JGI JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY

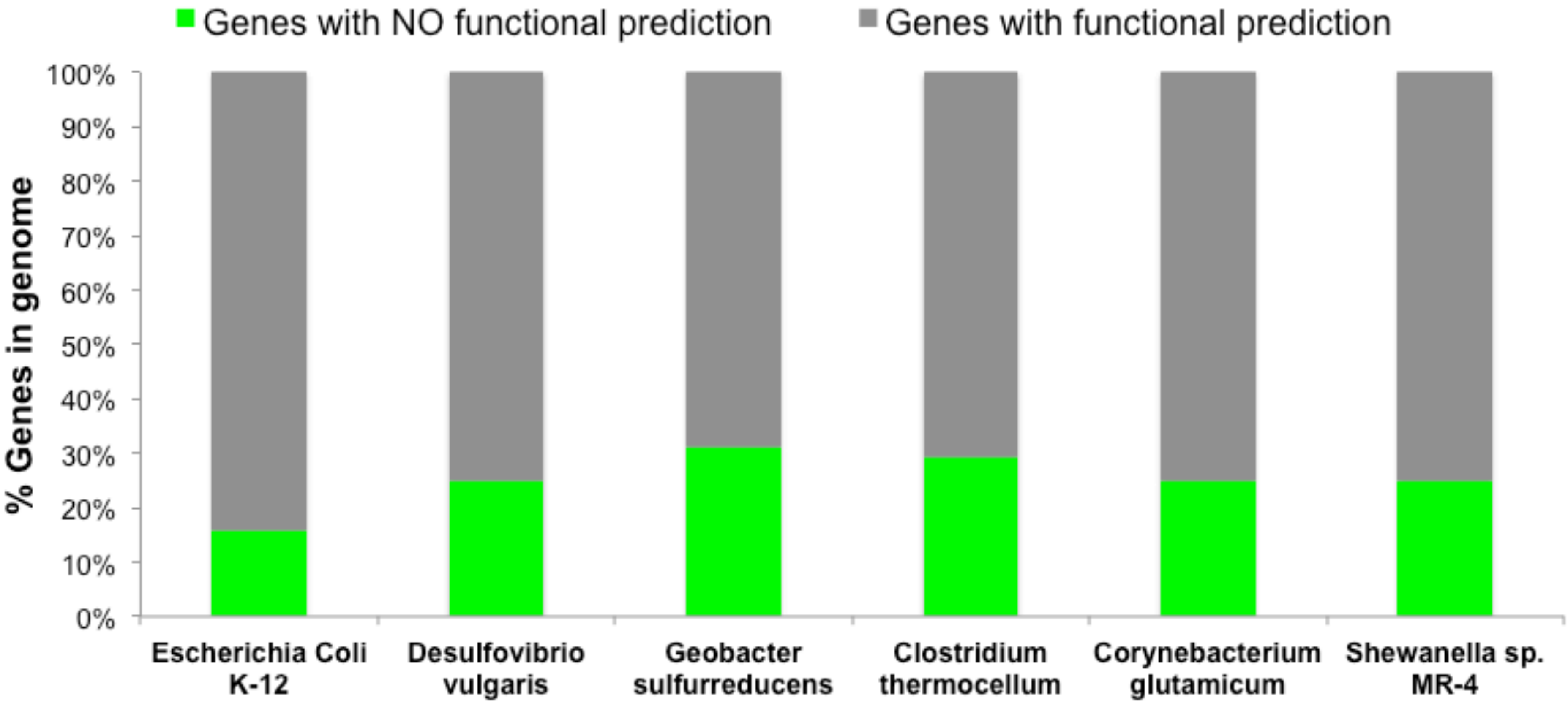
High-throughput gene function annotation using transposon mutagenesis

Adam Deutschbauer

Outline

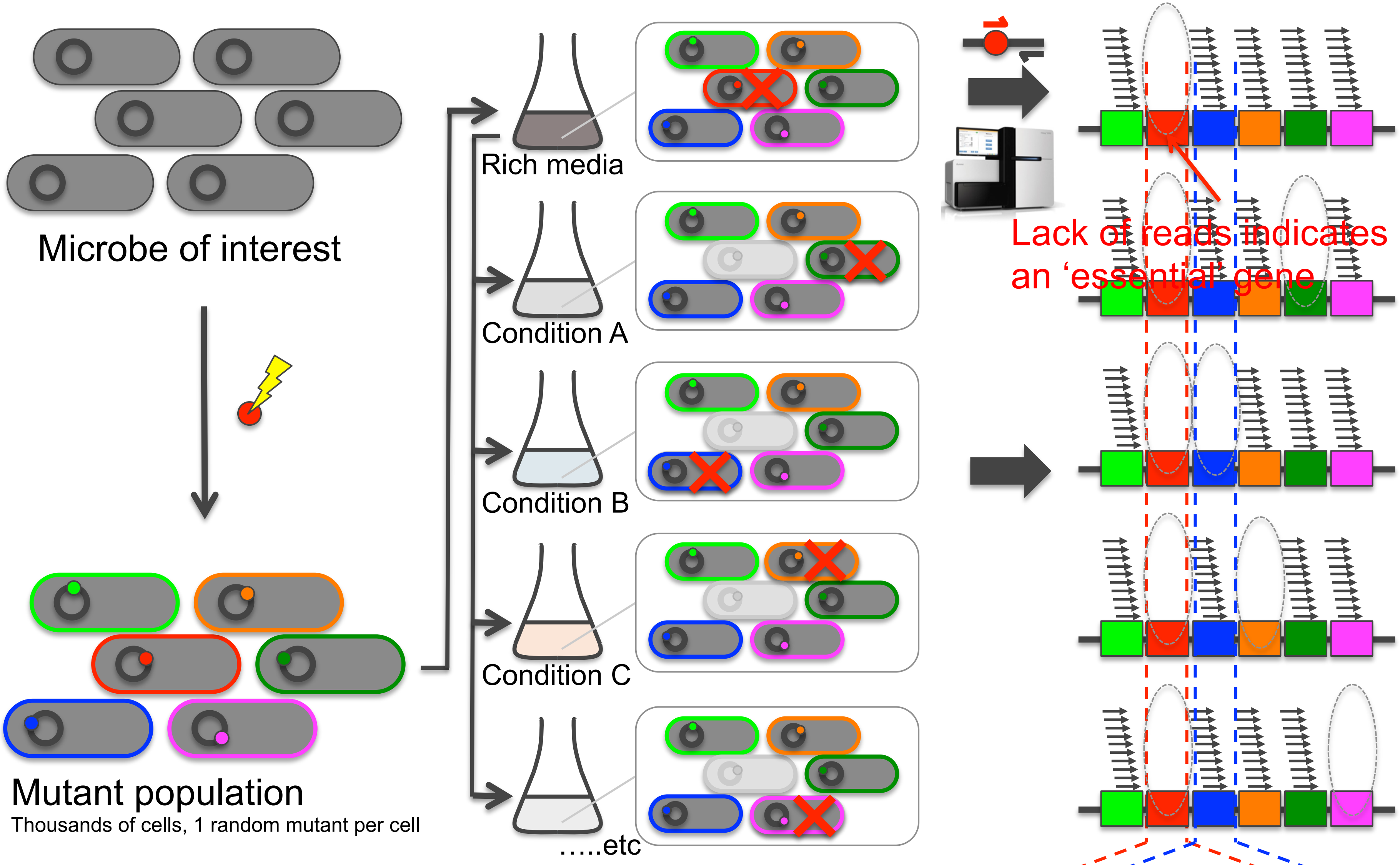
- Overview of transposon mutagenesis and sequencing (TnSeq)
- Applications and examples
- Development of RB-TnSeq for high-throughput gene function annotation in diverse bacteria

