Modeling Symbiotic Cross-Feeding in Synthetic Consortia of Auxotrophic *E. coli*

Will Crockett

May 10, 2022

1 Introduction

With the advent of modern sequencing technologies, we have found an amazing amount of microbial diversity everywhere we have looked, from the digestive system of termites to Antarctic permafrost [1, 2]. Many of the microbial strains in these systems are auxotrophic, meaning that they are not capable of producing all of the metabolites needed for growth. Studies of both synthetic microbial consortia as well as communities present in natural environments like gut microbiomes have shown how emergent cooperative behaviors between auxotrophic microbes lead to strains promoting the growth of their cooperative partners in order to increase their own growth rate [3]. Studies have shown that these cross-feeding communities can have an evolutionary advantage over wild-type strains capable of producing their own metabolites due to a division of labor in metabolic production [4].

Microbes and microbial communities play important roles in human health, global climate, and are being studied for use in industrial applications. Being able to accurately model microbial interactions will allow us to better understand the human microbiome and the role of soil and aquatic microbes in the biosphere, and will allow us to engineer synthetic microbial consortia to produce a variety of useful materials and medicines [5, 6, 7].

Modeling microbial consortia has traditionally taken one of two approaches: modeling the mechanistic interactions of chemicals at a metabolic level or modeling the species interactions at an ecosystem level. Modeling the mechanistic interactions at the chemical level offers a better way of understanding how metabolites are shared through members of the system, while ecological modeling is better suited to model how population sizes change over time.

To better understand microbial ecology, it is useful to construct synthetic communities to have more control over the strains present and be able to start off with simple systems. Mee et al. did that in their paper studying interaction in synthetic systems of auxotrophic *Escherichia coli* strains [8]. They used genetic engineering to 'knockout' the gene used to produce one of 14 essential amino acids in each strain. By removing the ability for the bacteria to produce one of these amino acids they make them auxotrophic, meaning that they are not capable of synthesizing a compound needed for growth [9]. This forces the strain of *E. coli* to find the necessary amino acid in the environment, such as through cross-feeding with strains capable of producing the amino acid. Mee et al. started by creating synthetic communities of two strains of auxotrophic *E. coli*, each lacking the ability to produce a different amino acid. The community was grown in a glucose solution to provide energy, but the only sources of the missing amino acids were through cross-feeding between the strains. They measured the fold-growth (the final population size divided by the initial population size) of the strains to understand how they grew, and repeated the process for every possible combination of two strains to understand how each strain interacted with the others.

After measuring the fold-growths of the strains in 2-member communities, the group created three member communities, where each strain was missing two essential amino acids. Each of the three members shared one missing amino acid with the other two, so that it was necessary for all three to exchange nutrients in order for them to grow. The researchers found that although some of these interactions were comparable to the two-member populations, other groupings led to far more growth, and some led to lower levels of growth than expected.

Being able to accurately model microbial interactions is a challenge due to the complexity that arises from the multitude of interactions between organisms at the molecular level, and things like non-uniform spacial distributions and evolution make the problem even more difficult. However, developing

better models for these systems could lead to great advances in medicine (such as in auto-immune diseases related to the gut microbiome), in understanding the environment, and in developing more robust biomanufacturing systems.

2 Model Formulation

2.1 Two-Strain Cohort Model

I based my modeling approach on the approach taken by Mee et al., in order to recreate their model and compare it to other models, as well as make use of the data they released along with their paper [8]. To do so we take an ecological approach to the system, modeling the population size of the strains and the interactions between the strains, rather than modeling exchange at a metabolic level. The model system was incubated in a liquid media, so there is assumed to be no spacial non-uniformities. In addition, the interaction between strains is assumed to be constant over time.

