

An integrative genome-scale *E. coli* model for Systems and Synthetic biology

Ilias Tagkopoulos



Assistant Professor
Department of Computer Science &
UC Davis Genome Center



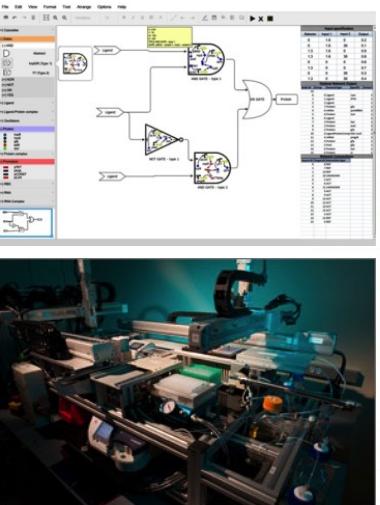
University of California, Davis

Systems and Synthetic Biology

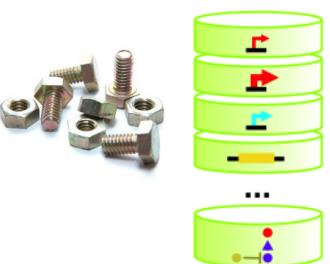


ENGINEERING FOR KIDS

Circuit design automation



Part construction and Characterization

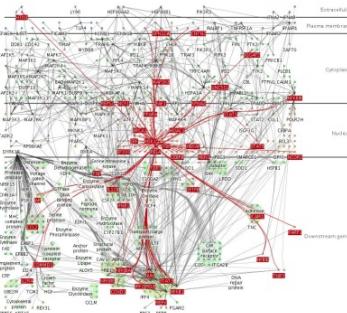


Machine Learning Algorithms Integration and Inference

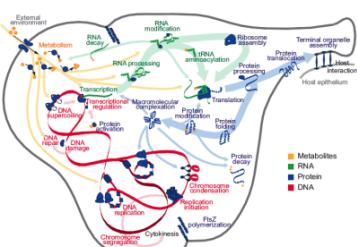


A large, shiny red heart is shown from a three-quarter perspective. A prominent, jagged vertical tear runs down the center of the heart, from top to bottom. The heart is set against a white background. To the left of the tear, the words "Synthetic Biology" are written in a bold, black, sans-serif font. To the right of the tear, the words "Systems Biology" are also written in a bold, black, sans-serif font. The overall image has a graphic, minimalist aesthetic.

