Accelerating Evolutionary Hill Climbs in Parallel Turbidostats

Christopher N. Takahashi,^{*,†} Luis Zaman,[‡] and Eric Klavins[¶]

[†]*School of Computer Science and Engineering, University of Washington, Seattle*

[‡]*Center for the Study of Complex Systems, University of Michigan, Ann Arbor*

[¶]*Department of Electrical Engineering, University of Washington, Seattle*

E-mail: cnt_at_uw_dot_edu

Abstract

Evolution has been used to address many engineering problems. Within the context of metabolic engineering and synthetic biology, directed evolution has natural applications. However, most research concerning optimizing microbial evolution has been focused on library generation and screening, while accelerating evolutionary hill climbs and been largely ignored. Here, we develop a model to explore how population structure can accelerate evolutionary hill climbs. We show that by adjusting the population size, environmental challenge, and meta-population dynamics that the rate of evolution can be accelerated in parallel turbidostats. Our analyses leads to two surprising results: small populations are favored over conventionally large microbial populations, and propagating modest fitness improvements is favored over propagating mutants with large beneficial mutations. When combined with rational design and other optimization techniques our theory can accelerate strain development for applications such as consolidated bioprocessing, and bioremediation systems.

Keywords

Continuous culture, Turbidostat, Optimal evolution, Evolution model

Introduction

Computational and mathematical models of evolution have helped uncovered fundamental patterns and processes occurring in nature (1–5). As our theoretical understanding of evolutionary dynamics improves, our ability to harness and direct microbial evolution towards desirable engineering goals increases. One such goal is to improve the production of economically relevant biosynthetic byproducts such as the antimalarial drug precursor, artemisinic acid(6). Realizing that cheating is a common evolutionary solution when selecting cells via indirect markers (e.g., producing the marker without paying the cost of making the desired byproduct), (7) devised a method of alternating selection and counter-selection that effectively curtails *de novo* cheats and ultimately improved byproduct yield. In other words, understanding likely evolutionary outcomes allowed the authors to short circuit evolution with carefully engineered environments.

Here we investigate how multiplexed turbidostats (*i.e.*, bioreactors that allow controlled microbial population sizes) we have previously developed (8) can be used to speed up adaptation during directed evolutionary *hill climbs*. We develop a mathematical model of evolutionary dynamics within a turbidostat. We then consider a multiplexed metapopulation of turbidostats and ask whether decisions about how populations are propagated (e.g., isolated populations vs propagating the most fit population to the rest of the turbidostats at various intervals) can optimize the rate of adaptation. Our analyses led to two surprising results: small populations are favored over conventionally large microbial populations, and propagating modest fitness improvements is favored over propagating mutants with large beneficial mutations.

We combine steady state continuous culture dynamics (*e.g.*, in turbidostats and chemostats)

with mutational events into a simple stochastic hybrid systems (SHS) (9) model. By considering a continuous culture environment much of the complexity associated with natural microbial populations much of the modeling complexity is removed. As a result, we are able to analytically compute the time to appearance of a more fit mutant and the time it takes that mutant to achieve majority in the population. Thus, we can query the model for optimal initial environments and population sizes for single and double step evolutionary hill climbs.

Our analysis demonstrates that uniformly sized steps are preferable to a large step followed by a small step due to the disproportionate amount of time required for small fitness improvements to sweep. This leads us to a counterintuitive control policy where large fitness jumps are rejected in favor of smaller but more uniform fitness increases. By applying this policy to a simulated metapopulation of bioreactors, we show that we can increase the rate of evolution. Finally, we examine the effects of population size and number of bioreactors in the metapopulation on the effectiveness of our control policy.

