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
Relevant projects in our lab

- Multi-scale modeling and HPC simulation
  - Automated CAD tools for Synthetic Biology
  - Decision support tools for agriculture and medicine
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
Challenges in modeling microbial evolution

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

• Tackling growth and evolution: **static and adaptive load balancing** for both fixed and non-fixed population sizes
  • 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

![Graph showing load balancing comparison]

- **Time step**
  - Effective computational time before balancing
  - Effective computational time
  - Ideal computational time

- **Global communication & population update phase**

- **Graph details**:
  - **Static load balancer** vs **Dynamic load balancer**
  - ~4x improvement for 4,096 processes

- **Axes**:
  - **X-axis**: Number of MPI processes
  - **Y-axis**: Time sec/epoch
**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
Large-scale simulations to explore evolutionary hypotheses

- Used to investigate **Horizontal Gene Transfer** and **evolutionary hypotheses**
The road ahead

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

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

**E. coli Multi-OMICs Dataset**

- **Transcriptome**: 4981 genes
- **Proteome**: 4451 proteins
- **Metabolome**: 1136 proteins

- **Expression Levels of Biomolecules**: Strain, Phase, Medium, Carbon, pH, Temperature
- **Phenome**: Genetic Perturbation, Iron Supply
- **Experiments**: Growth, Synthetic Lethality, Essentiality
- **Strain Phase Medium**: Genetic Perturbation, Iron Supply

**B**

- **Carbon Sources**: Medium supplements
- **Phosphate sources**: Metabolic process
- **Nitrogen sources**: Signal Transduction
- **Oxygen**: Upstream of Signal Transduction
- **Metals**: Evolutionary Consequences Processes
- **Acids**: Upstream of Signal Transduction
**SBROME: Synthetic Biology Reusable Optimization Methodology**

**Genome-scale models in Synthetic Biology**

1. Construct 3D model for WT
2. Solvate and equilibrate
3. Collect uncorrelated snapshots from an MD trajectory for WT
4. Mutate snapshots
5. Calculate ΔΔG for the mutant's snapshots (MMPB/SA)

Circuit specification

- Computationally-designed characterized parts
- Component and Module library
- Past characterized circuits

Circuit design and optimization framework

- Parts and topology selection
- Simulation and refinement

Update library with new modules

aTc

GFP

P_{tet}

tetR
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