Master's Thesis

Time-Implicit Mutant Population Dynamics: Luria-Delbrück Fluctuation Test Revisited

Soohyun Kim

Department of Biomedical Engineering

Graduate School of UNIST

2018

Time-Implicit Mutant Population Dynamics: Luria-Delbrück Fluctuation Test Revisited

Soohyun Kim

Department of Biomedical Engineering

Graduate School of UNIST

Time-Implicit Mutant Population Dynamics: Luria-Delbrück Fluctuation Test Revisited

A thesis
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Master of Science

Soohyun Kim

01 / 03 / 2018 of submission

Approved by

Advisor
Cheol-Min Ghim

Time-Implicit Mutant Population Dynamics: Luria-Delbrück Fluctuation Test Revisited

Soohyun Kim

This certifies that the thesis/dissertation of Soohyun Kim is approved.

01 / 03 / 2018 of submission

signature

Advisor: Cheol-Min Ghim

signature

Dougu Nam: Thesis Committee Member #1

signature

Semin Lee: Thesis Committee Member #2

Abstract

We applied novel time-implicit dynamical models to reassess the classic Luria-Delbrück Fluctuation Test (LDFT), which has established the spontaneous and random nature of genetic mutations. Despite the profound impact of the study, full quantitative features of the seemingly simple mutational population dynamics are still under active investigation. Particularly, with a recent surge of interest in Bayesian inference, there has been renewed interest in Luria-Delbrück experiment in relation to a potential support for the controversial concept of "adaptive mutation", by which cells increase the mutation rate corresponding to environmental stresses. Though majority of the detailed analyses showed that the mutations are not intended in any adaptive way but occur without any regard to their potential benefits, there is a possibility that random spontaneous mutation works along with Lamarckian mechanisms as exemplified by the CRISPR-Cas immunity in bacteria. Here we employed an alternative mathematical analysis, where the size of the wild-type population takes in place of time. By eliminating time, first of all, the usual but unjustified assumption of exponential growth could be easily relaxed. Furthermore, our analysis re-discovers all the stochastic versions of the original Luria-Delbrück model, and the resulting statistics could be re-examined on a more quantitative basis, shedding light on the recent calls for the extension of the Modern Synthesis.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Random *versus* adaptive mutations

Mutation is the ultimate source of innovation in biological evolution. It has been well known that microbes rapidly adapt to lethal environments such as antibiotics, toxic chemicals or viruses by quickly acquiring mutations that make them resistant to those lethal agents. In the early 20th century, such remarkable adaptability was attributed to a Lamarckian explanation, whereby the environmental stress specifically induces relevant mutations. That is to say, bacterial resistance to the infecting bacteriophage was largely held to be induced by exposure to phage, which made microbiology the “last bastion of Lamarckism" [1]. An alternative but equally plausible mechanism was then emerging (neo-)Darwinian modality of evolution. Neo-Darwinian hypothesis is based on two arguments: (1) Natural selection is the mechanism of evolution. (2) Adaptive traits appear through spontaneous mutation continuously over time.

It was not until the 1940s that biologists discarded the view that adaptations were acquired in direct response to a change in an environment. Salvador Luria and Max Delbrück observed that the two competing hypotheses would result in different distributions in the number of phage-resistant bacteria. They conceived the so-called fluctuation test to answer this question using a population of bacteria subjected to viral infection. Through meticulous experiments and elaborate statistical analyses, they showed that [2] mutations are not induced by an environment but occur in random and spontaneous manner before the selective pressure drives the change. Together with the elegant statistical analysis, Luria-Delbrück fluctuation test (LDFT) became a paradigm that epitomize the conventional view of biological evolution, or Modern Synthesis [3], which led to their 1969 Nobel prize in Physiology or Medicine (shared with Alfred Hershey).

Modern Synthesis eventually emerged as Mendel's inheritance, Darwin's natural selection and population-level thinking was integrated in a way to provide the well-accepted conceptual scheme for evolutionary biology [4]. It is arguably the standard theory of evolution which motivated widespread acceptance of some core assumptions. Nevertheless, there has been a continuing debate on whether organisms exert physiological control over their mutation rates. Though the Lamarckian modality alone is hardly acceptable as a sound theory of evolution, recent calls for the extended evolutionary synthesis (EES) opens a new window for the "reinstatement" of Lamarckism [5].

In the celebrated 1943 experiment, a small number of *Escherichia coli* (less than 500) were inoculated into individual culture tubes. After growing the bacteria for a few generations,

