Petascale Simulations of Complex Biological Behavior in Fluctuating Environments

Ilias Tagkopoulos, Ph.D.
Principal Investigator
Department of Computer Science and UC Davis Genome Center
University of California, Davis

Abstract

One of the central challenges in computational biology is the development of predictive multi-scale models that can capture the diverse layers of cellular organization. Even more scarce are models that encode biophysical phenomena together with evolutionary forces, in order to provide insight on the effect of adaptation at a systems-level. Over the last five years, our lab has created a multi-scale abstract microbial evolution model that unifies various layers, from diverse molecular species and networks to organism and population-level properties. With the help of Blue Waters and the NCSA team, we are able to scale up to hundreds of thousand cells, an unprecedented scale of simulation that is, however, well short of the billions of cells that are present in a bacterial colony. Here, we present our scalability results, the methods that we employed to achieve them and our current work on a data-driven, genome-scale, population-level model for *Escherichia coli.*

Specific Aims

The goal of this project is to create a scalable model and simulation framework to (a) investigate the dynamics of microbial evolution in complex environments and (b) assess its effect on microbial organization across the various biological layers. The simulation framework should be focused on the general principles governing evolution and microbial organization, so it can be generalized.

Significance

Microbes are the most abundant and diverse forms of life on Earth. Their impact in the human race and our ecosystem as a whole is difficult to exaggerate. They have been used extensively in industrial applications, ranging from bioremediation to production of organic compounds and they are relevant to human health, both as probiotics and as pathogens.

Over the past decades, we have studied extensively microbial organisms and have gained valuable insights on their system-level properties, as well as the mechanistic underpinnings of their complex behavior. Less is known about their potential to acquire new traits and become resilient to adverse environmental conditions through evolutionary forces, such as random mutations, horizontal gene transfer and genetic drift. Of particular interest is to elucidate the effect of such environments to their gene regulatory and biochemical networks that in turn can lead to a better understanding to what is possible, likely and potentially transformative to the environment they occupy. From antibiotic resistance, to stress-resistant biotechnological strains for recombinant protein production, such knowledge will have a tremendous impact on various industrial, agricultural and medical fields. While there have been many studies of adaptive laboratory evolution in the past couple years, these are limited to a few thousands generations that can hardly capture the vast phenotypic space that microbes can explore. Hence, the development of computational modeling and simulation tools that can capture these phenomena across multiple scales can lead to transformative advances in this field.

Challenges

There are a number of challenges that have to be addressed to achieve such feat. First, a model of biological organization that is both ***biological realistic and computationally feasible*** is paramount, incorporating the right level of biological abstraction. Second, the ***spatial and temporal scales*** of a model that encompasses genes, proteins, networks, cells and populations are very diverse, which create additional hurdles when

applying numerical methods to solve them. Third, since *evolution* is based on random mutations and natural selection, it ***is inherently hard to predict*** and can lead to imbalances in the distribution of active cells, and in its extension, computational tasks. Fourth, a typical microbial colony has ***billions of cells***, while current simulations are at most in the thousands, which leads to size-specific artifacts as ***size does matter***. Finally, ***storing and visualizing*** the fossil record of an evolutionary trajectory, especially since dozens of them are needed for assessing statistical significance for any hypothesis-testing experiment, is not an easy task, as a simulation can easily lead to terabytes of complex data for analysis.

Accomplishments

We have created the Evolution in Variable Environments (EVE) v3.0 synthetic ecology framework that is currently the most sophisticated, abstract simulator for microbial evolution with the capacity to scale up to 8,000 MPI processes and 128,000 organisms. To compare, our previous work (before the PRAC award) scaled up to 200 organisms with a less complex underlying model (1). To cope with unforeseen computational load due to the emergence of complex phenotypes, we have developed both static and adaptive load balancers (TG'11 Best Paper Award) that can account for both fixed and non-fixed population sizes (2-3). We developed intuitive visualization tools (4), HDF5 storage solutions, and novel analysis algorithms based on network flows (5) to efficiently project data to accelerate biological discovery. The EVE simulator has since been used to investigate the effect of horizontal gene transfer (6), distribution of fitness effects and the hypothesis of accelerated evolution through guided, step-wise adaptation (7) with interesting results that drive biological experimentation (8-9). Future work includes pushing the limits of microbial simulations to break the million-cell barrier, parallelization of organism-specific, data-driven models that integrate omics layers, starting from our recent work in the model bacterium *Escherichia coli* (10) and integration with Synthetic Biology computer-aided design tools for targeted, chassis-aware genome engineering (11-14).

References

1. I.Tagkopoulos, Y.Liu, S. Tavazoie, "Predictive Behavior Within Microbial Genetic Networks", **Science**, 320:1313-7, 2008

2. 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

3. 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, DOI:10.1145/2016741.2016749

4. 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

5. 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

6. V.Mozhayskiy, I.Tagkopoulos, "Horizontal gene transfer dynamics and distribution of fitness effects during microbial In silico Evolution", doi: 10.1186/1471-2105-13-S10-S13, 13:S13, **BMC Bioinformatics**, 2012

7. 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

8. V. Mozhayskiy, I. Tagkopoulos, "Microbial evolution *in vivo* and *in silico*: methods and applications", DOI:10.1039/C2IB20095C, 5(2):262-77, **Integrative Biology**, 2012

9. M. Dragosits, V. Mozhayskiy, S. Quinones-Soto, J. Park, I.Tagkopoulos, "Evolutionary potential,cross-stress behavior, and the genetic basis of acquired stress resistance in *Escherichia coli*", doi:10.1038/msb.2012.76, 9:643, Nature/EMBO **Molecular Systems Biology**, 2013

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

11. L. Huynh, J. Kececioglu, M. Köppe, I.Tagkopoulos, "Automated Design of Synthetic Gene Circuits through Linear Approximation and Mixed Integer Optimization", 7(4):e35529, **PLoS ONE**, 2012

12. L. Huynh, I.Tagkopoulos, "A robust, library-based, optimization-driven method for automatic gene circuit design", **2nd IEEE International Conference on Computational Advances in Bio and Medical Sciences** (ICCABS), Las Vegas, NV, pp.1-6, 24-26, 2012

13. L. Huynh, A. Tsoukalas, M. Köppe, I.Tagkopoulos, "SBROME: A scalable optimization and module matching framework for automated biosystem design", DOI: 10.1021/sb300095m, pp 263-273, **ACS Synthetic Biology**, 2013

14. L. Huynh, I. Tagkopoulos, "Optimal part and module selection for synthetic gene circuit design automation", doi:10.1021/sb400139h, **ACS Synthetic Biology**, 2014