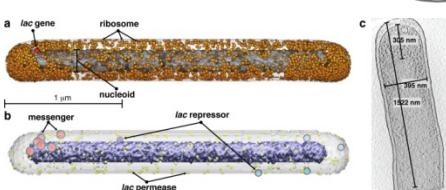
Biological Networks



Whole cell modeling

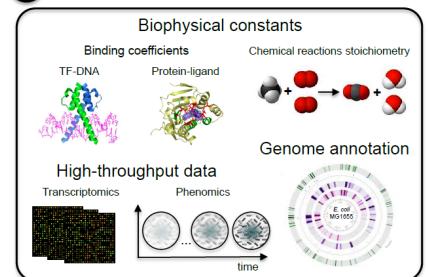


Multi-scale simulation

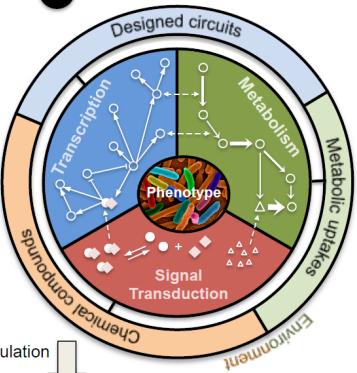


Relevant projects in our lab

A Training data



B Computer model



D Validation and extension

C Experimentally testable predictions

Decision support tools for agriculture and medicine



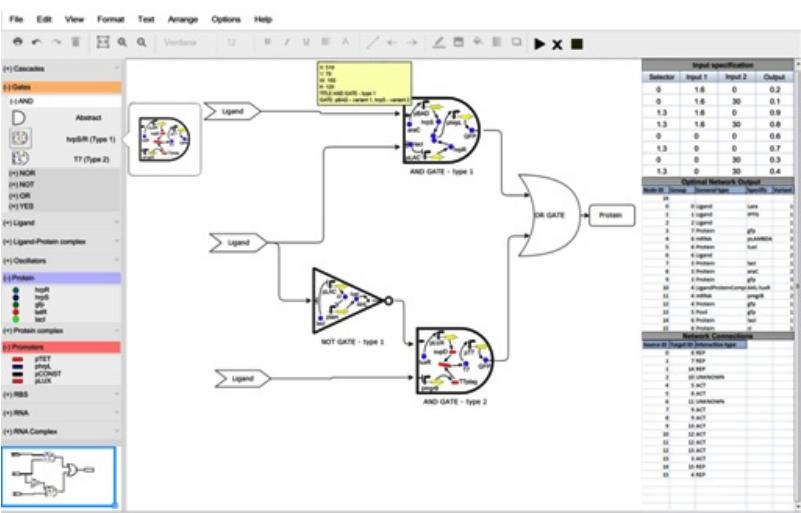
Multi-scale modeling and HPC simulation

XSEDE

Extreme Science and Engineering Discovery Environment



Automated CAD tools for Synthetic Biology

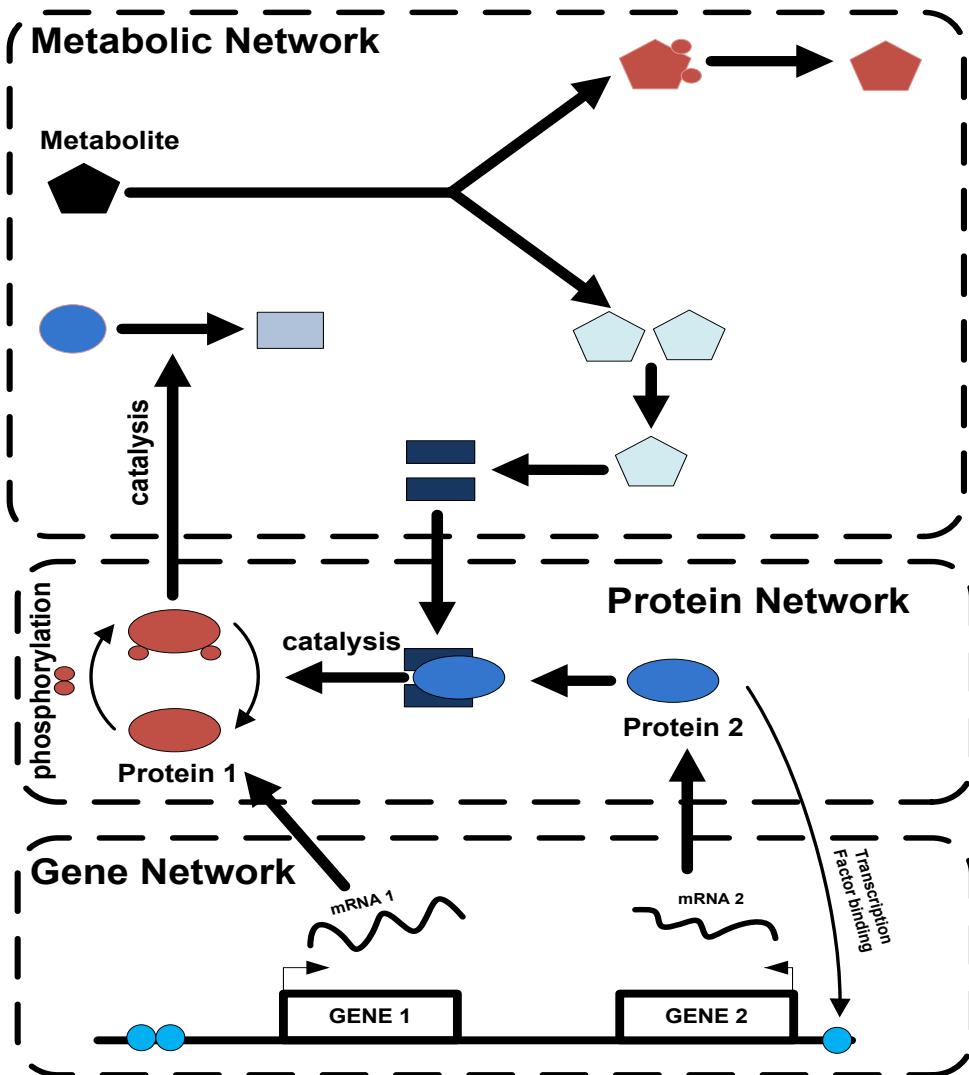




Project goal Create an **abstract, multi-scale microbial evolution model** to:

- investigate the **dynamics of microbial evolution** in complex environments
- assess its effect on **microbial organization** across the various biological layers.

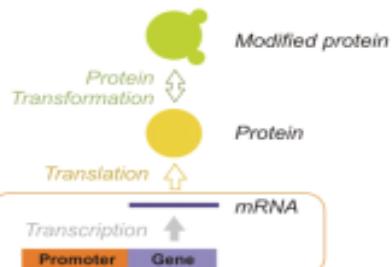
Focus on **general evolutionary principles** and **microbial organization**