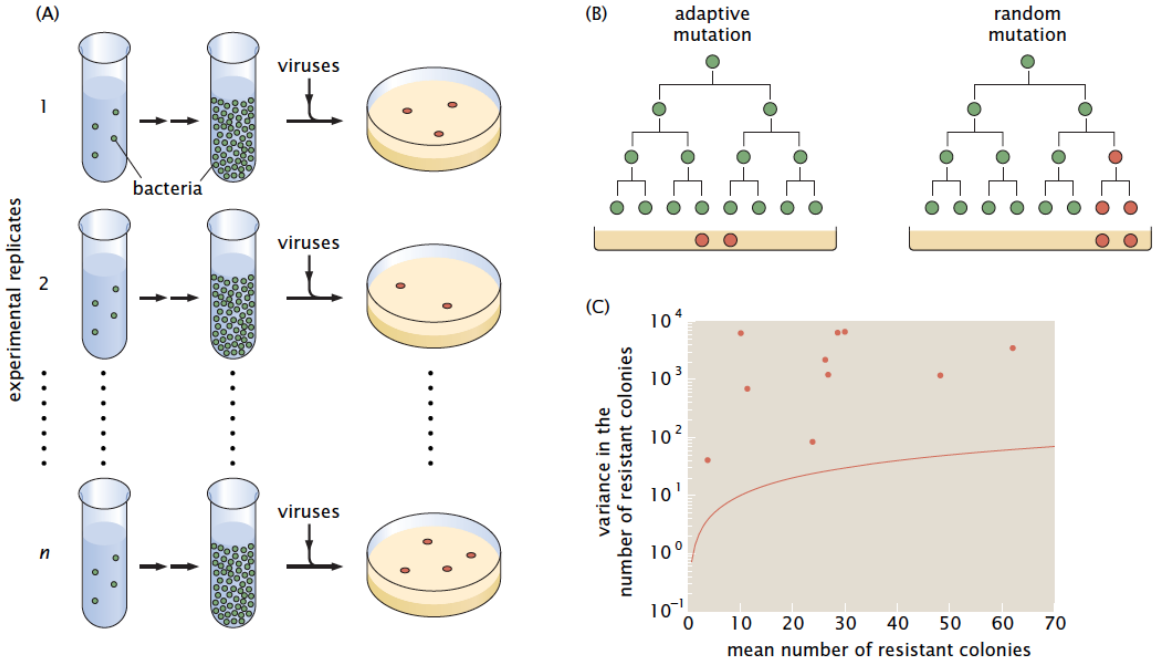

Fig. 1.1 Schematic of the Luria-Delbrück experiment excerpt from [6]. (A) Experiment starts with small number of bacteria (less than 500) grown for approximately 15 generations. Then bacteria are plated on plates that contain T1 bacteriophage. The growth of some colonies on the plate tells that phage-resistant bacteria exist. By repeating the experiment many times, both the mean number of mutants and the variance can be estimated. (B) Following the Lamarckian scenario (adaptive mutation hypothesis), the mutant bacteria (red) do not appear until bacteriophage is introduced. In contrast, according to the Darwinian scenario (random mutation hypothesis), mutation can occur ahead of bacteriophage introduction and propagate their resistance to their offspring. (C) The variance in the number of resistant colonies is shown as a function of the mean number of resistant colonies. The results show that the observed fluctuations are larger than those predicted by the Lamarckian scenario (adaptive mutation model) which predicts that the variance should be equal to the mean.

equal volumes of each culture were then plated on agar containing T1 phage viruses which would infect and kill the bacteria. In the Lamarckian scenario, where resistance to the virus is caused by induced activation in bacteria, mutations will occur only after phage infection and each plate should contain approximately the same number of phage-resistant bacteria. In other words, the distribution of the resistant (mutant) cells in each culture should be Poissonian and thus the variance in the number of resistant bacteria is more or less the same as mean.

On the other hand, in the Darwinian scenario, where resistance is due to heritable genetic components in bacteria, mutations will occur randomly prior to phage exposure and there should be a large variance in the number of phage-resistant bacteria. The two alternative scenarios in a toy model are shown in Fig. 1.2. From the results they discovered that the distribution of phage-resistant bacteria was rather heavy-tailed, therefore not consistent with the Poisson distribution which led to the conclusion that bacteria evolve through the Darwinian mechanism.

While the Luria-Delbrück experiment was successful in showing that mutations in bacteria are random rather than directed, it was only a semi-quantitative analysis they presented as

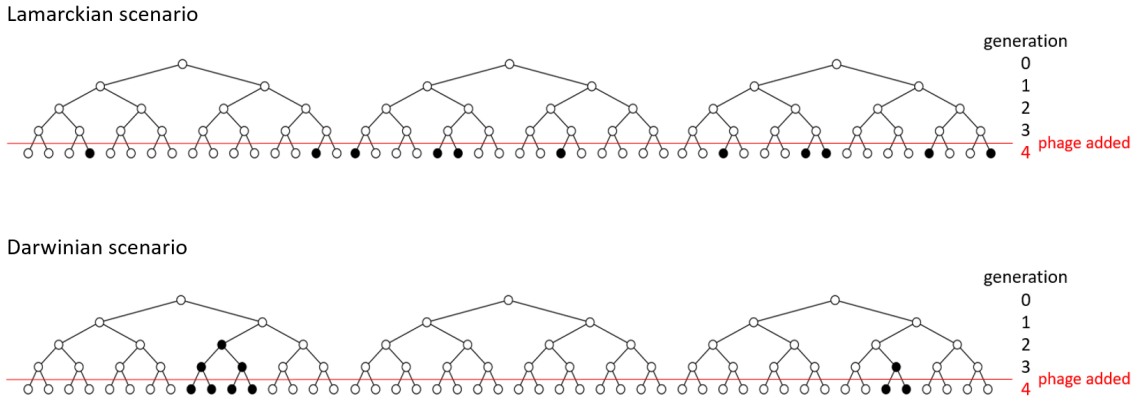

Fig. 1.2 Schematic of the two scenarios tested by the Luria-Delbrück fluctuation test. The black dots illustrate phage-resistant bacteria and the white dots are bacteria susceptible to the T1 phage. An experiment starts with a single bacterium which is at the top of each tree. The single bacterium grows a few number of generations and the T1-phage is then introduced into the plated culture (the fourth generation of each tree in the figure). The Lamarckian case predicts that mutations happen only after the phage is introduced. This will result in a standard deviation of the number of resistant bacteria that scales with the square root of their mean. In other words, there will be a low variance in the number of resistant bacteria among the trees. The Darwinian case predicts that mutations occur spontaneously. Therefore, a broad distribution of the number of resistant bacteria will be produced from a number of experiments, which means that the variance will be large compared to the Lamarckian case.

they were not able to quantify how much the data differs from the Lamarckian (Poissonian) model. Other concerns regarding the Luria-Delbrück experiment include that the model assumes deterministic growth of mutant bacteria and that the experiment considers only pure Darwinian and pure Lamarckian mechanisms. To resolve such concerns and to correctly estimate the mutation rate based on Luria-Delbrück fluctuation tests we discuss more details in Sections 1.2 and 1.3. Nonetheless, the LDFT is meaningful in that the experiment showed that exquisite statistical analysis can be capable of revealing biological processes which cannot be easily discovered through real observation.

