## Evolution through programming

This is the 1st time we are doing this course. In this course we will try to explore the relationship between evolution and programming.

Types of programming tasks:

1. Simulations
2. Data related

The regular structure of the course is going to be on all lectures except the 1st and the last – today will be different.

**Nothing in biology makes sense – unless in the light of the theory of evolution.**

We have several topics in this course – they appear in the syllabus. Each week we will do programming homework upcoming to the class, we will discuss it in the 1st half of the class and then there will be a new theoretical topic.

We are expected to submit the programming task, in each class only 2 teams present (2-3 people in each team). We must decide about the order of presentations.

**Topics in the course:**

1. Today – talk about evolution, source&rate of mutations.
2. Fitness landscape and fitness scape.
3. Evolution of cooperation.
4. Neutral selection
5. The genetic code (codons) – why was this code selected
6. Regulation of gene expression – data analysis and programming
7. The notion of phylogeny – phylogenetic trees etc. try to reconstruct models of common ancestors. Over the course of 100 million years that mammals exist, we still have 80-90% similarity between mammals.
8. The evolutionary processes of cancer, B cells, T cells – evolution within the body
9. Genetic algorithms – a subbranch in computer science that tries to solve everyday problems by representing possible solutions as genes and applying the laws of evolution on them.
10. Not everything that evolves is biological – talking about evolution of language, ideas etc. talk about evolution of culture.

Today we will start introducing evolution.

## Lecture 1 – 27Mar2025

Although Darwin did not use the terms genotype and phenotype, he had these notions in mind. Note that the environment cares only about the phenotype and not about the genotype – only what is presented happens. Genotypes instruct the phenotype, they can be introduced with mistakes and the selected phenotypes have by chance the instructions to the phenotype that suits the current conditions best.

Prior to Darwin there was another scientist – LaMarck, but he believed that the mechanism of evolution is inheritance of changes that happen by phenotypes only (if you broke the arm, your child's arm would be broken. All giraffs were born with same length of neck, but some were able to extend their necks to get reward – food and passed it on to their progeny). The model of Lamarck is called inheritance of acquired traits.

This is not entirely wrong – epigenetics can be considered soft Lamarckian evolution, and some things such as language are hard- Lamarckian because the changes in language propagate into dictionaries. CRISPR is an example of Lamarckian evolution – the events that happen to the bacterium (infection by a phage) are acquired into its genome as spacers and is inherited to offsprings.

If we move a bit away from Darwinian evolution we get to a problem – according to Darwin mutations happen purely at random, but there is another option – that mutations occur when and where they are needed the most.

Luria & Delbruck experiment – transfer liquid bacteria into petri plate to get colonies. The bacteria grew in good conditions, but the plate has antibiotics, so we expect no bacteria on the plate but more bacteria on a non-antibiotic plate. In real life we do get antibiotic-resistant colonies. We assume that the resistance is a genetic property (although the Lamarckian possibility is interesting). They could have acquired the mutation in the liquid, or only in the presence of the bacteria – these mutations often come with a trade off so are selected against in non-antibiotic conditions. Luria and Delbruck set to answer the question when do mutations occur (randomly/on demand), and this will be the center of our home assignment this week. We should try to solve the problem in the terms that were available at their experimental time, but we can try to implement modern solutions as well.

Note – some resistance mechanisms are drug-specific, some are more broad; for example there is a phenomenon of multi drug resistant bacteria (such as pumps that pump the drugs back out of the cells).

A solution that was used by microbiologists Ester & Joshua Lederberg and started from a plate that has no drugs. They then took 2 or more plates that do have the drug (same drug) and blotted the bacteria to these plates – if the resistance developed on the 1st dish (not on demand) the same colonies should always be resistant (number & location) and if the mutations happen on demand the number & position would be different. This experiment was conducted after the original experiments of Luria & Delbruck and the results are that the mutations indeed happen randomly and not on demand (pure Darwinian evolution). This mechanism is called "replica plating".

Evolvability – the capacity to evolve (the center of research at the Pilpel lab).

The Luria & Delbruck solution – they decided to solve the problem mathematically and then perform an experiment. They used a single drug, single concentration and single timepoint but they repeated the experiment many times. In each of the replications they grew cells in a liquid culture (for ~10 generations), took one plate and plated the cells on it, and counted the number of resistant colonies they got. Therefore, they can check the properties of the distribution of the number of colonies achieved – mean, variance etc. they tried to characterize the distribution expected in both options (mutation in random or on demand). Rise of mutations upon demand will get the Poisson distribution – the probability to get a "saving" mutation at the timeframe before you die, assuming a constant rate of mutations. K is the number of colonies

K is the number of colonies (saving events in the timeframe) by the chances to get an event in such timeframe. We do not know lambda but we know the probability of k=0 (because in some experiments we did not get any colonies) and therefore we can calculate lambda from the equation.

In a Poisson distribution, the mean and the variance is equal to lambda. If the mean is small the variance would be small too, so the distribution would not be wide and will get a relatively strong peak. But they checked and the mean and variance of the actual distribution were not the same, so they came to conclusion that mutations happen randomly and not on demand. The option of random mutations would generate a different distribution, because we choose a number of cells out of the entire population to plate and X% of the population is already resistant, so we get only a percentage of our population resistant. The percentage depends on the time in which we got the mutation, so the earlier the mutation occurs the more resistant cells we would get. The events are dependent on each other, so we get a distribution that allows outliers, the possibility of 1/2 of the cells being resistant is essentially 0. This generates a probability distribution with a very long tail. Luria & Delbruck made a mathematical description of their non-Poisson distribution and when they made the experiment their equation fit the data.

In the assignment we will try to simulate the Luria & Delbruck experiment. The homework would try to deal also with assigning a cost to the mutation (so the growth rate is lower). Assume:

* P = probability of mutation (same or different for getting mutation and losing mutation?)
* H = depth (number of generations)
* R = repeats of experiment (how many times to we run the simulation)
* Simulate what happens in either scenario (mutations can occur at any timepoint or only on the last generation)

We need to convince ourselves that we get a difference between the two modes?

2nd part of the assignment – can we use this to simulate real life processes? Get processes that are both Darwinian and non-Darwinian? Can we run a simulation that has the two options combined and get a distribution that will indicate if we get the 1st option, the 2nd option or the two combined? We have to show that the distributions are distinct from one another.

Note – we are prompted to use our assignments to further explore other questions if they are more interesting to us. (e.g. epigenetic effect, horizontal gene transfer etc.)

Epigenetics – a level of regulation on the expression of genes, that sits on the genome itself and controls the expression. Mutations that are epigenetic in nature are with a memory time sorter than h (the height of the tree), so epigenetic mutations that appear too early in the tree are forgotten before we exam. If K is the epigenetic memory time, only mutations that occur up to K generations ago will be present in the last generation.

## Lecture 2 – 03Apr2025

**Part 1: presentations about the last week's homework**

It looks complicated to distinguish between the random model (pure Darwinian) and the combined model (both random and induced mutations), because the distributions are similar. Today we would sequence everything but back then they did not know what DNA is, let alone sequence. Can we solve this with sequencing? Hypothetically you could sequence the bacterial population right before plating the cells on antibiotics. One problem is that we do not necessarily know the nature of the beneficial mutation. Another problem is that the error rate is higher than the rate of natural mutation, so our measurement is not adequate to answer such question.

**Part 2: basic models for mathematical and computer analysis of evolution.**

In England even prior to Darwin people stated to think of evolution. Martus thought about the mathematical process of an exponential process (replicating bacteria), with a differential equation (see notebook for further notes).