Indian Institute of Technology, Bombay

CL717: Evolutionary Dynamics
Mid-semester Project Report

Name: Abhishek Behera Roll No.: 174076006

September 16, 2018

1 Introduction

In this report, I discuss the questions that I find to be interesting and would eventually want to work towards answering, for the final course project. The areas of interest are: Growth rates in Auxotrophs, NK Model for Fitness Landscape and Cascades of Chemical Reaction Networks.

2 Growth Rates of Auxotrophs

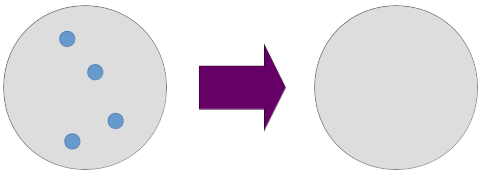
2.1 Preliminaries

Definition 1. An *auxotroph* is a mutant organism (especially a bacterium or fungus) that requires a particular additional nutrient which the normal strain does not. A *prototroph* is an organism that can utilize simple carbon and nitrogen substrates to produce all the essential metabolites.

Notation 1. If the auxotroph requires the additional nutrient X (which could be an amino acid such as arginine or histidine), then we say it is an *auxotroph for* X.

What makes auxotrophs interesting, is their puzzling growth rates, that they exhibit when they are allowed to *cross-feed*. Before we get on to the main experiment, due to *S. Pande et. al.* [5], we mention the following preliminary experiments first.

Experiment 1. If an auxotroph for X is grown in an enviroment deficient in X then the auxotroph would perish (Figure-1).



Auxotroph for X in environment deficient in X

Figure 1: Experiment 1

Experiment 2. However if an auxotroph for X is grown in an environment rich in X then the auxotroph outgrows the prototroph in the same environment (Figure-2a and Figure-2b).

(a) Auxotroph for X in environment rich in X

(b) Prototroph in environment rich in X

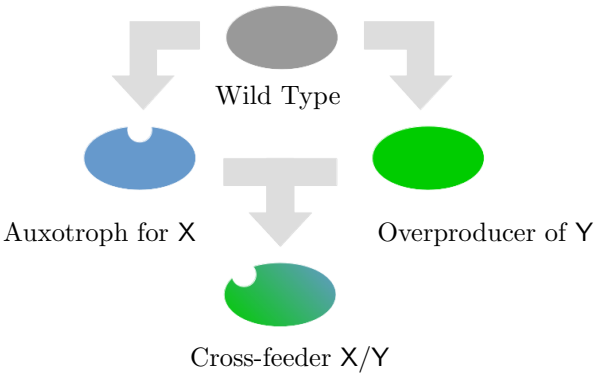
Figure 2: Experiment 2

Remark 1. One may imagine that an explanation for the results of Experiment-2 is that: the auxotroph for X in an envionment rich in X saves energy required for synthesizing X and thus has a growth advantage over the prototroph.

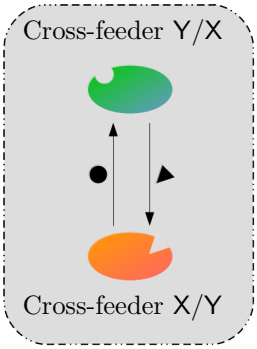
2.2 Main Experiment

Experiment 3. Here we presents the results of the experiments due to *S. Pande et. al.* [5]. In the experiment, the authors, synthetically modify *Escherichia coli* by introducing paired mutations. The two mutations are chosen in such a manner that, if one of the mutations makes the new genotype an auxotroph for X then the other mutation makes the new genotype an overproducer for Y (figure-3a).

Notation 2. A genotype which is an auxotroph for X and is an overproducer for Y will be denoted as *cross-feeder* X/Y.



