PROJECT DRAFT: THE GEOMETRY OF ANTIBIOTIC RESISTANCE

1. Overview

Project deadline: April 14

Second meeting: March 12 1:30pm-5pm, Conference room in SE370K

Journal: Bulletin of Mathematical Biology (?)

Miriam Barlow   mbarlow@ucmerced.edu
Kristina Crona   kcrona@ucmerced.edu
Devin Greene   devin.greene169@gmail.com
Juan Meza   jcmeza@ucmerced.edu
Portia Mira   pmira@ucmerced.edu
Bernd Sturmfels   bernd@berkeley.edu

Kristina 209-2339601
Portia 209-819-0582

2. Project overview

This draft reflects discussions January 30.

2.1. Background: Population genetics, Devin.
Everything we do should be conventional or correct (or both), and we rely on population genetics. First important question:

How do we determine probabilities of mutational trajectories? The established method is straight forward, but may not be completely correct in our case (see the next section).

2.2. Background: Tem data and lab results, Miriam.
Detail: Would be nice to squeeze in some TEM-85 data where TEM-85 has optimal fitness, if possible.

2.3. Optimal antibiotic cycling, Bernd.
In contrast to the previous study [when we had fitness ranks only], we will work with probabilities for mutational trajectories, and try to find optimal strategies for cycling antibiotics (see the next section for background). The goal is to find a sequence of landscapes which will "force" the bacteria to mutate in a cycle starting from the wild-type.

We have discussed the following simplification of the problem:

When cycling, assume that exactly one mutation will occur before we switch to the next antibiotic. Find the sequence of (at most 15) antibiotics without any repeats, which is optimal [maybe associated with minimal escape risk].

2.4. **Properties of the fitness landscapes, Juan.** Analysis of ruggedness, additivity (?), and other properties of the landscapes.

2.5. **Discrete analysis of fitness landscapes, Kristina.** Analyze good and poor combiners among beneficial mutations using graphs and shapes.

3. Background

The introduction of beta-lactam antibiotics in 1944 began with penicillin and was a triumph of modern medicine. Among antibiotics, beta-lactams were of particular interest, because of their reliability in treating bacterial infections, and their minimal side effects on humans. However, shortly after their introduction, bacteria started expressing enzymes from mobile plasmids that could hydrolyze and inactivate the beta-lactams. In 1963, the TEM beta-lactamase emerged among gram negative bacteria, and it rapidly increased in frequency to become the most frequent beta-lactamase in most pathogenic gram negative populations. The term TEM was derived from Temoneira, the name of the patient from whom the enzyme was first isolated. TEM beta-lactamases have been found in Escherichia coli, Klebsiella pneumoniae and other gram-negative bacteria.

The length of a particular beta-lactamase? can be defined by the number of amino acids it has, where the amino acids are represented by letters in the 20-letter alphabet. For example, TEM-1, which is considered the wild-type, has a length of 287. Over 170 TEM variants have been found clinically, where 41 are single mutants, i.e., they have exactly one amino acid substitution, and the majority (90 %) have at most 4 amino acid substitutions. One such example is the TEM-50 variant, which contains four mutations. In this paper, we will study 16 possible variations of these mutations, all of which we created using site directed mutagenesis.

To study the effects of various antibiotics, we determined the growth rate of the strains expressing each variant in one of 14 antibiotics including: AM (amoxicillin), AMC (amoxicillin clavulate), AMP (Ampicillin), CAZ (ceftazidime), CEC (ceflacor), CPD (cefpodoxime), CPR (Cefprozil), CRO (cefuroxime), CTT (cefotetan), CTX(cefotaxime), FEP (cefepime), SAM (Ampicillin sulbactam), TZP (pipercillin tazobactam) and ZOX (ceftizoxime). In each case, we employed 12 replicates for each sample and for each antibiotic. We computed the mean and variance of each growth rate and computed significance between adjacent variants that differ by one mutation using one-way ANOVA analysis.

