	0	0	1	−1	1	0	0	0	= s
fdp	0	0	0	1	−1	−1	0	0	
pi	0	0	0	0	1	0	0	0	
h2o	0	0	0	0	−1	0	0	0	
g3p	0	0	0	0	0	1	1	0	
dhap	0	0	0	0	0	1	−1	0	

UCSD Bioengineering

systems biology research group

Genome Scale Metabolic Reconstructions

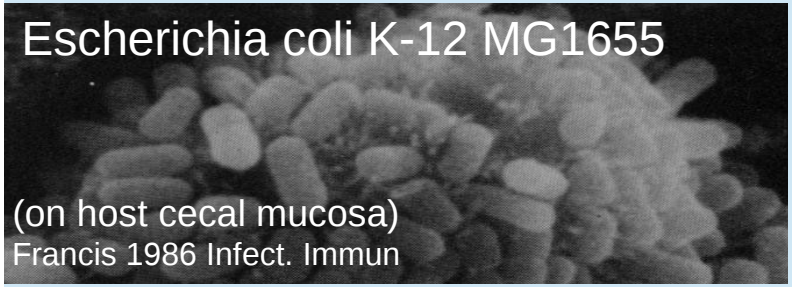
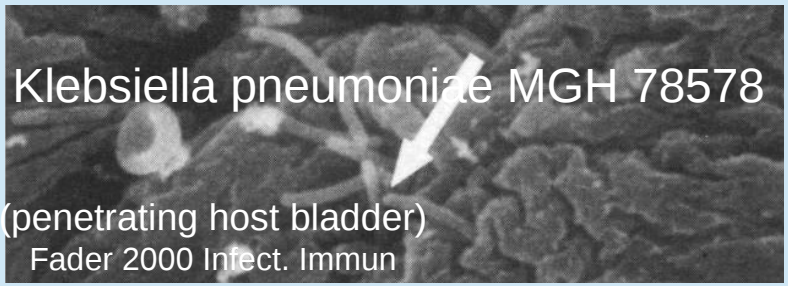
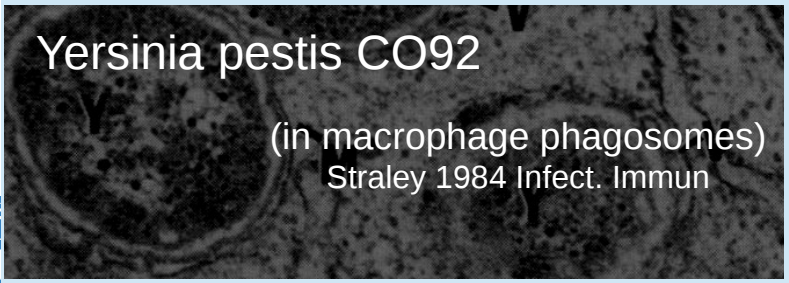
- Take into account all known metabolic reactions in an organism
- Flux Balance Analysis can be used to examine this networks:
  - Allows calculation of phenotypic states
  - Bridges genotype with phenotype through GPRs (gene-protein-reactions relation)



