Large Deviations Analysis for Stochastic Models of Bacterial Evolution

Robert Azencott,* Brett Geiger† Ilya Timofeyev‡

August 30, 2024

Abstract

Radical shifts in the genetic composition of large cell populations are rare events with quite low probabilities, which direct numerical simulations generally fail to evaluate accurately. In this paper, we develop a theoretical large deviations framework for a class of Markov chains modeling the genetic evolution of bacteria such as E. coli. In particular, we develop the cost function for discrete-time Markov chains which describe the daily evolution of histograms of bacterial populations. We also develop explicit formulas that can be used to numerically quantify the most likely evolutionary trajectories connecting an initial histogram and the target histogram.

AMS Classification: 60F10, 60J20, 92D15

1 Introduction

We focus this theoretical study on stochastic models for the genetic evolution of bacterial populations. Stochastic modeling of evolutionary dynamics has been an active area of research for several decades (e.g. [6, 17, 26, 27, 37]) and we utilize tools from stochastic analysis to study rare events that correspond to the emergence of non-dominant genotypes in long-term evolutionary experiments. In particular, we recast a large class of such models as discrete-time Markov chains in the large-dimensional space of population histograms for bacteria with multiple genotypes. These models, sometimes called “locked-box” models, have been developed to describe in the context of laboratory experiments, the daily evolution of finite-size bacterial populations [23, 30, 39]. On day $n$, the bacterial population is characterized by the histogram $H_n$ of all genotype frequencies. Due to daily selection, in typical laboratory contexts, the size $N$ of the population remains roughly constant. Bacterial genotypes are characterized by their fitness, or equivalently by their growth factors, which remain fixed throughout the evolution of the bacterial colony. Random mutations are assumed to be roughly Poissonian and to occur at a fixed very small rate $m$. For a mutating cell of genotype $i$ the probability $q_{i,j}$ of generating a cell of genotype $j$ is assumed to be constant in time. Recall that many estimation techniques based on intensive simulations and experimental data have been implemented and tested (see [13, 23, 24, 31, 33, 39]) to evaluate the mutation rate $m$ and the growth factors (or selective advantages) of genotypes of interest.

*University of Houston, Mathematics Dept, razencot@math.uh.edu
†High Point University, Dept of Mathematical Sciences, bgeiger@highpoint.edu
‡University of Houston, Mathematics Dept, ilya@math.uh.edu
Our analysis was initially motivated by long-term laboratory experiments on the genetic evolution of bacteria such as *Escherichia coli*. In these experiments (see [11,12,19,23,29,36,38]), on day \( n \), the current cell population has size \( N \) and grows freely until nutrient exhaustion. One then extracts (by dilution for instance) a random sample of approximately \( N \) cells, which constitutes the next-day population. Genetic evolution models for such experiments implement a succession of “daily” cycles comprising three steps: growth phase, mutations, and random selection of a sub-sample of fixed size \( N \).

In this context, our study develops a rigorous mathematical *large deviation theory* for the analysis of rare genetic events, with an emphasis on the rare “fixation” of a genotype \( j \) which does not have maximum fitness. Recall that fixation is said to occur when the frequency of such a genotype \( j \) becomes unusually large. Fixation events have probabilities vanishing at exponential rates for large population sizes \( N \). The computation of these exponential rates involve large deviation functionals defined on the path space of all sequences of population histograms. We actually prove fairly explicit theoretical formulas for these large deviations functionals, enabling the numerical analysis of fixation events as well as the numerical computation of the most likely evolutionary histogram trajectory linking an initial population histogram \( H \) to a terminal histogram \( G \). The present paper deliberately focuses on proving the theoretical large deviation results and the associated explicit formulas needed for pragmatic evaluation of large deviations functionals.

But extensive numerical results based on our theoretical developments are presented in our companion paper [34], where we discuss the numerical implementation of our algorithms for computing the most likely evolutionary trajectory linking two histograms \( H \) and \( G \), and where we explore how to handle numerical complexity when the number of genotypes increases.

