Statistical mechanics approaches for studying microbial growth (Lecture 1)

Ariel Amir, 10/2019

In the following are included:

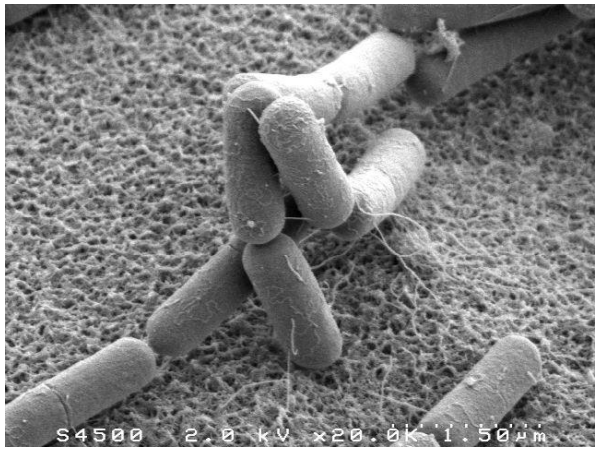
1) PDF of lecture slides

2) Calculations related to the Luria-Delbruck experiment. (See also the original paper referenced therein)

3) Calculation of the fixation probability in a serial dilution protocol, taken from SI of: Guo, Vucelja and Amir, *Science Advances, 5(7), eaav3842* (2019).

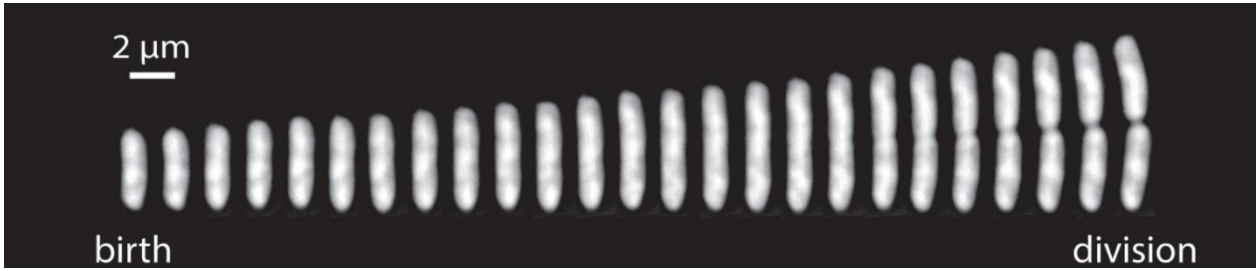
Statistical mechanics approaches for studying microbial growth

Ariel Amir



HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

E. coli



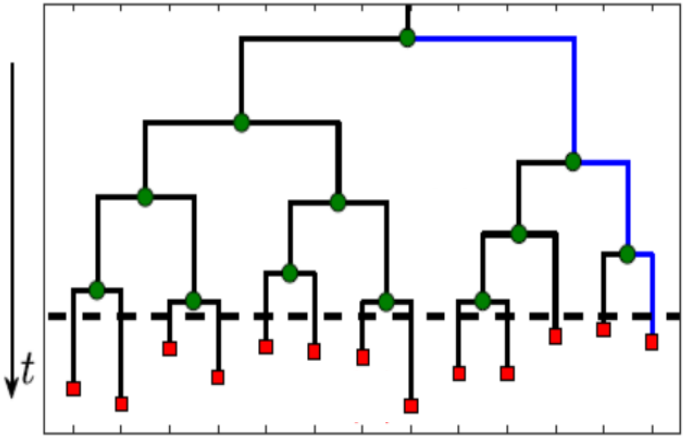
Mother cell

Daughter cells

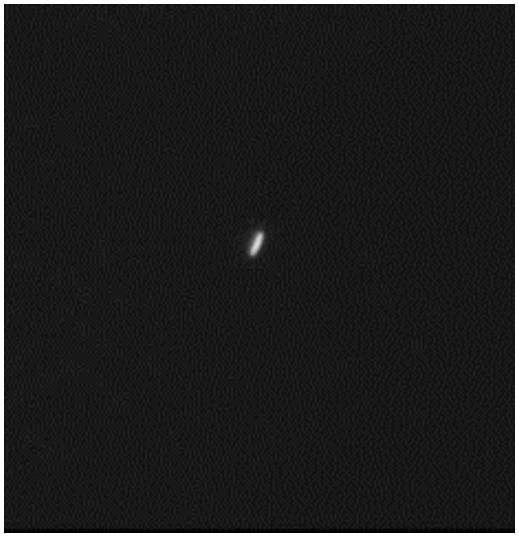
Time

Doubling time/ generation time/growth/replication

Taheri-Araghi et al. (2015)