EVE: Evolution in Variable Environments



Capturing “Central dogma” of Molecular Biology



mobile genetic element

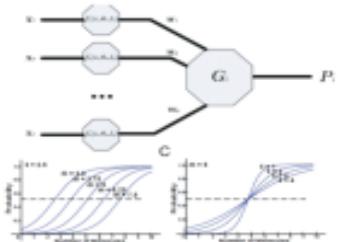
Transformation

Apoptosis

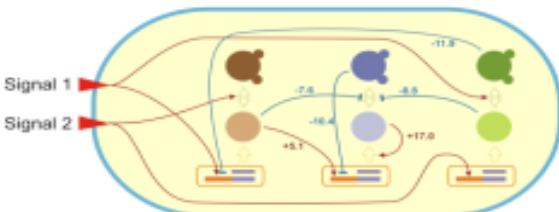
Conjugation

Division

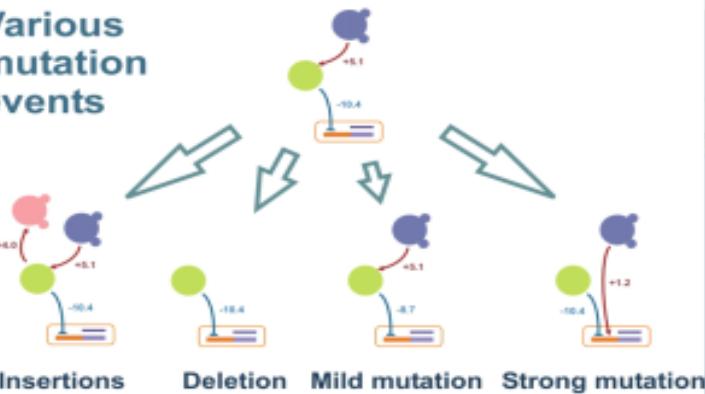
Expression model



Cell and its gene regulatory network



Various mutation events





Challenges:

- Balancing **biological realism** and **computational feasibility**
- Different **spatial and temporal scales** across various phenomena (genes, proteins, networks, cells, populations)
- Modeling evolution, inherently **hard to predict**, leads to different cell sizes, growth rates, etc., which translates to **load imbalance**
- **Size does matter**: a bacterial colony may have **billions of cells** and small population size leads to biases (clonal interference, genetic drift)
- **Storage, fast access and visualization** of the ***evolutionary fossil record*** from replicate simulations.

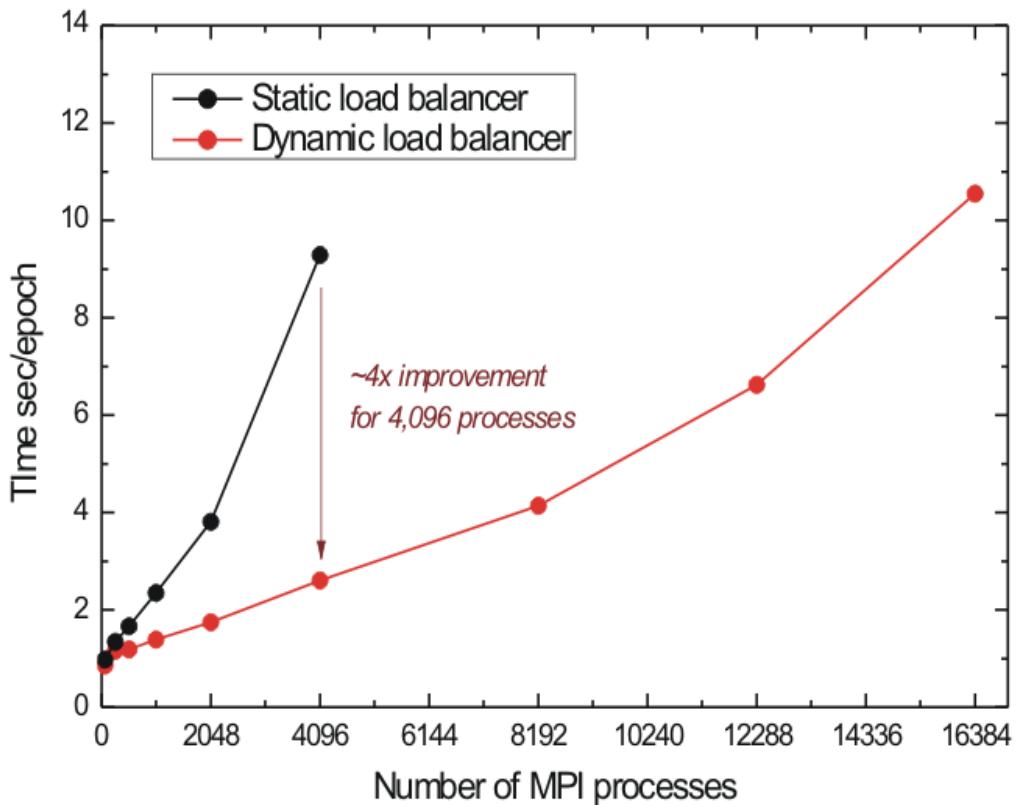
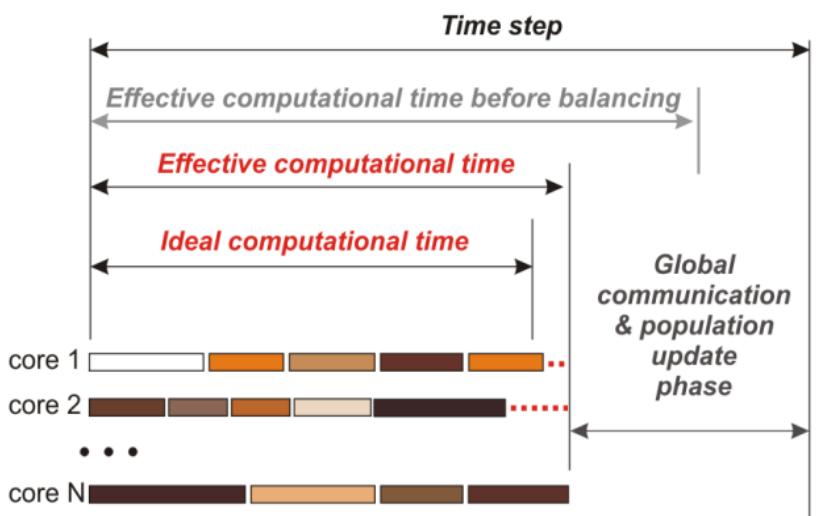