**Rare events and genetic evolution.** Large deviations results have been obtained for the trajectories of a wide range of vector-valued stochastic processes (see, for example, [1–3, 15, 20, 35]), including Markov chains, Gaussian processes, stochastic differential equations with a small diffusion coefficient, etc. Here we extend the large deviations approach to discrete-time Markov chains modeling genetic evolution of bacterial populations.

The numerical applicability of rare events analysis has not often been exploited in concrete models of cell populations. Previous large deviations results for stochastic population evolution have involved theoretical asymptotic studies such as [7,10]. These papers have studied very general asexual population evolution in the space of phenotypic traits vectors. In such models, Darwinian evolution of asexual populations is driven by birth and death rates, which are themselves dependent on phenotypic traits. These traits are approximately transmitted to offspring with rare but important variations due to gene mutations. Competition for limited resources forces a permanent or roughly periodic selection. Mutations are assumed to be rare enough so that most of the time, only one currently dominant trait vector can coexist with the trait vectors of emerging mutants.

Evolutionary models considered here are quite different because they combine two random steps - mutations and dilution. So observed rare events can be triggered by unexpected combinations of these two steps, and it is important to understand the joint impact of these steps for experiments on bacterial evolutionary dynamics [1,16,22,23,28].

Here, we focus our large deviations study on developing rigorous and computationally implementable formulas to quantify rare genetic events for a discrete Markov chain which describes the genetic evolution of bacterial populations in the space of population histograms. In that space, we develop an explicit “rate function” defined for all paths connecting an initial histogram \( G \) to
a target histogram $H$ in a fixed number of time steps. Minimizing the rate function over the set $S(G,H)$ of all histograms paths linking $G$ to $H$ then yields formulas to quantify the log-probability of rare events, for instance when the target histogram $G$ describes the "fixation" of a non-dominant genotype. We also develop an explicit second-order reverse recurrence equation satisfied by optimal paths minimizing the rate function over $S(G,H)$. These recurrence equations essentially solve the optimization problem of finding the most likely path maximizing the probability of fixation for any specific genotype.

Manuscript Organization. In Section 2 we describe the Markov chain dynamics that serves as the stochastic model of bacterial evolution. In Section 3, we establish a large deviations framework for these Markovian successions of daily cycles, in order to derive an explicit formula for the one-step cost function given by (3.60) in Theorem 3.24. These results are applied in Section 4 to formulate a large deviations theory for evolutionary trajectories taking values in the space of population histograms. The main application of our large deviations framework is considered in Section 5 where we develop theoretical results for computing the most likely path connecting an initial population state $G$ to a fixed given terminal state. We conclude our main results in Section 6 by deriving an explicit reverse recurrence relation which can be used to numerically compute the most likely path connecting the initial $G$ and the final $H$ in the space of histograms.

2 Stochastic model for bacterial evolution experiments

To model the main features of random bacterial evolutions, we focus on a class of Markov chains often used in this context [23,30,39]. We assume that all possible cell genotypes belong to a finite set $\{1,2,\ldots,g\}$. Cells of genotype $j$ are called $j$-cells here and have a fixed growth factor $F_j > 0$. We always order genotypes by increasing fitness so that $F_1 < F_2 < \ldots < F_g$. The genotype $g$ with the highest fitness is called dominant. We also denote the vector of growth factors as $F := [F_1, \ldots, F_g]$.

When a $j$-cell divides, its genotype is typically inherited by the two daughter cells, unless a rare random genotype mutation from $j$ to $k \neq j$ occurs. Mutations occurrences approximately have a Poisson distribution with a very small mutation rate, $m$, with typical values in the range $10^{-9} \leq m \leq 10^{-6}$.

Definition 2.1. When a mutation occurs during the division of a given $j$-cell, the conditional probability that a mutant daughter cell will be a $k$-cell is denoted $q_{j,k} \geq 0$. The transition mutation matrix $Q \in \mathbb{R}^g \times \mathbb{R}^g$ with entries $q_{j,k} \geq 0$ naturally verifies $\sum_k q_{j,k} = 1$ and $q_{j,j} = 0$.