Results

Continuous Culture Dynamics

Here, we consider the case of continuous culture with microbes growing in a constant volume of well mixed liquid media. We begin with a standard ODE model for mixed populations in continuous culture (10)

$$\begin{aligned}\dot{x}_i &= \varphi_i(s)x_i - ux_i \\ \dot{s} &= -\sum_i \frac{1}{\gamma} \varphi_i(s)x_i - us + uC.\end{aligned}\tag{1}$$

Here, x_i denotes the total amount of cells of genotype i . Individual cells divide at a rate $\varphi_i(s)$ for a limiting substrate concentration s . Cells and substrate are washed out with fresh

media containing substrate at concentration C at a rate u . Finally, substrate is consumed at a rate of γ cells per unit of substrate.

Turbidostat Evolutionary Dynamics

A turbidostat is a continuous culture device where the total cell population is held constant using feedback control(11). The resulting fixed population size allows us to make simplifying assumptions and enables us to apply previous work on fixed size populations by Moran (12). We begin by assuming the dilution rate $u(t)$ is set by a feedback controller that ensures that the cell density is constant, which simplifies the resulting dynamics. While a complete model would include turbidostat feedback dynamics, such a system would only differ in the moments soon after inoculation as any reasonable controller will act on timescales much faster than the growth of microbes. We then extend the ODE model to a stochastic hybrid systems model to capture the random nature of mutations.

Mathematically, a fixed population size implies that

$$\sum_i x_i(t) \approx \bar{x}, \forall t > t_{ss} \quad (2)$$

for some desired population size $\bar{x}$. Differentiating equation (2) we get $\sum_i \dot{x}_i(t) = 0$ (note that this differs from the steady state of the system where $\dot{x}_i = 0$ for all i) implying that

$$u(t) = \bar{\varphi}(t) := \frac{\sum_i \varphi_i x_i(t)}{\bar{x}}, \forall t > t_{ss}, \quad (3)$$

which is the weighted average growth rate.

Substituting (3) into (1) yields

$$\begin{aligned} \dot{x}_i &= x_i(\varphi_i(s) - \bar{\varphi}(s)) \\ \dot{s} &= (-\bar{x}\frac{1}{\gamma} + C - s)\bar{\varphi}(s). \end{aligned} \quad (4)$$

Furthermore, since cell population and substrate are a conserved quantity, we can restrict analysis to the attractive manifold $s = -\bar{x}\frac{1}{\gamma} + C$, further simplifying our continuous model to

$$\dot{x}_i = x_i(\varphi(\bar{x}, C)_i - \bar{\varphi}(\bar{x}, C)). \quad (5)$$

When considering basic bacterial growth in large (greater than 1000 individuals) populations of cells, mass action models such as (4) are accurate. However, when considering rare events such as mutation mass action models become inaccurate and stochastic models are preferable. To model growth dynamics and mutational dynamics together we use a stochastic hybrid systems(SHS) model (9). In this model, equation (4) comprises the continuous state of a SHS that has reset maps

$$(x_i, x_j) \mapsto (x_i - 1, x_j + 1) \quad (6)$$

with transition intensities $\lambda_{ij}(x_i)$. In other words, at rate $\lambda_{ij}(x_i)$ a member of genotype i mutates to genotype j , removing one cell of type i and adding one cell of type j . In order to model $\lambda_{ij}(x_i)$, we make the following assumptions:

1. Mutations are sufficiently rare that only individual mutants arise and either fix or are displaced (i.e., there is no clonal interference).
2. The population is large ($\bar{x} > 1000$). This is true for most microbial populations of interest, excluding some microfluidic devices.
3. The population size is constant.
4. Mutations happen at a constant rate per reproduction event.
5. Cell and nutrient mass are at steady state.

Assumptions 1 and 2 imply the standard strong selection weak mutation (SSWM) assumption used in literature (13). While assumptions 1 and 2 are stricter than the SSWM

assumption, assumption 2 is motivated by the experimental setup we are considering and is therefore more reasonable than simply taking SSWM outright. Assumptions 2 and 3 imply that substitution happens via a Moran process and therefore the probability that a new mutant i fixes is a function of its fitness relative to the population average fitness $r_i := \varphi_i/\bar{\varphi}$,

$$\rho(r_i) = \frac{1 - 1/r_i}{1 - 1/r_i^x} \approx \begin{cases} 1 - 1/r_i, & \text{if } r_i > 1 \\ 0, & \text{otherwise} \end{cases}. \quad (7)$$

For more detail see Nowak Chapter 6 (14) and references within.

