How Quorum Sensing Interactions Affect Population Structures

02-712 Final Project

Siddharth Reed*[1], Evan Trop[1], Neel Mehtani[1], Deepika Yeramos[1], and Sarah Wenger[1]

[1]Computational Biology Department, Carnegie Mellon University

December 12, 2021

*slreed@andrew.cmu.edu

Background

Quorum sensing (QS) involves the ability of bacteria to change their population behavior. QS systems consist of a signal molecule that is constitutively expressed and a complementary receptor. When the population density of the bacteria reaches a certain threshold, the increased concentration signal bound receptor triggers expression of a public good such as a biofilm, surfactin, or enzyme to break down a complex nutrient.

These QS systems resemble some form of kin recognition as only bacteria that produce the unique signal - receptor pair take part in the production of the public good. However, bacteria that do not produce the signal - receptor pair, also called "cheaters," can take advantage of the public good without incurring the expense of signal or public good production. Cheaters help maintain diversity in QS signaling systems as diverged incompatible QS systems are still maintained at the population level through the indiscriminate public good (Pollak et al. 2016).

These QS interactions naturally give rise to a matrix where row R_i represents strain i's receptor and column S_i represents strain i's signal. Matrix values are one if receptor - signal binding is possible and zero otherwise. Our project aims to understand how QS interactions amongst different bacterial strains affect the population structure. Specifically, we are interested in answering how different biologically relevant matrix patterns and matrix sparsity affect population statistics such as total population growth rate, population diversity, time until fixation. Developing an understanding of the significance of these QS interactions might allow one to manipulate microbial environments. This could prove to be a useful tool in medical applications such as gastrointestinal diseases or in improving environmental states in wastewater treatment facilities.

Previous studies of QS systems have explored how and why evolutionary divergence of QS pathways occurs as well as how population diversity is maintained in the presence of obligate cheaters. These studies do not investigate how specific patterns and level of interactions between QS systems affect population structure.

Methods

We first note that all code, results and figures are available on GitHub here.

In order to experiment and observe how QS interactions affect population structure we first needed to first obtain a representative model of a multi strain QS system. The model outlined in Eldar (2011) provided us with a good basis to test our questions. This model is defined by a system of ordinary differential equations which explain how the level of cell densities, signal molecules, enzyme, and public good change with respect to time. Note that we have

have a separate equation for each strain and signaling molecule $i \in 1 \dots N$.

$$\frac{dn_i}{dt} = n_i(\frac{P_d}{P_d+1}(1 - rf(R^{active}i)) - n_{tot} - \gamma_n)$$
$$\frac{dS_i}{dt} = \beta_S(n_i - S_i)$$
$$\frac{dE}{dt} = -\beta_E E + \sum_i f(R^{active}i)n_i$$
$$\frac{dP_d}{dt} = J_{P_d} + V_{max}E - \beta_{P_d}(\frac{P_d}{P_d+1})n_{tot}$$

To investigate how different QS interactions affect the population levels of each strain we simulate the model with a constant set of parameters (see Table 1). To carry out simulations of a QS system given this model, parameter set, and a given K_ac matrix, the `solve_ivp` function from the scipy library is used to solve the above ODE system for a specified time range (200 time steps). The solver specifically implements the 4th order Runge Kutta approximation method as well as adaptive time steps to ensure stability and accuracy. Our final resulting simulator function takes as input a vector of initial cell density for x number of strains, stopping time, and the QS interaction matrix between the all strains.

Interaction matrices K_{ac} are square binary matrices with dimension equal to the number of strains. We preformed simulations with both pre-defined matrices (e.g. identity) or with randomly generated matrices (start with a 0 matrix and randomly set entries to 1 up to a specified threshold). To see specifically how this is implemented see `src/matrix.py` in the GitHub repository for details.

Parameter	Value	Description
r	0.5	growth cost of producing the public good
γ_n	0.01	spontaneous cell death rate
β_S	0.1	density dependant cell death rate
β_E	0.2	spontaneous enzyme degradation rate
J_{P_d}	0.20	spontaneous usable nutrient production rate
V_{max}	20	enzyme activity rate
β_{P_D}	100	usable nutrient consumption rate
m	1	exponent in activation function $f(x) = x^m$
K_{RS}	0.025	receptor-signal binding constant
K_{ac}	varies	receptor-signal activation matrix

Table 1: Parameters and values in the ODE model defined by Eldar, 2011.

Using this basic model we examined how using different K_{ac} matrices and using microbiome data to specify initial conditions would change population trajectories and the final population

Model Statistics

Here we list out the various model statistics calculated, the specific code is in `src/analysis.py`. Note that t_0 is the initial time step and t_n is the final time step.