UCSD Bioengineering

systems biology research group

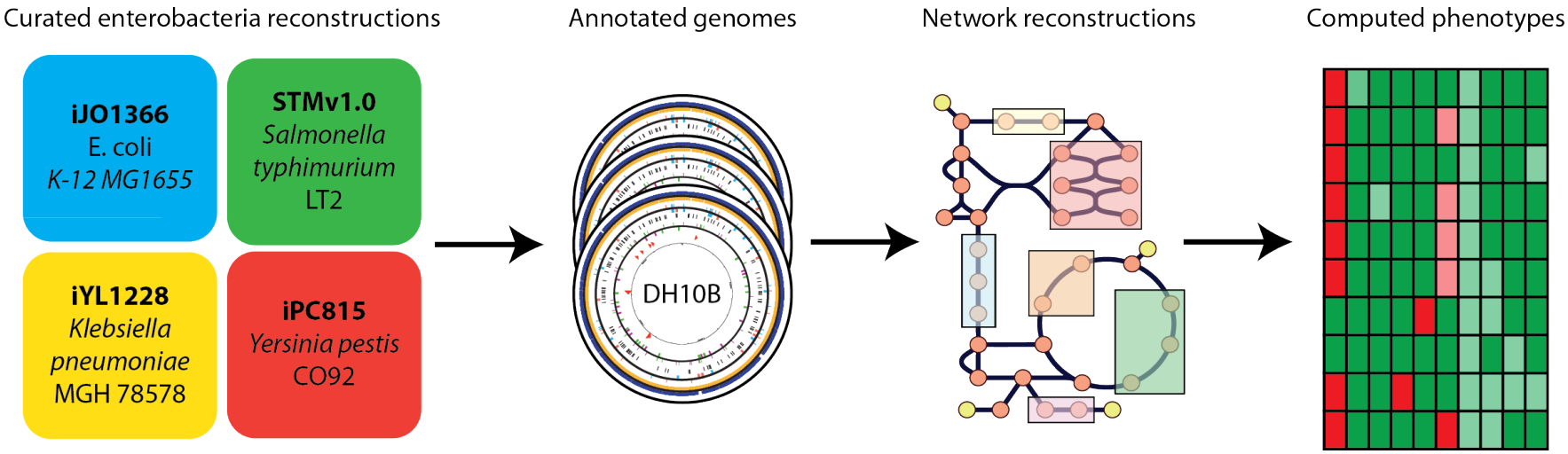
Summary of curated enterobacteria reconstructions

Organism	Reconstruction
	**iJO1366** Orth et. al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011 1366 genes, 2259 reactions, 1805 metabolites
	STM_v1.0 Thiele et. al. A community effort towards a knowledge-base and mathematical model of the human pathogen SalmonellaTyphimurium LT2 1271 genes, 2205 reactions, 1802 metabolites
	**iYL1228** Liao et. al. An Experimentally Validated Genome-Scale Metabolic Reconstruction of Klebsiella pneumoniae MGH 78578, iYL1228 1228 genes, 1973 reactions, 1658 metabolites
	**iPC815** Charusanti et al. An experimentally-supported genome-scale metabolic network reconstruction for Yersinia pestis CO92 815 genes, 1687 reactions, 1562 metabolites