To begin constructing the mathematical model of the system, we start with the Lotka-Volterra predator-prey model, with a few key differences. The Lotka-Volterra model is defined by the following:

$$\frac{dx_1}{dt} = a_1 x_1 + c_{12} x_1 x_2 \tag{2.1}$$

$$\frac{dx_2}{dt} = a_2 x_2 + c_{21} x_1 x_2 \tag{2.2}$$

Where x_1 is the population of prey and x_2 is the population of predator. a_1 and a_2 are the reproductive and death rate in the absence of one another, and c_{12} and c_{21} define the interactions between species (for the predator-prey model, it would be positive for the predator, and negative for the prey). In our case, both strains benefit from interaction, so both c_{12} and c_{21} are positive. In addition, due to them being auxotrophic, they cannot increase in the population without the presence of the other so both a_1 and a_2 are 0 (note: glucose is available, so they do not die, but they cannot reproduce because they cannot build new proteins without the missing amino acid), and their auxotrophy means that their growth is reliant on the size of the other population. This leaves us with the following simple model:

$$\frac{dx_1}{dt} = c_{12} x_2 \tag{2.3}$$

$$\frac{dx_2}{dt} = c_{21} x_1 \tag{2.4}$$

However, using this model results in exponential growth, as long as the initial populations are non-zero. The synthetic system is made in a liquid glucose media with a limited amount of nutrients, so to accurately model this we need to introduce a limit in total population. To do so, we need something with the behavior of the logistic growth model:

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}) \tag{2.5}$$

To fully model our system, we can combine these models to get:

$$\frac{dx_1}{dt} = c_{12} x_2 \left(1 - \frac{x_1 + x_2}{K}\right) \tag{2.6}$$

$$\frac{dx_2}{dt} = c_{21} x_1 \left(1 - \frac{x_1 + x_2}{K}\right) \tag{2.7}$$

This model takes into account the growth rate of each strain being limited by the population of the other times the interaction factor c_{ij}, and the size of the total population being limited by the carrying capacity K. A diagram of the model is shown in figure 1.

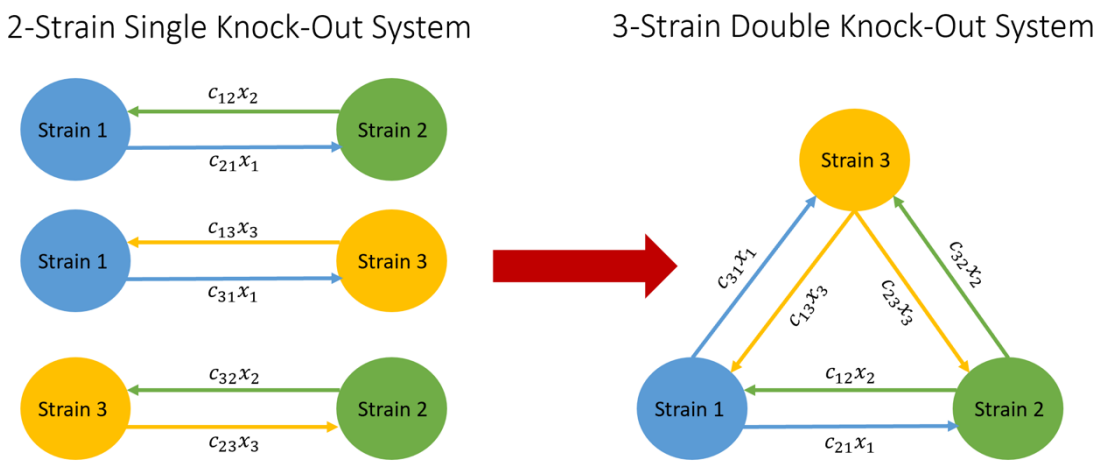


Figure 1: Diagrams of the models using 2 and 3 strains. The 3 strain system is composed of double knock-out auxotrophs, and was built using the interaction coefficients that were fitted to the 2-strain fold-growths.

2.2 Three-Strain Cohort Models

The interaction between a pair of microbes does not reflect the complexity in actual microbial ecosystems, where there are both more species and more interactions. Ultimately we would like to accurately model and predict the behavior of these more complex systems, and ideally we would like our models to provide insight into the processes that drive these complex ecological networks.