Microbial genomes are poorly characterized



Up to 30% genes in microbial genomes have no known function

Gene function annotation by transposon mutagenesis and sequencing (TnSeq)



Transposon Mutagenesis

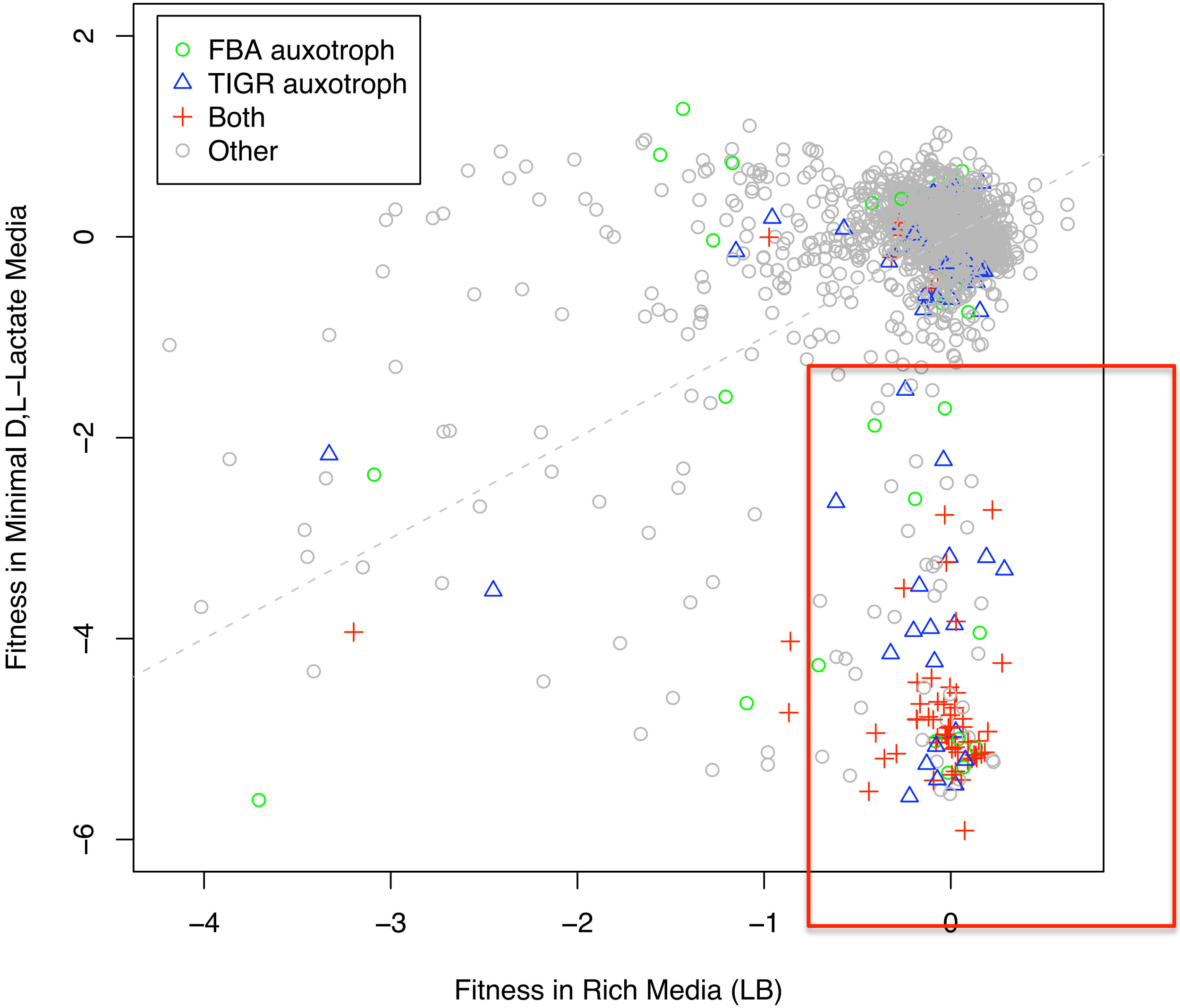
Grow mutants under many conditions

PCR and Sequencing

Gene essential in all conditions

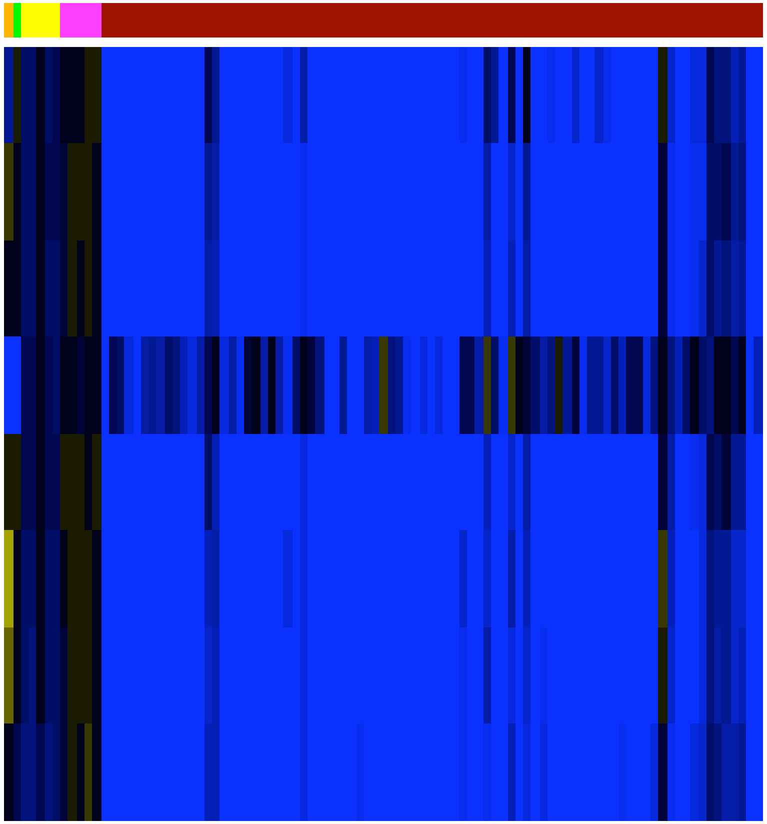
Gene essential in Condition B

Fitness in rich versus minimal media



Fill holes in biosynthetic pathways