Definition 2.2. The population size $N$ and the process parameters $\mathcal{P} = \{m,g,F_1,\ldots,F_g,Q\}$ define our stochastic evolutionary model for finite-size bacterial populations.

In all proofs throughout the manuscript, we assume that these parameters are fixed. We will also assume a large population size $N \gg 1$ and a small mutation rate $m \ll 1$ and utilize expansions to obtain the leading-order terms in $N$ and $m$ in some of the proofs. For typical bacterial populations experiments, one has $m < 10^{-6}$ and $N > 10^5$.

We detail below the three successive phases (growth, mutation, random selection) implementing each daily evolutionary cycle and we define the space of population histograms which quantify the concentrations of $j$-cells for each $j \in \{1,\ldots,g\}$. Then, we will describe the Markov transition kernel associated with each daily cycle.
2.1 The space of genetic histograms

Each bacterial population is described by a histogram of genotype frequencies

\[ H := [H(1), \ldots, H(g)]. \]

We often denote

\[ H_n := [H_n(1), \ldots, H_n(g)] \]

as the histogram of bacterial frequencies on the \( n \)-th day. Next, we introduce some basic definitions.

Definition 2.3. A matrix \( A \) will be called \( N \)-rational if for a positive integer \( N \) all the coefficients of \( NA \) are non-negative integers.

Definition 2.4. Denote \( \mathcal{H} = \{ H \in \mathbb{R}^g \} \) the set of all possible population histograms. Each histogram \( H \) is a vectors of length \( g \) such that \( 0 \leq H(j) \leq 1 \) and \( \sum_j H(j) = 1 \). Note that \( \mathcal{H} \subseteq \mathbb{R}^g \) is compact and convex.

Definition 2.5. The subset \( \mathcal{H}_N \) of \( \mathcal{H} \) is the set of \( N \)-rational vectors \( H \in \mathcal{H} \). Note that \( \mathcal{H}_N \) is finite, with \( card(\mathcal{H}_N) \leq (N + 1)^g \).

In a cell population of size \( N \), denote \( N_j \) the number of \( j \)-cells, and \( H(j) = N_j/N \) their concentration. This ”genetic histogram” \( H = [H(1), \ldots, H(g)] \) then belongs to \( \mathcal{H}_N \).

The sets \( \mathcal{H} \) and \( \mathcal{H}_N \) are endowed with the \( L_\infty \)-distance

\[ \| H - G \| = \max_j |H(j) - G(j)| \text{ for all } H, G \in \mathcal{H}. \]

Definition 2.6. The ”boundary” of \( \mathcal{H} \) is the set of histograms \( H \) for which at least one \( H(j) = 0 \). To quantify closeness to the boundary for any \( H \in \mathcal{H} \) we define the support \( supp(H) \) and the essential minimum \( b(H) > 0 \) by

\[ supp(H) = \{ j \mid H(j) > 0 \} \quad \text{and} \quad b(H) = \min_{j \in supp(H)} H(j). \]  

Note that \( b(H) \geq 1/N \) for all \( H \in \mathcal{H}_N \). Genetic evolution during \( T \) days is then described by the histogram trajectory \( \mathbf{H} := \{ H_1, H_2, \ldots, H_T \} \), where \( H_n \in \mathcal{H}_N \) is the population histogram on day \( n \) for \( 1 \leq n \leq T \). We now present the three phases of each ”daily” cycle.

Our ”one-day” time unit refers to the fixed duration of each single evolutionary cycle. In laboratory experiments, single-cycle durations are roughly constant but may be shorter than 24 hours in real time. In the T.Cooper experiments for instance (e.g. [12]) the growth phase typically takes 8 – 12 hours, after which the nutrients are exhausted and cells become dormant.