Here, I attempt to model a three microbe system based on the interactions measured for the two strain communities. For the two strain communities, each strain was unable to produce a single amino acid, while in the three strain system each strain is unable to produce two amino acids. Each member of the three strain community shares one of its missing amino acids with each of the others, so all three must cooperate in order for the population to grow (for example strain 1 could be missing amino acids I and K, strain 2 I and F, and strain 3 K and F).

To model this system, I created two models, one a re-creation from Mee et al., and one based on a generalized form of the Lotka-Volterra model. In both models I use the interaction coefficients found for the two strain model, and use them to model the 3 strain system as shown in figure 1 [8, 10].

2.2.1 Limiting Interaction Model

The auxotrophic nature of the *E. coli* strains mean they are reliant on the cooperation of the other members in order to grow. Due to the design of the system, the strains need cooperation from both of the others to grow, so if there is a very low level interaction between two of the strains the growth will be small, even if the other interactions are much higher.

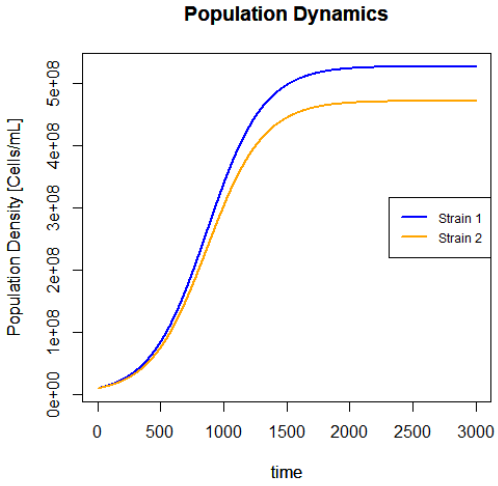
To model this, we can start with our two-strain model, and replace the interaction coefficient between the strains with the minimum interaction coefficient between the given strain and the two others:

$$\frac{dx_1}{dt} = min(c_{12}x_2, c_{13}x_3)\left(1 - \frac{x_1 + x_2 + x_3}{K}\right) \tag{2.8}$$

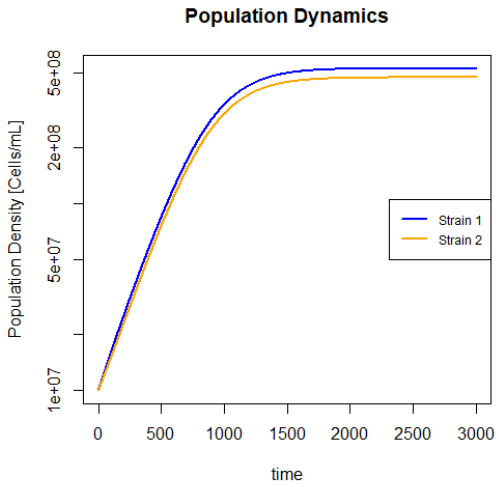
$$\frac{dx_2}{dt} = min(c_{21}x_1, c_{23}x_3)\left(1 - \frac{x_1 + x_2 + x_3}{K}\right) \tag{2.9}$$

$$\frac{dx_3}{dt} = min(c_{31}x_1, c_{32}x_2)\left(1 - \frac{x_1 + x_2 + x_3}{K}\right) \tag{2.10}$$

This model limits the growth of each strain by the lowest contribution from the other strains, which reflects the reliance that each strain has on receiving amino acids from its partners.



(a) An example of the growth curves of two strains using the two strain model. The curves plateau when the sum of the populations is equal to the carrying capacity. In this case $c_{12} > c_{21}$, causing strain 1 to reach a higher level than strain 2.



(b) Plot of the same simulation as figure 2a with a logarithmic y-axis to show growth from initial populations of 10^7 to a final total population of 10^9.

Figure 2: Figures showing growth under the two-strain model with non-zero interaction coefficients.