These results can be represented as a directed graph where the vertices denote a particular allele and the edges connect pairs of alleles that differ by one mutation. For each pair of adjacent alleles, we indicate the one whose expression resulted in the fastest growth by directing the edge towards that variant. This can be interpreted to mean that evolution would proceed in that direction if the two variants occurred simultaneously in a population. In other words, the vertex indicated by the arrowhead would increase in frequency and reach fixation in the population, while the other variant would be lost. We also used red arrows to indicate significance, and black arrows to indicate differences that were not statistically significant by ANOVA. We should note however that more sensitive assays might alter the significance levels.

4. FITNESS, GROWTH RATE AND SELECTION COEFFICIENT

For a population with non-overlapping generations, the population size, N, can be given by

$$(1) \qquad N = N_0 W^g,$$

where N_0 the the initial population and g is the number of generations.

The *absolute fitness* W is a measure of the expected reproductive success. Let us now consider a population where the wild-type has absolute fitness W_0, and a single mutant has absolute fitness W_1. The relative fitness w_1 is defined as

$$w_1 = \frac{W_1}{W_0}.$$

One parameter that will be of interest is the *selection coefficient*, which can be defined by

$$s = w_1 - 1.$$

If s is small we can use the estimate $s \approx \ln(s+1)$, from which it follows that

$$(2) \qquad s \approx \ln(s+1) = \ln\frac{W}{W_0}.$$

Now let us consider the equation

$$N = N_0 e^{\alpha t},$$

with continuous growth rate α (also known as the *Malthusian parameter*) and t is time. The continuous version of (1) can be written as

$$N = N_0 W^{\frac{t}{T}},$$

where T is the generation time, from which it follows that

$$(3) \qquad \alpha = \frac{\ln W}{T}.$$

Using (2) and (3) it then follows that

$$(4) \qquad s \approx \ln\frac{W}{W_0} = \ln W - \ln W_0 = T(\alpha - \alpha_0).$$

The importance of equation(4) is that we can approximate the selection coefficient through experimentally computed growth rates. However, one should keep in mind that this approximation holds true only when the selection coefficient is small; an assumption that we will see does not hold true in some cases.

5. PROBABILITIES OF MUTATIONAL TRAJECTORIES

Need to have a short introduction here to explain where we're going .... e.g.
Computing the probabilities of mutations has generated much discussion in the literature. Some methods widely useful make simplifying assumptions .... Other methods ...

In this section, we will present two/three? of the more widely used methods and discuss some of the advantages and disadvantages of them.

5.1. **Estimation of probabilities.** *M*ethod 1: Suppose that the population is dominated by a particular genotype. If the genotype has k beneficial mutational neighbors, with selection coefficients $s_1 \ldots s_k$, then one method for estimating the probability for the mutation j can be given by

$$(5) \qquad p_j = \frac{s_j}{s_1 + \cdots + s_k}.$$

For instance, if $w_{00} = 1, w_{10} = 1.1, w_{01} = 1.2$, then the probability for the mutation $00 \mapsto 10$ is $\frac{1}{3}$, and the probability for the mutation $00 \mapsto 01$ is $\frac{2}{3}$. Equation(5) is well established, but strictly speaking only correct if the fitness differences between genotypes are relatively small. **do we need a citation to back this claim up?**

If one now applies (4), the probability for the mutation j equals

$$\frac{\alpha_j - \alpha_0}{\alpha_1 - \alpha_0 + \cdots + \alpha_k - \alpha_0}.$$

Example 5.1. For Drug 1 and TEM-50 (see the next section) there are exactly two beneficial single mutations,

$$\alpha_{0000} = 0.0017775, \quad \alpha_{0010} = 0.002041667 \quad \alpha_{0001} = 0.001781667.$$

The probability that 0010 goes to fixation is

$$\frac{0.002041667 - 0.0017775,}{0.002041667 - 0.0017775 + 0.001781667 - 0.0017775} = 0.9844708$$

and the probability that 0001 goes to fixation is

$$\frac{0.001781667 - 0.0017775,}{0.002041667 - 0.0017775 + 0.001781667 - 0.0017775} = 0.01552915$$

*M*ethod 2: According to a different model, (**citation**??) the probabilities are equal for all beneficial mutations, so that one needs the fitness graphs only for computing the probabilities.

5.2. **Limitations and alternative approaches.** For accurately determining probabilities for mutational trajectories one may need to know population size, generation length and other parameters. As far as we know, existing theory applies only to a few ideal cases. In either case, both Method 1 and Method 2 have disadvantages in the current setting. As described earlier, Model 1 depends on small fitness differences between genotypes.