Statistic	Description
total	The total abundance of all strains at t_n
growth_rate	The total groth rate of all strains
has_grown	The number of strains with a larger abundance at t_n than at t_0
euclidean	The euclidean distance between abundance vectors at t_n and t_0
shannon index	The absolutve value of the difference in Shannon Entropy between abundance vectors at t_n and t_0 (α-diversity metric)
bray_curtix	The Bray-Curtis distance between abundance vectors at t_n and t_0 (β-diversity metric)

Results

Comparing K_{ac} Matrices

Figure 1 illustrates the growth rates of different bacteria strains (N1 through N6) as determined by the eight different K_{ac} matrices. We see that in most cases, a larger initial abundance is associated with a larger growth rate. This is most apparent in the cycle and complete matrix patterns. It is also evident that regardless of initial abundance, different interaction matrices largely affect population structures.

It is interesting to note between the initial and final timesteps we can get different relative abundances and largely different trajectories depending on the matrix specified. We still see "cheating" behaviours specified by Eldar (2011) using the `ident` matrix as even though the strains start at quite different abundances the least abundant strains (6 times smaller) still end up at similar final abundances to the most abundant initial strain. This is because the most abundant strain reaches quorum quickly, produces the public good and allows all other strains to benefit and grow without cost i.e. cheat.

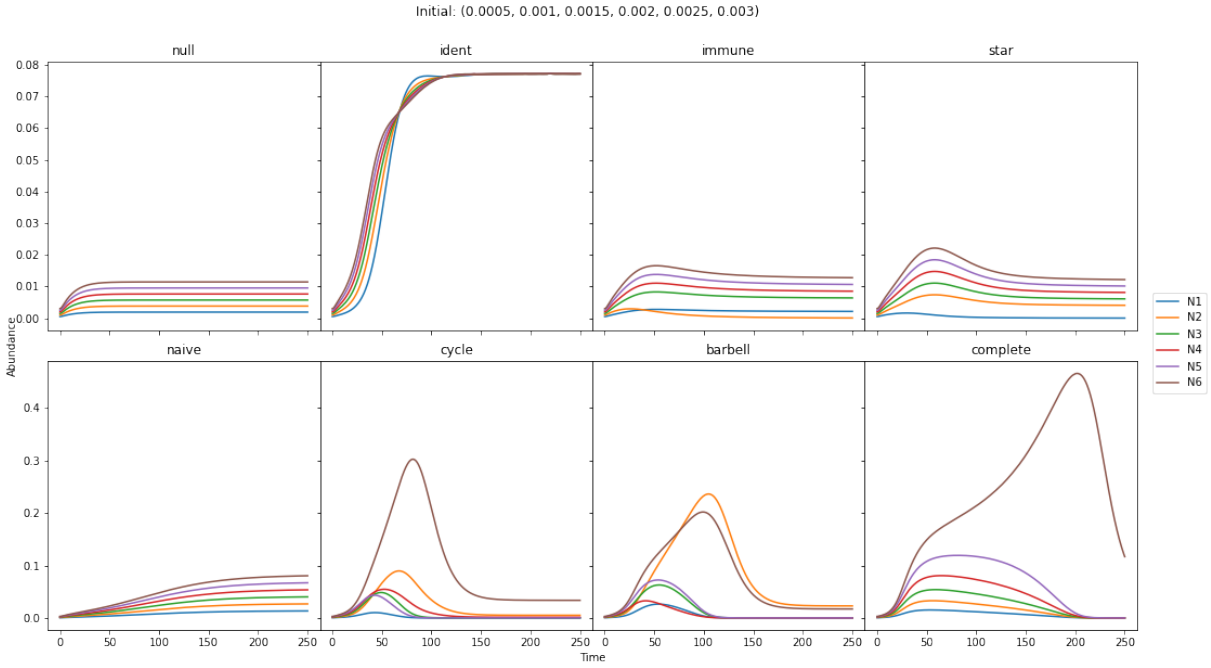

Figure 1: Population trajectories for 6 differnt strains using the ODE with different K_{ac} matrices specified. The initlal abundances for each strain are $(0.0006, 0.001, 0.0015, 0.002.0.0025, 0.003)$, chosen arbitrarily.