To compute the rate that new mutants appear, we suppose that the current population is homogeneous with fitness φ_1 . The chance a new mutant that will eventually fix in the population arises in the next Δt seconds is proportional to the number of children born, $\varphi_1 x_1 \Delta t$, times the chance it has of fixing, $\rho(\varphi_2/\bar{\varphi})$. i.e.,

$$P(T_m < t + \Delta t | T_m > t, \varphi_2) \approx K_{1,2} \varphi_1 x_1 \rho(\varphi_2/\bar{\varphi}) \Delta t, \quad (8)$$

where T_m is the mutant's time of arrival and $K_{1,2}$ is some independent constant to represent proportionality as an equality. It follows from the definition of transition intensity (9) that

$$\begin{aligned} \lambda_{i,j}(x_i) &= \lim_{\Delta \rightarrow 0} P(T_m < t + \Delta t | T_m > t, \varphi_j) \\ &= K_{i,j} \varphi_i x_i \rho(\varphi_j/\bar{\varphi}). \end{aligned} \quad (9)$$

Evolutionary Rate in a Turbidostat

With the continuous dynamics (4) and the transition intensities (9) we can now compute the time it takes a small sub-population with greater than average fitness to sweep the population.

Here, we define the *substitution time*, T_s , as the time it takes a cell of genotype i to reach

a population fraction of 50% starting from a single cell. This definition differs from what is typically used when considering a birth death processes. Since the dynamics are continuous, it is impossible to reach a population fraction of 1 in finite time when starting from an initial fraction of less than 1. Instead we choose a threshold where the novel cell genotype could be easily detected.

To compute the substitution time, we begin with our simplified continuous model in equation (5), where s is treated as constant and cell densities are at steady state. We also note that x may be normalized by $\bar{x}$, in which case units are population fraction. For the remainder of this paper we will use population fraction when referring to x .

Equation (5) has solution

$$x_i(t) = \frac{\exp(\varphi_i t)}{\sum_j x_j(0) \exp(\varphi_j t)} x_i(0), \quad (10)$$

which for the case of genotype 2 superseding genotype 1, may be used to solve $x_2(T_s(\varphi_1, \varphi_2, \bar{x})) = 0.5$ for $T_s(\varphi_1, \varphi_2, \bar{x})$ yielding

$$T_s(\varphi_1, \varphi_2, \bar{x}) = \frac{\log(\bar{x} - 1)}{\varphi_2 - \varphi_1} \quad (11)$$

for $x_1(0) = 1 - 1/\bar{x}$, $x_2(0) = 1/\bar{x}$, and $\varphi_2 > \varphi_1$.

Figure 1: Rate of substitution (per hour) for a homogeneous population of genotype 1 to genotype 2 assuming $k = 10^{-7}$ and $\varphi_2 = 0.6$. For small $\bar{x}$ the penalty grows close to $1/\bar{x}$ while for large $\bar{x}$ the penalty grows with $\log(\bar{x})$.

Next, we compute the time to appearance of the next mutant T_m , which is captured by a Poisson process(9). For a homogeneous population, as we have assumed, we can compute the expected time to mutation by

$$E\{T_m|x_i = \bar{x}\} = \frac{1}{\sum_j \lambda_{i,j}}. \quad (12)$$

A turbidostat allows us to pick the population size $\bar{x}$, while a chemostat allows us to pick the average growth rate $\bar{\varphi}$. A turbidostat with mixture control enables us to control both $\bar{x}$ and $\bar{\varphi}$ by choosing a fixed dilution rate and then adjusting the ratio of media sources until nutrient limitation is reached. This allows the population size $\bar{x}$ and the average growth rate $\bar{\varphi}$ to be free variables. The time to appearance and substitution of a new mutant with fitness φ_2 given a homogeneous population with fitness φ_1 is

$$\begin{aligned} T(\bar{x}, \varphi; 1, 2) &= \mathbb{E}\{T_m|x_1 = \bar{x}\} + T_s(\bar{x}, \varphi_1) \\ &= \frac{1}{K\varphi_1\bar{x}\rho(\varphi_2/\varphi_1)} + \frac{\log(\bar{x} - 1)}{\varphi_2 - \varphi_1}. \end{aligned} \quad (13)$$

The time to appearance and substitution given in equation (13) has no closed form minimum over $(\bar{x}, \varphi_1)$ but may be optimized numerically (Fig. 1).