2.2 Path Space of histograms sequences

Genetic evolution is modeled as a sequence of daily cycles indexed by day \( n \). To simplify our evolution model, we assume that all daily random mutations occur nearly simultaneously, after the growth phase. One can develop analogous models where random mutations can occur at any time during the growth phase (see for instance [32]); but these models involve stochastic differential equations driven by Poisson processes, so that rigorous large deviations frameworks similar to the theory developed here require more complicated mathematical proofs. We consider that the simplified models studied here are sufficient for quantitative analysis of genetic evolutionary pathways in realistic experimental setups.

We now describe the stochastic dynamics of each daily cycle. At the beginning of day \( n \), the population \( pop_n \) is always of size \( N \) and is identified by its genetic histogram \( H_n \in \mathbb{R}^g \) recording
the frequencies $H_n(j)$ of $j$ cells in population $pop_n$. The day $n$ cycle initiated by $pop_n$ generates $pop_{n+1}$ (and hence its genetic histogram $H_{n+1}$) in three successive phases.

**Phase 1:** purely deterministic growth, with growth factor $F_j$ for the $j$-cells colony.

**Phase 2:** random mutations occur simultaneously at the end of the growth phase.

**Phase 3:** after growth and mutations, a random sub-sample of fixed size $N$ is extracted from the current population, and will constitute $pop_{n+1}$.

### 2.3 Phase 1 - Deterministic growth

In actual experiments on bacterial evolution (see e.g. [11,23]), the daily multiplicative growth factors $F_j$ for observable genotypes are typically in the range $[20, 300]$. These growth factors can be computed as $F_j = \exp(\Delta \times a_j)$, where $\Delta$ is the duration of the growth phase and $a_j > 0$ is the "selective advantage" of $j$-cells. Detectable selective advantages over the ancestor genotype are typically larger than 0.01.

On day $n$, the initial population $pop_n$ has genetic histogram $H_n$, and the $j$-cell colony has initial size $NH_n(j)$. The $j$-cell colony then grows according to its growth factor, $F_j$, and reaches the size $NF_jH_n(j)$. At the end of the growth phase, the population reaches a much larger size $\lceil N \langle F, H \rangle \rceil = \lceil \sum_j NF_jH(j) \rceil$ where $\langle F, H \rangle$ is the inner product of vectors $F, H$ and $\lceil u \rceil$ is the smallest integer $\geq u$. The population genetic histogram after growth, $\varphi$, is then given by $\varphi_j(H) = \lceil NF_jH(j) \rceil / \lceil N \langle F, H \rangle \rceil$. We naturally approximate $\varphi(H)$ by $\Phi(H) \in \mathcal{H}$, which we define as

$$\Phi_j(H) = F_jH(j) / \langle F, H \rangle. \quad (2.2)$$

For $N > 20$ and all $H \in \mathcal{H}$, the approximation error is

$$\|\Phi(H) - \varphi(H)\| \leq \frac{4F_g}{NF_1}. \quad (2.3)$$

### 2.4 Phase 2 - Random mutations

As outlined earlier, in our simplified model, we consider that on day $n$ all random mutations occur simultaneously right after the growth phase.

#### 2.4.1 Random Mutation matrices and mutation rates

**Definition 2.7.** During the mutation phase of day $n$, a random number $R_n(j, k)$ of $j$-cells mutate into $k$ cells. All the $R_n(j, k)$ are non-positive integers and $R_n(j, j) = 0$ for all $j$. Denote $R_n$ the corresponding $g \times g$ random matrix.

On day $n$, at the end of the growth phase, each $j$-cells colony has just reached the finite size $F_jNH_n(j)$. This forces the random numbers $R_n(j, k)$ of $j$-cells mutating into another genotype $k \neq j$ to verify a set of natural linear inequalities which we explicitly develop below, before describing technically the joint probability distribution of the $R_n(j, k)$.