Complete *E. coli/Shigella* sequences examined

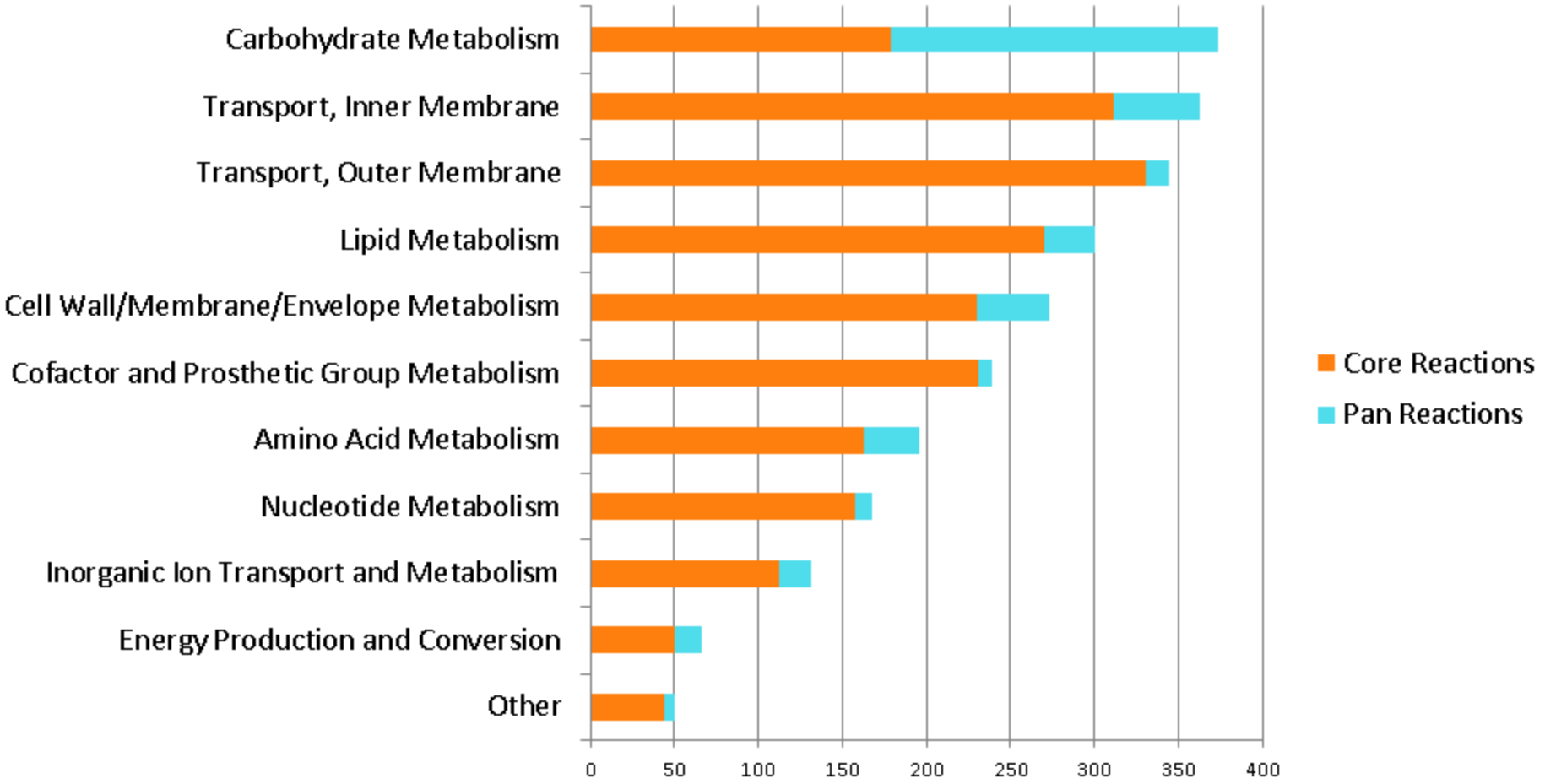
Genus	Species	Subspecies/Pathotype/Example	Count
Escherichia (48)	Coli (47)	Commensal (e.g. E. coli K12 MG1655, E. coli BL21, etc.)	18
		EHEC (e.g. E. coli O157:H7, E. coli EDL933, O157 Sakai, etc.)	8
		UPEC (e.g. UTI89, CFT073, etc.)	6
		Other (e.g. ExPec, APEC, EAEC and more)	15
	Fergusonii	Escherichia fergusonii ATCC 35469 (Ancestral)	1
Shigella (8)	Flexneri	e.g. Shigella flexneri 5 str. 8401, Shigella flexneri 2a str. 2457T	4
	Boydii	e.g. Shigella boydii Sb227, Shigella boydii CDC 3083-94	2
	Dysenteriae and Sonnei		2

Reconstruction Process



- Reconstruction content mapped to complete annotated genomes using GPRs
  - Using BBH and genetic context
  - Supplemented with information from Model Seed and Ecocyc/Metacyc
  - Manually curated