2.2.2 Generalized Lotka-Volterra Model

To see whether the limiting model was an accurate reflection of the three strain communities, I created a generalized Lotka-Volterra model of the system that used a linear sum of the interactions in place of the minimum used by the limiting model:

$$\frac{dx_1}{dt} = (c_{12}x_2 + c_{13}x_3)\left(1 - \frac{x_1 + x_2 + x_3}{K}\right) \tag{2.11}$$

$$\frac{dx_2}{dt} = (c_{21}x_1 + c_{23}x_3)\left(1 - \frac{x_1 + x_2 + x_3}{K}\right) \tag{2.12}$$

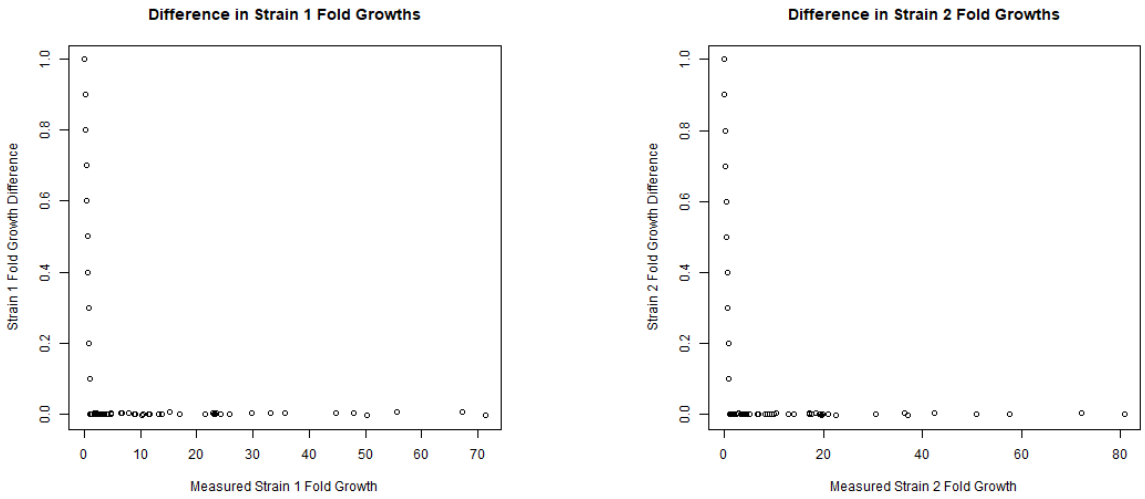
$$\frac{dx_3}{dt} = (c_{31}x_1 + c_{32}x_2)\left(1 - \frac{x_1 + x_2 + x_3}{K}\right) \tag{2.13}$$

Rather than defining the growth of each strain by the limiting interaction, it defines the growth as the sum of each of the strain's interactions. This model is commonly used when constructing ecological networks based on pair interactions, and serves as an alternative hypothesis. By comparing the two models to the measured fold growths, we can get a better understanding of the relationship between the two and three strain growths.

3 Results

3.1 Fitting the Two-Strain Model

In their supplemental data, Mee et al. provided the fold growths of each strain for every possible 2-strain combination. The fold-growth is defined as the final population size divided by the initial population size, and is used as a simple measurement of population growth. Following the values used by Mee et al., the initial population of each strain was set to $x_1(0) = x_2(0) = 10^7$, the carrying capacity was set to $K = 10^9$, and the final population was measured at $t = 5500$ [8]. An example of the growth curves over time is shown in figure 2a, with figure 2b showing the growths on a logarithmic scale. In this example, $c_{12} > c_{21}$, allowing for strain 1 to grow faster than strain 2, and make up a larger portion of the total population when they reach the carrying capacity.



(a) (b)

Figure 3: Plots showing the difference between simulated fold-growths for the best fit interaction coefficients and the measured fold-growths. Figure 3a shows the relationship for strain 1, and figure 3b shows it for strain 2.