**Definition 2.8.** The mean emergence rate $m$ of mutants is assumed to have the same very small value for all $j$-cell colonies. For bacterial populations, one typically has $10^{-9} \leq m \leq 10^{-6}$. Recall that $Q = \{q_{j,k}\}$ is the fixed conditional mutation transition matrix defined in (2.1). So for $N$ large, we want the conditional expectation $E(R_n(j, k)|H_n)$ to be very close to $mF_jNH_n(j)q(j,k)$.
2.4.2 Mutations constraint set $K(H)$

Given $H_n$, the total number of mutants $\sum_k R_n(j,k)$ emerging from the $j$-cell colony must be smaller than $NF_jH_n(j)$, which is the number of $j$-cells after growth. This imposes $g$ linear constraints on $R_n$, namely

$$\sum_k R_n(j,k) < NF_jH_n(j) \text{ whenever } H_n(j) > 0$$

Hence the random matrix $r_n = R_n/N$ must belong to a convex set of matrices $K(H_n)$ which we now describe.

**Definition 2.9** (Constraint Sets). Let $Z$ be the set of all $N$-rational $g \times g$ matrices, and note that $R_n/N \in Z$. For each histogram $H$ and each $j$, define $K(j,H)$ as the set of all $g \times g$ matrices $r$ with non-negative coefficients verifying for all $j,k$

$$\begin{cases} \sum_{k=1}^g r_{j,k} < F_jH(j) & \text{if } H(j) > 0; \\ r_{j,k} = 0, & \text{if } H(j) = 0 \text{ or } q_{j,k} = 0. \end{cases} \quad (2.4)$$

Then define the convex set of matrices $K(H) = \bigcap_{j=1}^g K(j,H)$ and let $K_N(H) = Z \cap K(H)$ be the subset of all $N$-rational matrices in $K(H)$.

The matrices $r \in K(H)$ are potential values for all the random matrices $r_n = R/N$, which are $N$-rational and must, therefore, belong to $K(H_n)$.

Note that $K_N(H) \subset K(H)$ by definition. Due to (2.4), we have $r_{j,k} \leq F_j$ and $\text{card}(K_N(H)) \leq [(1 + F_j)(N + 1)]^g$ for all $H \in H$, $r \in K(H)$, $(j,k)$, and $N$. The next technical lemma quantifies how accurately $K_N(H)$ approximates $K(H)$ for $N$ large.

**Lemma 2.10.** Fix $a > 0$. For $N > \frac{a^2}{\delta^2}$, any $H \in H$ with $b(H) \geq a$, and any $r \in K(H)$, there is an $N$-rational matrix $s = \{s_{j,k}\}$ such that

$$s \in K_N(H) \subset K(H), \quad \text{supp}(s) = \text{supp}(r), \quad ||s - r|| \leq g/N. \quad (2.5)$$

**Proof.** See Appendix [B] \hfill \square

We will build our stochastic mutations model to enforce $P(R_n/N \in K(H) \mid H_n = H) = P(R_n/N \in K_N(H) \mid H_n = H) = 1$. We also want to ensure that for large $N$, the conditional joint distribution of the $R_n(j,k)$ given $H_n$ will become extremely close to a product of independent Poisson distributions with means $mq_{j,k}NF_jH_n(j)$. To this end, we first introduce companion matrices $Z_n$.

2.4.3 The Poisson companion matrices $Z$

Given $H_n = H$, let $Z$ be a matrix of independent random variables $Z_{j,k}$ having Poisson distributions with respective means $\mathbb{E}[Z_{j,k} \mid H_n = H] = mq_{j,k}NF_jH(j)$. We will show that the conditional probability $P(Z/N \in K(H) \mid H_n = H)$ tends to 1 very fast for large $N$ (see Theorem [3.4]).