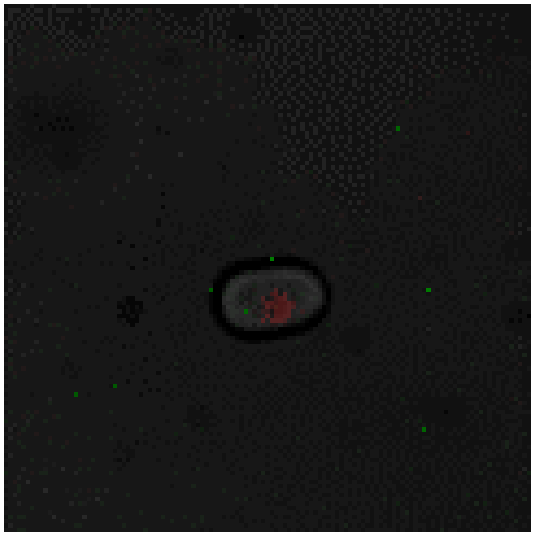


Microbial growth



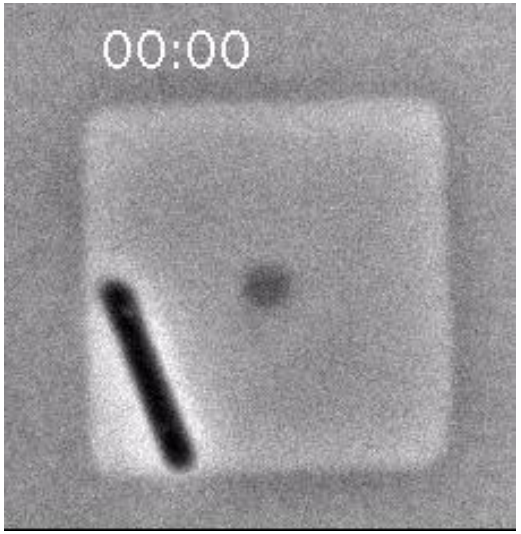
Escherichia coli

Stewart et al.,
PLoS Biol (2005),



Saccharomyces cerevisiae

Soifer al.,
Current Biology (2016)



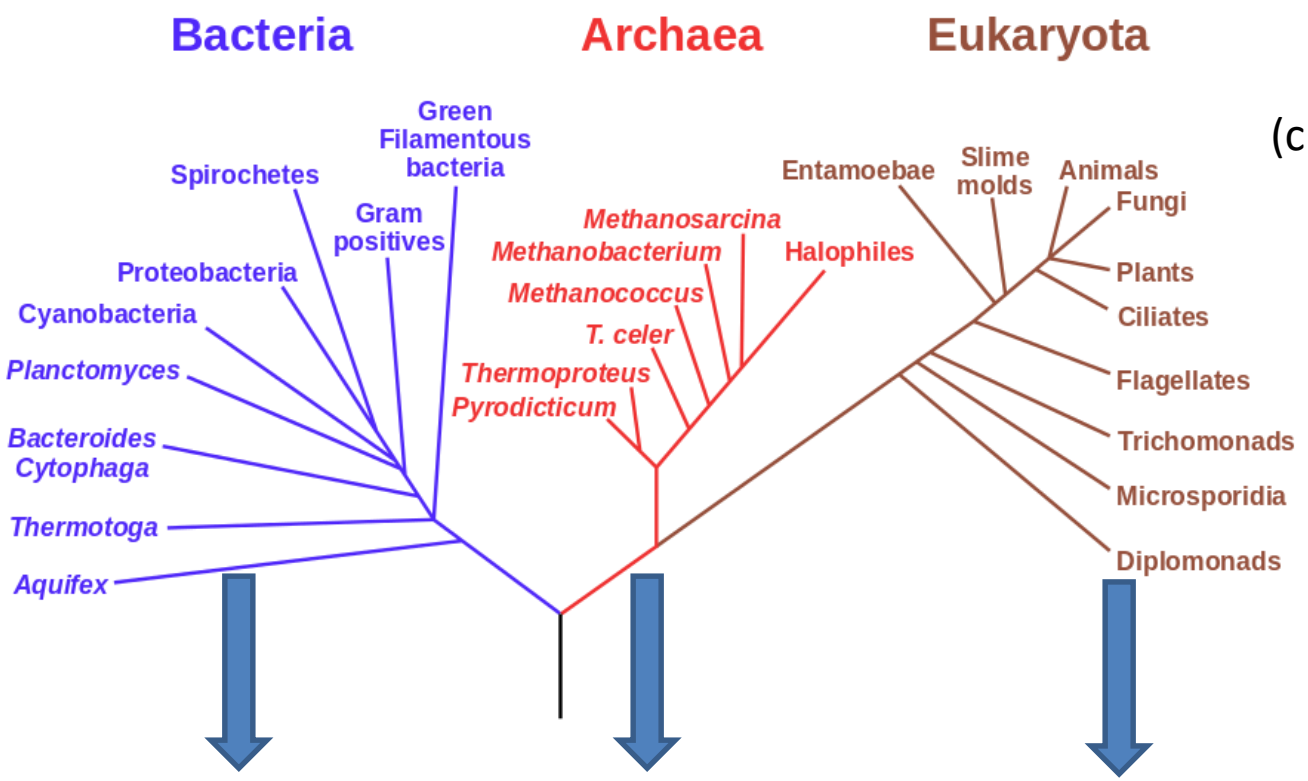
Halobacterium salinarum

Eun et al.,
Nat. Micro (2018)