What have we accomplished

Accomplishments:

- EVE microbial evolution simulator scaling up to **8,000 MPI processes** and **128,000 cells**.
 - To compare: our previous work (run on a Blue Gene machine) scaled up to **200 organisms** and had a less complex underlying model
 - I. Tagkopoulos, Y. Liu, S. Tavazoie, "Predictive Behavior Within Microbial Genetic Networks", *Science*, 320:1313-7, 2008
- Tackling growth and evolution: **static and adaptive load balancing** for both fixed and non-fixed population sizes
 - V. Mozhayskiy, I. Tagkopoulos, "In silico Evolution of Multi-scale Microbial Systems in the Presence of Mobile Genetic Elements and Horizontal Gene Transfer", ISBRA2011, *Lecture Notes in Bioinformatics*, LNBI 6674, pp.262-273, Springer, 2011
 - V. Mozhayskiy, R. Miller, KL. Ma, I. Tagkopoulos, "A Scalable Multi-scale Framework for Parallel Simulation and Visualization of Microbial Evolution", *TeraGrid2011*; Salt Lake City, Utah, 2011 (**Best of Science** and **Best of Show** awards)
- Other extensions: HDF5 storage, AMPI

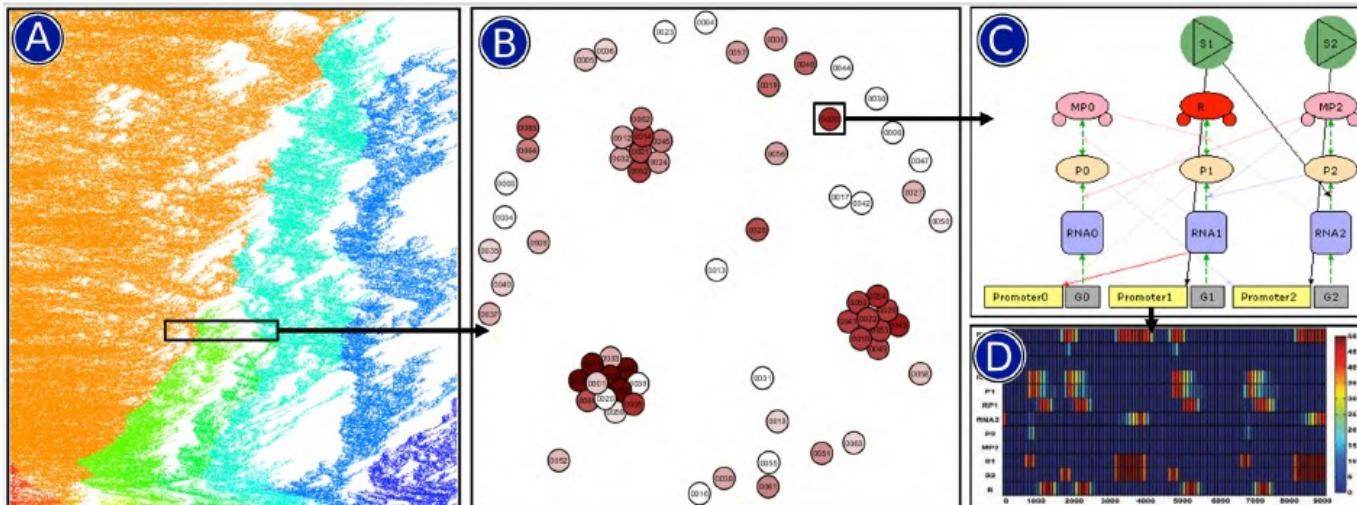
Load Balancing in simulations of cell populations