So, for all histograms $H$ and matrices $z$ with non-negative coefficients, we define the conditional distribution of $R_n$ given $H_n = H$ by $P(R_n/N = z \mid H_n = H) = P(Z/N = z \mid Z/N \in K(H))$. Since $Z/N$ is $N$-rational, this forces $R_n/N$ to be $N$-rational with

$$P(R_n/N \in K_N(H) \mid H_n = H) = P(R_n/N \in K(H) \mid H_n = H) = 1.$$
We now complete the analysis of the mutations phase by computing the population histogram at the end of Phase 2.

2.4.4 Population histogram after mutations

Accounting for all the random mutations that occurred during phase 2, each \( j \)-cell colony has lost \( \sum_k R_n(j, k) \) outgoing mutants and gained \( \sum_k R_n(k, j) \) incoming mutants. But the total population size has not changed since the end of phase 1 and is given by \( \lceil N\langle F, H_n \rangle \rceil \). Hence the random number of \( j \)-cells at the end of Phase 2 is given by

\[
\lceil N\langle F, H_n \rangle \rceil - \sum_k R_n(j, k) + \sum_k R_n(k, j).
\]

(2.6)

Note that \( R_n(k, j) = 0 \) for all \( j \notin \text{supp}(H) \). Hence, given \( H_n = H \in \mathcal{H}_N \) and \( R_n/N = r \in K_N(H) \), the population histogram \( J_n \) at the end of Phase 2 is a deterministic function \( J_n = \mathcal{J}(H_n, R_n/N, N) \) of \( H_n, R_n/N \) and \( N \). The histogram-valued function \( (H, r, N) \rightarrow \mathcal{J}(H, r, N) \) does not depend on \( n \) and is defined for all \( H \in \mathcal{H} \) and all \( r \in K(H) \). We can rewrite (2.6) to get

\[
J_n(j) = \mathcal{J}_j(H, r, N) = \frac{\lfloor NF_jH(j) \rfloor}{N\langle F, H \rangle} + \frac{N}{N\langle F, H \rangle} \left[ -\sum_k r_{j,k} + \sum_k r_{k,j} \right].
\]

(2.7)

One has \( \frac{\lfloor NF_jH(j) \rfloor}{N\langle F, H \rangle} - \frac{F_jH(j)}{\langle F, H \rangle} \leq \frac{13F_j}{NF_1} \) due to (2.3). Therefore, given \( H_n = H \in \mathcal{H}_N \) and \( R_n/N = r \in K_N(H) \), the histogram \( J_n(j) = \mathcal{J}_j(H, r, N) \) is well-approximated by

\[
\Psi_j(H, r) = \frac{1}{\langle F, H \rangle} \left( F_jH(j) - \sum_k r_{j,k} + \sum_k r_{k,j} \right).
\]

(2.8)

The function \( \Psi_j(H, r) \) is well-defined for all \( H \in \mathcal{H} \) and \( r \in K(H) \). Moreover, one has \( \Psi_j(H, r) \geq 0 \) and \( \sum_j \Psi_j(H, r) = 1 \), so that \( \Psi(H, r) \) is always a histogram belonging to \( \mathcal{H} \).

One can directly verify that when \( N > 20 \),

\[
\|\mathcal{J}(H, r, N) - \Psi(H, r)\| \leq \frac{13F_j}{NF_1}
\]

(2.9)

for all \( H \in \mathcal{H} \) and \( r \in K(H) \). In particular, one has with probability 1

\[
\|J_n - \Psi(H_n, R_n/N)\| \leq \frac{13F_j}{NF_1}.
\]

(2.10)

Since the histogram \( J_n \) is \( \tilde{N} \)-rational with \( \tilde{N} = \lfloor N\langle F, H \rangle \rfloor \), its essential minimum \( b(J_n) \) must verify

\[
b(J_n) \geq \frac{1}{\tilde{N}} \geq \frac{1}{N\langle F, H_n \rangle} \geq \frac{1}{NF_j}.
\]

(2.11)

For \( H \in \mathcal{H} \) and \( r \in K(H) \), equations (2.4) and (2.8) yield for each \( j \)

\[
\Psi_j(H, r) = 0 \text{ if and only if } H(j) = r_{j,k} = r_{k,j} = 0 \text{ for all } k.
\]