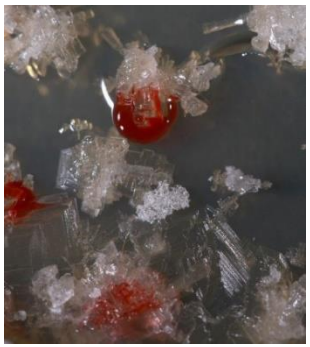
Doubling time

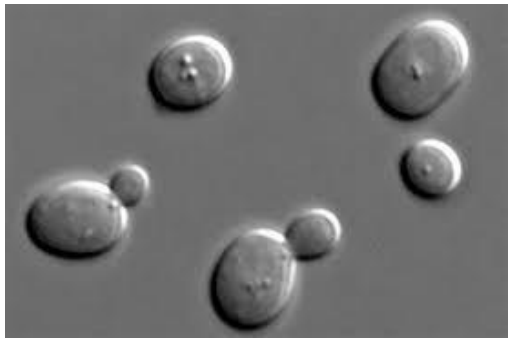
~ tens of minutes to several hours

Phylogenetic Tree of Life



(credit: wikipedia)





Outline

(Lecture I)

- Why study microbes? Luria-Delbruck experiment, Evolution experiments
- Introduction to microbial growth, with focus on cell size regulation

(Lecture II)

- Size control and correlations across different domains of life
- Going from single-cell variability to the population growth

(Lecture III)

- Bet-hedging
- Optimal partitioning of cellular resources

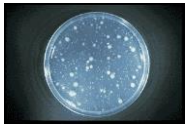
Why study microbes? Luria-Delbruck experiment

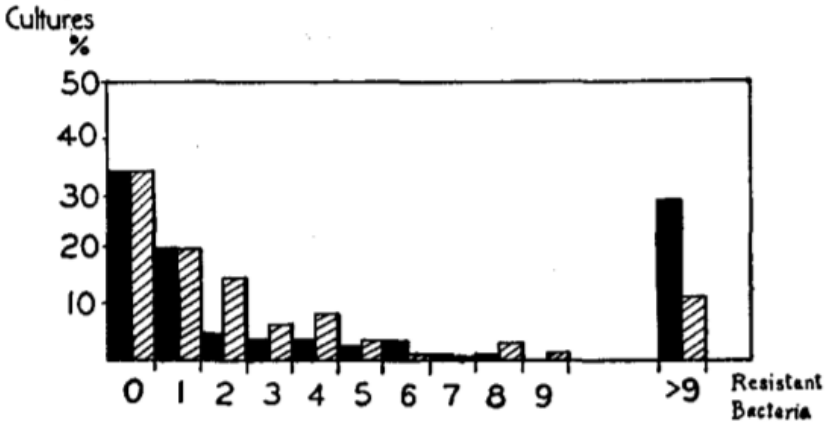
Protocol

- Grow culture to ~10^8 cells.

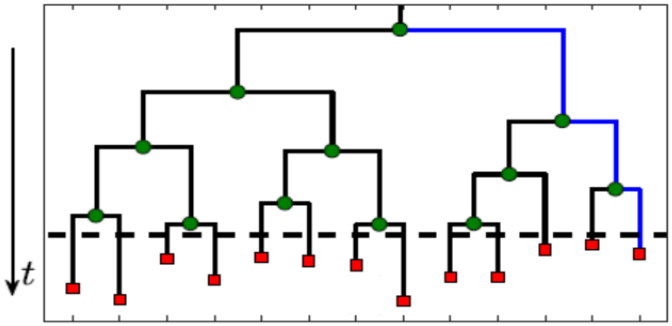
- Expose to bacteriophage.

- Count survivors (plating).



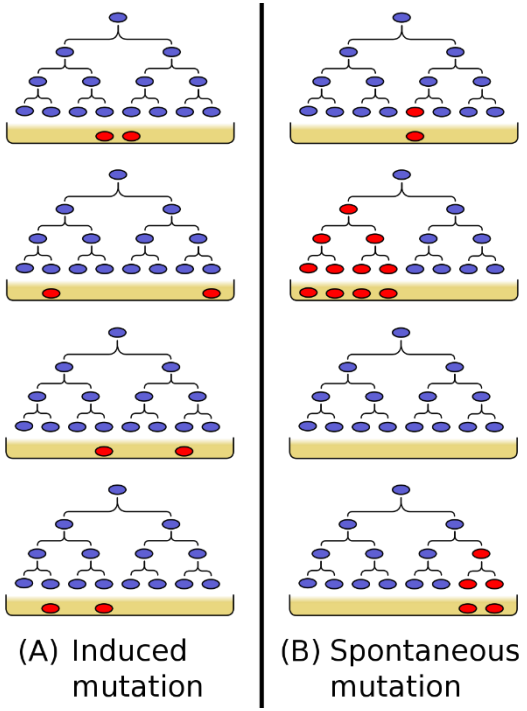


Luria and Delbruck (1943)



What to do when the experiment
is **not** reproducible?
(variance, standard deviation >> mean)

Why study microbes? Luria-Delbruck experiment



(Wikipedia)