Visualization



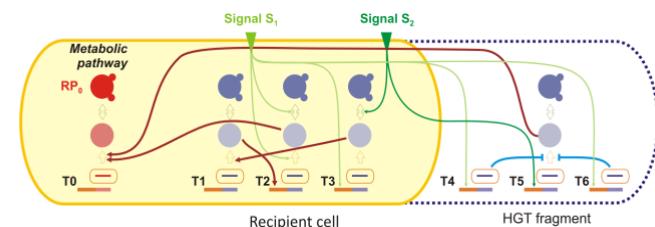
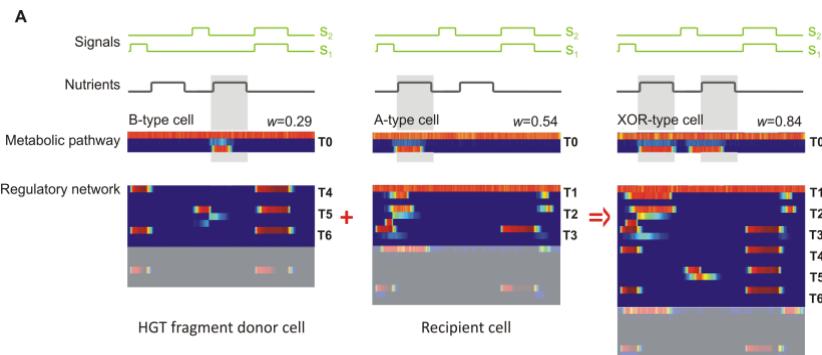
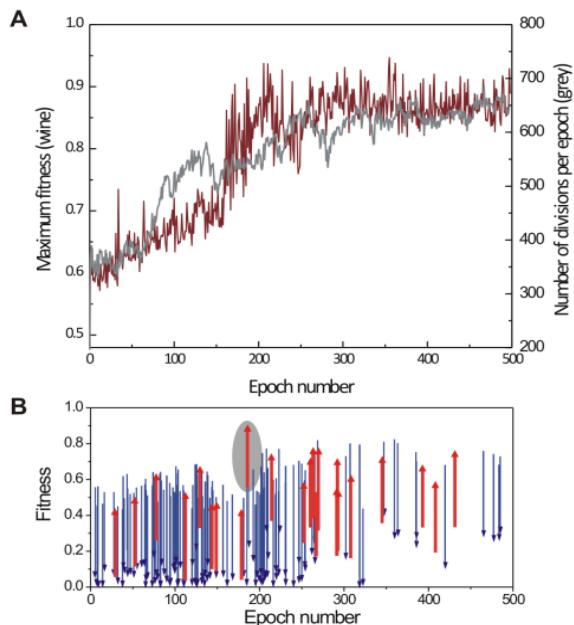
- **Visualization and information flow analysis tools** to cluster cells based on phenotype, view hierarchical organization, mutations and evolutionary trajectory.
 - R. Miller, V. Mozhayskiy, I. Tagkopoulos, KL. Ma, "EVEVis: A Multi-Scale Visualization System for Dense Evolutionary Data", *1st IEEE Symposium on Biological Data Visualization*, pp. 143-150, Rhode Island, 2011
 - A. Pavlogiannis, V. Mozhayskiy, I. Tagkopoulos, "A flood-based information flow analysis and network minimization method for bacterial systems", 14:137 DOI: 10.1186/1471-2105-14-137, *BMC Bioinformatics*, 2013



Large-scale simulations to explore evolutionary hypotheses



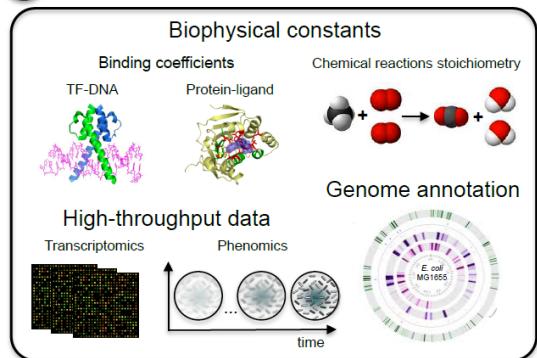
- Used to investigate **Horizontal Gene Transfer** and **evolutionary hypotheses**
 - V.Mozhayskiy, I.Tagkopoulos, "Horizontal gene transfer dynamics and distribution of fitness effects during microbial In silico Evolution", 13:S13, *BMC Bioinformatics*, 2012
 - V.Mozhayskiy, I.Tagkopoulos, "Guided evolution of in silico microbial populations in complex environments accelerates evolutionary rates through a step-wise adaptation", 13:S10, *BMC Bioinformatics*, 2012
 - V. Mozhayskiy, I. Tagkopoulos, "Microbial evolution *in vivo* and *in silico*: methods and applications", 5(2):262-77, *Integrative Biology*, 2013



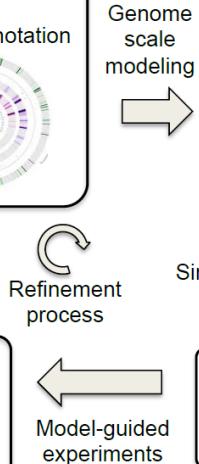
The road ahead

A data-driven probabilistic graphical model for microbial evolution (*E. coli* as a proof-of-concept case)