Carbon/Nitrogen sources:
Arginine (N), Gelatin (C), CAS (C), LB, Other defined



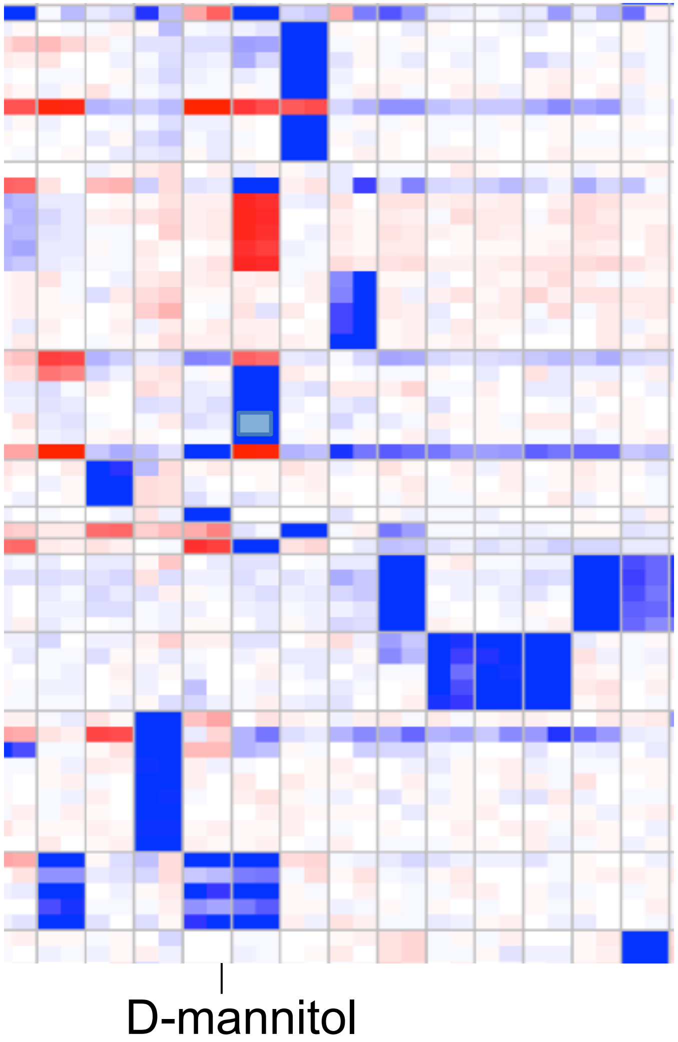
Fitness: **-3** to **0**

L-glutamate
argA (SO_4245) ↓
N-acetyl-glutamate
argB (SO_0276) ↓
N-acetylglutamyl-P
argC (SO_0275) ↓
N-acetyl-glu 5-semialdehyde
argD (SO_0617) ↓
N-acetyl-ornithine
SO_3749 ↓
ornithine
argF (SO_0277) ↓
citrulline
argG (SO_0278) ↓
arginino-succinate
argH (SO_0279) ↓
L-arginine

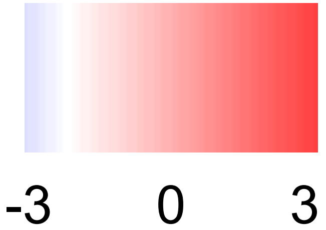
Example from *Shewanella oneidensis* MR-1