Model 1: adaptation

- Survivors# must be follow Poisson distribution

→ Variance = Mean

Model 2: Random mutations lead to resistance

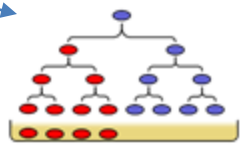
- Leads to broad ("jackpot") distribution since mutations can occur early/late
- Mathematically non-trivial: Landau distribution (Levy stable)
- Simulation straightforward (as is variance:mean estimate)

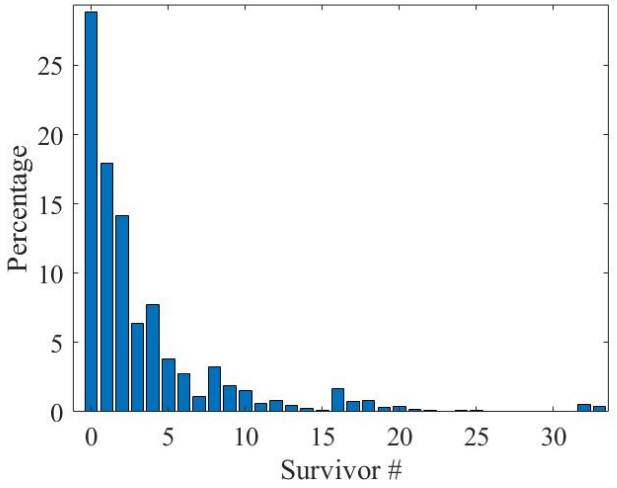
Why microbes? Luria-Delbruck experiment

```
mutant_rate=6*10^-8;
trial_num=10000;

for k=1:trial_num
    Mutant_num=0;
    Non_mutant=10;
    for j=1:20
        Total_1=2*Non_mutant;
        Total_2=2*Mutant_num;

        tmp1=poissrnd(mutant_rate*Total_1); %new mutants

        Mutant_num= Total_2+tmp1;
        Non_mutant = (Total_1-tmp1);

    end;
    Res(k)=Mutant_num;
end;
q=hist(Res);
```

# of Experiments

# of generations

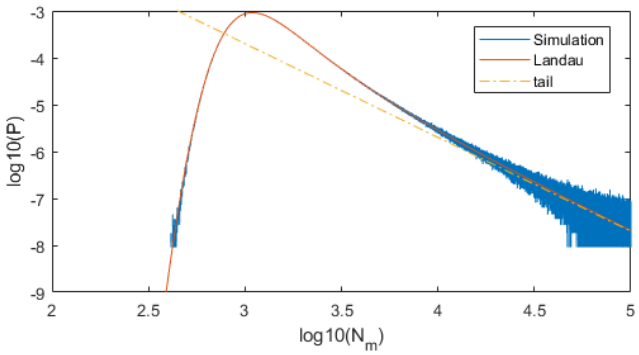




- In case of large number of generations:

Convergence to a Levy-stable distribution (Mandelbrot, Journal of Applied Probability, 1974)

See also: Kessler and Levine, PNAS 2013



Quantitative evolution

Imagine that you have discovered a well-preserved and clearly stratified fossil bed that provides a record of evolution extending thousands of generations for the particular organism that you study….

And the fantasy continues. Imagine that you could resurrect these organisms (not merely bits of fossil DNA but the entire living organisms) and reconstruct their environment exactly as it was during the thousands of generations preserved in the fossil bed….

From: Lenski and Travisano, PNAS (1994)

Why microbial evolution?

Yet this fantasy is not fiction; it is fact. We have many such "fossil beds" preserved, and we have "traveled in time" to manipulate populations with respect to their history and environment. The fossil beds are preserved in a freezer and contain populations of the bacterium Escherichia coli. Our time travel thus far extends over 5 years, representing >10,000 generations in this system, and we have manipulated many populations each comprising millions of individual organisms. In essence, our approach might be called experimental paleontology.

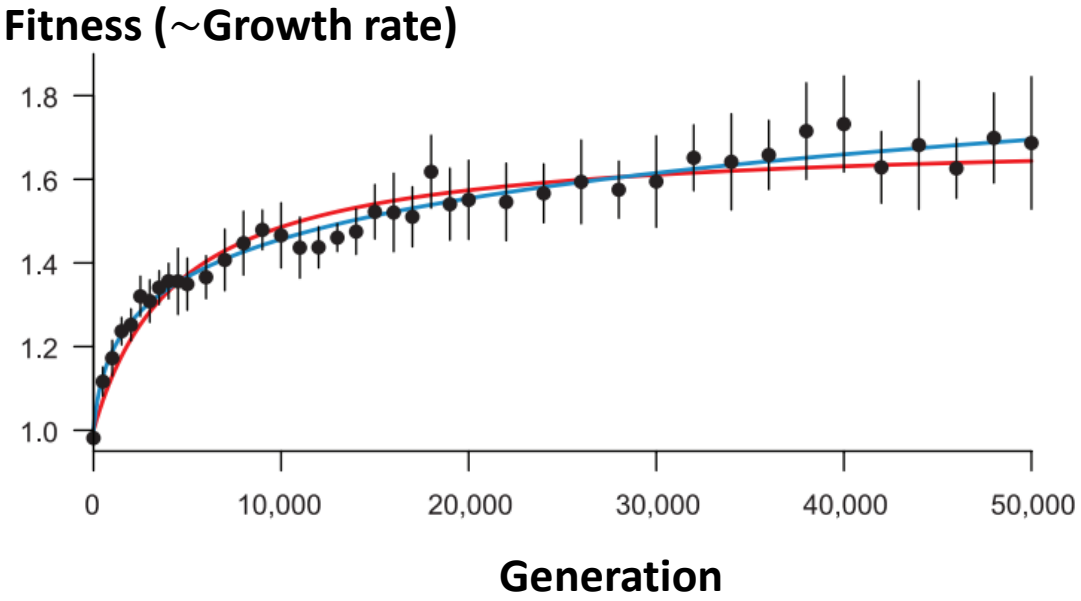
- “Replaying the tape of life”
- Fast doubling time
- Physiology better understood