E. coli Core and Pan Reactions by System



UCSD Bioengineering

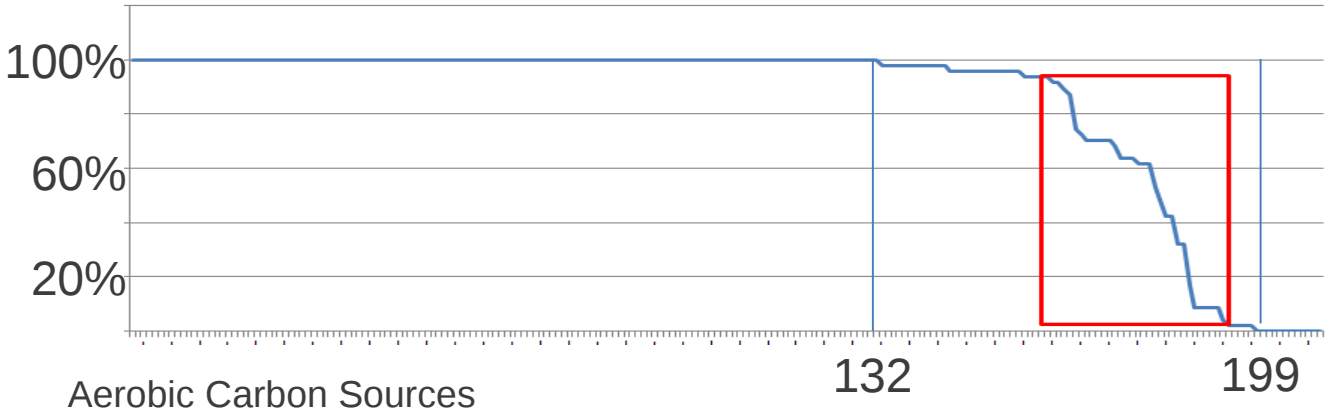
systems biology research group

Flux Balance Analysis

- Allows simulation of growth capabilities in different conditions:
  - Profiled growth *in-silico* in more than 650 conditions that support growth in at least one strain:
    - C-sources: 199 aerobic, 163 anaerobic
    - N-Sources: 96 aerobic, 79 anaerobic
    - P-Sources: 12 aerobic, 12 anaerobic
    - S-Sources: 12 aerobic, 1 anaerobic

