

1 实验目的

Luria-Delbrück 实验诞生于上世纪 40 年代，当时人们虽然还没有认识到 DNA 遗传物质，但已经建立起遗传和突变的概念。

本实验是为了探究当外界环境压力出现时，存活下来的细菌是因为在环境压力到来之前产生的随机突变让它们躲过筛选 (此观点之后简称为“突变假说”)，还是因为在压力到来时一些随机因素让种群中的部分个体幸存 (此观点之后简称为“获得性假说”)。

最初实验的实验对象是大肠杆菌，环境压力由噬菌体造成。本次实验使用的模型是酵母菌，环境压力来自含有刀豆氨酸的选择培养基。

通过本次实验课，我们学习了一些实验室常用的无菌操作，尝试使用理论模型去分析实验数据，同时使用计算机对生物系统的演化过程进行模拟。

2 实验过程

湿实验部分，助教提供了较为详尽的实验方案，在此不再赘述其细节，仅列举其核心环节。

2.1 配置培养基

配置过量的含有刀豆氨酸固体选择培养基，以及 SD 培养基。

2.2 酵母培养

制备 400 ml 细胞浓度在 188.75/ml 量级的培养液于锥形瓶中，向 96 孔板中各加入 1ml 的含酵母菌的培养液，剩余留在原锥形瓶中，在合适的条件下培养。(约 65 小时)

2.3 涂板计数

在深孔板中随机取 3 个孔，稀释 100 倍后测定其浓度，也测定锥形瓶中酵母菌的浓度。从深孔板的 96 个孔 (理论上，实际上由于之前的实验操作出现失误，最终有效的是 93 个孔) 中各取 50 μl 涂在固体选择培养基的一个网格 (3x3) 中，此操作重复一次，对照组的锥形瓶中也涂相同数量。

培养两到三天后收取计数 (由于疫情原因，此过程为助教完成)，计数的结果中的每一个菌落，认为是在涂到选择培养基上时活下来的单个细菌。

3 实验结果以及数据分析

3.1 计算突变率

有 2 种计算突变率的方法：

3.1.1 根据 $p_0 = e^{-m}$ 计算

认为在刚刚涂布到培养基上时，存活下来的细菌数目符合泊松分布，酵母菌初始状态为对外界选择没有抗性的野生型酵母（由 N_0 很小保证的），没有发生任何突变（本实验不区分有助于酵母菌逃过筛选的突变的种类）就无法逃过筛选（发生了抗性突变则不会反向突变），可以得到由泊松分布预测的下的酵母菌数。通过所得结果中 0 所占比例，可以得到平均突变率：

$$p_0 = e^{-m}, m = a(N_t - N_0), a = \frac{m}{N_t}$$

m 为突变数目， a 为平均突变率， N_0 为刚开始培养时每个 culture 的酵母菌数约为 188.75（由于其相比于 N_t 很小，在计算时可以忽略）

3.1.2 根据 $r = aN_t \ln(N_t Ca)$ 计算

由于 N_0 取得很小，实际的突变率也很低，因此平均而言在第一个突变发生前可能会过一段比较长的时间，我们将此时间记作 t_0 。 r 为观测到的 culture 中平均抗性菌的个数，有 $r = (t - t_0)aN_t$ ，对于 C 个酵母菌 culture，依据 t_0 定义，满足 $1 = aC(N_{t_0} - N_0)$ 。又因为， $N_{t_0} = N_0 e^{-(t-t_0)}$ （这里的时间单位为酵母菌的细胞周期/ $\ln 2$ ），有 $t - t_0 = \ln(N_t Ca)$ ，将 $t - t_0$ 消去，得到 $r = aN_t \ln(N_t Ca)$ ，可以据此计算平均突变率 a 。