From: Lenski and Travisano, PNAS (1994)

Evolution in the lab

- Since 1988 Lenski et al. grow bacteria in culture tube,

  X100 fold diluting into new media every day.
- Surprisingly, growth-rate ("fitness") keeps increasing without saturation!



Credit: *E. coli* long-term evolution experiment, wikipedia

Wiser, Ribek and Lenski, Science (2013)

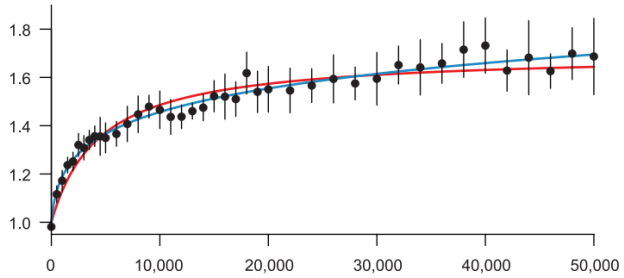
Evolution in the lab

- If beneficial mutations along genomes are independent:

  this maps (more or less) to “coupon collector problem”

- Then fitness should rapidly plateau as power-law

$$F = \sum a_i S_i \rightarrow \; F(t) \sim Fmax - C/t^2$$

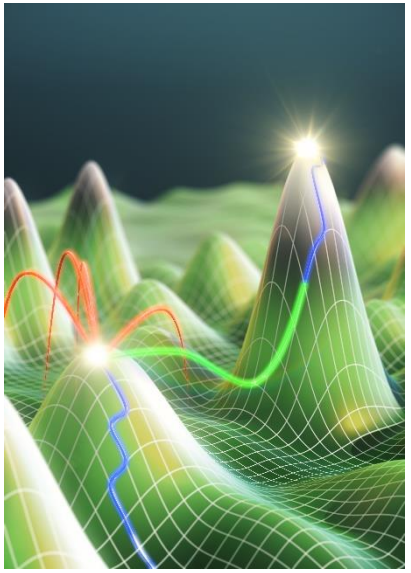


- Slow relaxations hint at role of **epistasis**: interactions between genes

- Adding *random interactions between genes* (similar to spin-glass model)

  slows down dynamics and can give rise to logarithmic trajectories.

$$F = \sum a_i S_i + \sum I_{ij} S_i S_j \rightarrow \; F(t) \sim a + b\log(t)$$



Yipei Guo, Marija Vucelja and AA,
Stochastic tunneling across fitness valleys can give rise
to a logarithmic long-term fitness trajectory (*Science Advances, 2019)*

Why microbes?

- We can learn fundamental biology – applicable to higher organisms – by the quantitative study of microbes.

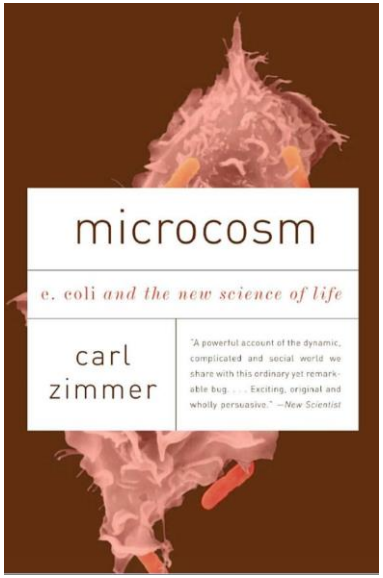
Example I: establishing the role of mutations in evolution.

Example II: Quantifying evolutionary dynamics and the processes involved in it.

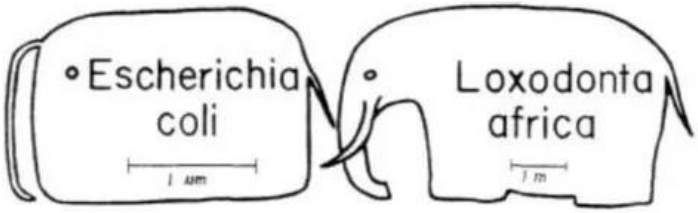
- Additional lesson:

Quantitative analysis of the data can lead to novel, qualitative insights

- So far… no physical forces etc., but a physicist's approach



Why microbes?