Fill holes in catabolic pathways

Phaeobacter inhibens BS107



Gene Fitness



gene is important

gene is detrimental

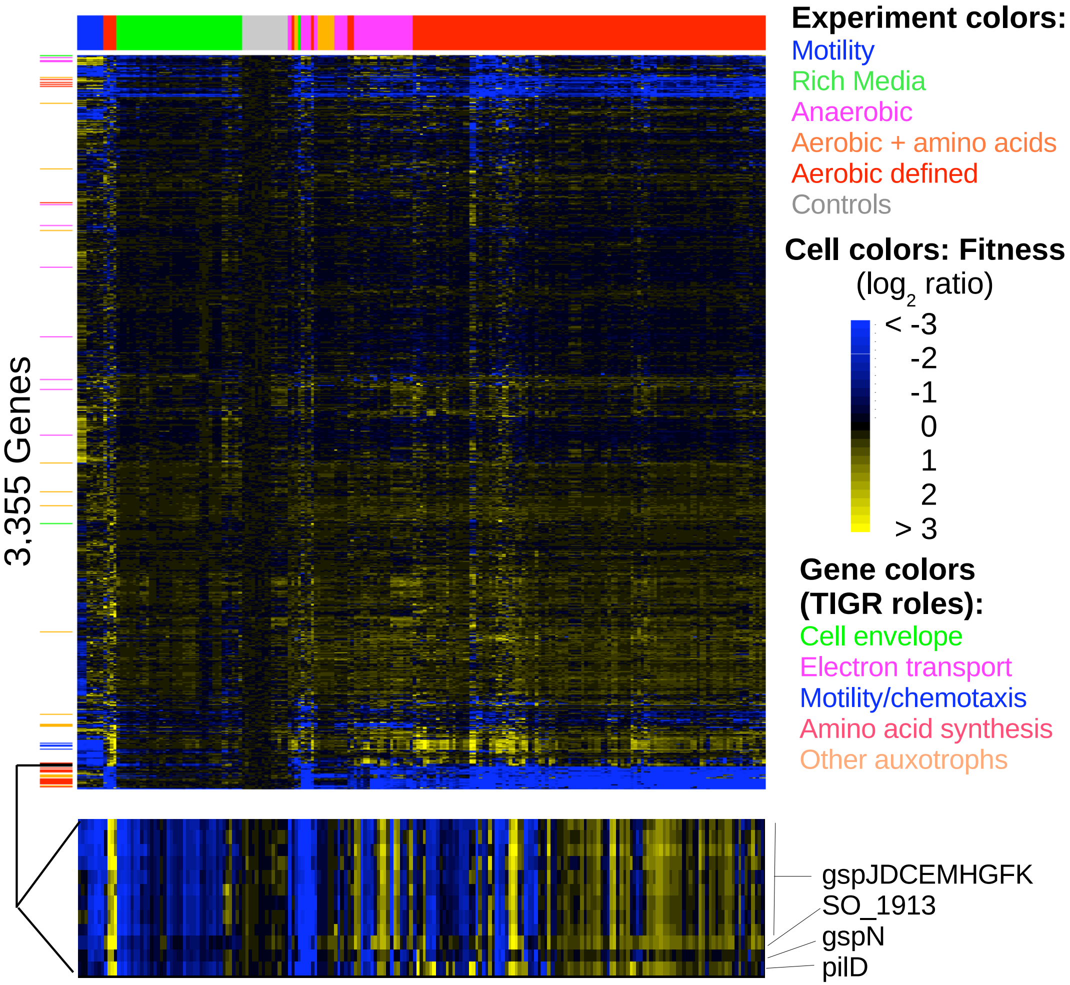
PGA1_c16670
“putative N-acylglucosamine 2-epimerase”
Actually, D-mannose isomerase

Predicted for *Shewanella amazonensis* (Rodionov et al. 2010) and also confirmed by our data.

Also identified roles for transporters and regulators

Global annotation of gene function

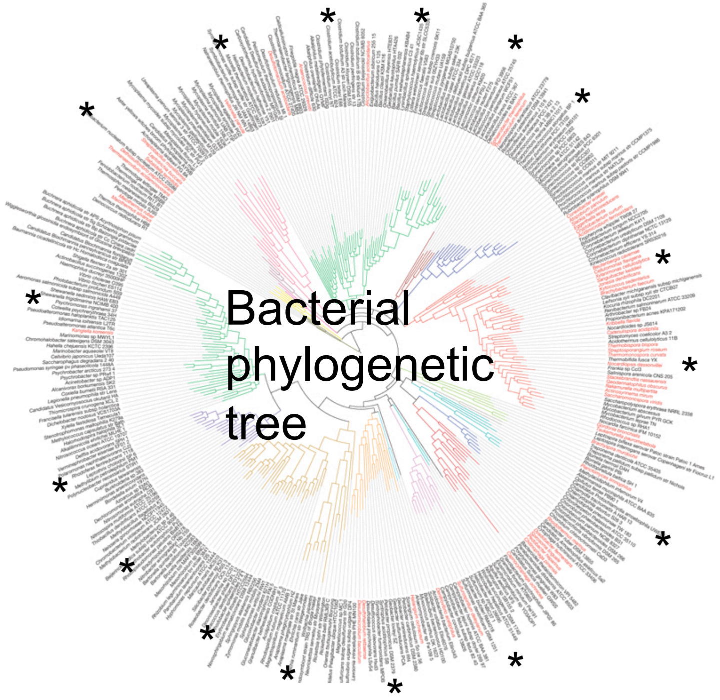
Example from *Shewanella oneidensis* MR-1

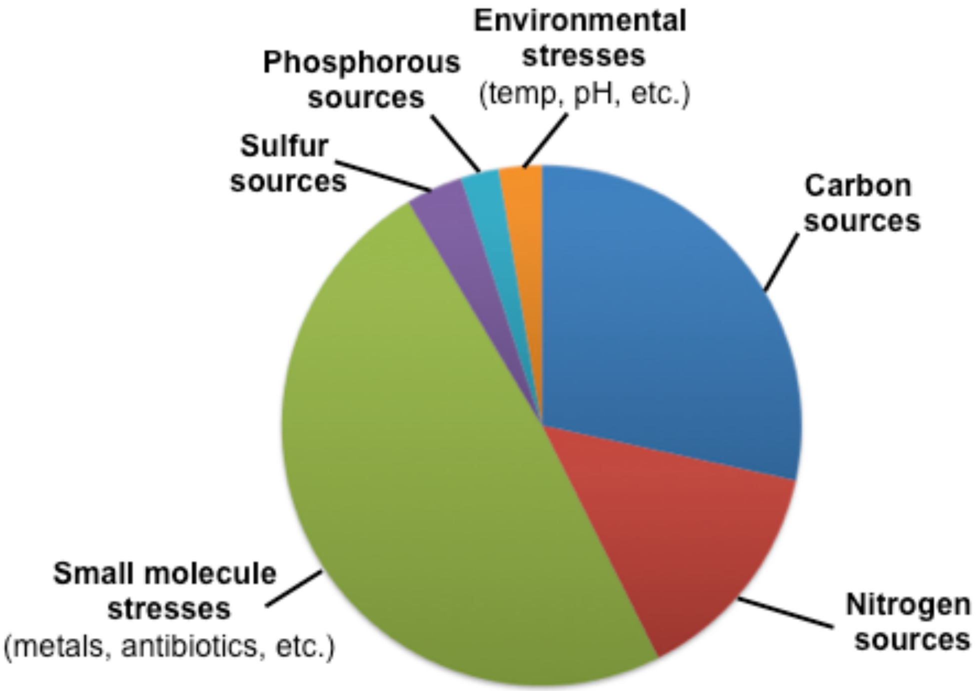


195 Experiments & 19 Controls

- **Phenotypes for over 2,000 genes including hundreds of previously uncharacterized genes**
- **Used fitness data to infer functions for 41 genes or operons**