突变速率 *:mutations per bacterium	WL	ZSJ&ZGX1	ZSJ&ZGX2	TYX	ZTY
方法 1 *per time unit	7.04E-09	6.82E-08	8.75E-08	2.02E-08	1.17E-08
方法 1 *per division cycle	4.88E-09	4.73E-08	6.06E-08	1.40E-08	8.09E-09
方法 2 *per time unit	7.13E-08	3.62E-07	3.66E-07	1.02E-07	7.03E-08
方法 2 *per division cycle	4.94E-08	2.51E-07	2.54E-07	7.04E-08	4.87E-08

第二种方法计算出的突变率大于第一种。可能的原因是第二种方法计算 r 均值时，受到了数值较大的数据点的影响，导致该方法计算得出的突变率偏高。

3.2 实验数据方差、均值分析

据“获得性假说”，当压力来临时，每个酵母菌的随机存活，存活下来酵母菌数应服从泊松分布。

对照组的样品取自同一个 culture，可以认为对照组数据的分布情况近似于泊松分布，实验组的每个数据点都来自于不同的 culture。

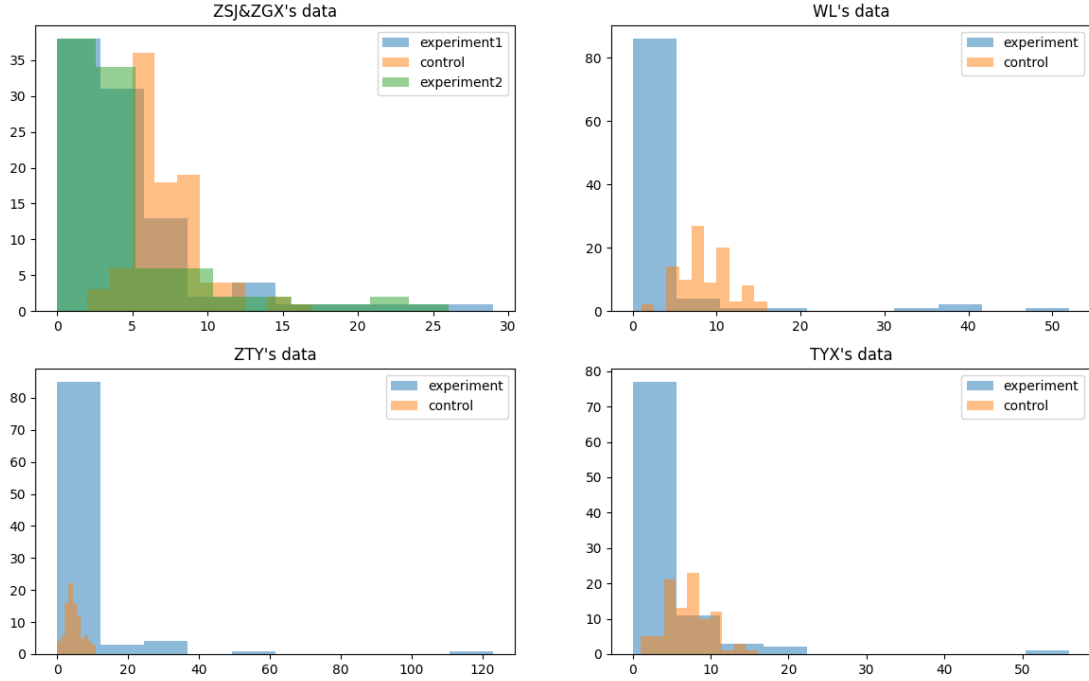


图 1: 各自数据的频数分布直方图

直观上可以看出，各个实验组突变数较低组数量多，但也存在一定的突变数大的数据点。对照组分布较为集中，数据的方差小。进一步计算：

实验组方差 var	73.085829	27.9551393	24.7346514	44.7071073	206.233137
实验组均值 mean	3.63541667	4.62365591	4.67741935	4.12765957	7.85106383
实验组方差/均值	20.1038383	6.04611153	5.28809789	10.8311033	26.2681774
实验组理论方差/均值	-0.6065288	2.76608571	2.12422974	-13.301805	-0.2882251
对照组方差 var	9.046875	6.048098046		8.91081938	4.83567225
对照组均值 mean	8.375	6.924731183		6.93617021	4.80851064
对照组方差/均值	1.08022388	0.873405463		1.28468868	1.00564865