Circumventing Clonal Interference

We can now use our evolutionary model to explore a more complex landscape involving two alleles and negative epistasis (Fig 2). Applying our previous result to two substitution events with the optimal population size we get that the time to substitution of AB along the path $ab \rightarrow Ab \rightarrow AB$ is $\min_{\bar{x}}\{T(\bar{x}, \varphi; ab, Ab) + T(\bar{x}, \varphi; Ab, AB)\} \approx 177$ h while the substitution time along the path $ab \rightarrow aB \rightarrow AB$ is approximately 907 h. Clearly the former path is preferable from an engineering standpoint, however due to clonal interference, if phenotype aB arises at any point before phenotype AB, final fixation of AB will progress more slowly

Figure 2: Example fitness landscape suffering from negative epistasis. In this example, two loci are considered each with two alleles, yielding four genotypes, ab , Ab , aB , and AB . The length of each “lolipop” denotes growth rate per hour. Edge weights and direction denote the expected substitution time for an optimal population size, given a mutation rate of 10^{-7} per cell division.

due to type aB ’s superior growth rate with respect to type Ab and AB ’s small improvement over aB (15). Indeed, regardless of whether AB arises from Ab or aB , the population dynamics (10) imply that the progress of type AB ’s sweep relative to aB happens at a rate $\varphi_{AB} - (\varphi_{aB}x_{aB} - \varphi_{AB}x_{AB}) / (x_{aB} + x_{AB})$, which is very small for small differences growth rate and will delay the fixation of AB when the fraction x_{aB}/x_{AB} is significant.

To explore conditions under which evolution is optimized in this environment we compute the probability that an evolutionary trajectory will progress along $ab \rightarrow Ab \rightarrow AB$, which can be approximated by

$$\begin{aligned} \mathbb{P}[Ab] &:= \mathbb{P}[T_m(ab, Ab) + T_s(\varphi_{ab}, \varphi_{aB}, \bar{x}) < T_m(ab, aB)] = \\ &= \frac{\varphi_{aB}(\varphi_{ab} - \varphi_{Ab})(\bar{x} - 1)^{\frac{k\bar{x}\varphi_{aB}(\varphi_{aB} - \varphi_{ab})}{\varphi_{aB}(\varphi_{ab} - \varphi_{Ab})}}}{\varphi_{ab}(\varphi_{aB} + \varphi_{Ab}) - 2\varphi_{aB}\varphi_{Ab}}. \end{aligned} \quad (14)$$

Equation (14) allows us to optimize the rate of two step evolution for known fitness land-

scapes. We get that the expected time to substitution of AB starting from ab is

$$T_{AB} := (T(\bar{x}_0, \varphi; ab, Ab) + T(\bar{x}_A, \varphi; Ab, AB)) \mathbb{P}[Ab] + \\ + (T(\bar{x}_0, \varphi; ab, aB) + T(\bar{x}_B, \varphi; aB, AB)) (1 - \mathbb{P}[Ab]). \quad (15)$$

T_{AB} can be minimized over $\bar{x}_0$, $\bar{x}_A$, and $\bar{x}_B$ to find an optimal population size for each point in our hill climb. Applying this strategy to our example gives 837 h at an initial population size of approximately $\bar{x} = 7 \times 10^5$.