1.2 Review of the previous studies

Luria and Delbrück assumed in their model that both the mutant bacteria and normal bacteria grow deterministically while mutations occur randomly. According to the proportion of cultures without mutants and the mean number of mutants they estimated the mutation rate. Furthermore, Luria and Delbrück used an exponential function $e^{\beta(t-t')}$ as a continuous approximation for the number of mutant bacteria at time t, resulting from a mutation which occurred at an earlier time t' where β is the bacteria growth rate. As a result, the model induces continuous mutant distribution. The Luria-Delbrück model was considered incomplete due to such exponential function since the number of mutant bacteria resulting from an earlier mutation would be generally small.

Since the number of normal bacteria was underestimated compared to the number of mutations by time t in the Luria-Delbrück formulation, Lea and Coulson came up with a different model in 1949, where the normal bacteria grow deterministically while the mutant bacteria grow stochastically [7]. The exponential function $e^{\beta(t-t')}$ in the Luria-Delbrück model is replaced by the growth rate b, also interpreted as the relative fitness of mutant bacteria ($0 < b < \infty$) compared to a normal cell of fitness 1. In this model the relative fitness is unity, which means that a same growth rate is assumed for both mutants and normal bacteria. The result is a probability generating function (PGF) in the form of $\exp[m(1/z-1)\log(1-z)]$ where m is the expected number of mutations per plate and z is a dummy variable for defining a generating function ($0 \leq z \leq 1$). Such PGF is utilized for obtaining an approximate value of the mean and variance, which in turn is used for estimating the probability distribution of the mutant bacteria. After 3 years in 1952, Armitage came up with an idea that not every mutant existing prior to plating will be capable of growing a colony after plating [8]. Such idea taken into account in his work by defining plating efficiency ε, which represents the probability that a mutant grows a colony after a plating process.

Table 1.1 Significance, key ideas and the representative characteristics of the previous models. LD stands for Luria-Delbrück; LC, Lea-Coulson; MK, Mandelbrot and Koch.

Model	Normal bacteria growth	Mutant bacteria growth	Key Ideas
LD	Deterministic	Deterministic	Exponential mutant growth
LC	Deterministic	Stochastic	Relative fitness
Armitage	Stochastic	Stochastic	Plating efficiency
Bartlett	Stochastic	Stochastic	Yule's birth process
MK	Stochastic	Stochastic	Differential growth rates

In another three years, Bartlett suggested a fully stochastic formulation by utilizing Yule's birth process on the growth of both normal bacteria and mutants [9]. Assuming that both the mutant bacteria and normal bacteria grow stochastically allowed one to derive an exact PGF by considering the population size of both the mutant bacteria and normal bacteria at time t. This fully stochastic formulation was later improved by Mandelbrot (1974) and Koch (1982) by allowing differential growth rates between normal bacteria and mutant bacteria [10, 11]. Studies on the formulation of Luria-Delbrück distribution introduced in this section are summarized in Table 1.1.

1.3 Luria-Delbrück distribution and beyond

Mutations are an essential element of the genetic heritage of living organisms allowing life to evolve and adapt to new environments. In the context of evolutionary biology for phage-resistant bacteria, it was shown through the Luria-Delbrück experiment that bacteria follows Darwinian modality of evolution. However, the results from this well known "classic experiment" has been continuously challenged as acquired resistance was discovered in the CRISPR-Cas bacterial immune system. Such discovery was a clear evidence of a bacteria following the Lamarckian modality of evolution. Additionally, it was pointed out that induced or directed mutations and undirected mutations cannot be distinguished in the Luria-Delbrück experiment. Such stress can actually increase the undirected mutation rates which is also a Lamarckian mechanism as more mutations will occur at the time of stress input [12]. Furthermore, in the view of extended evolutionary synthesis where the theory is organism centered, the physical development process directs the generation of variation along guided pathways, the form and traits of organisms are shaped by the environment and in turn the organisms modify environmental conditions in a way that it enforces biases on future generations [5].

More recently, Holmes quantitatively analyzed the original data from the Luria-Delbrück experiment to compare and evaluate the performance of three different models: Darwinian, Lamarckian and Combined models using the Bayesian model selection [13, 14]. Based on their analysis of the fits of the three models (Lamarckian, Darwinian and Combined), they estimated a nonzero Lamarckian component combined with Darwinian mutation and thus concluded that the Lamarckian contribution to phage-resistance should not be ruled out [15].

However, almost all of the analytic computations done in the researches subsequent to Luria-Delbrück assumed exponential growth (either stochastic or deterministic) of both normal and phage-resistant bacteria. Technically, for estimating the statistical traits of the number of resistant bacteria, the number of wild-type bacteria must be expressed explicitly as a function of time. From a simple idea that an initial number of N_0 wild-type cells grows to a final number of N, Houchmandzadeh managed to simplify the formulation of Luria-Delbrück distribution by using the wild-type population (WTP) size n as an independent variable to replace real time [16].

Luria and Delbrück discovered that the evolution of phage-resistant bacteria is Darwinian only after the real experiment was carried out. Similarly, Holmes was able to claim that the Lamarckian mechanism should not be ruled out by analyzing the quantitative models. Utilizing Houchmandzadeh's general treatment for the population dynamics in terms of WTP and quantitatively analyzing the resulting outcomes, we aim to re-evaluate the performance of the pure-Darwinian model. Additionally, we set up a Combined model by convolving Darwinian and Lamarckian to further compare the performance with the previous two models as only pure-Darwinian was assumed throughout the Houchmandzadeh formulation.