实验组均值与方差的差异巨大，这符合突变假说的预期（有些 culture 中存在突变得比较早的酵母菌，在观测之前已经经历了许多代繁殖，导致该数据点数值很大，导致数据方差大）。对照组中均值与方差差距很小，近似于泊松分布的特点。

表中给出了实验组方差与均值的理论计算值 $\frac{var}{r} = \frac{N_t Ca}{\ln(N_t Ca)}$ ：

在某一时间点的平均新增的突变菌为 $d\rho = aN_t d\tau$ ，其方差为 $var_{d\rho} = aN_t e^\tau d\tau$ ，将 τ 从 t_0 积分到 t ，得到 $var_\rho = aN_t(e^{t-t_0} - 1)d\tau$ ，将之前得到的 $t - t_0 = \ln(N_t Ca)$ 带入，得到 $var_r = Ca^2 N_t^2$ ，再带入 $r = aN_t \ln(N_t Ca)$ ，得到 $\frac{var}{r} = \frac{N_t Ca}{\ln(N_t Ca)}$ 。

关于 var/r : 实际数据比计算值大

从理论层面上分析, 实际方差比理论预测值大的原因:

- a. 有个别 culture 中第一个突变菌出现得很早, 导致大数据点的产生, 使得方差显著增大。 t_0 的引入只是平均意义上的, 并不能完全消除这种情况带来的实际影响。
- b. 理论模型简单地假设了“突变率恒定不变”。如果酵母菌在快速分裂期新陈代谢旺盛, 突变率高, 接近平台期时突变率降低, 那么在早期就发生分裂, 突变数大的 culture 就更容易出现, 从而增大方差。

在实际操作中, 有若干可能引入误差, 导致方差增大:

- a. 向深孔板中加样时, 每个深孔板板中的液体体积不一 (虽然是移液枪加样理论上不会出现很大的偏差, 但可能是由于操作的原因, 实际上还是比较明显地观察到了这一点);
- b. 在向深孔板中加酵母菌时, 没有及时将培养液摇匀导致每个深孔板中的 N_0 存在差异;
- c. 从深孔板中取出酵母菌时, 酵母菌发生沉底等因素导致的取样方差;
- d. 在涂平板时, 液体溅出或者洒到其他邻近区域;
- e. 计数时, 有部分幸存的酵母菌由于距离很近连成一个菌落, 或因加样时戳坏了培养基导致其中的菌落不好辨认, 以及人在计数时犯错带来的误差;
- f. N_t 的测量动作时产生的误差, 这可能带来不小的偏差, 本人测量 N_t 时发现对于同样的样品, 细胞计数仪每次的读数很不稳定。

对于取样误差产生的方差, 文中也提到了其理论修正方法:

- (1) 平均的抗性菌数 r 乘以采样体积占总样品体积的比例 $\frac{\text{volume of culture}}{\text{volume of sample}}$
- (2) 由上文提到的方法 1 ($p_0 = e^{-m}$) 计算突变率
- (3) 由 $\text{var}_r = Ca^2 N_t^2$ 计算 r 的方差
- (4) 将 (3) 中得到的方差除以 $(\frac{\text{volume of culture}}{\text{volume of sample}})^2$ 得到由采样产生的方差

采样方差	WL	ZSJ&ZGX1	ZSJ&ZGX2	TYX	ZTY
	0.000577	0.00394	0.00403	0.000589	0.000101

理论上可对实验数据中直接计算的方差减去此部分, 即完成对采样动作产生的方差的修正, 但如此计算得到的修正幅度过小, 几乎可以忽略不计。

对照组 r 的均值普遍大于实验组

可能的原因:

- a. 虽然所用的培养液相同, 二者的所处环境有差异 (大锥形瓶和深孔板), 导致某些物理因素产生变化, 酵母菌的突变速率可能因此有所差别。
- b. 在零时刻, 对照组中 N_0 数量巨大, 已经有了突变体。

3.3 实验数据相关性分析

此部分分析所用数据: ZSJ&ZGX 的 2 个实验组和 1 个对照组。

假设酵母菌的突变符合获得性假说，当外界环境压力来临时，样品中的酵母菌随机存活。对照组的所有数据来自于同一个 culture，从此数据截取 93 个（与实验组等长），进行 5000 次打乱，模拟从同一个 culture 中取样 5001 次，涂布在选择性培养基上，共获得 5001 组数据。

计算原始数据与这 5000 组数据的相关性，得到皮尔逊相关系数分布：

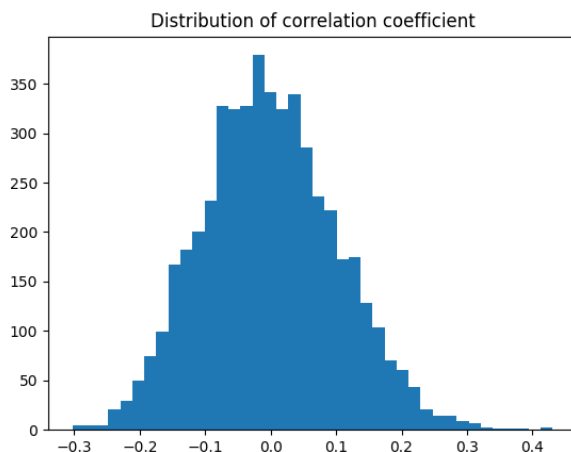


图 2: 相关系数分布统计

其 99% 的置信区间为：(-0.23562349455171097, 0.27819661632575043)

两个实验组数据的皮尔逊相关系数为 0.742，落在拒绝域，拒绝原假设，否定了获得性假说的预测。而从同一个 culture 中取出的两个实验组数据有显著的相关性，符合突变假说的推演。

3.4 计算机模拟

根据实验条件设定相关参数，在计算机中模拟此实验。

以与实验完全相同的参数模拟出的抗性菌数目频数分布直方图为（实验数据为 ZSJ&ZGX 的第一个实验组），simulation1, 2 的区别为突变率是采用上文所述的方法 1 还是 2 得到：