A Training data



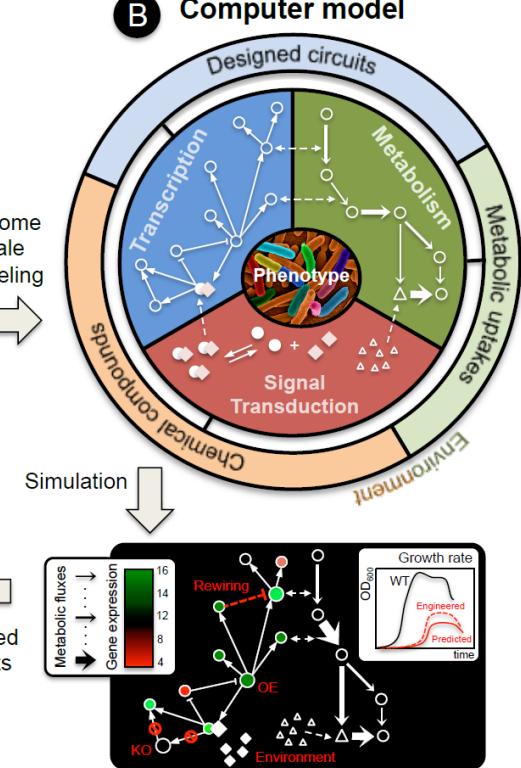
Integration



Refinement process

Model-guided experiments

B Computer model



Genome scale modeling

Simulation

C Experimentally testable predictions

Knowledgebase

Signal Transduction Systems
EcoST

Phenomics
EcoPhe

Transcriptomics
EcoMAC

Databases & Models

Transcriptional Regulatory Network
RegulonDB

Metabolic reactions
Orth et al., 2011

Integrative Genome-Scale Models

Input

Genetic perturbations (KO, OE, rewiring)
Environmental perturbations (nutrients)

EBA

$\min(\text{gene expression error})$
given:
Phenomenological constraints
Capacity constraints
Genetic constraints
Environmental constraints

FVA

Identify flux boundaries,
given:
Stoichiometric constraints
Capacity constraints
Wild-type growth rate

TAME

Transcription-based metabolic fluxes
Network inference and reconstruction

FBA

$\max(\text{growth rate}),$
given:
Stoichiometric constraints
Capacity constraints