Chapter 2

Theory

2.1 Population dynamics with explicit time-dependence

2.1.1 The Luria-Delbrück model

We begin with listing the fundamental assumptions of the Luria-Delbrück formulation for future reference, as the subsequent formulations introduced in the following are basically variations of the Luria-Delbrück model.

1. At time $t = 0$, the process begins with a single normal bacterium and no phage-resistant bacteria.

2. At a constant rate β_1, normal bacteria deterministically grow and the number of normal bacteria at time t is
$$N(t) = e^{\beta_1 t} \quad . \tag{2.1}$$

3. At a constant rate β_2, mutant bacteria grow deterministically. Hence, if a mutant is spawned by normal bacteria at time $s > 0$, the clone of this mutant will be of size $e^{\beta_2(t-s)}$ for $t \geq s$.

4. Mutations occur in accordance with a Poisson process with an intensity function such as,
$$\nu(t) = \mu e^{\beta_1 t} \tag{2.2}$$
where μ is the mutation rate per cell per unit time. Hence, the expected number of mutations in the time interval $[0, t)$ is
$$m(t) = \int_0^t \nu(s)\, ds = \frac{\mu}{\beta_1}(e^{\beta_1 t} - 1) \tag{2.3}$$

Denoting $X(t)$ as the number of mutant bacteria existing at time t and from the assumption 1–4, $X(t)$ can be mathematically expressed as

$$X(t) = \begin{cases} 0, & M_t = 0 \\ \sum_{i=1}^{M_t} \exp\{\beta_2(t-\tau_i)\}, & M_t \geq 1 \end{cases} \tag{2.4}$$

where M_t is the mutations process which is fundamentally a Poisson process with the intensity function in Eq.(2.2) and τ_i is the ith epoch at which mutations occur.

As a result of Campbell's theorem, the nth cumulant of $X(t)$ becomes

$$\begin{aligned} \kappa_n(t) &= \int_0^t \mu e^{\beta_1 s} e^{n\beta_2(t-s)} \mathrm{d}s \\ &= \begin{cases} \dfrac{\mu}{\beta_1 - n\beta_2}(e^{\beta_1 t} - e^{n\beta_2 t}), & (\beta_1 \neq n\beta_2) \\ \mu t e^{n\beta_2 t}, & (\beta_1 = n\beta_2) \end{cases} \end{aligned} \tag{2.5}$$

The special case $\beta_1 = n\beta_2$ is well known as it was solved first by Armitage [8]. The mean and the variance of $X(t)$ for such case are

$$\mathrm{E}[X(t)] = \mu t e^{\beta_1 t} \tag{2.6}$$

$$\mathrm{Var}[X(t)] = \frac{\mu}{\beta_1} e^{\beta_1 t}(e^{\beta_1 t} - 1) \tag{2.7}$$

and by letting $\beta_1 = 1$ these two identities were first derived by Luria and Delbrück [2]. Later, Koch worked on the differential growth case where $\beta_1 \neq \beta_2$. Assuming $\beta_1 = 1$ and defining $\beta_2 = b$, Koch derived the mean and variance respectively in the following form

$$\mathrm{E}[X(t)] = \frac{\mu N(t)(1 - e^{-(1-b)t})}{(1-b)} \qquad (b \neq 1) \tag{2.8}$$

$$\mathrm{Var}[X(t)] = \begin{cases} \mu N(t)(e^{(2b-1)t} - 1)/(2b-1) & (b \neq 0.5) \\ \mu t N(t) & (b = 0.5) \end{cases} \tag{2.9}$$

where the case of $b = 0.5$ is from the work of Zheng [17].

2.1.2 The Holmes model

Based on Haldane's modeling hypotheses, Holmes model estimates the mutation rate with slightly more detailed assumptions that are listed as follows,

1. Normal bacteria and mutant bacteria have the same fitness prior to bacteriophage introduction.
2. All bacteria go through synchronous divisions.
3. No bacteria die before the phage is introduced.
4. Mutation occurs only during synchronous divisions with each of the offspring independently turning into a mutant (Darwinian) or only after bacteriophage introduction (Lamarckian).

Pure Lamarckian

Denoting N_0 as the initial number of wild-type bacteria which are susceptible to phage and and g as the total number of epochs before bacteriophage introduction, the total number of bacteria N after the final level of divisions is,

$$N = 2^g N_0 \tag{2.10}$$

and the total number of bacteria that have existed is $2N - N_0$. Now having θ_L for the probability of a Lamarckian mutation after phage introduction and decreasing the probability of an additional mutation occuring in the post-resistant progeny, the mean number of Lamarckian mutations at epoch g is

$$m_L = \theta_L N \tag{2.11}$$

and finally the number of resistant bacteria in the Lamarckian model is a Poisson distribution with such parameter m_L

$$P_L(k|\theta_L, N_0) = \frac{e^{-m_L} m_L^k}{k!} \tag{2.12}$$

Pure Darwinian

For the Darwinian model there can be a number of ways to estimate a certain number of mutant bacteria k in the population of size N prior to bacteriophage introduction. As an example, in order to obtain five resistant bacteria ($k = 5$): (i) One mutation can occur two generations prior to the bacteriophage introduction (with the total living population $N/4$ at that generation) or (ii) Two mutations can occur one generation prior to introducing T1-phage, and an additional mutation in the final generation. Example (i) is a case with probability

$$P_5^{(i)} = (1 - \theta_D)^{(2N - N_0) - 8} \theta_D^2 \binom{N/4}{1} \binom{N-4}{1} \tag{2.13}$$

where $(2N-N_0)-8$ is the total number of bacteria that have been active without mutating and θ_D^2 is a factor that indicates a total of two mutations have happened in the population. The factor first to choose is the number of independent mutational chances two generations before the bacteriophage is introduced and the second factor to choose is the number of mutational chances during the last generation. Similarly, the case illustrated in Example (ii) is less likely to happen with probability