E. coli growth capabilities

Percentage of *E. coli* strains that grow on substrate

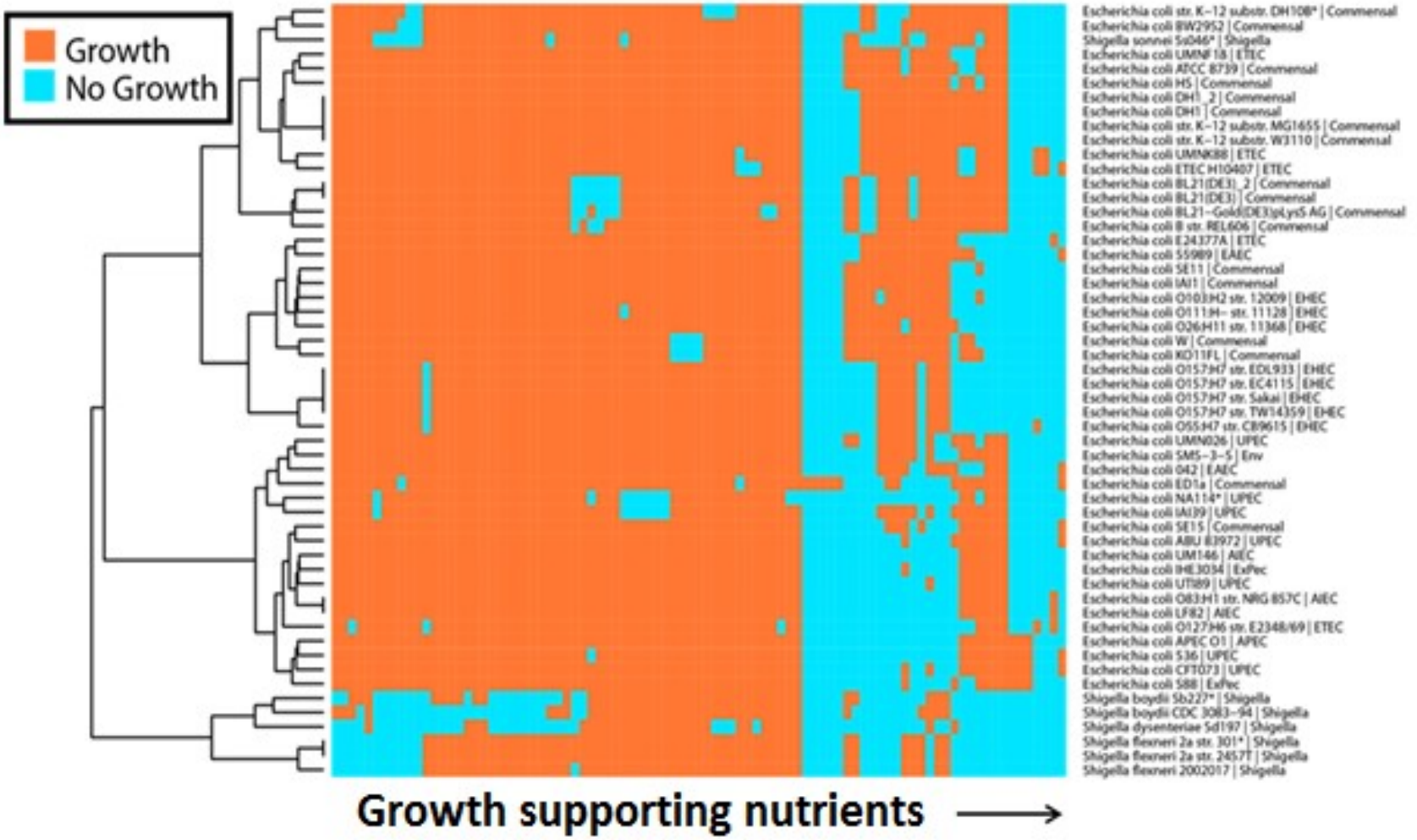


20 most variable carbon sources (aerobic)