Devin performed a few simulations of fixation probabiliites using the so called Wright-Fisher model. The result indicated that Model 2 (equal probabilities) is more accurate in the event all beneficial mutations have a strong effect. Indeed, in such a case the probabilitiesfor fixation is almost equal for the beneficial mutations (whereas the more fit mutant seems to have a considerably greater chance to go to fixation if the fitness differences are small).

We believe the reason is that the stochastic element [for instance, even a mutant of high fitness and its descendants may go extinct because of "bad luck"] is of less importance if the all available mutants have very high fitness.

Fitness data:
Key:
0000 TEM-1
1000 M69L
0100 E104K
0010 G238S
0001 N276D
1100
1010
1001
0110
0101
0011
1110
1101
1011
0111
1111

We used 15 drugs, where Drug 1-Drug 15 are (in order):
AM, AMC, AMP8X, CAZ, CEC, CPD, CPR, CRO, CTT, CTX, CXM, FEP, SAM, TZP, ZOX
The mean fitness for each genotype [using the order described] for Drug 1 - Drug 15 are as follows (see the next page):

Drug 1:

0.0017775, 0.00172, 0.001448333, 0.002041667, 0.001781667, 0.001556667, 0.001799167, 0.002008333, 0.001184167, 0.001751667, 0.001544167, 0.0017675, 0.002246667,0.002005, 0.0000625, 0.002046667

Drug 2:

0.001435, 0.001416667, 0.001671667, 0.001060833, 0.001573333, 0.001376667, 0.0015375, 0.001350833, 0.0000733, 0.001456667,0.001625, 0.001306667, 0.001914167, 0.00159, 0.0000675, 0.0017275

Drug 3:

0.001850833, 0.00157, 0.002024167, 0.001948333, 0.002081667, 0.002185833, 0.0000508, 0.002165, 0.0020325, 0.002434167, 0.0021975, 0.0000875, 0.002321667, 0.0000825, 0.0000342, 0.002820833

Drug 4:

0.002134167, 0.000288333, 0.002041667, 0.002618333, 0.002655833, 0.00263, 0.001604167, 0.000575833, 0.002924167, 0.0026875, 0.002755833, 0.002893333, 0.002676667, 0.001378333, 0.000250833, 0.0025625

Drug 5:

0.002258333, 0.000234167, 0.002395833, 0.002150833, 0.001995833, 0.00215, 0.002241667, 0.000171667, 0.00223, 0.0026475, 0.001845833, 0.00264, 0.000095, 0.0000933, 0.000214167, 0.000515833

Drug 6:
0.000595, 0.000431667, 0.001760833, 0.002604167, 0.000245, 0.0006375, 0.002650833, 0.000388333, 0.00291, 0.0030425, 0.001470833, 0.0009625, 0.000985833, 0.0011025, 0.003095833, 0.003268333

Drug 7:
0.001743333, 0.001553333, 0.0020175, 0.0017625, 0.001661667, 0.0002225, 0.000165, 0.000255833, 0.002041667, 0.001785, 0.00205, 0.001810833, 0.000239167, 0.000220833, 0.002175, 0.000288333

Drug 8:
0.001091667, 0.00083, 0.00288, 0.002554167, 0.000286667, 0.001406667, 0.0031725, 0.00054, 0.002731667, 0.003041667, 0.000655833, 0.00274, 0.000750833, 0.0011525, 0.000435833, 0.003226667

Drug 9:
0.002125, 0.003238333, 0.003290833, 0.002804167, 0.001921667, 0.000545833, 0.0028825, 0.002965833, 0.003081667, 0.0005875, 0.0028875, 0.0031925, 0.003180833, 0.00089, 0.0035075, 0.002543333

Drug 10:
0.00016, 0.000185, 0.001653333, 0.001935833, 0.000085, 0.000225, 0.001969167, 0.00014, 0.002295, 0.0023475, 0.0001375, 0.000119167, 0.0000917, 0.000203333, 0.002269167, 0.002411667