$$P_5^{(ii)} = (1-\theta_D)^{(2N-N_0)-7}\theta_D^3 \binom{N/2}{2}\binom{N-4}{1} \tag{2.14}$$

where a total of 7 mutants have ever lived and three mutational events occurred prior to phage introduction. Following Haldane's approach [18], we write the probability of having k resistant bacteria in the Darwinian model as

$$P_D(k|\theta_D,N_0) = \sum_{\{a_s\}\in\prod_K} (1-\theta_D)^{(2N-N_0)-\sum_{s=0}^{\infty} a_s(2^{s+1}-1)}\theta_D^x \times \prod_{s=0}^{\infty}\binom{2^{-s}N-\sum_{n=s+1}^{\infty} a_s\cdot 2^{n-s}}{a_s}, \tag{2.15}$$

where $x \equiv \sum_{\{a_s\}}$ and $P_D(k|\theta_D,N_0)$ are summed over all sequences $\{a_s\} \in \prod_K$ which produces the number k.

Combined Model

Finally for the Combined model, both the Darwinian and Lamarckian mechanisms contribute when generating resistant bacteria. Hence the probability in the Combined model is a convolution

$$P_C(k|\theta_L,\theta_D,N_0) = \sum_{k'=0}^{k} P_D(k'|\theta_D,N_0)P_L(k-k'|\theta_L,N_0). \tag{2.16}$$

where k denotes the number of resistant bacteria in the Combined model and k' is the number of survivors in the Darwinian case.

2.2 Implicit-time formulation

The quantitative features of the mutagenesis in microbial populations and the resulting statistics over the numerous experimental cultures could be obtained by an explicit solution to the population dynamics. To this end, the simplifying assumption of the constant exponential growth had been employed for most of the previous analyses of Luria-Delbrċk experiments.

Despite efforts to extend the exponential growth to more general Monod or Gompertz growth [19, 20], the generality of those approaches are still limited and restrictive.

Recently, Houchmandzadeh observed that using real-time as an independent variable is not mandatory in terms of which the evolving population is described. Instead, came a mathematical formulation of population dynamics of asexual populations, where the population size of the wild-type cells is the independent variable, playing the role of time [16]. If we use the size of the wild-type population (WTP) as the proxy of time, we do not have to assume any specific growth dynamics, such as Monod, Gompertz, Richards, or Stannard [21].

■ Model I: Deterministic population growth

When a wild-type cell divides, regardless of the duration of the cell cycle, there is a small probability μ that a mutant daughter cell will appear. Assuming "no death" in a culture of wild-type bacteria growing from the initial size N_0 to size N, we denote X_n as the random variable that describes the number of the newborn mutant cells while WTP stays at n.

In case of the deterministic growth of a cell, probability comes into play only for the time point at which a cell division gives rise to a mutant cell. Thus, letting Γ_n^N be the contribution of a mutant cell to the final number of the mutant cells M_N when WTP size reaches N,

$$M_N = \sum_{n=N_0}^{N} \Gamma_n^N X_n \,. \tag{2.17}$$

For large N, such derivation is valid (which is the general case in experiments dealing with bacteria). Since the occurrences X_n of the mutant cells are independent random variables, the moment generating functions (MGF) and the cumulant generating function (CGF) is respectively given by

$$\mathscr{M}(s) = \langle e^{sM_N} \rangle = \prod_{n=N_0}^{N} \langle e^{s\Gamma_n^N X_n} \rangle \tag{2.18a}$$

$$\mathscr{K}(s) = \log\langle e^{sM_N} \rangle = \sum_{n=N_0}^{N} \log\langle e^{s\Gamma_n^N X_n} \rangle \tag{2.18b}$$

Now from $\langle e^{s\Gamma_n^N X_n} \rangle = 1 - \mu + \mu e^{sN/n}$, where μ denotes the mutation probability, the CGF can be expressed as

$$\begin{aligned} \mathscr{K}(s) &= \sum_{n=N_0}^{N} \log(1 - \mu + \mu e^{sN/n}) \\ &= N \int_{x_0}^{1} \log(1 - \mu + \mu e^{sx})\, \mathrm{d}x \end{aligned} \tag{2.19}$$

where we use the continuous approximation for the sum, $x = n/N$ and $x_0 = N_0/N$ in the second expression. Expanding Eq. (2.19) to the first order in μ and restricting the domain to $s \lesssim -x_0 \log \mu$,

$$\mathscr{K}(s) = -\theta\phi + \theta \int_{x_0}^{1} e^{s/x} \mathrm{d}x + O(\mu^2) \,, \tag{2.20}$$

where $\theta = N\mu$ and $\phi = 1 - x_0$. Hence the mean κ_1 and the variance κ_2 are

$$\kappa_1 = \mu N \log(N/N_0) \tag{2.21a}$$

$$\kappa_2 = \mu N \left(\frac{N}{N_0} - 1 \right) \,, \tag{2.21b}$$

which coincides with the expression originally derived by Luria and Delbrück for the exponential growth of bacteria.