Output

$\text{Growth rate} = \text{Benefit} - \text{Cost}$

Biological Modules

Transcription

Signal Transduction

Metabolism

Integration

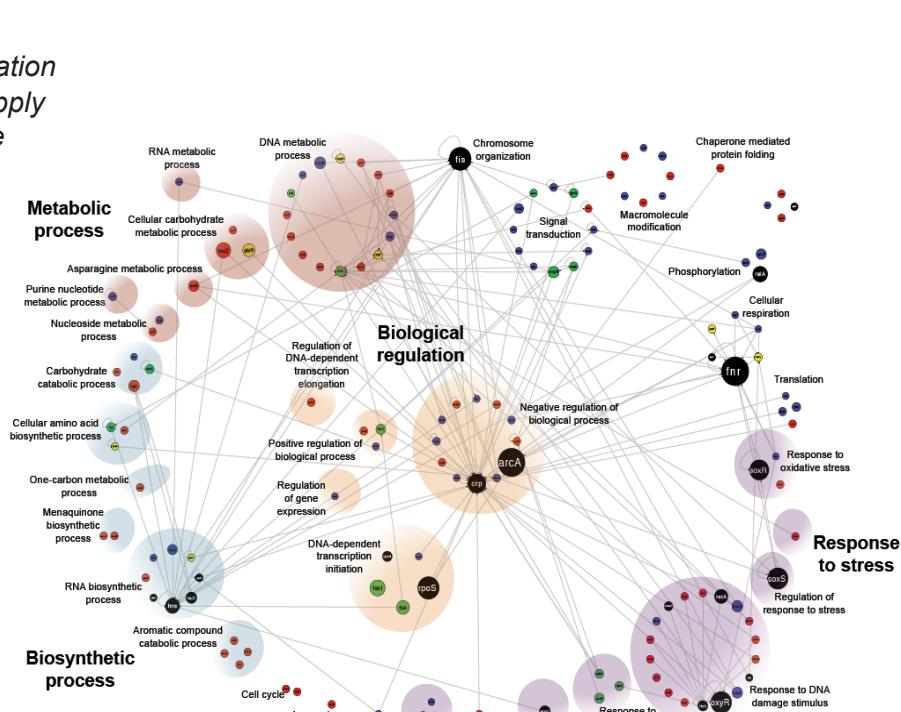
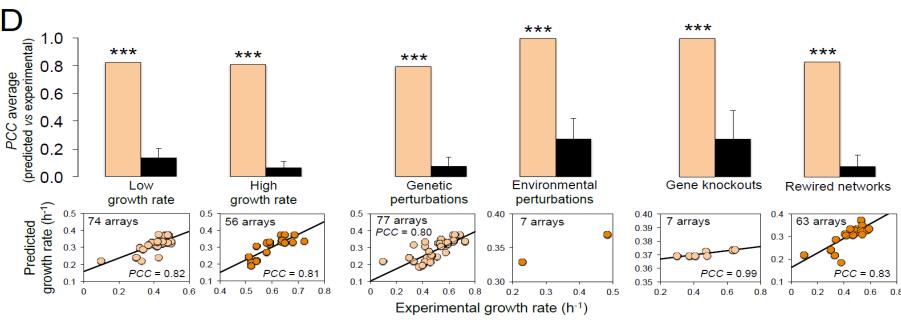
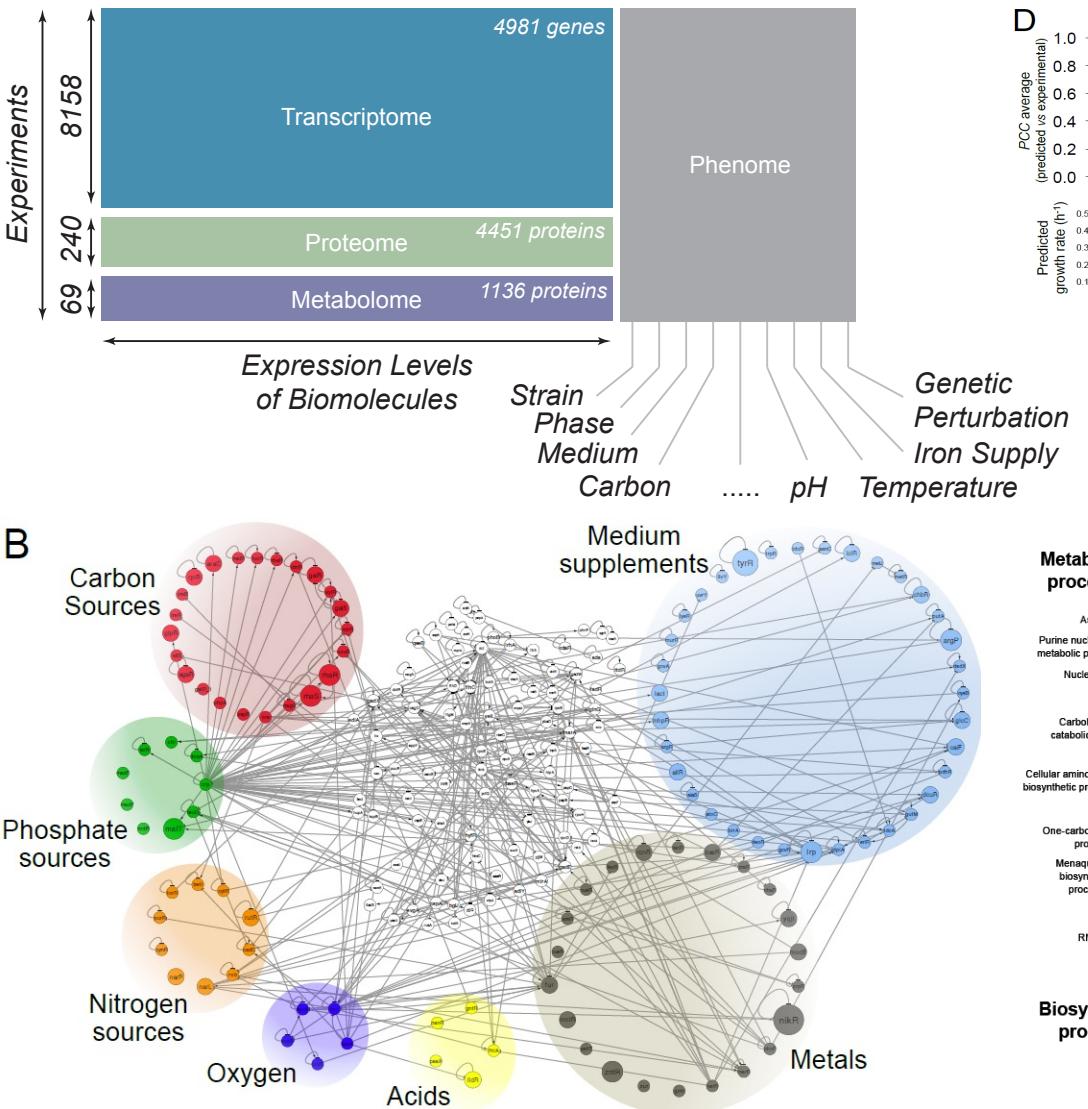
D Validation and extension

J. Carrera, R. Estrela, J. Luo, N. Rai, A. Tsoukalas, I. Tagkopoulos, "An integrative, multi-layer, genome-scale model reveals the phenotypic landscape of *Escherichia coli*", accepted, *Nature/EMBO Molecular Systems Biology*, 2014