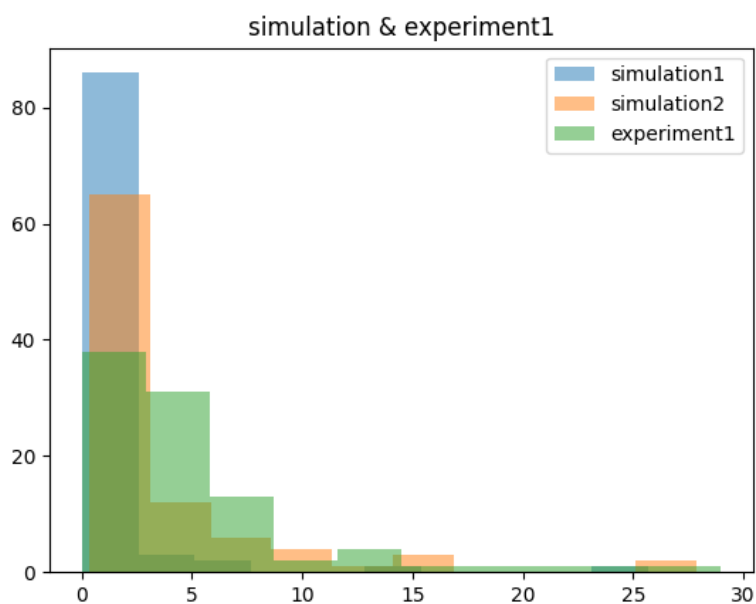


图 3: 模拟结果与实际数据对照

它们分布相似度很高（其相关系数将在下文分析），说明了此理论模型的正确性。给出相同条件下，根据获得性假说进行模拟，所预测的分布结果为：

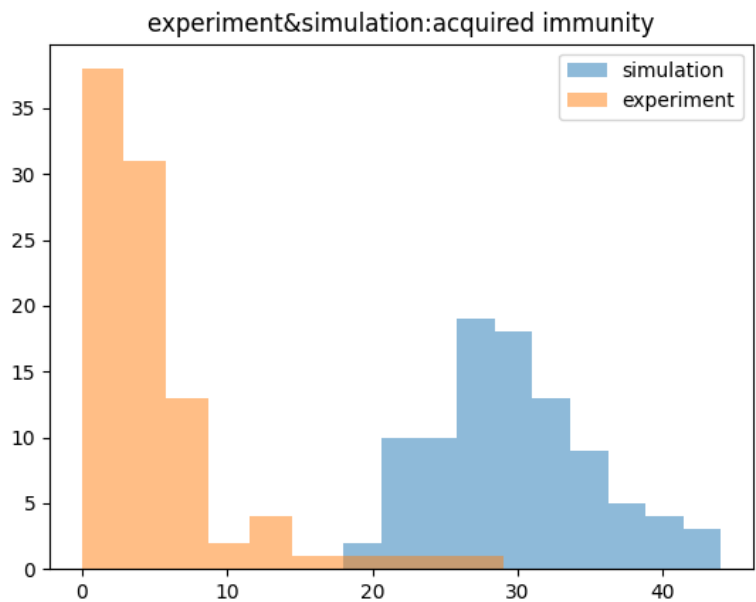


图 4: 获得性假说模拟结果与实际数据对照

看出二者分布差异巨大，与“实验数据相关性分析”中讨论的相似。

讨论

(1) 0 的数目

进行 500 次模拟统计，发现在 simulation2 中几乎从来不会出现 0，这和方法 2 得到的突变率比实际值偏高有关。对更为接近实际情况的 simulation1，平均每一次模拟有 8.94 个 0，实验数据为 8 个，说明实际上实验组中 0 的数目并没有显著偏小。

WL 数据中的 0 较多有 34 个，对其中 0 的情况进行同样的模拟分析，500 次模拟的平均结果为 36.53 个，实验组中 0 的数目也并没有显著偏小。

(2) t_0 的计算

t_0 是为了修正第 1 次突变发生时间较长而出现，实际上，它直接限制了长尾分布的范围。可以通过数据中最大的数字来推断从 t_0 到当前，经过了多少个细胞周期，从而反推出 t_0 。

以下是一个模拟举例 $N_0 = 200$, 繁殖 20 代, 300 组平行实验。

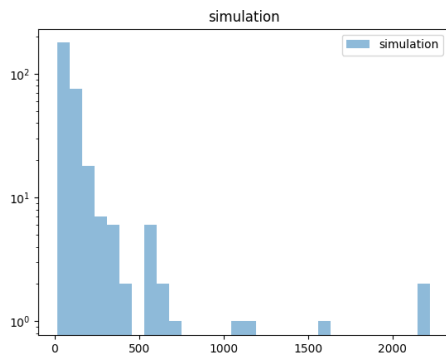


图 5: 模拟举例

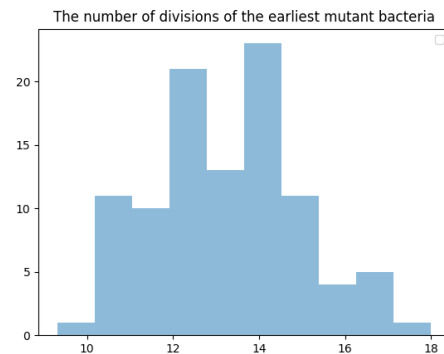


图 6: 最早突变的细菌到目前为止的分裂次数

可见最早突变的细菌到目前为止的分裂次数分布比较集中，方差较小为 2.62（这也验证了此方法估计 t_0 的可行性），平均值为 12.67，因此第一次突变大约发生在第 8 次分裂。