To derive the probability distribution, it is more convenient to compute the PGF defined as

$$\mathscr{G}(s) = \langle z^X \rangle = e^{\mathscr{K}(s)} = e^{-\mu(N-N_0)+O(\mu^2)} \exp\left[\int_{x_0}^{1} e^{s/x} \mathrm{d}x \right] \tag{2.22}$$

$$\simeq e^{-\mu(N-N_0)} \Big[e^s - x_0 e^{s/x_0} + s\big(\Gamma(0,-s) - \Gamma(0,-s/x_0)\big) \Big], \tag{2.23}$$

where $\Gamma(0,\cdot)$ is the upper incomplete gamma function. Once PGF is computed with respect to s, the inverse z-transform gives the probability distribution for the number of mutant cells via

$$P_D(k) = \frac{1}{k!} \left(\frac{\mathrm{d}}{\mathrm{d}s} \right)^k \Bigg|_{s=0} \mathscr{G}(s). \tag{2.24}$$

Now considering the case where wild-type and mutant bacteria have different growth rates respectively denoted as

$$\frac{\mathrm{d}}{\mathrm{d}t} n(t) = \alpha(n,t) n(t) \tag{2.25a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} m(t) = c\alpha(n,t) m(t) \tag{2.25b}$$

where the constant c is the ratio of the instantaneous growth rate of mutant cells to that of wild type cells. While the growth rate is not specified in any particular form, we assume that the mutant bacteria follow the same law as the wild-type bacteria within a constant multiplicative factor.

Eliminating time from the relations (2.25a) and (2.25b), $\mathrm{d}m/\mathrm{d}n = c(m/n)$. A resistant bacteria from one copy when the population size is n will contribute $\Gamma_n^N = (N/n)^c$ to the final

number of mutant bacteria. Through the same derivation process introduced in the previous subsection, we obtain

$$\psi(s) = N \int_{x_0}^{1} \log(1 - \mu + \mu e^{s/x^c}) dx \tag{2.26}$$

and leaving only the leading term in x_0 and μ for $c \neq \frac{1}{p}$, the pth cumulant coefficient is given by

$$\kappa_p = \frac{\theta}{cp-1}\left(x_0^{1-cp} - 1\right) . \tag{2.27}$$

■ Model II: Stochastic population growth

The contribution of the mutant that appears in WTP size n to the final population N was denoted as $\Gamma_n^N X_n$ where the deterministic propagator Γ_n^N was assumed as N/n for the deterministic case and the MGF of $\Gamma_n^N X_n$ was expressed as $\langle e^{s\Gamma_n^N X_n} \rangle = 1 - \mu + \mu e^{sN/n}$.

Now for a stochastic propagator Γ_n^N, the MGF is $\langle e^{s\Gamma_n^N X_n} \rangle = 1 - \mu + \mu e^{s\Gamma_n^N}$ since X_n can be only 0 or 1. Following the same logic as in 2.19 and 2.20, the first two cumulant coefficients are

$$\kappa_1 = \mu \sum_{n=N_0}^{N} \langle\, \Gamma_n^N \,\rangle \,, \tag{2.28a}$$

$$\kappa_2 = \mu \sum_{n=N_0}^{N} \langle\, (\Gamma_n^N)^2 \,\rangle - \mu^2 \sum_{n=N_0}^{N} \langle\, \Gamma_n^N \,\rangle^2 \,. \tag{2.28b}$$

Then expressing the second moment of Γ_n^N as a function of its mean and variance V_n^N we have

$$\langle\, (\Gamma_n^N)^2 \,\rangle = V_n^N + \langle\, \Gamma_n^N \,\rangle^2 \,. \tag{2.29}$$

which gives

$$\kappa_2 = \mu(1-\mu) \sum_{n=N_0}^{N} \langle\, \Gamma_n^N \,\rangle^2 + \mu \sum_{n=N_0}^{N} V_n^N \,. \tag{2.30}$$

The first term on the right-hand side of 2.30 is what we already derived for the deterministic growth case. The following term is the contribution of the stochasticity of Γ_n^N on the variance of the number of mutants.

■ Time-implicit formulation of the Combined Model

According to the Combined Model, the probability distribution for the number of mutant cells is given by the convolution of the two independent contributions, i.e. the Poisson distribution

from Lamarckian mechanism and that of Darwinian mechanism given by Eq. (2.24). That is,

$$P(k) = e^{-\lambda} \sum_{k'=0}^{k} P_D(k') \frac{\lambda^{k-k'}}{(k-k')!} , \tag{2.31}$$

where λ is the mean that is to be determined by regression.

Chapter 3

Numerical Methods

3.1 Statistical inference on the origin of mutation

For a quantitative evaluation of the Lamarckian contribution to mutagenesis, we consider the Combined Model. The probability of the number of mutant bacteria is estimated by the convolution of the two independent contributions, i.e. the Poisson distribution from Lamarckian mechanism and that of Darwinian mechanism given by Eq. (2.24). Since

$$P(k) = e^{-\lambda} \sum_{k'=0}^{k} P_D(k') \frac{\lambda^{k-k'}}{(k-k')!} = e^{-\lambda} \sum_{k'=0}^{k} \frac{\lambda^{k-k'}}{(k-k')! \cdot k'!} \left(\frac{\mathrm{d}}{\mathrm{d}s}\right)^k \Bigg|_{s=0} \mathscr{G}(s) \,, \tag{3.1}$$

it suffices to compute the PGF with various requirements. In order to numerically calculate such probabilities, Eq. (3.1) is programmed in a simulation model. Within Eq. (3.1), the only linear term of θ is for $m_k = 1$ and $m_{i<k} = 0$. Hence, for $\theta \ll 1$, the probabilities reduce to a generalized form of

$$P(0) = e^{-\theta} \tag{3.2}$$

$$P(k) = e^{-\theta} \frac{\theta}{k(k+1)} \text{ for } k > 0. \tag{3.3}$$

Explicit formulation for the probabilities in such linear birth models has been profoundly studied by a number of researchers and thoroughly reviewed by Zheng [17]. Note that the Eq. (3.2) and Eq. (3.3) are already known for the constant linear birth model. Recently, Kessler and Levine [22, 23] found an explicit expression of the mutant population distribution in terms of Landau distribution in both of the deterministic and stochastic growth model.