Functional Encyclopedia of Bacteria and Archaea (FEBA)

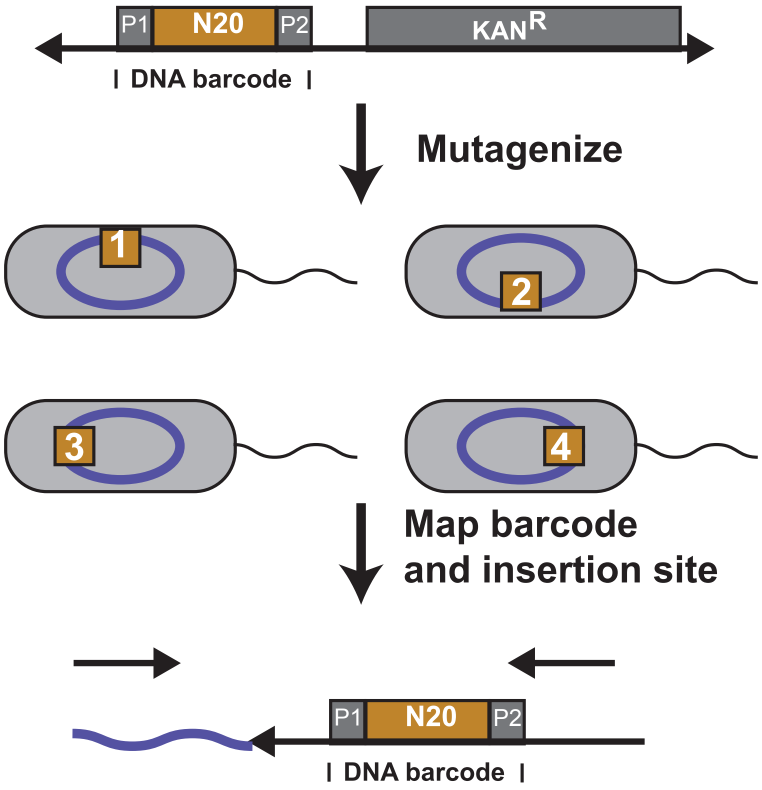




Mutagenesis of 20 diverse organisms × ~100 growth conditions

Random barcode transposon site sequencing (RB-TnSeq)