- We can learn fundamental biology – applicable to higher organisms – by the quantitative study of microbes.

  "Anything found to be true of *E. coli* must also be true of elephants."
  Jacques Monod, 1954

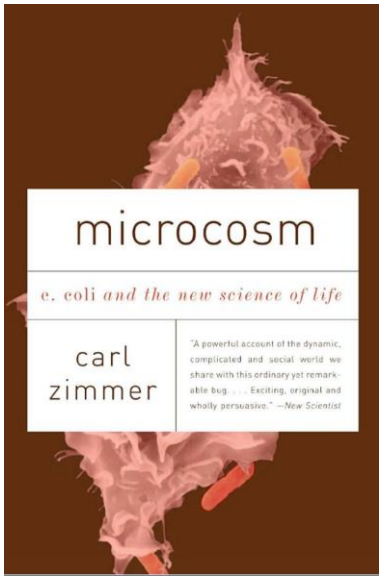
Example I: establishing the role of mutations in evolution.

Example II: Quantifying evolutionary dynamics and the processes involved in it.

- Additional lesson:

Quantitative analysis of the data can lead to novel, qualitative insights

- So far… no physical forces etc., but a physicist's approach



Chapter 3

Delbruck-Luria experiment: learning from stochasticity

Imagine doing an experiment a large number of times, and finding a completely different outcome on every run. How would you respond to such scenario? Most likely you will discard the data, and attempt to improve the experimental setup – perhaps attempt to get a more precise measurement setup, or try to ensure that the initial conditions are identical in all runs. In this chapter, we will describe a classical experiment in biology, where this happened. Specifically, the standard deviation of the measurement far exceeded the mean. Instead of being a nuisance or an unwanted result, this observation by itself led to the most important results of the experiments. Delbruck and Luria were examining how many bacteria in a culture survive a viral attack. The combined work of Luria – who did the experiment – and the modeling work of Delbruck, led the pair to an understanding that it is mutations (rather than adaptation) that provide bacteria with resistance to bacteriophages, a finding for which they were awarded the Nobel prize. You are encouraged to read their original paper on this from 1943 [1], which is a beautiful case of modeling and simple maths leading to profound results.

3.1 The experimental setup and results

The experiment done by Luria is simple to define: we start with a few *E. coli* bacteria (50-500 in the original experiment), and let them grow. They reproduce by asexual reproduction, with each one growing and dividing into two daughter cells within about $\tau_d = 20$ minutes, i.e., at time t the number of cells will be:

$$N(t) = N_0 \cdot 2^{t/\tau_d}. \tag{3.1}$$

After getting about 10^9 cells, Luria exposed the bacteria to a virus (a bacteriophage), that killed nearly all of the cells. Nevertheless, several bacteria survived in most of the experiments. By plating the cells and counting the number of individual colonies that emerged, he could know approximately how many cells managed to survive the viral attack. This number was the main result of his experiment. Annoyingly, this number varied enormously from experiment to experiment, no matter how hard Luria tried to control things!

We also note that once the cells developed resistance to the virus, all of their offspring would also be resistant: to prove this, Luria showed that exposing the progeny of cells from the surviving colonies to the virus does not kill them.

3.2 What's going on?

As of 1943, there were two possible hypotheses as to the mechanism through which bacteria develop resistance to viruses: in the first, "Lamarckian" approach, the cells attempt to adapt to the virus, after exposure. The majority do not succeed in adapting, and die, but a small fraction adapts and survives. Note that in this scenario, the cells only start to adapt in the *final* generation, once they are exposed to the virus.

In the second hypothesis, resistance comes from a beneficial mutation that provides the cell with immunity against the virus. In this case, if the mutation happened early on in the experiment, all of the progeny of that cell will also be resistant.

In the next section we will show that if the Lamarckian adaptation hypothesis is correct, the probability distribution of the number of surviving cells approximately follows a Poisson distribution – since every cell has some finite probability p to adapt, and the adaptation of every cell is an independent process. We will also show that the variance to mean ratio of the number of survivors over many trials would be 1, and hence this model cannot account for the high variance:mean ratio observed experimentally. Because of this, the adaptation hypothesis can be ruled out!

In the mutation hypothesis, the crucial thing to note is that the final survivor population is exponentially sensitive to the generation in which the mutation occurred - in the unlikely event that it happened early on, we will have a huge number of resistent cells at the end; while if it happened very late in the experiment only a handful of cells will survive. Therefore the distribution of the number of survivors is very broad, with an enormous variance:mean ratio, which we will quantify in the next section.

Together, the modeling of Delbruck and Luria allowed them to rule out the adaptation hypothesis and accept the mutation hypothesis. Remarkably, it is the stochasticity and fluctuations in the results of these experiments which proved to be the "smoking gun" of the model. Only via their quantitative analysis could they support their biologically relevant conclusion.

Note: also today the question of how bacteria develop resistance (to antibiotics rather than bacteriophages) is highly relevant to society, with exciting ongoing research unravelling some surprising strategies used by the cells. In the next chapter we will discuss how "gambling" may help the bacteria be more resistant, a phenomenon known as "bet hedging".