推算 ZSJ&ZGX 第一组实验中的 t_0 :

首先从 N_t/N_0 得到酵母菌分裂的总次数约为 17.54，用上述方法得出最早的突变细菌的分裂了约 9.17 次，那么第一个突变发生的时间约在第 8 次分裂。

(3) 平台期的影响

由于此次实验酵母菌培养时间较长，最终得到的酵母菌浓度较大，应已到达生长的平台期，细胞代谢并不如快速分裂期活跃，认为其突变率比其旺盛分裂时有所下降。

假设第 10 代开始，每一代突变率变成原来的 0.9 倍，其余参数与 ZSJ&ZGX 的第一个实验相同，并与实验数据比较。

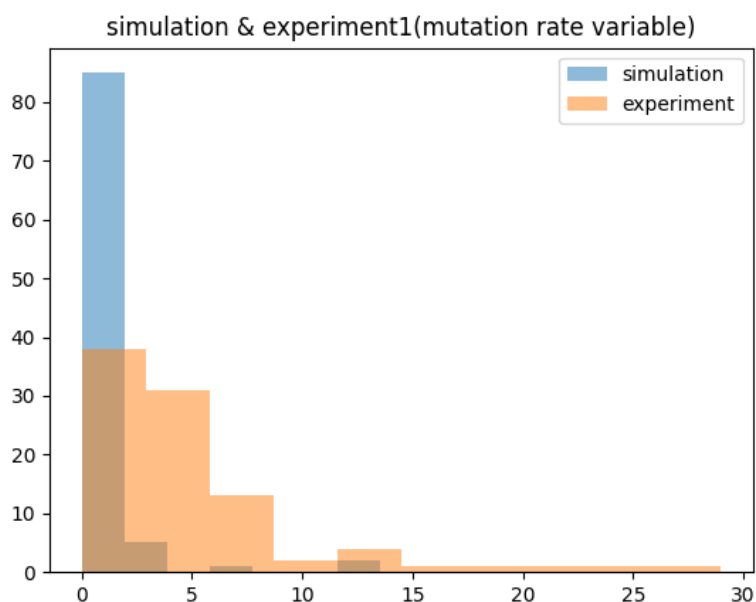


图 7: 考虑到平台期对突变率的影响

仅观察分布,难以体会其差异。计算 100 次模拟数据和实验数据的皮尔孙相关系数发现从 0.792 升高到 0.885,可见突变率的衰减符合真实情况。

4 结论

当菌落遇到外界压力时,单次试验结果不能区分突变假说和获得性假说的正确与否。但对于一系列相似的 culture,它们受到筛选后,关于突变菌的分布,两种理论给出的预测相差甚远。设计实验验证,发现结果符合突变假说的预言。

在细菌(以及本次实验所用的酵母菌)中,基因突变是在没有选择压力的情况下产生的,而不是对选择压力的反应。由达尔文的自然选择学说演绎出的随机突变的理论适用于细菌以及更复杂的生物。Max Delbrück 和 Salvador Luria 部分因这项工作获得了 1969 年诺贝尔生理学或医学奖。

5 致谢

衷心感谢助教学姐为我们准备相关实验用具并指导我们完成实验,为我们答疑解惑,让我们领略了定量模型的风采。

衷心感谢与我合作的周书珺同学,我们进行了许多讨论;也感谢本组其余同学,我从与他们交流中受益颇丰;特别感谢和我有很多交流,并一同感染奥密克戎的张天晔同学。

衷心感谢汤超老师实验室中所有相关人员提供良好的实验条件,让我们的实验在疫情笼罩下得以顺利进行。

6 参考文献

[1]Luria, S. E.; Delbrück, M. (1943). "Mutations of Bacteria from Virus Sensitivity to Virus Resistance". *Genetics*. 28 (6): 491–511.

实验代码整理后将被上传至：https://github.com/ElijahChang929/2022_ISP_Experiment