Examining Specific K_{ac} Matrices

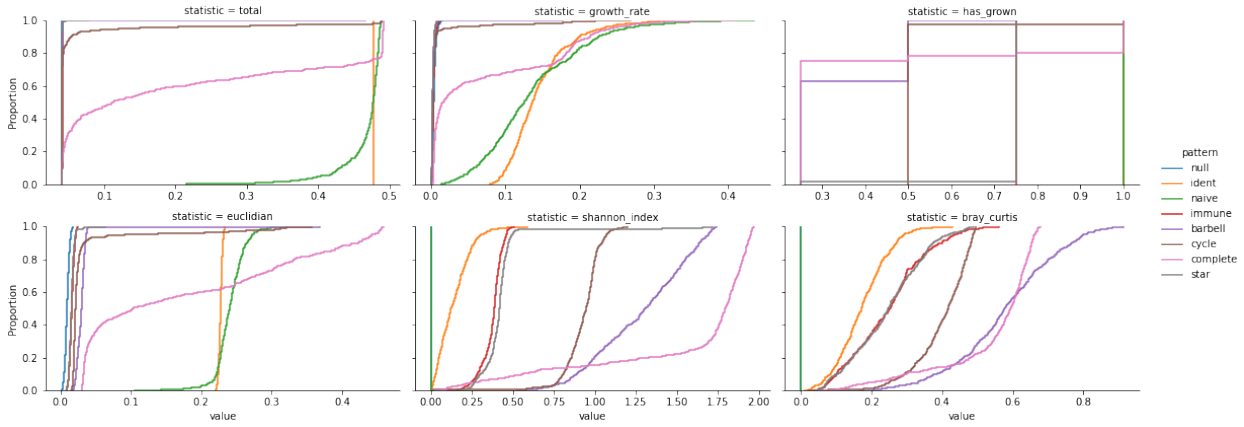

Figure 2: How different K_{ac} matrices can affect how a microbial population evolves. Each line represents the empirical cumulative distribution function of each statistic. Each color represents a different K_{ac} matrix. Each distribution is generated from 500 simulations with random initial conditions.

Figure shows what proportion of simulations had statistics below that value for the statistic. While there is significant diversity between matrices of the different statistics, it seems that across all statistics the complete matrix (all entries 1) appears the most different and has the largest variance (curves that are not steep). This is especially interesting as even though all strain contribute to the public good when any of them reach quorum we still get a wide range of values in both the total population size and the euclidean distance. In contrast for the identity matrix we see almost no diversity among the random initial conditions. So though every strain "cheats" of of all strains at quorum, this appears to lead to very little variance in both the growth rate of the entire population and the difference between the initial and final populations.

How K_{ac} Density Affects Populations

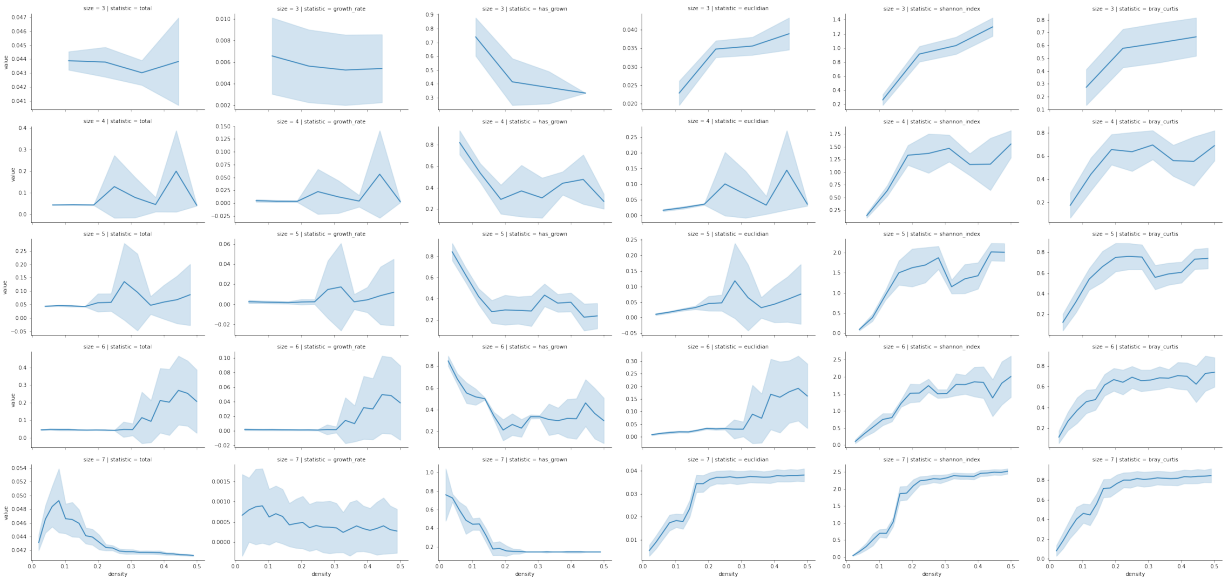

Figure 3: Evaluation of different model statistics (columns) with different numbers of strains N (rows). For each number of strains and statistics 250 simulations were run with random initial conditions. Shaded areas represent standard deviation of each statistic across the random initial conditions. Density specifically refers to $\frac{\text{sum}(K_{ac})}{|N|\times|N|}$, i.e. the fraction of all entries in K_{ac} that are non-zero.