Rapid Evolution in an N-plex turbidostat

In the previous section, we showed that for the case of epistasis there is a fastest evolutionary path. For the case of negative epistasis, clonal interference ensures that the probability that the fastest path will be taken is never more than 50%. Recent developments have made small multiplex bioreactors possible (8, 16) enabling parallel evolutions. With parallel bioreactors an experimenter only needs a single population of many to complete a hill climb before the evolved strain can be collected. This multiplex experimental setup then allows us to model the evolution of our meta-population (the population of individual communities within each bioreactor) as a N individual Bernoulli trials. For the two locus example, the probability of success (taking the faster path) is given in equation (14) and the probability that at least one of N cultures evolves along the fastest path is

$$P[\text{number fastest path} > 0] = 1 - (1 - \mathbb{P}[Ab])^N. \quad (16)$$

We can now subject equation (16) to either specification (e.g. Find $\bar{x}$ so that there is a 95% chance of at least one population taking the fastest path) or optimization. Fig. 3 shows the expected time and optimal per bioreactor population size as a function of number of bioreactors. Adding multiple bioreactors significantly decreases the wait time for an evolved

strains with returns diminishing beyond 15 bioreactors.

Figure 3: The relationship between the number of bioreactors and optimal rate of evolution ($\tau = 0$). Plot corresponding to $\tau > 0$ are given a linear penalty for the number of bioreactors, representing the additional labor required to setup and maintain additional bioreactors in units of culture time. For example, if each additional bioreactor takes 1 hour of labor and 1 hour of labor is worth 10 hours of waiting then $\tau = 1 \times 10$ hours per bioreactor. The points in red indicate the optimal number of bioreactors for each value of τ .

A control strategy for parallel turbidostats

In prior sections, our optimization strategies have required exact knowledge of the fitness landscape. In practice this information is rarely known. Instead, estimates of optimal fitness may be available by applying knowledge of the metabolic load of expressing enzymes as well as the fitness benefit of downstream metabolites. With an estimate of the optimum fitness and insights gained from the previous sections, we can devise a heuristic control strategy to accelerate evolutionary hill climbs in parallel turbidostats.

We believe that the following heuristics apply in general and can be used to generate a control strategies that accelerate hill climbs of metapopulations.

1. Due to clonal interference, populations on the same order (or greater) as the mutation rate tend to take the largest fitness jumps available.
2. Larger jumps provide diminishing returns and should be avoided.

Figure 4: Performance of Algorithm 1 vs null algorithm for eight turbidostats. **a**, Simulated time to reach fitness peak over 500 trials. **b**, Mean time to reach fitness peak of Algorithm 1 and the null algorithm for varying population sizes. Algorithm 1 still performs well when the population is chosen smaller than is optimal and performs no worse than null for large populations.

Using the stated heuristics, we have devised the control strategy outlined in Algorithm 1. To evaluate our control strategy we compared its performance in simulation to that of the null strategy where N turbidostats are inoculated but left without further manipulations.

Algorithm 1

- 1: Inoculate N turbidostats
 - 2: $\varphi_m \leftarrow$ growth rate of inoculum.
 - 3: $\varphi_M \leftarrow$ hypothesized maximal growth rate.
 - 4: **while** $\varphi_m < \varphi_M$ **do**
 - 5: **for all** chambers c **do**
 - 6: $\varphi_c \leftarrow$ growth rate of chamber c
 - 7: $\Delta \leftarrow \varphi_M - \varphi_m$
 - 8: **if** $\varphi_M - 0.25\Delta > \varphi_c > \varphi_m + 0.25\Delta$ **then**
 - 9: Restart all chambers from chamber c
 - 10: $\varphi_m \leftarrow \varphi_c$
-

Our simulation implements the SHS with continuous state as in equation (5) and reset map given in equation (6) and terminates when a single chamber reaches 50% population fraction of the most fit type. We choose a randomly generated fitness landscape with three epistatically interacting loci (e.g. binding affinity, expression, and importation), each with a single alternate allele conferring a fitness benefit regardless of order of fixation (i.e. no sign epistasis). Fig 4 shows the mean time reach the fitness peak for Algorithm 1 vs the null algorithm. Comparing Algorithm 1 with the null algorithm we see an improvement of nearly

40%. In addition to the improved optimal time, Algorithm 1 is also much more robust to the choice of an overly small population and is no worse than the null algorithm for large populations. The robustness of Algorithm 1 is an important feature because the a precise value for the optimal population size is not known a priori.