(a) Synthetically engineering cross-feeders



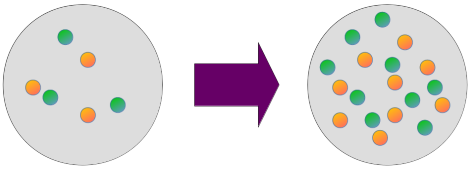
(b) Interaction of complementary cross-feeder pairs

Experiment 3 (continues)**.** Similarly a complementary pair of the mutation is used to create the cross-feeder Y/X (i.e., a genotype which is an auxotroph for Y but is also an overproducer for X). If complementary cross-feeders are grown in the same media then they can cooperate and exchange metabolites (figure-3b).

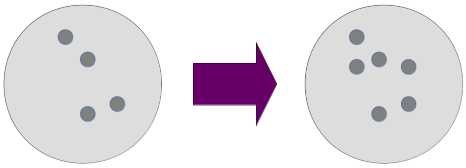
The most interesting aspect of their experiment was that **complementary cross-feeders outgrew the prototrophs** in the media deficient in nutrients needed by the crossfeeders. This experiment was done under two settings:

1. The complementary cross-feeders and prototrophs were grown separately and their growth rates were compared.
2. The complementary cross-feeders were put in direct competition against the prototrophs by growing them together and their growth rates were compared.

Although in the direct competition, certain pairs of cross-feeders lost some of their competitive advantages, but overall, the cross-feeders performed better (figure-4a and figure-4b).



(a) Crossfeeder X/Y and Crossfeeder Y/X in environment deficient in X and Y



(b) Prototroph in environment deficient in X and Y

Figure 4: Experiment 3

2.3 Attempts at Explanation

Economic Analogies

One may posit that cells operate like factories with a cascade of parallel and sequential reactions. The central carbon metabolic cycle is such an example.

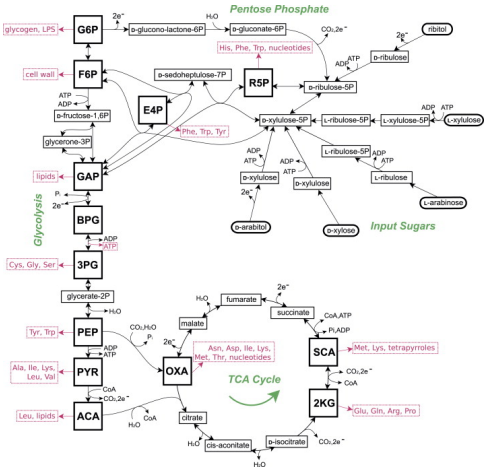


Figure 5: Central carbon metabolic cycle (Courtesy:[3])

So one may speculate that, as in a factory, perhaps there is cost-benifit advantage in mass-production. However, if one supposes a linear model to the cost per unit of production, then the cross-feeders could have at best matched the growth rate of prototrophs and not outgrown them! So the model for cost benifit has to be non-linear. But such explanations can only serve as an inspiration and can not be the full explanation, as they have no connection with biological reality.

Leaky Functions

Benjamin Kerr et. al. in [2], develop a framework to analyze emergence of depedency across organisms. In their model:

- They consider a community with S strains and n functions.
- A strain $i \in \{1, \ldots, S\}$ may or may not perform a function $j \in \{1, \ldots, n\}$ and this choice is denoted by the variable $a_{ij} \in \{0, 1\}$, taking value 1 if the strain performs the function and taking value 0 otherwise. The strain that performs in a function is called a *producer* strain.

- Then in their model, they account for the cost of performing a function by hindering the growth rate. The growth of strain i is given by

$$g_i = g_{\max} - c\left(\frac{\sum_{j=1}^{n} a_{ij}}{n}\right)^{\theta}$$

where $g_{\max}$ is the growth rate of the genotype that doesn't perform any function and c is the cost incurred by a genotype that performs all the functions. They make the assumption that the cost function obeys a power law and the exponent of the power law is given by θ.

- And in their model, they account for the benefit of performing a function by decreasing the death rate. The death rate for strain i is given by

$$d_i = d_{\min} + (d_{\max} - d_{\min}) \exp\left[-\prod_{j=1}^{n}\left(a_{ij}z + \gamma(1-z)\sum_{k=1}^{S} a_{kj}N_k\right)\right]$$

where z is the fraction of production that is for private consumption and $1-z$ is the fraction of production that is given away for public welfare and N_k is the density of strain k.

