

論文の内容の要旨

Deciphering Evolutionary Constraints through Microbial Laboratory Evolution combined with Machine Learning

(大腸菌進化実験と機械学習を用いた進化的拘束の探究)

氏名 岩澤 諄一郎

The prediction and control of evolution is a crucial topic for both evolutionary biology and tackling antibiotic resistance. Although the lack of sufficient data has long hindered the mechanism of evolution, laboratory evolution experiments equipped with high-throughput sequencing/phenotyping are now gradually changing this situation. The emerging data from recent laboratory evolution experiments revealed repeatable features in evolutionary processes, suggesting the existence of constraints which could lead to actual predictions of evolutionary outcomes. These results also paint an upbeat picture of evolution: biologically feasible states and evolutionary trajectories could be distributed on a low-dimensional manifold within the high-dimensional space spanned by biological features. However, previous laboratory evolution experiments were performed on a limited number of environments and we thus lack a systematic investigation of evolutionary constraints leading to the low-dimensional dynamics. In addition, fine-grained approaches for predicting and controlling evolutionary trajectories were out of the reach. This dissertation is dedicated to solve these problems through the utilization of laboratory evolution combined with machine learning based techniques.

First, we study the evolutionary constraints of *Escherichia coli* through the multi-omics data acquired from a large-scale laboratory evolution experiment. In biological data, it is often the case that the dimensionality p is much larger than the number of samples N which makes it difficult to perform statistical analyses such as covariance estimation. We will show how the the utilization of machine learning such as random forest regression and supervised principal component analysis could overcome this $p \gg N$ problem and contribute to probe the low-dimensional manifold of *E. coli*'s phenotypes. Our analyses also identify the genotype-phenotype map, revealing the mutations that lead to the different strategies of stress resistance.

We further discuss how our analyses could decipher the evolutionary constraints of *E. coli*.

We next develop a novel method for predicting and controlling stress resistance evolution by inferring an empirical fitness landscape based on phenotypes of *E. coli*. The concept of the fitness landscape has been influential in many areas of research on evolution since they provide information on the predictability of evolution. However, the high-dimensionality of the genotypic space kept us from constructing an empirical genotype-fitness landscape capable of predicting evolution. Focusing on the fact that evolution leads to low-dimensional phenotypes rather than genotypes, we infer the phenotype-fitness landscape based on the stress resistance profiles. To do so, unlike typical laboratory evolution experiments, we monitor the resistance levels to multiple stresses during the course of evolution which allows the dense sampling of phenotypes and the corresponding fitnesses along evolutionary trajectories. We show that the structure of the inferred landscapes corresponds with the resistance acquiring mechanisms of *E. coli* and provide information of the directionality of evolution. We discuss how the inferred landscapes could be utilized for predicting and controlling evolution.