Drug 11:
0.0017475, 0.0004225, 0.00294, 0.00207, 0.0017, 0.002024167, 0.001910833, 0.001578333, 0.002918333, 0.0019375, 0.002173333, 0.001590833, 0.0016775, 0.002754167, 0.003271667, 0.002923333

Drug 12:
0.00259, 0.002066667, 0.00244, 0.002393333, 0.002571667, 0.002735, 0.002956667, 0.002445833, 0.002651667, 0.002831667, 0.0028075, 0.002795833, 0.002863333, 0.0026325, 0.000610833, 0.0032025

Drug 13:
0.001879167, 0.0021975, 0.002455833, 0.000133333, 0.0025325, 0.002504167, 0.002308333, 0.00257, 0.0000833, 0.0000942, 0.002436667, 0.002528333, 0.003001667, 0.002885833, 0.0000942, 0.003453333

Drug 14:
0.002679167, 0.002709167, 0.0030375, 0.002426667, 0.002905833, 0.002453333, 0.000171667, 0.0025, 0.0025275, 0.000140833, 0.003309167, 0.000609167, 0.002739167, 0.0000933, 0.0001425, 0.000170833

Drug 15:
0.000993333, 0.001105833, 0.0016975, 0.002069167, 0.000805, 0.001115833, 0.001894167, 0.001170833, 0.0021375, 0.0026825, 0.00201, 0.001103333, 0.001105, 0.000680833, 0.002688333, 0.002590833

6.1. **Fitness ranks.** The fitness ranks are listed for each genotype (from 0000 to 1111, with the same order as before) for the 15 drugs.

8, 11, 14 , 3, 7, 12, 6, 4, 15, 10, 13, 9, 1, 5, 16, 2

9, 10, 3, 14, 6 , 11, 7, 12, 15, 8, 4, 13, 1, 5, 16, 2

11, 12, 9, 10, 7, 5, 15, 6, 8, 2, 4, 13, 3, 14, 16, 1

10, 15, 11, 8, 6, 7, 12, 14, 1, 4, 3, 2, 5, 13, 16, 9

4, 12, 3, 7, 9, 8, 5, 14, 6, 1, 10, 2, 15, 16, 13, 11

13, 14, 7, 6, 16, 12, 5, 15, 4, 3, 8, 11, 10, 9, 2, 1

7, 9, 3, 6, 8, 13, 16, 11, 2, 5, 1, 4, 12, 14, 15, 10

10, 11, 4, 7, 16, 8, 2, 14, 6, 3, 13, 5, 12, 9, 15, 1

12, 3, 2, 10, 13, 16, 9, 7, 6, 15, 8, 4, 5, 14, 1, 11

11, 10, 7, 6, 16, 8, 5, 12, 3, 2, 13, 14, 15, 9, 4, 1

11, 16, 2, 7, 12, 8, 10, 15, 4, 9, 6, 14, 13, 5, 1, 3

10, 15, 13, 14, 11, 7, 2, 12, 8, 4, 5, 6, 3, 9, 16, 1

12, 11, 8, 13, 5, 7, 10, 4, 16, 14, 9, 6, 2, 3, 15, 1

6, 5, 2, 10, 3, 9, 12, 8, 7, 15, 1, 11, 4, 16, 14, 13

14, 11, 8, 5, 15, 10, 7, 9, 4, 2, 6, 13, 12, 16, 1, 3

Mean fitness ranks for the 16 genotypes:
9.9, 11, 6.4, 8.4, 10, 9.4, 8.2, 10.5, 7, 8.5, 6.5, 6.9, 7.5, 10.5, 10.7, 4.7

Maximal fitness rank for the 16 genotypes:
14, 16, 14, 14, 16, 16, 16, 15, 15, 15, 13, 14, 15, 16, 16, 13

Observation The drugs have very different effect, so that at least one drug works well for any genotype:

A. For any genotype, there exists one drug so that the fitness rank of the genotype is at least 13.

B, The least mean fitness rank for the 16 genotypes is 4.7. (Specifically, the genotype 1111 [TEM-50] has mean fitness rank 4.7).

The situation seems favorable for cycling, since the drugs have very different effect.

sectionOptimal antibiotic cyclic

When is cycling a good strategy, and what strategy would be optimal?