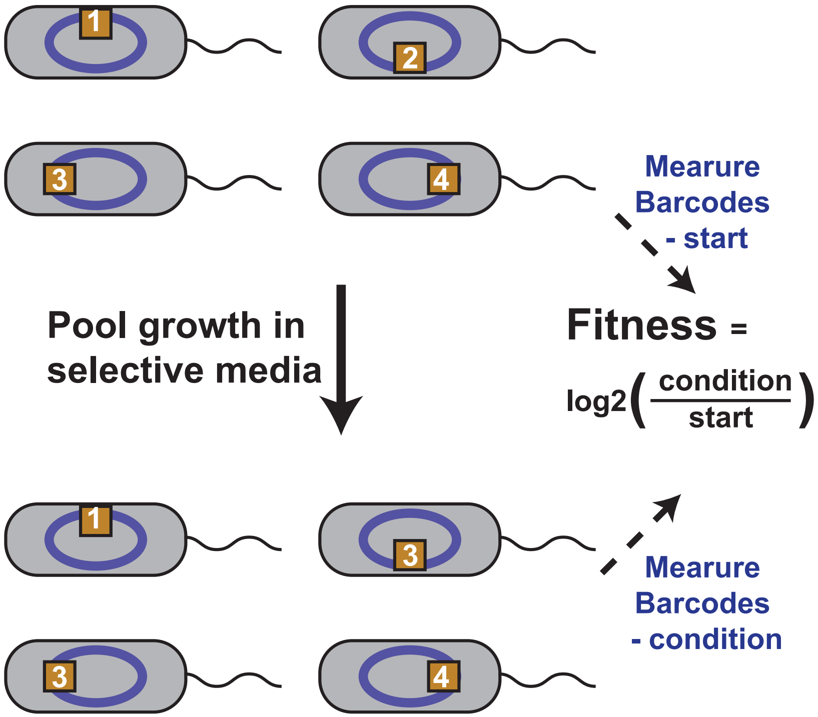
A. Characterize mutant population



Mutagenize

Map barcode and insertion site

B. Assay mutant fitness with DNA barcodes



Pool growth in selective media

Mearure Barcodes - start

Fitness = $\log_2\left(\frac{\text{condition}}{\text{start}}\right)$

Mearure Barcodes - condition

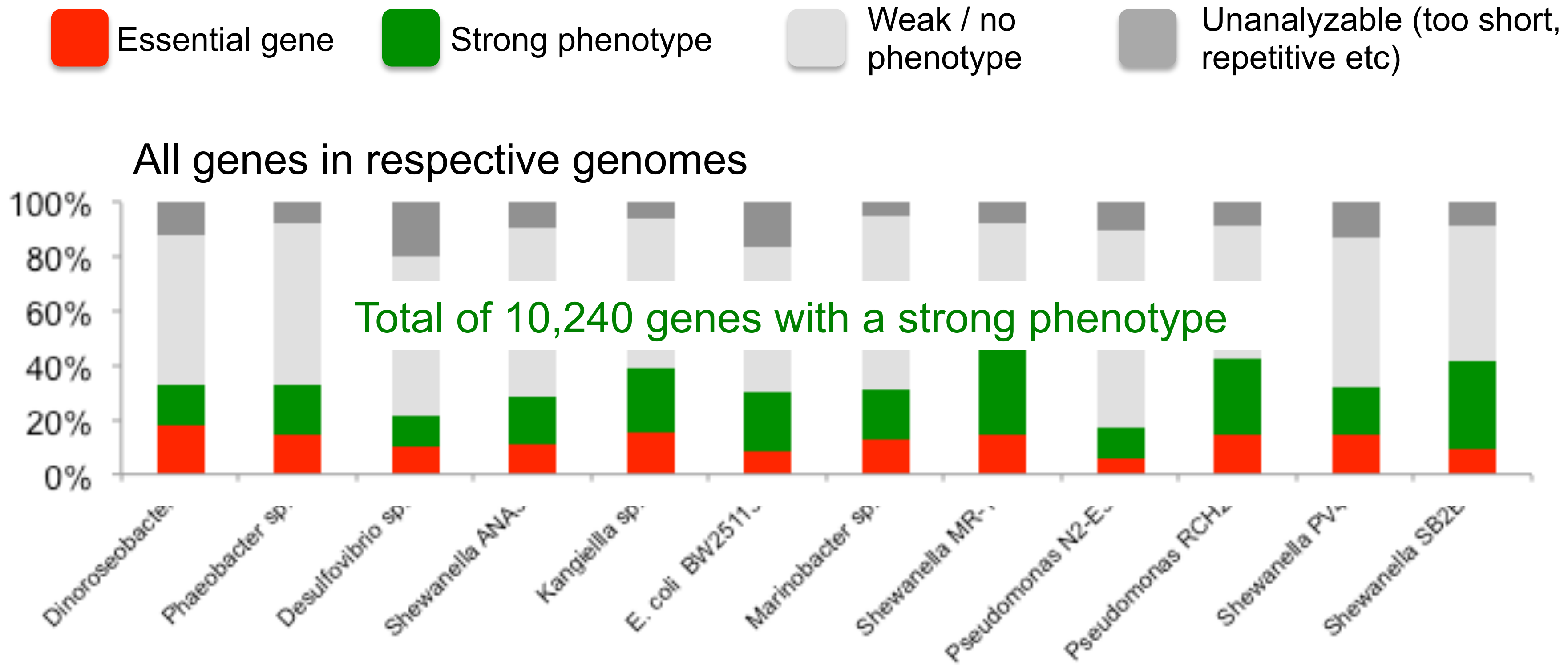
See Wednesday poster – Kelly Wetmore

Properties of 20 FEBA mutant libraries

Phylum / Class	Species	# Mutants
alpha-proteobacteria	Phaeobacter inhibens	215,858
	Dinoroseobacter shibae	119,703
	Sphingomonas koreensis	226,420
beta-proteobacteria	Azospira suillum PS	241,691
	Cupriavidus sp. 4G11	112,124
gamma-proteobacteria	Marinobacter adhaerens	85,718
	Pseudomonas sp. N2-F3	510,225
	Shewanella oneidensis	181,569
	Pseudomonas sp. N1-B4	176,258
	Pseudomonas stutzeri RCH2	166,448
	Escherichia coli	152,018
	Pseudomonas sp. N2-E2	110,065
	Kangiella aquimarina	100,069
	Shewanella sp. ANA3	69,557
	Shewanella sp. PV4	51,094
delta-proteobacteria	Desulfovibrio miyazaki	28,252
cyanobacteria	Synechococcus elongatus	141,062