Discussion

We introduce a simplified stochastic hybrid systems model that captures the general features of evolving populations in a turbidostat while remaining analytically tractable. With our model we show that the expected time to reach an evolutionary fitness peak can be computed analytically when the fitness landscape is known. From the analytical expression for expected time to hill climb we can optimize the population size to minimize the time to fixation of the most fit genotype. Our optimization then leads to a set of heuristics that can be used to devise control strategies to further optimize the time to evolve by directly manipulating meta-populations of turbidostats. We show that one such control method enables acceleration of evolution by nearly 40% while also making the evolutionary time more robust to choice of population.

Recent work has enabled the real world application of these control policies(8, 16, 17). Applications could include metabolic engineering problems where optimal growth rates can be extrapolated from known systems, and synthetic biology where it may be used to predict how “unhealthy” a cell may be made before it is likely to mutate in a given time span. Systems where engineered performance can be tied to metabolism may also benefit by showing experimenters how to optimize conditions, which in turn will increase research productivity.

In addition to direct applications, future work may focus on alternate algorithms for controlling meta-populations of turbidostats. Algorithms that allow a wider breath of evolutionary paths simultaneously may work better, especially in rugged landscapes. The application of stochastic remixing of populations may also provide more robust of faster control

strategies. We believe that further exploration will reveal a rich space of control methods that have implications for both engineering and experimental evolution.

References

1. Ofria, C., and Wilke, C. O. (2004) Avida: A Software Platform for Research in Computational Evolutionary Biology. *Artificial Life* 10, 191–229.
2. Ray, T. S. Synthetic life: Evolution and optimization of digital organisms. Scientific Excellence in Supercomputing: The 1990 IBM Contest Prize Papers. 1992; pp 489–531.
3. Kauffman, S., and Levin, S. (1987) Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology* 128, 11–45.
4. GERITZ, S., KISDI, E., MESZENA, G., and METZ, J. (1997) *Evolutionary Ecology* 12, 35–57.
5. Otto, S. P., and Day, T. *A biologist’s guide to mathematical modeling in ecology and evolution*; Princeton University Press, 2007; Vol. 13.
6. Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943.
7. Raman, S., Rogers, J. K., Taylor, N. D., and Church, G. M. (2014) Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences* 111, 17803–17808.
8. Takahashi, C. N., Miller, A. W., Ekness, F., Dunham, M. J., and Klavins, E. (2015) A Low Cost, Customizable Turbidostat for Use in Synthetic Circuit Characterization. *ACS Synthetic Biology* 4, 32–38.
9. Hespanha, J. P. (2006) Modelling and analysis of stochastic hybrid systems. *IEEE Proceedings – Control Theory & Applications* 153, 520 – 535.

10. Smith, H. L., and Waltman, P. *The Theory of the Chemostat*; Cambridge University Press, 1995.
11. BRYSON, V., and SZYBALSKI, W. (1952) Microbial selection. *Science* 116, 45–51.
12. Moran, P. A. P. *The statistical processes of evolutionary theory*; Clarendon Press, 1962.
13. Gillespie, J. H. (1983) Some Properties of Finite Populations Experiencing Strong Selection and Weak Mutation. *The American Naturalist* 121, pp. 691–708.
14. Nowak, M. A. *Evolutionary Dynamics: Exploring the Equations of Life*; Harvard University Press, 2006.
15. Gerrish, P., and Lenski, R. (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* 102-103, 127–144.
16. Miller, A. W., Befort, C., Kerr, E. O., and Dunham, M. J. (2013) Design and Use of Multiplexed Chemostat Arrays. *Journal of Visualized Experiments*
17. Toprak, E., Veres, A., Michel, J. B., Chait, R., Hartl, D. L., and Kishony, R. (2012) Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* 44, 101–105.