UCSD Bioengineering

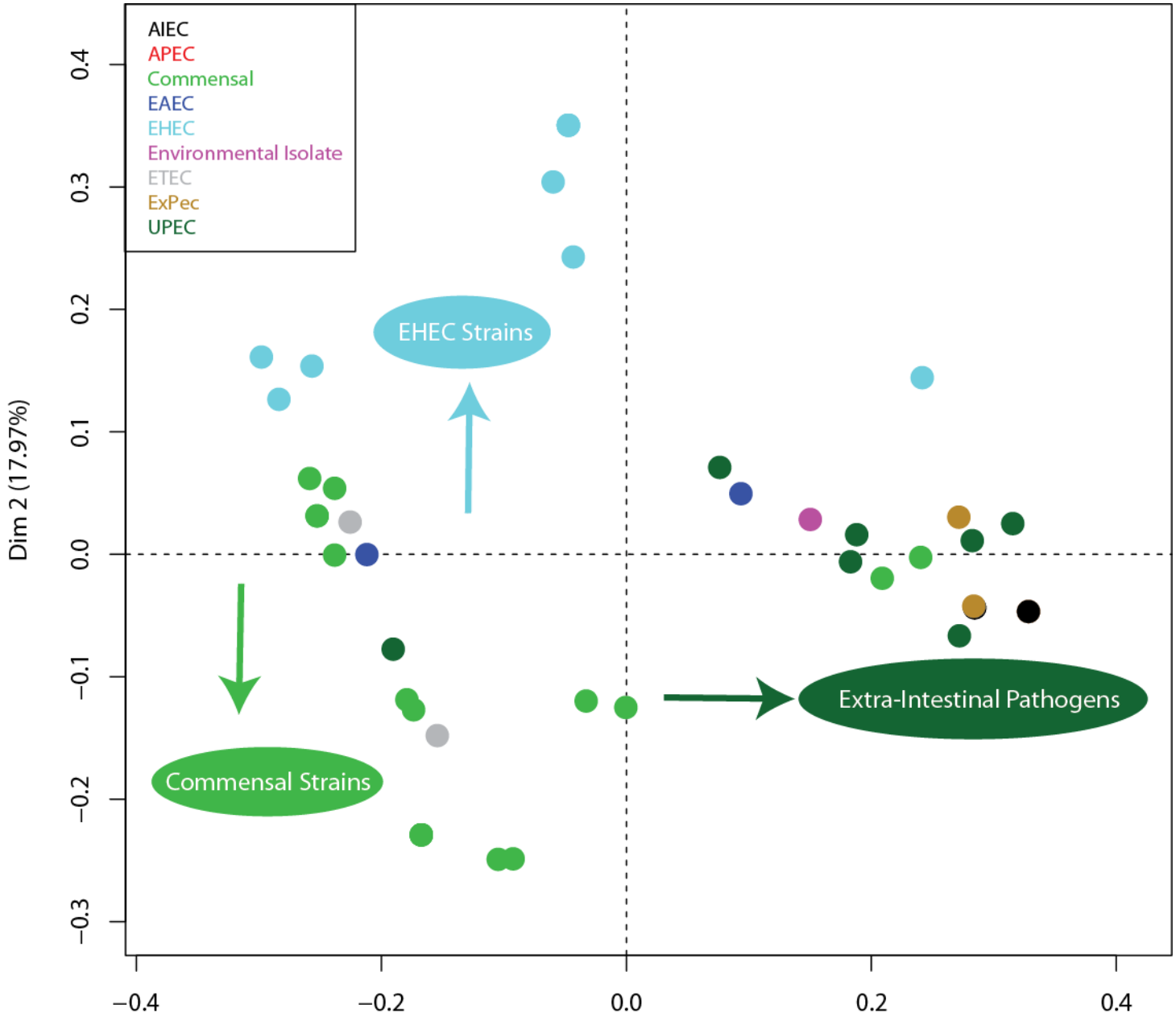
systems biology research group

Growth predictions for *E. coli/Shigella*



UCSD Bioengineering

systems biology research group


AIEC
APEC
Commensal
EAEC
EHEC
Environmental Isolate
ETEC
ExPec
UPEC
EHEC Strains
Commensal Strains
Extra-Intestinal Pathogens
Dim 2 (17.97%)
Dim 1 (34.34%)
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.2
0.0
0.2
0.4

Experimental Validation

- Obtained 12 strains from diverse pathotypes
  - EHEC, UPEC, DAEC, Shigella, Commensal
- Purchased difference driving carbon and nitrogen sources to test
- 4 possible outcomes
  - Correct Model Predictions:
    - True Positives, True Negatives
  - False Model Predictions:
    - False Negative: No pathway present
      - Drives discovery of new biology
    - False Positive: Pathway is present, but strain doesn't grow.
      - Could be explained by regulation
- Biolog technology is perfect for this study!

  - Except for some documented inconsistencies!

Example of Inconsistencies

- Compared Biolog datasets from two studies:
  - The evolution of metabolic networks of E. coli
    - David J Baumler1*, Roman G Peplinski1, Jennifer L Reed2, Jeremy D Glasner1 and Nicole T Perna1,3
  - The decoupling between genetic structure and metabolic phenotypes in Escherichia coli leads to continuous phenotypic diversity
    - V Sabarly,*†‡ O Bouvet,† J Glodt,† O Clermont,† D Skurnik,† L Diancourt,§ D de Vienne,‡ E Denamur,† and C Dillmann
- Three strains overlap: K-12MG1655, EDL933 and CFTO73
  - Large number of growth/no growth inconsistencies

ECOCYC: inconsistent results

Plate ID: Biolog PM1 - Carbon Sources | No growth/respiration | Low growth/respiration | Growth/respiration | Inconsistent results | No data

Conditions: wildtype at 37°C (aerobic); 5 Datasets; Growth: 68; Low Growth: 2; No Growth: 20; Inconsistent results: 5.