For this we wanted to examine how the sparsity of the K_{ac} can affect the population. Specifically, if a system has more QS interactions (more entries equal to 1) does this affect population sizes or relative abundances over time. In order to do so we generate K_{ac} matrices randomly, starting with the null matrix (all 0s) and randomly set some entry $K_{ac}[i,j]$ to 1 and continue to set a new non-zero entry to 1 until some fraction f of all entries in the matrix are filled. So if we have a 4×4 matrix and $f = 0.5$ then we will generate 8 random matrices, with $1, 2, \ldots, 8$ entries randomly chosen to not be zero. Note that the plots are quite jagged in some cases due to the limited size of the matrix. For example, filling 3×3 up to $f = 0.5$, you can only fill 4 of 9 entries so we only get results for 4 specific sparsity values. This is less apparent with larger populations (i.e. larger matrices). Again the specific function for is `random_matrix_generator()` in `src/matrix.py`.

So for each K_{ac} 250 simulations with random initial conditions are run and model statistics are calculated as shown in Figure 3. It is interesting to note that as we increase the density (number of non-zero entries) we can observe specific trends, like the Bray-Curtis distance increasing and then plateauing at near 0.2. We see a similar but nosier correlation between density and euclidean distance, although this may be explained by the fact that Bray-Curtis was designed specifically to compare the diversity of populations while euclidean is a generic measure of distance. It may be that despite the increasing density ultimately less than half of all possible QS interactions are present and this results in certain populations largely growing or shrinking over the course of the simulation. The growth rate and total abundance do not seem to obey obvious trends but this may be the result of "flattening" all of the information

from each strain and looking at these statistics over all strains.

Regardless it does appear that just the amount of QS interactions alone can strongly influence population trajectories, specifically how "different" the final population is from the original.

Simulations with OTU data

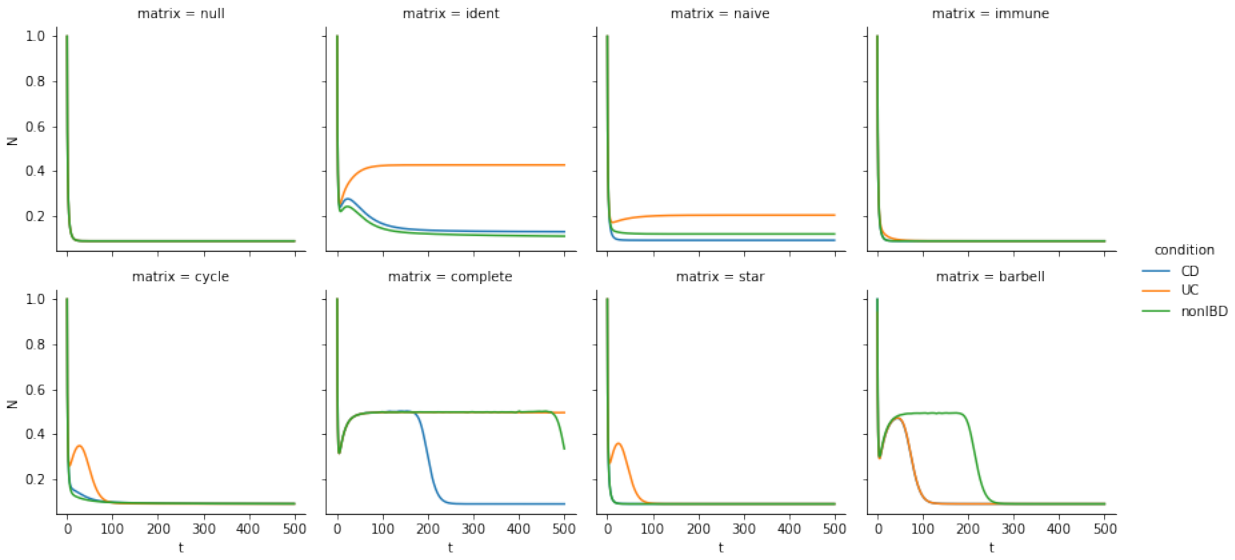

Figure 4: We plot the total abundance of all strains across time for each condition. In each case we use the specified K_{ac} matrix with 133 strain abundances (normalized) from a single patient with each condition.