3.3 Quantitative analysis

Adaptation Consider first the adaptation model. A single cell present when the virus is applied either adapts and survives with some small probability p (hence contributing $x_i = 1$ to the number of surviving cells) or dies (and contributes $x_i = 0$ to the total number of surviving cells). This is also known as a Bernoulli process, or coin-flipping. Therefore, to find the probability of having X survivors out of the entire population of N cells, we have to consider N "coin-flips" with a probability of success p for every flip. Hence:

$$P(X) = \binom{N}{X} p^X (1-p)^{N-X}, \tag{3.2}$$

i.e., it is a binomial distribution. When p is very small, which is the case in the Delbruck-Luria experiments, we can approximate the distribution for $X \ll N$ as:

$$P(X) \approx \frac{1}{X!} N^X p^X (1-p)^N, \tag{3.3}$$

where we used $\binom{N}{X} = N(N-1)...(N-X+1)/X! \approx N^X/X!$. Note that using the Taylor expansion we previously used for $\log(1+\epsilon)$, we have:

$$N \log(1 - X/N) \approx -X \rightarrow (1 - X/N)^N \approx e^{-X}. \tag{3.4}$$

Therefore we find:

$$(1-p)^N = (1 - \frac{Np}{N})^N \approx e^{-Np}, \tag{3.5}$$

Defining $\lambda \equiv Np$ we find that:

$$P(X) \approx \frac{\lambda^X}{X!} e^{-\lambda}. \tag{3.6}$$

This is the Poisson distribution. It is easy to check that the distribution is indeed normalized, using the following Taylor expansion formula to sum over the X's:

$$e^{\lambda} = 1 + \lambda + \lambda^2/2! \ldots. \tag{3.7}$$

Note that in our example from week 1 where we consider buses being dispatched randomly from the station, this distribution would describe the number of buses arriving in a large time interval T – also there, there is a constant probability p for a bus to arrive at every time interval.

Consider a single cell present when the virus is applied, corresponding to a single "coin-flip". The expectation value of its contribution to the number of living cells at the end of the experiment, X, is clearly $\mathbf{E}(x_i) = p$, and the variance (for the single cell contribution) is:

$$Var[x_i] = p(1-p)^2 + (1-p)(0-p)^2 = (1-p)p((1-p)+p) = p(1-p) \approx p = \mathbf{E}[x_i]. \tag{3.8}$$

Therefore, for a *single* coin-flip, the variance and mean are approximately the same for small p!

Furthermore, the total number of survivors X is simply the sum of N such independent variables, each having approximately the same variance and mean. Thus we find:

$$\mathbf{E}[X] = Np \approx Var[x], \tag{3.9}$$

as we have asserted previously. In fact, since the Poisson distribution arises from the binomial one in the limit where $p \to 0$, this relation then becomes *exact* for the Poisson distribution. This can also be verified directly by calculating its two first moments.

Mutation

Let us now assume that the mutation hypothesis is correct, and consider the cells of the nth generation. Their number is $N_0 \cdot 2^n$, so the expected number of mutations occurring in that generation is:

$$\langle M_n \rangle = N_0 \cdot 2^n \cdot p, \tag{3.10}$$

where the $\langle\rangle$ denotes averaging over many trials.

Since the probability of mutation is small, the probability of two mutations occurring in a single experiment is small, and the expected number of survivors originating from a mutation in the nth generation is:

$$\langle S_n \rangle = N_0 \cdot 2^n \cdot p \cdot 2^{g-n} = N_0 \cdot 2^g \cdot p, \tag{3.11}$$

where g is the total number of generations. The fact that this number is independent of n is crucial for our understanding of the DL experiment. It means that the survivors are equally likely to come from early or late generation, since although the odds are small for the mutation to occur earlier (due to the much smaller number of cells), this rare event will result in a huge number of survivors.

To see this more formally, let us calculate the mean and variance of the number of survivors. We have:

$$S_{total} = \sum_{j=1}^{g} S_j. \tag{3.12}$$

From Eq. (3.11) we can write:

$$\langle S_{total} \rangle = \sum_{n=1}^{g} N_0 \cdot 2^n p \left(2^{g-n}\right) = N_{total} \cdot g \cdot p, \tag{3.13}$$

where $N_{total} = N_0 \cdot 2^g$, the growth of the initial population over g generations. Since mutations are rare, we can write the variance as a sum of the square number of survivors arising from each particular mutant. Pooling these together by generation leads to:

$$\langle S^2_{total} \rangle \approx \sum_{n=1}^{g} N_0 \cdot 2^n p \left(2^{g-n}\right)^2, \tag{3.14}$$

where $N_0 \cdot 2^n p$ is the expected number of mutations in the nth generation and $\left(2^{g-n}\right)^2$ is their contribution to the variance. This sum can readily be evaluated:

$$\langle S^2_{total} \rangle \approx p N_{total} \sum_{i=1}^{g} 2^{g-n} \approx p N_{total} 2^g. \tag{3.15}$$

It is easy to see that the second moment is much larger than the mean squared, and hence it is also approximately equal to the variance:

$$Var[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \approx \langle S^2_{total} \rangle. \tag{3.16}$$

We therefore conclude that the variance:mean ratio is ridiculously high. In fact, the mean and variance calculation reflect the expected result if we repeat the experiments an enormous – and impractical – number of times. In order to find typical variance:mean ratio for the experiments, we have to consider the finite number Q of experiments performed. In your problem set you will simulate the experiment, which can be used to provide precisely this estimate.