A1 carbon negative control	A2 L-Arabinose	A3 N-Acetyl-D- Glucosamine	A4 D-Saccharic acid	A5 Succinic acid	A6 D-Galactose
B1 D-Serine	B2 D-Sorbitol	B3 Glycerol	B4 L-Fucose	B5 D-Glucuronic acid	B6 D-Gluconic acid
C1 D-Glucose- 6-Phosphate	C2 D-Galactonic acid-γ-Lactone	C3 DL-Malic acid	C4 D-Ribose	C5 Tween 20	C6 L-Rhamnose
D1 L-Asparagine	D2 D-Aspartic acid	D3 D-Glucosaminic acid	D4 1,2-Propanediol	D5 Tween 40	D6 α-Ketoglutaric acid
E1 L-Glutamine	E2 M-Tartaric acid	E3 D-Glucose- 1-Phosphate	E4 D-Fructose- 6-Phosphate	E5 Tween 80	E6 α-Hydroxyglutaric acid-γ-Lactone
F1 Gly-Asp	F2 Citric acid	F3 M-Inositol	F4 D-Threonine	F5 Fumaric acid	F6 Bromosuccinic acid
G1 Gly-Glu	G2 Tricarballylic acid	G3 L-Serine	G4 L-Threonine	G5 L-Alanine	G6 Ala-Gly
H1 Gly-Pro	H2 p-Hydroxyphenyl Acetic acid	H3 m-Hydroxyphenyl Acetic acid	H4 Tyramine	H5 D-Psicose	H6 L-Lyxose

http://ecocyc.org/ECOLI/NEW-IMAGE?object=Growth-Media

UCSD Bioengineering

systems biology research group

Biolog Growth Comparisons

Baumler
PM1 plate
5 strains

Sabarly
GN2 Plate
12 strains

95 C sources each

Overlap:
57 C sources
3 Strains: K-12 MG1655, EDL933 and CFT073

- Baumler predicts growth on many more C sources
- Different protocols:
  - Major difference:
  - Baumler grew for 48 hours
  - Sabarly grew for 18 hours

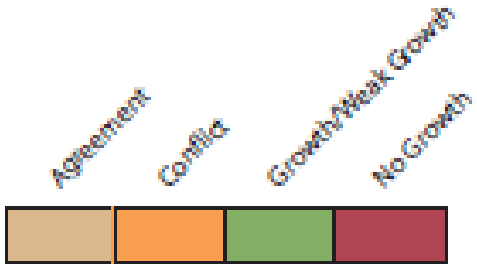
Difficult to compare results to each other

EDL933	CFT073
35% (20/57) Agreement	42% (24/57) Agreement

UCSD Bioengineering

systems biology research group

Inconsistent Results: K-12 MG1655



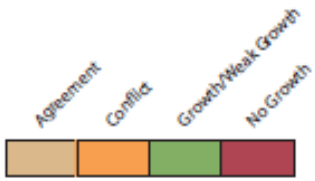
MG1655 Substrate	Baumler et. al.	Sabarly et. al.
2-Aminoethanol		
Acetic Acid		
Adonitol		
α-Hydroxy Butyric Acid		
α-Keto-Butyric acid		
α-Keto-Glutaric Acid		
α-D-Lactose		
β-Methyl-D-Glucoside		
Bromosuccinic acid		
D-Cellobiose		
Citric Acid		
D,L-α-Glycerol-Phosphate		
L-Lactic Acid		
D-Alanine		
D-Fructose		
D-Galactose		
D-Gluconic Acid		
α-D-Glucose		
D-Mannitol		
D-Mannose		
D-Melibiose		
D-Serine		
D-Sorbitol		
D-Trehalose		
Formic Acid		
Glycerol		
Glycyl-L-glutamic acid		
Inosine		
L-Alanine		

L-Alanyl-Glycine		
L-Arabinose		
L-Asparagine		
L-Aspartic Acid		
L-Fucose		
L-Glutamic Acid		
L-Proline		
L-Rhamnose		
L-Serine		
L-Threonine		
Maltose		
m-Inositol		
Mono Methyl Succinate		
N-Acetyl-D-Glucosamine		
p-Hydroxyphenyl Acetic Acid		
D-Psicose		
Succinic Acid		
Sucrose		
Tween 40		
Tween 80		