Average 170,000 mutants per genome
Average 60 mutants per gene

Summary of results from 12 bacteria



Total of 10,240 genes with a strong phenotype

Summary of results from 12 bacteria

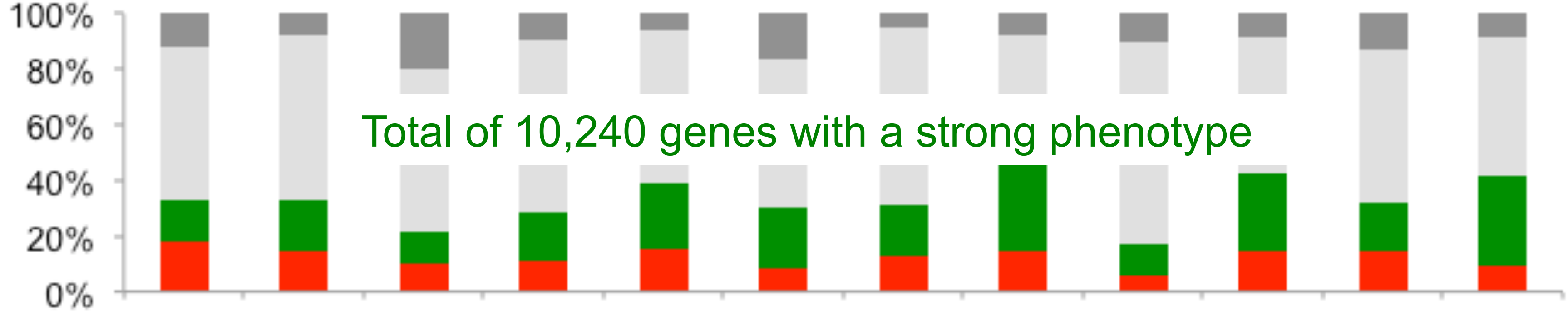
🟥 Essential gene

🟩 Strong phenotype

⬜ Weak / no phenotype

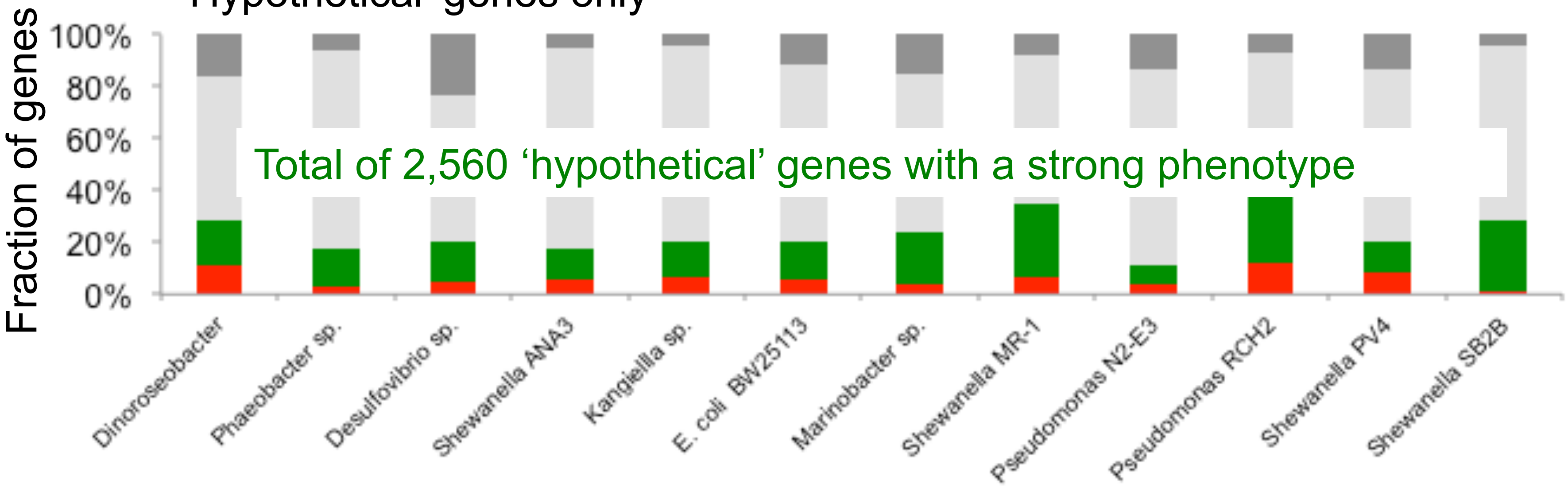
▩ Unanalyzable (too short, repetitive etc)

All genes in respective genomes



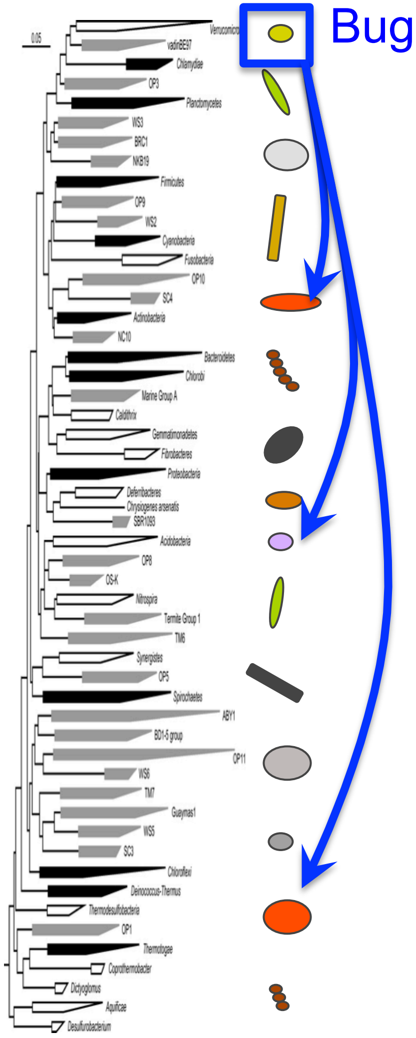
Total of 10,240 genes with a strong phenotype

‘Hypothetical’ genes only



Total of 2,560 ‘hypothetical’ genes with a strong phenotype

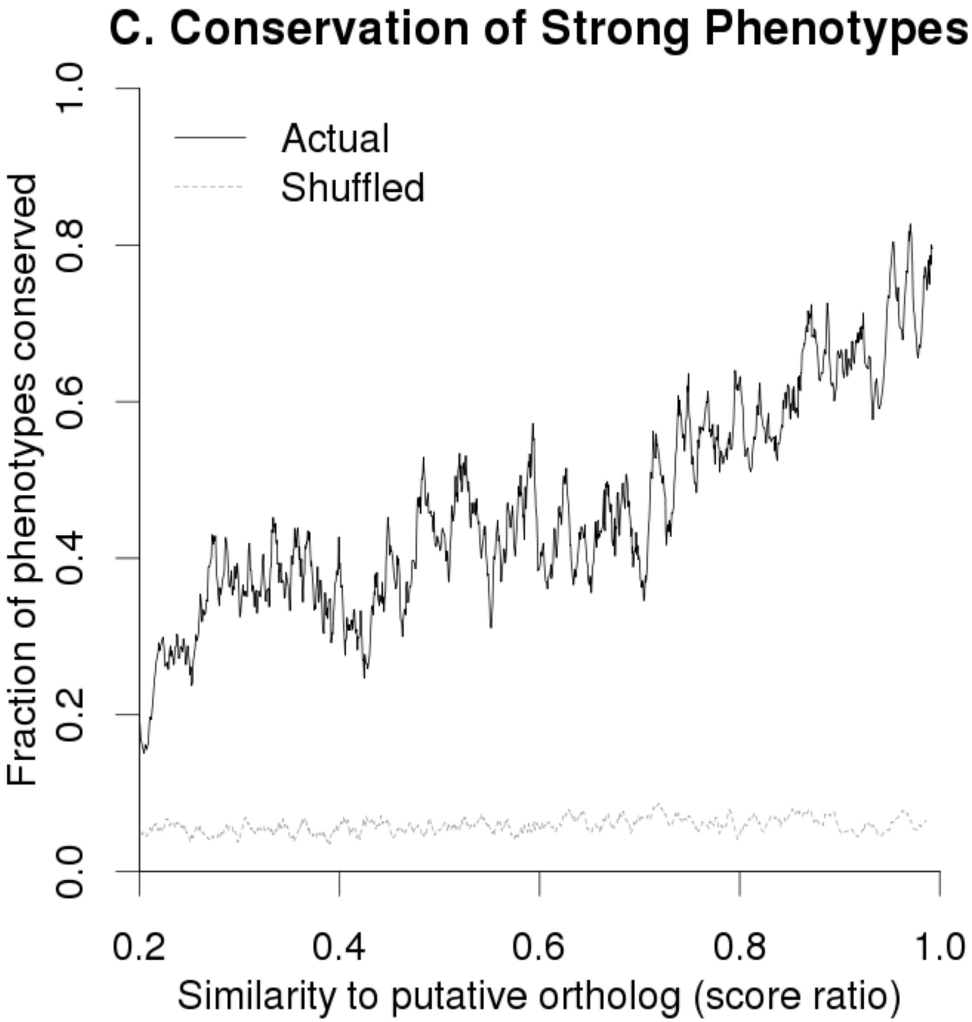
Can mutant fitness annotations be propagated across diverse genomes?