It is not of a practical use to obtain the mutant population distribution from experiments due to the fact that the probabilities require many sets of parallel cultures. It is assumed in

the simulation that with a mean number of 135, each culture is initially made of a Poisson distribution of bacteria number similar to the classical LDFT paper [2]. The bacterial cells are modeled to divide in distinct and discrete 21 generations. Taking into account the already known final cell density and growth rate, such numbers above can be easily inferred. We make assumptions of synchronous division of the cells and that each of the offsprings are prone to experience a mutation at the time of division with probability θ_D. Such probability is nonzero in Combined and Darwinian models. Offsprings of the mutant bacteria are considered resistant and a bacteriophage that promotes Lamarckian mechanism with the probability of θ_L is introduced to the nonresistant cells in the final generation. Note that the total number of mutated cells through Lamarckian modality is a Poisson distribution with the mean θ_L multiplied by the number of wild type cells.

In order to reduce the simulation time of the simulations, the total number of bacteria that ever existed is $N_t = 2N_0 2^g - N_0$. Therefore, the total number of mutation attempts through Darwinian modality is a Poisson distribution with mean $N_t\theta_D$. Then with a single Poisson draw, we estimate the number of such mutations and in turn distribute the mutations over the tree of bacteria in a random manner. Note that every daughter of the mutated bacteria is marked as "mutated". Now for the correction of overestimating the probability of mutations resulting from the decrease in mutant population size of each generation when there existed pre-mutated bacteria, with the chance that is equal to the ratio of mutated bacteria when the mutation occurred to the total number of bacteria in a generation of interest, we randomly remove the original mutations. This approach can be more efficient comparing to the simulation of the mutations performed one generation at a time as such removal of the original mutations is not common and the mutation events itself does not occur very often.

To compute $P_C(k_p|\theta_D, \theta_L)$ on a group of the ordered pair of (θ_D, θ_L), we perform numerous runs of simulation starting with a seed population of bacterial cells which are Poisson-distributed. Then, we plate the culture of just a fraction in silico and measure the number of bacteria cultures which survive in each simulation run. From that, we now can estimate $P_C(k_p|\theta_D, \theta_L)$ which is the normalized frequency of k_p.

3.2 Moment generating functions

The generating functions are an auxiliary means of expressing the probability distribution, which provides the basis of an alternative method to analytical results. In particular, compared with the probability density functions or cumulative distribution functions, the moment generating function (MGF) defined by the weighted sums of random variables, provides anaytical results for the cases even when analytical results are impractical.

A single mutant that appeared when the WTP size is n expands to the size K_n^N when the wild-type population reaches N. The chemical master equation for the probabilities of the propagator $\Pr(K_n^N = m)$ is then given by the master equation

$$\frac{\partial P(m,N)}{\partial t} = T_{m,m-1}P(m-1,N) - T_{m+1,m}P(m,N) \ , \tag{3.4}$$

where $T_{m,m-1}$ is the transition rate or propensity of the event where the mutant cell population increases from $m-1$ to m. Therefore, the MGF of the propagator $\varphi(s,N)$ follows

$$N\frac{\partial \varphi}{\partial N} + (1-e^s)\frac{\partial \varphi}{\partial s} = 0 \tag{3.5}$$

with the boundary conditions

$$\varphi(s,N=n) = e^s \tag{3.6}$$

$$\varphi(s=0,N) = 1 \tag{3.7}$$

The equation above is a linear first-order partial differential equation that can be solved by the method of characteristics. That is,

$$\varphi(s,N) = \frac{(n/N)e^s}{(n/N-1)e^s+1} \tag{3.8}$$

Now we change the notation to denote N as the wild-type population size of the mutant occurrence, and $\mathscr{N}$ the final size of the wild-type population, we obtain

$$\left\langle e^s K_N^{\mathscr{N}} \right\rangle = \frac{Ne^s}{(N-\mathscr{N})e^s + \mathscr{N}} \ . \tag{3.9}$$

If the relative growth rate of the mutant is given by $c \neq 1$, the transition probability for the mutant once it has appeared is

$$T(m \to m+1) = c\frac{m}{N} \tag{3.10}$$

and $\varphi(s,n)$ follows

$$\frac{1}{c}\frac{\partial \varphi}{\partial N} + (1-e^s))\frac{\partial \varphi}{\partial s} = 0 \tag{3.11}$$

This equation can be re-written as Eq. (3.8) by a simple scaling $N \to N^c$ for the ratio of the growth rates of wild-type and mutant cells. MGF is therefore given by

$$\varphi(s,N) = \frac{(n/N)^c e^s}{[(n/N)^c - 1]e^s + 1} \ . \tag{3.12}$$

wherefrom the CGF of the number of mutants can be deduced.

Chapter 4

Summary

The origin of mutation is not merely a matter of academic debate. How the mutation arises can be asked by clinicians who are fighting against the antibiotic resistance crisis. The Luria-Delbrück experiment confirmed the validity of the conceptual framework of the Modern Synthesis. Their data and analyses clearly falsified the claim that bacteria became only resistant after being exposed to the bacteriophage and legitimately supported that the virus-resistant mutations could be described by a constant probability of occurring in each cell division.

While LDFT provides clear evidence for the random pre-exposure mutations, it does not rule out the possibility that mutation occurs in response to the exposure to the selective environment. To get over the limitations in which the lethal agent (e.g. canavanine) wipes out parent cells prior to the chance of adaption, LDFTs modified in a manner that selective medium grants the survival of parent cells, but not to generate colonies without acquiring a new growth-permitting mutation. Those tests discovered the mutations that are post-exposure by rigorous criteria [24]. It is now well established that the induced mutations are generally possible by the stress imposed by the selective medium. For an example, bacteria exposed under stress has a possibility of adopting a hypermutable state in a way of expressing error susceptible DNA polymerases and DNA repair systems. Furthermore, the competence of the conceptual scheme of Modern Synthesis to keep up with the rapid change in developmental biology, genomics and ecology has been continuously questioned.