Desirable goals:

1. Maximal probability that the sequence of mutants starting with the wild-type ends with the wild-type.

2. Avoid escape genotypes (if such genotypes exist).

3. If a genotype has very high fitness for one drug in the cycle, then it should have low fitness for the next drug in the cycle.

Example 6.1. Consider the landscape defined by TEM-50 and the 15 drugs.

The sequence : Drug 1- Drug 5 - Drug 9- Drug 1 ... seems like a good choice if for the third goal.

The rank 1 genotype for Drug 1 has rank 15 for Drug 5.

The rank 1 genotype for Drug 5 has rank 15 for Drug 9.

The rank 1 genotype for Drug 9 has rank 16 for Drug 1.

7. OPTIMAL ANTIBIOTIC CYCLIC

We are considering 15 antibiotics, labeled AM, AMC, AMP, CAZ, CEC, CPD, CPR, CRO, CTT, CTX, CXM, FEP, SAM, TZP, and ZOX. For each of these 15 antibiotics, we select exactly one TEM *fitness landscape*. Such a landscape is a real $2 \times 2 \times 2 \times 2$ tensor $\mathbf{f} = (f_{ijkl})$. The indices i, j, k, l are 0 or 1. We can identify f with a vector whose coordinates are indexed by $\{0, 1\}^4$.

A *mutation model* is a function $M : \mathbb{R}^{16} \to \mathbb{R}^{16 \times 16}$ that assigns a transition matrix to each fitness landscape. Recall that a *transition matrix* has nonnegative entries and its rows sum to 1. The rows and columns of $M(f)$ are labeled by $\{0, 1\}^4$, in some order that is fixed throughout. We require that our transition matrices respect the adjacency structure of the 4-cube, that is, $M(f)_{a,b} = 0$ unless a and b are vectors in $\{0, 1\}^4$ that differ in at most one coordinate. Thus each row of $M(f)$ has at most 5 non-zero entries.

Two mutation models are described in Section 3.2 of `projectdraft0204.pdf`. The second model is obtained by simply considering the directed graph on $\{0, 1\}^4$ where $a \to b$ means that $f_a < f_b$. The non-zero diagonal entries of $M(f)$ are $M(f)_{a,a} = 1$ if a has no outgoing edges, and the non-zero off-diagonal entries are $M_{a,b} = 1/\text{outdegree}(\text{a})$ for every directed edge $a \to b$.

Let $f_1, f_2, \ldots, f_{14}$ denote our 14 given fitness landscape, with derived transition matrices $M(f_1), M(f_2), \ldots, M(f_{14})$. Let $\mathcal{W}$ denote a finite set of words on the alphabet $\{0, 1, \ldots, 14\}$. These words represent the feasible *treatment plans* we are considering. Every word $w = w_1 w_2 \cdots w_k$ represents a new transition matrix, namely the corresponding product of 16×16-matrices

$$M[w] = M(f_{w_1}) \cdot M(f_{w_2}) \cdot \cdots \cdot M(f_{w_k}).$$

Our task is to solve the following discrete optimization problem:

Maximize the entry upper left entry $M[w]_{0000,0000}$ over all words w in $\mathcal{W}$.

The methodology used to solve this problem will depend on the choice of $\mathcal{W}$. For instance, it would be natural to take $\mathcal{W}$ as the set of all words of length exactly k, for some small positive integer k. Then $\mathcal{W}$ has 14^k elements. For $k \leq 5$ we can solve our problem by brute-force enumeration, but for $k \leq 6$ something more clever will be needed. At this point, I do not know whether a polynomial-time algorithm exists. The problem is reminiscent of the MAP inference

problem for Hidden Markov Models, which can be solved efficiently by a dynamic programming approach known as as the *Baum-Welch algorithm*. Our problem seems to be more difficult. We will need to do some literature search and talk to some experts to find more efficient algorithms. But, for starters, let's run the bruce-force computation for some small sets $\mathcal{W}$ of treatment plans that make sense from a biomedical perspective.

8. EPISTASIS

Let f_g be the fitness of the genotype g. For genotypes

$$00, 10, 01, 11,$$

no epistasis means that

$$f_{00} - f_{10} - f_{01} + f_{11} = 0.$$