Using these assumptions along with the fold-growth data, I fit the interaction coefficients c_{12} and c_{21} for each two-strain population using the Nelder-Mead optimization algorithm. To verify the fit I compared the simulated fold-growths to the measured fold-growths, a summary of which is shown in figures 3a and 3b. These graphs show that for the majority of the fits, the modeled fold-growth matches the measured fold-growth for both strains very well (the difference between them is 0). However, as the measured fold-growth goes to zero, there is an uptick in the difference due to the limit on the interaction coefficients to be equal to or greater to 0. The Nelder-Mead optimization algorithm calculates very low values for the interaction coefficients (e.g. on the scale of 10^{-12}), but due to the model's exponential growth there is still a noticeable growth in population over the long time-span the model is run for. Given that fold-growths can reach 100, a difference between simulated and measured fold-growths of 1 is still a relatively good fit.

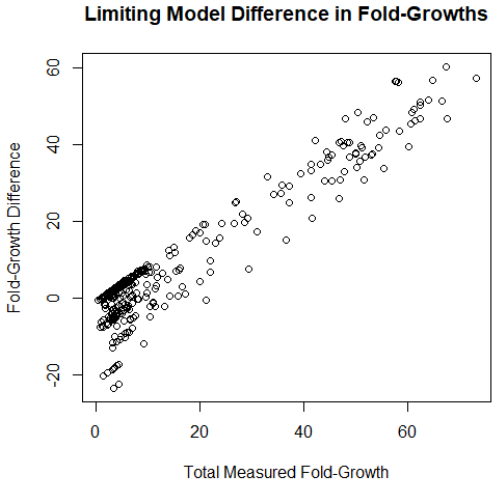
3.2 Modeling the Three-Strain System

The fits found for the two-member systems were then used to model the three-member systems using the limiting interaction model and the generalized Lotka-Volterra model, as shown in figure 1. Figures 4a and 4b compare the simulated total fold-growths to the measured fold-growths. The y-axes show the difference between the measured fold-growth and the modeled fold-growth for each possible 3-strain group, defined as:

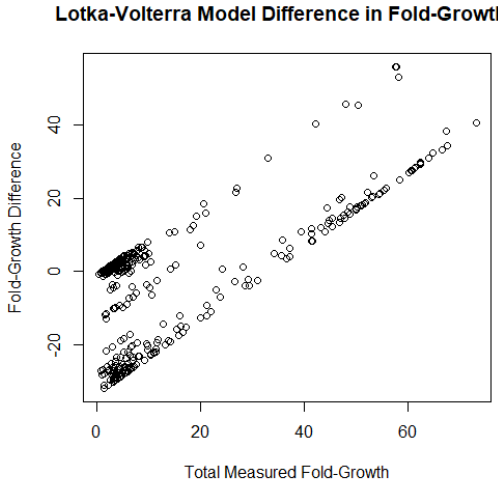
$$Y = FG_{meas} - FG_{model} = \frac{\sum_{i=1}^{3} x_{i,meas}(t_f)}{\sum_{i=1}^{3} x_{i,meas}(t_i)} - \frac{\sum_{i=1}^{3} x_{i,model}(t_f)}{\sum_{i=1}^{3} x_{i,model}(t_i)} \tag{3.1}$$

Meaning that where the difference is negative the model is predicting a total fold-growth larger than the measured fold-growth, and where the difference is positive the model is predicting a smaller fold-growth than the actual one. A perfect model for the three-strain groups would have a difference of zero.

In the limiting interaction model (figure 4a) there is a grouping around (0,0), showing that this model does a relatively good job modeling groups where there was little to no fold-growth, although there are still some where it predicts a larger fold-growth. As the measured fold-growth increases, the difference also increases, showing that the limiting interaction model under-predicts the size of the fold-growth for three-strain systems that were able to successfully grow.



(a)



(b)

Figure 4: Plots showing the difference between simulated fold-growths for the best fit interaction coefficients and the measured fold-growths. Figure 3a shows the relationship for strain 1, and figure 3b shows it for strain 2.