- And finally assuming a logistic model for growth, the population dynamics of strain i is given by

$$\frac{dN_i}{dt} = \left[g_i\left(1 - \frac{\sum_{k=1}^{S} N_k}{\mathcal{C}} - d_i\right)\right] N_i$$

where $\mathcal{C}$ is the carrying capacity of the environment.

Using this model they are able to identify tipping points between mutual dependencies and one-way dependencies and other qualitative behaviours. The soundness of this of the model follows from the rich class of qualitative behaviour it is able to capture. The other arguement for soundness is that it under a limiting case reduces to the *Lotka-Volterra* model. My criticism of the model is that, it seems ad-hoc and is not throughly justified. For example, I am not convinced why the exponent in the death rate must be a product and not a sum. Another problem with this model, is that is too abstract, to explain something as specific as cross-feeding in auxotrophs. For example, how would one know what θ or z should one use.

Greedy Scheduling

Since here are we are trying to explain the growth rates in microbes, I came across the paper by *Remy Pugatch* [6]

...to be completed...

2.4 Course Project Proposal

1. For the first phase of the project I would like to write code to simulate the cell division time given the reaction graph. The code verification is to be done by checking if the cell division time is indeed distributed according to Log-Frechet distribution or not.

2. In the second phase of the project for the auxotrops, we would use a reaction graph that will strictly be a sub-graph of the reaction graph for the prototrophes. I am hoping, that the Log-Frechet distribution for the auxotrophs' cell division time will have lower mean than the corresponding mean for the prototrophes. That would be a step towards the explanations.

3. The final step would be understanding how the cascades of the reaction in a micro-organisms can be realistically encoded as reaction graphs. I am hoping doing this would help us predict the outcome of competitions between different cross-feeding pairs.

3 NK Model for Fitness Landscapes

Here we do a review of the **NK Model** due to *Stuart A. Kauffman* which are very popular in modelling fitness landscapes.

...to be completed...

3.1 Course Project Proposal

I am interested in the question, how would one fit a NK model to some qualitative data that one may have about the fitness landscape. I am particularly interested in how does one go from the genetypic landscape to the phenotypic landscape. I imagine that informatino about phenotypic landscapes would be more prevalent. So one of the questions I have is, if one believes that the genotypic landscape is given by a NK model, and has the phenotypic landscape, then what value of K should one choose. I found a github library available as mentioned in this paper due to *Obolski et. al.* [4]. ...to be completed...

4 Cascades of Chemical Reaction Networks

Here we do a review of the paper *Cascades of transition in molecular information theory* by *Madan Rao et. al.* [1].

...to be completed...

4.1 Course Project Proposal

...to be completed...

References

[1] Suman G. Das, Madan Rao, and Garud Iyengar. Cascade of transitions in molecular information theory. *arXiv:1804.03827*, 2018.

[2] Sylvie Estrela, J. Jeffrey Morris, and Benjamin Kerr. Private benefits and metabolic conflicts shape the emergence of microbial interdependencies. *Environmental Microbiology*, 18(5):1415–1427, oct 2015.

[3] Elad Noor, Eran Eden, Ron Milo, and Uri Alon. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Molecular Cell*, 39(5):809–820, sep 2010.

[4] Uri Obolski, Yoav Ram, and Lilach Hadany. Key issues review: evolution on rugged adaptive landscapes. *Reports on Progress in Physics*, 81(1):012602, dec 2017.

[5] Samay Pande, Holger Merker, Katrin Bohl, Michael Reichelt, Stefan Schuster, Luís F de Figueiredo, Christoph Kaleta, and Christian Kost. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *The ISME Journal*, 8(5):953–962, nov 2013.

[6] Rami Pugatch. Greedy scheduling of cellular self-replication leads to optimal doubling times with a log-frechet distribution. *Proceedings of the National Academy of Sciences*, 112(8):2611–2616, feb 2015.