Bug with FEBA annotations

Does: sequence conservation = function conservation ?

Strong phenotypes are often conserved between distant homologs

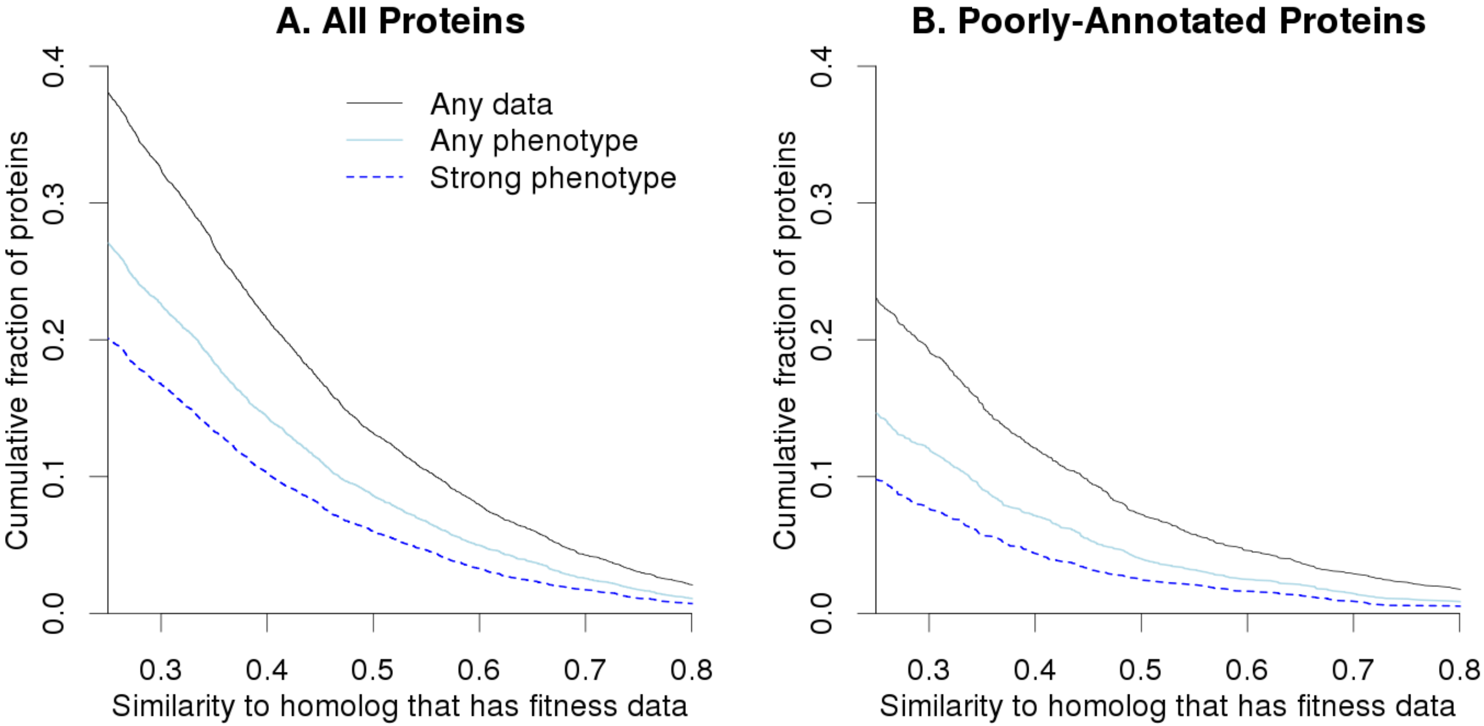


Shows why annotation by homology (mostly) works

Can objectively define “functional orthologs” to study evolution?

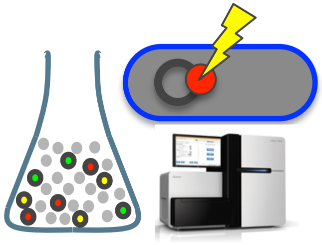
Coverage of all bacterial proteins

Take a random sample of bacterial proteins:
do we have phenotypes for their homologs?

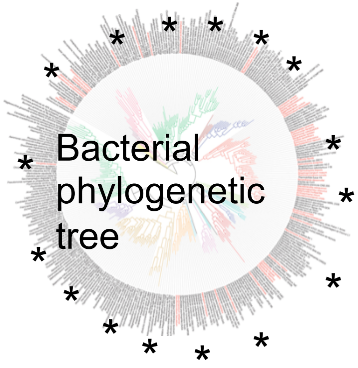


$$similarity(A, B) = \frac{protein\ BLAST\ score}{self\ score}$$

Summary



We have developed RB-TnSeq, a high-throughput approach to annotate gene function in bacteria



The ‘FEBA’ project will provide functional annotation for 20 diverse organisms / 1000s novel genes

JGI
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE
User Programs

Future ‘product’ of JGI? Keen to target bugs of interest to DOE and to JGI user community

Thanks to

Morgan Price
Kelly M. Wetmore
R. Jordan Waters
Jayashree Ray
Jennifer V. Kuehl
Jacob S. Lamson
Cindi A. Hoover
Matthew J. Blow
Ryan A. Melnyk
Ben E. Rubin
Michael W. Adams
Jim Bristow
Adam P. Arkin

and funding from

- ENIGMA – Ecosystems and Networks Integrated with Genes and Molecular Assemblies
- LBL Laboratory Directed Research and Development (LDRD)
- A Community Science Proposal at JGI
  (US Department of Energy's Joint Genome Institute)