An integrative genome-scale model for *E. coli*



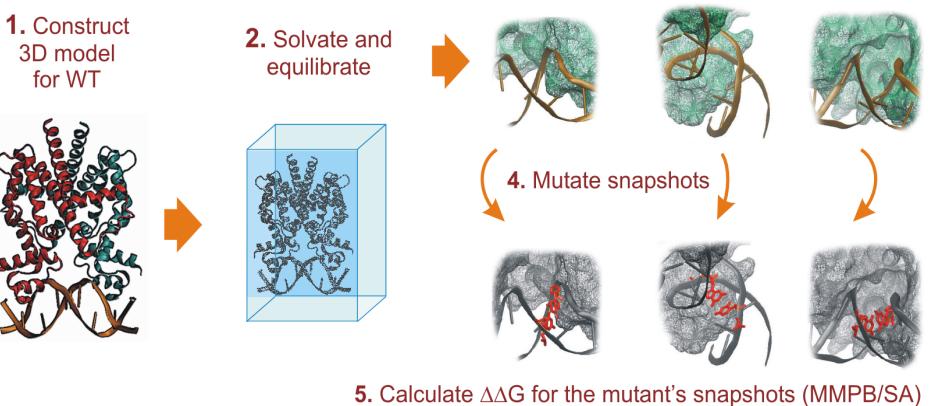
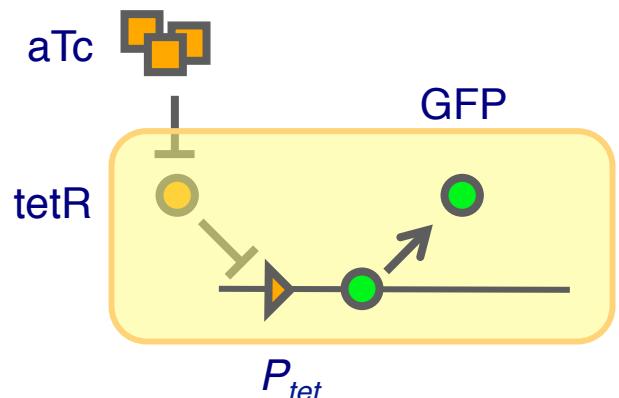
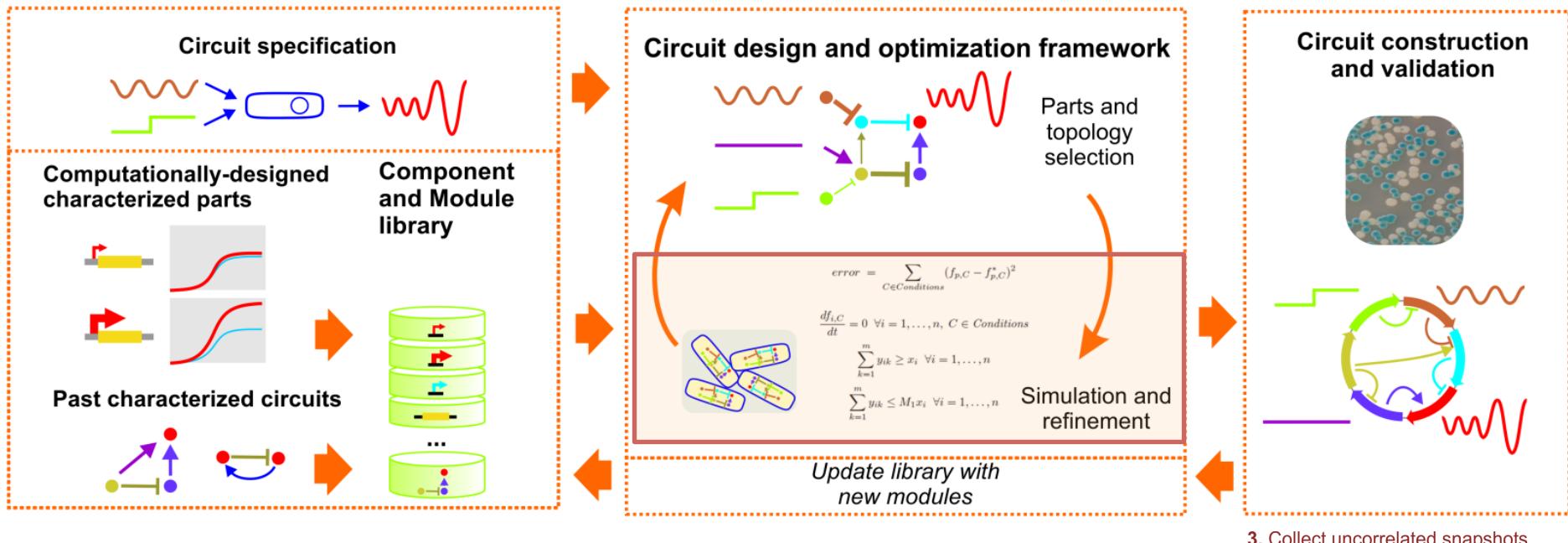
***E. coli* Multi-OMICs Dataset**



Genome-scale models in Synthetic Biology



SBROME: Synthetic Biology Reusable Optimization Methodology



Acknowledgements



Lab members:

Postdocs

- Violeta Zorraquino
- Navneet Rai
- Nasos Tsoukalas
- Dina Zhabinskaya
- Javier Carrera (now with Markus Covert, Stanford)
- **Vadim Mozhayskiy** (now with Life Technologies)



Graduate students

- Linh Huynh
- Minseung Kim
- Runyu Shi
- Matt Meisner
- Beatriz Pereira

Funding



Links

<http://tagkopouloslab.ucdavis.edu>



- Dragosits, V. Mozhayskiy, S. Quinones-Soto, I. Tagkopoulos, “**Evolutionary potential, cross-stress dependencies, and the genetic basis of acquired stress resistance in of *E. coli***”, doi:10.1038/msb.2012.76, 9:643, *Molecular Systems Biology*, 2013
- J. Carrera, R. Estrela, J. Luo, N. Rai, A. Tsoukalas, I. Tagkopoulos, ” **An integrative genome-scale model reveals the phenotypic landscape of *Escherichia coli*.**”, accepted, *Molecular Systems Biology*, 2014.