Bibliography

[1] Luria, Salvador E., and Max Delbrck. "Mutations of bacteria from virus sensitivity to virus resistance." Genetics 28.6 (1943): 491.

Section SF. Batch culture

Our findings also hold for the case where cells are grown in batch culture (subject to the standard dilution protocol), after making the corresponding changes to the expression for fixation probabilities derived below.

Average fixation probabilities

Single mutants For a standard dilution protocol, we assume there are N_0 cells at the start of the day, N_f cells at the end of the day, after which a dilution of $D = \frac{N_f}{N_0}$ fold is performed.
The probability $p_{ext,k}$ of k single mutants with fitness effect s at the start of a day eventually going extinct over many dilutions can be found from requiring that all daughter mutants in the next day must eventually go extinct. This can be represented by the following equation:

$$p_{ext,k} = \sum_{j=0}^{min(N_{m,k},N_0)} Hyge(j, N_{m,k}, N_f, N_0) p_{ext,j} \tag{S7}$$

where $N_{m,k} \approx kD^{1+s}$ is the number of mutants at the end of the day if we started the day with k mutants, and $Hyge(j, N_m, N_f, N_0) = \frac{\binom{N_m}{j}\binom{N_f - N_m}{N_0 - j}}{\binom{N_f}{N_0}}$ is the distribution for the number of mutants drawn during the dilution process. We also demand the boundary conditions $p_{ext,0} = 1$ and $p_{ext,N_0} = 0$. Solving this set of linear equations, the fixation probability of a single mutant emerging at the start of a day can be obtained from $p_f = 1 - p_{ext,1}$. For $s = 0$, this would give $p_f = 1/N_0$.
To obtain analytical expressions for the fixation probability, we assume that $p_{ext,k} \approx p_{ext,1}^k$, which is a good approximation for the typical case where $k \ll N_0$ since in this regime the mutant cells can be considered to behave independently from one another. In the limit where $N_m, N_f, N_0 \gg 1$, $Hyge(j, N_m, N_f, N_0)$ can be approximated by the Poisson distribution with mean $\Lambda = N_0 \frac{N_m}{N_f}$. Therefore, p_f can be found approximately from the following self-consistent

equation (*45*):

$$\begin{aligned} p_f &= 1 - \sum_{j=0} \frac{e^{-\Lambda}\Lambda^j}{j!}(1-p_f)^j \\ &= 1 - e^{-\Lambda p_f} \end{aligned} \tag{S8}$$

For $|s|\log(D) \ll 1$, which is typically the case, this gives $p_f \approx 2\text{log}(D)\text{max}(s,0)$.

However, in general the mutation can occur after some fraction of the day $\tilde{t}$ has passed. Given that a mutation has occurred, the probability that it emerged at $\tilde{t}$ is given by the probability density

$$P(\tilde{t}) = \frac{D^{\tilde{t}}\text{log}(D)}{D-1} \tag{S9}$$

For $\tilde{t} > 0$, there is a reduction in the number of mutants at the end of the first day, such that $N_m(\tilde{t}) = D^{(1+s)(1-\tilde{t})}$, and following the same argument as in Eqn.S8,

$$\begin{aligned} p_f(s,\tilde{t}) &= 1 - \sum_{j=0} \frac{e^{-\tilde{\Lambda}}\tilde{\Lambda}^j}{j!}(1-p_f(s,0))^j \\ &= 1 - e^{-\tilde{\Lambda} p_f(s,0)} \end{aligned} \tag{S10}$$

where $\tilde{\Lambda}(\tilde{t}) = N_0\frac{N_m(\tilde{t})}{N_f} \approx D^{-\tilde{t}}$ for $|s| \ll 1$. This gives $p_f(s,\tilde{t}) = 2\text{log}(D)D^{-\tilde{t}}\max(s,0)$ for $\Lambda(\tilde{t})p_f(s,0) \ll 1$ (*45*).

Given that a mutation of selection coefficient s arises, its average fixation probability is then the average over all possible times within a day the mutation could have occurred:

$$\begin{aligned} \overline{p_f}(s) &= \int_0^1 p_f(s,\tilde{t})P(\tilde{t})d\tilde{t} \\ &= 2\frac{(\text{log}(D))^2}{D-1}\max(s,0) \end{aligned} \tag{S11}$$

Probability of successful double mutants In the large population limit, $p_f(s \leq 0) = 0$ (Eqn. S11), which implies that the population can only successfully accumulate beneficial mutations. For the population to escape a metastable state through beneficial double mutations, the first mutant (with a single deleterious mutation) must gain a second mutation before going extinct, and the effective mutant (with two mutations) must fix in the population.