There is a similar increase in the generalized Lotka-Volterra model, with some slight differences. First, although there is a grouping at (0,0), there also appears to be a larger amount of groups that the model predicts a large fold-growth for where the measured fold-growth was close to 0. Although the Lotka-Volterra predicts higher fold-growths than the limiting model, the linear change between measured fold-growth and the fold-growth difference is still present.

The figures plot this difference versus the measured fold-growths, and show a somewhat linear pattern- suggesting that both models are not able to predict the higher fold-growth systems. This suggests that these systems arise from combinations of two-strain pairs with relatively lower fold-growths, and that adding a third member makes the total population grow faster than the sum of each 2-member pair.

4 Discussion and Conclusion

The failure of the 3-strain models to accurately predict the fold-growths shows the nonlinear nature of microbial interactions. The relationship between the two-strain interactions and three-strain interactions appear to be different from group to group, as seen in the width of the linear trend for both models. For example, the Lotka-Volterra model does a good job predicting some of the low fold-growth groups, but for others it predicts a much higher fold-growth than what truly occurred. This suggests that for some groups going from 2-member systems to a 3-member system reduced interactions and inhibitted growth. Both models failed to predict the high fold-growths that were attained by some 3-strain communities, suggesting that these groups were able to promote interactions and have higher levels of cooperation and growth.

To truly understand and predict these interactions, it may prove necessary to move to the metabolic level for modeling. Although it appears somewhat arbitrary that the sharing of some amino acids is more successful than others, this approach could be used to find which amino acids and other metabolites are more easily shared between strains. It is possible that there is molecular machinery not yet understood that promotes the cross-feeding of some metabolites and inhibits the cross-feeding of others. Studying which metabolites are cross-fed at a higher level in synthetic systems could help us understand common trends in natural microbial communities. Understanding the interactions between different metabolites and cross-feeding mechanisms could also help us better predict which groups would grow better or worse together, helping us both understand communities in nature and design new ones.

References

[1] A. Brune, "Symbiotic digestion of lignocellulose in termite guts," *Nature Reviews Microbiology*, vol. 12, no. 3, pp. 168–180, 2014.

[2] J. K. Jansson and N. Taş, "The microbial ecology of permafrost," *Nature Reviews Microbiology*, vol. 12, no. 6, pp. 414–425, 2014.

[3] E. H. Wintermute and P. A. Silver, "Emergent cooperation in microbial metabolism," *Molecular systems biology*, vol. 6, no. 1, p. 407, 2010.

[4] S. Pande, H. Merker, K. Bohl, M. Reichelt, S. Schuster, L. F. De Figueiredo, C. Kaleta, and C. Kost, "Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria," *The ISME journal*, vol. 8, no. 5, pp. 953–962, 2014.

[5] K. Z. Coyte, J. Schluter, and K. R. Foster, "The ecology of the microbiome: networks, competition, and stability," *Science*, vol. 350, no. 6261, pp. 663–666, 2015.

[6] H. W. Paerl and J. Pinckney, "A mini-review of microbial consortia: their roles in aquatic production and biogeochemical cycling," *Microbial Ecology*, vol. 31, no. 3, pp. 225–247, 1996.

[7] T. L. Weiss, E. J. Young, and D. C. Ducat, "A synthetic, light-driven consortium of cyanobacteria and heterotrophic bacteria enables stable polyhydroxybutyrate production," *Metabolic engineering*, vol. 44, pp. 236–245, 2017.

[8] M. T. Mee, J. J. Collins, G. M. Church, and H. H. Wang, "Syntrophic exchange in synthetic microbial communities," *Proceedings of the National Academy of Sciences*, vol. 111, no. 20, pp. E2149–E2156, 2014.

[9] R. LaRossa, "Nutritional mutations," 2013.

[10] M. R. Antoniewicz, "A guide to deciphering microbial interactions and metabolic fluxes in microbiome communities," *Current opinion in biotechnology*, vol. 64, pp. 230–237, 2020.

A Code and Data Availability:

The R scripts and data used for this project are available on Github.