We obtained microbiome data from HMP2, including OTU (Operational Taxonomic Unit) abundance data from stool samples and patient metadata (Research Network Consortium 2019). Specifically we care about the disease status as patients were either healthy (nonIBD), have ulcerative colitis (UC) or Crohn's Disease (CD). We had abundance data from 174 samples for 982 OTUs with 44 nonIBD, 44 UC and 86 CD samples. We use these abundance vectors as initial populations to our QS model and analyze the population trajectories.

We were interested in observing if different biologically relevant interaction matrix patterns (K_{ac}) would play a role in differentiating the different disease states. For each set of participants with the same diagnosis, we created a strain abundance list which consisted of every bacterial strain and its corresponding average abundance derived from the taxonomic profiles of the participants. Simulations were then performed for each interaction matrix pattern with the initial cell density parameter set as the normalized strain abundance list for each diagnosis. Although there is a loss of information pertaining to the growth dynamics of each strain over time, graphs corresponding to the total cell densities over time were generated for each matrix pattern.

The results below show that for most matrix patterns, given enough time, total cell densities

are very similar across all diagnosis types. We observe differentiation between non-IBD and disease states for the case where all strains have the same receptor and are considered naive cooperators. Differentiation in the trajectory of total cell density is also observed for the identity matrix pattern which signifies that each bacteria has an independent QS system. An interesting result is that the complete matrix pattern, where every strain's receptor binds every other strain's signal, shows different trajectory between CD and the UC/nonIBD cases.

Discussion

It is clear that differences in QS interactions related to producing a public good can results in significantly different population dynamics. Further examinations of more biologically relevant scenarios may yield more interesting results, especially examining other sets of parameters and interactions.

It is hard to determine whether the specific type of interaction matrix from our set of biologically relevant matrices can significantly differentiate the trajectory of total cell densities between the microbiome environments of IBD vs non-IBD. It would be interesting to use interaction matrices derived from real receptor-signal pairs found in nature for simulations (see (Aggarwal et al. 2020)). This could provide more insight into if these interaction matrices are important in determining disease states such as strep throat.

Contributions

Contribution	People
Building the model	Evan, Neel
Analysis	Sid, Evan, Neel, Deepika, Sarah
Presentation	Sid, Evan, Neel, Deepika, Sarah
Writing the report	Sid, Evan, Neel, Deepika, Sarah

Bibliography

Aggarwal, Surya D., Hasan Yesilkaya, Suzanne Dawid, and N. Luisa Hiller. 2020. "The Pneumococcal Social Network." *PLOS Pathogens* 16 (10). https://doi.org/10.1371/journal.ppat.1008931.

Calle, M. Luz. 2019. "Statistical Analysis of Metagenomics Data." *Genomics &Amp; Informatics* 17 (1). https://doi.org/10.5808/gi.2019.17.1.e6.

Dimitriu, Tatiana, Frances Medaney, Elli Amanatidou, Jessica Forsyth, Richard J. Ellis, and Ben Raymond. 2019. "Negative Frequency Dependent Selection on Plasmid Carriage and Low Fitness Costs Maintain Extended Spectrum Beta-Lactamases in Escherichia Coli." *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-53575-7.

Eldar, A. 2011. "Social Conflict Drives the Evolutionary Divergence of Quorum Sensing." *Proceedings of the National Academy of Sciences* 108 (33): 13635–40. https://doi.org/10.1073/pnas.1102923108.

Pérez-Escudero, Alfonso, and Jeff Gore. 2016. "Selection Favors Incompatible Signaling in Bacteria." *Proceedings of the National Academy of Sciences* 113 (8): 1968–70. https://doi.org/10.1073/pnas.1600174113.

Pollak, Shaul, Shira Omer-Bendori, Eran Even-Tov, Valeria Lipsman, Tasneem Bareia, Ishay Ben-Zion, and Avigdor Eldar. 2016. "Facultative Cheating Supports the Coexistence of Diverse Quorum-Sensing Alleles." *Proceedings of the National Academy of Sciences* 113 (8): 2152–57. https://doi.org/10.1073/pnas.1520615113.

Research Network Consortium, iHMP. 2019. "The Integrative Human Microbiome Project." *Nature* 569 (7758): 641–48. https://doi.org/10.1038/s41586-019-1238-8.

Waters, Christopher M, and Bonnie L. Bassler. 2005. "Quorum Sensing: Cell-to-Cell Communication in Bacteria." *Annual Review of Cell and Developmental Biology* 21: 319–46. https://doi.org/10.1146/annurev.cellbio.21.012704.131001.