With the advancement in the study of developmental plasticity, inclusive inheritance, and niche construction, there has been the proposal of Extended Evolutionary Synthesis (EES), differentiating the role of constructive processes in development and evolution. The developmental processes, operating through niche construction, inclusive inheritance and developmental bias, share responsibility for the rate and direction of evolution, and the origin of character variation.

For example, under the stress of estrogen destitution, breast cancer cell lines with pre-exposure mutations conferring genetic instability can acquire new mutations that facilitates the growth in low estrogen [25]. Stress-induced mutations are an adaptive response to the stress, but there is still no convincing evidence that the mutations are specifically targeted to a group of genes that is responsible for the growth in a negatively selective media. To qualify for "Lamarckian", the mutations should be specifically directed to genes that can relieve the stress. Thus this example still points to "Darwinian" evolution.

Another kind of heritable change is the epigenetic modifications, which are chromatin marks altering the gene expression with no involvement of DNA sequence changes. Along with the reversible modifications, epigenetic changes are often adaptive and targeted to a specific gene(s). In those cases, what have evolved is not the individual responses but the cellular pathways that mediate the responses, and the mediating pathways have evolved over time by Darwinian variation and natural selection, which is the point of debate.

In this thesis, we explored the possibility of including induced mutagenesis as a new component of the evolutionary synthesis. Throughout analytical and numerical approaches, we find that the "Lamarckian" component may not be ruled out.

References

[1] H. F. Judson. Activating esr1 mutations in hormone-resistant metastatic breast cancer. *Nature Genetics*, 45:1446–1451, 2013.

[2] S. E. Luria and M. Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28:491–511, 1943.

[3] Vassiliki Betty Smocovitis. *Unifying Biology: The Evolutionary Synthesis and Evolutionary Biology*. Princeton University Press, 1996.

[4] Ernst Mayr. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Harvard University Press, 1982.

[5] K. Laland et al. Does evolutionary theory need a rethink? *Nature*, 514:161–4, 2014.

[6] Rob Phillips, Jane Kondev, Julie Theriot, and Hernan Garcia. *Physical Biology of the Cell*. Garland Science, 2012.

[7] D. E. Lea and C. A. Coulson. The distribution of the numbers of mutants in bacterial populations. *Genetics*, 49:264, 1949.

[8] P. Armitage. The statistical theory of bacterial populations subject to mutation. *J. Royal Statist. Soc. B*, 14:34–40, 1952.

[9] M. S. Bartlett. *An introduction to stochastic processes*. Cambridge University Press, 1955.

[10] B. Mandelbrot. A population birth-and-mutation process i: Explicit distributions for the number of mutants in an old culture of bacteria. *J. Appl. Prob*, 11:437, 1974.

[11] A. L. Koch. Mutation and growth rates from luria-delbrück fluctuation tests. *Mutation Res.*, 95:129–143, 1982.

[12] R. Barrangou et al. Crispr provides acquired resistance against viruses in prokaryotes. *Science*, 315:1709–1712, 2007.

[13] D. J. MacKay. Bayesian interpolation. *Neural Comput.*, 4:415–447, 1992.

[14] D. J. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[15] C. M. Holmes et al. Luria-delbrück, revisited: the classic experiment does not rule out lamarckian evolution. *Phys Biol*, 14, 2017.

[16] B. Houchmandzadeh. General formulation of luria-delbrück distribution of the number of mutants. *Phys. Rev. E*, 92, 2015.

[17] Q. Zheng. Progress of a half century in the study of the luria-delbrück distribution. *Math. Biosci*, 162:1–32, 1999.

[18] S. Sarkar. Haldane's solution of the luria-delbrück distribution. *Genetics*, 127:257, 1991.

[19] S. H. Moolhavkar A. Dewanji, E. G. Luebeck. A generalized luria-delbrück model. *Math. Biosci*, 197:140–152, 2005.

[20] J. Monod. The growth of bacterial cultures. *Annu. Rev. Microbiol.*, 3:371–394, 1949.

[21] M. H. Zwietering et al. The growth of bacterial cultures. *Appl. Environ. Microbiol.*, 56:1875–1881, 1990.

[22] H. Levine D. A. Kessler. Large population solution of the stochastic luria–delbrück evolution model. *PNAS*, 110:11682, 2013.

[23] H. Levine D. A. Kessler. Scaling solution in the large population limit of the general asymmetric stochastic luria–delbrück evolution process. *Journal of Statistical Physics*, 158:783–805, 2015.

[24] S. M. Rosenberg. Evolving responsively: adaptive mutation. *Nature Reviews Genetics*, 2:504–515, 2001.

[25] D. R. Robinson et al. Activating esr1 mutations in hormone-resistant metastatic breast cancer. *Nature Genetics*, 45:1446–1451, 2013.

Acknowledgements

I would like to acknowledge my advisor, Professor Cheol-Min Ghim for his thoughtful guidance and unconditional support not only throughout my master's course but also my undergraduate years. It was an honor to work and study with professor Ghim. I would also like to thank the thesis committee members, Professor Dougu Nam and Professor Semin Lee for giving precious comments and feedbacks. My family members, Dad, Mom, Grandmother, Jintae, and Jiyoung has given me the uttermost support. They believed in me and encouraged me to keep up with my research. If it were not them, I would not have been able to successfully finish the course. Thank you and I love you. I want to mention Eunhee Kim, Soobin Kim, Jiwon Choi, Junseon Min and Yongin Choi who has been great friends and I truly enjoyed all the time we shared together. I also want to mention Hyun-A Lee. Outside campus, she has been a true mentor and more than a friend. It was a great pleasure to work with the PBBP group members, Hyunmo Yang and Juneil Jang. I wish you all the best on your current and future studies. At last, I would like to especially thank Samuel for his heartful advice and encouragement on this work.