UCSD Bioengineering

systems biology research group

Inconsistent Results: EDL933



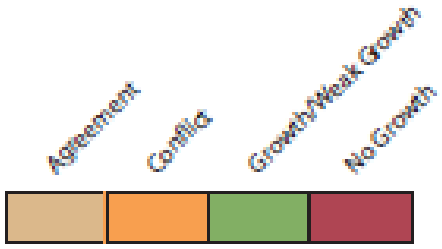
EDL 933 Substrate	Baumler et al.	Sabarly et al.
2-Aminoethanol		
Acetic Acid		
Adonitol		
Bromo Succinic Acid		
Citric Acid		
D,L-a-glycerol-phosphosphate		
D-Alanine		
D-Fructose		
D-Galactonic acid-y-lactone		
D-Galactose		
D-Galacturonic Acid		
D-Gluconic Acid		
D-Glucosaminic Acid		
D-Glucuronic Acid		
D-Mannitol		
D-Mannose		
D-Melibiose		
D-Saccharic Acid		
D-Serine		
D-Sorbitol		
D-Trehalose		
Formic Acid		
Glucuronamide		
Glycerol		
Glycyl-L-Aspartic Acid		
Glycyl-L-Glutamic Acid		
Inosine		
Lactulose		
L-Alanine		

L-Alanyl-Glycine		
L-Arabinose		
L-Asparagine		
L-Aspartic Acid		
L-Fucose		
L-Glutamic Acid		
L-Proline		
L-Rhamnose		
L-Serine		
L-Threonine		
Maltose		
Methyl Pyruvate		
M-Inositol		
Mono Methyl Succinate		
N-Acetyl-D-Glucosamine		
Phenylethylamine		
p-Hydroxy Phenyl Acetic Acid		
Propionic Acid		
Succinic Acid		
Sucrose		
Uridine		
α-Hydroxy Butyric		
α-Keto-Butyric acid		
α-Keto-Glutaric Acid		
β-Methyl-DGlucoside		
Tween 40		
Tween 80		
D-Psicose		

UCSD Bioengineering

systems biology research group

Inconsistent Results: CFT073



CFT073 Substrate	Baumler et. al.	Sabarly et. al.
2-Aminoethanol		
Acetic Acid		
Adonitol		
Bromo Succinic Acid		
Citric Acid		
D,L-a-glycerol-phosphate		
D-Alanine		
D-Fructose		
D-Galactonic acid-y-lactone		
D-Galactose		
D-Galacturonic Acid		
D-Gluconic Acid		
D-Glucosaminic Acid		
D-Glucuronic Acid		
D-Mannitol		
D-Mannose		
D-Melibiose		
D-Saccharic Acid		
D-Serine		
D-Sorbitol		
D-Trehalose		
Formic Acid		
Glucuronamide		
Glycerol		
Glycyl-L-Aspartic Acid		
Glycyl-L-Glutamic Acid		
Inosine		
Lactulose		
L-Alanine		

L-Alanyl-Glycine		
L-Arabinose		
L-Asparagine		
L-Aspartic Acid		
L-Fucose		
L-Glutamic Acid		
L-Proline		
L-Rhamnose		
L-Serine		
L-Threonine		
Maltose		
Methyl Pyruvate		
M-Inositol		
Mono Methyl Succinate		
N-Acetyl-D-Glucosamine		
Phenylethylamine		
p-Hydroxy Phenyl Acetic Acid		
Propionic Acid		
Succinic Acid		
Sucrose		
Uridine		
α-Hydroxy Butyric		
α-Keto-Butyric acid		
α-Keto-Glutaric Acid		
β-Methyl-DGlucoside		
Tween 40		
Tween 80		
D-Psicose		

UCSD Bioengineering

research group

Inconsistencies in protocol?

- Different shaking/aeration?
- Growth time 18 hrs (Sabarly) vs 40hrs (Baumler)
- Evaporation?
- Different plates?? GN2 vs PM plates
- Possibly different strains?
- Pre-culture conditions?
- Growth calling threshold

Thank you

- Josh Lerman, Jeff Orth, Adam Feist
- Ned Premyodhin
- Pep Charusanti, Ramy Aziz
- Bernhard Palsson
- Funding Source:
  - NIH: GM057089-15

UCSD Bioengineering

systems biology research group