(2.12)

The relations (2.6) and \( R_n/N \in K_N(H_n) \) imply \( J_n(j) = 0 \) if and only if \( H_n(j) = R_n(k, j) = 0 \) for all \( k \). Similarly, for all \( N \)-rational \( H \) and \( r \in K(H) \) equation (2.7) shows that \( \mathcal{J}_j(H, r, N) = 0 \).
if and only if \( H(j) = r_{k,j} = 0 \) for all \( k \). Hence, one has \( \text{supp}(J_n) = \text{supp}(\Psi(H_n, R_n/N)) \) for all \( H_n \) and \( R_n/N \in K(H_n) \). In addition, \( \text{supp}(J(H, r, N)) = \text{supp}(\Psi(H, r)) \) for all \( N \)-rational \( H \) and \( r \in K(H) \). Equation (2.12) proves that for \( H, H' \in \mathcal{H}, r \in K(H), \) and \( r' \in K(H') \),

\[
\text{supp}(\Psi(H, r)) = \text{supp}(\Psi(H', r')) \quad \text{whenever} \quad \text{supp}(H) = \text{supp}(H') \quad \text{and} \quad \text{supp}(r) = \text{supp}(r').
\]

(2.13)

We conclude the study of Phase 2 with a couple of relations that will be useful later on. For all \( H \) and \( r \in K(H) \), the partial derivatives of \( \Psi(H, r) \) are given by

\[
\begin{align*}
\frac{\partial}{\partial H_{ij}(r)} \Psi_j &= -\frac{F_j F_i H(j)}{(F, H)} + \frac{1\{i=j\} F_j}{(F, H)}, \\
\frac{\partial}{\partial r_{ij}} \Psi_j &= -\frac{F_j H(j)}{(F, H)}, \\
\frac{\partial}{\partial r_{k,j}} \Psi_j &= \frac{F_j H(j)}{(F, H)},
\end{align*}
\]

(2.14)

where the indicator \( 1\{i=j\} \) equals 1 if \( i = j \) and 0 otherwise. Since \( F_j H(j) \leq (F, H) \) and \( (F, H) \geq F_1 \), we have

\[
\| \Psi(H', r') - \Psi(H, r) \| \leq \frac{3gF_g}{F_1} (\| H' - H \| + \| r' - r \|)
\]

(2.15)

for all \( H', H \in \mathcal{H}, r \in K(H), \) and \( r' \in K(H') \).

### 2.5 Phase 3 - Random selection

At the end of Phase 2 on day \( n \), the current population \( POP_n \) has large size \( \tilde{N} \simeq N(F, H) \) and histogram \( J_n \simeq \Psi(H_n, R_n/N) \) with \( \Psi \) given by (2.3). During Phase 3 one extracts from \( POP_n \) a random sample of fixed size \( N \). This sample becomes the new initial population \( pop_{n+1} \) on day \( n + 1 \) and has genetic histogram \( H_{n+1} \). Phase 3 is thus a simplified emulation of natural selection.

The multinomial distribution \( \mu_{N,J}(V) \) parameterized by \( N \) and histogram \( J \) is defined for all vectors \( V \in \mathbb{N}^g \) having integer coordinates \( V(j) \geq 0 \) such that \( \sum_j V(j) = N \) by

\[
\mu_{N,J}(V) = N! \prod_{j \in \text{supp}(V)} \frac{J(j)V(j)}{V(j)!}.
\]

(2.16)

When \( J_n(j) = 0 \), no mutant of genotype \( j \) is present before selection so that \( H_{n+1}(j) = 0 \). Hence \( \text{supp}(H_{n+1}) \subset \text{supp}(J_n) \) with probability 1. For \( G \in \mathcal{H}_N \), all coordinates of \( V = NG \) are integers, and one has

\[
P(H_{n+1} = G \mid H_n; R_n) = P(H_{n+1} = G \mid J_n) = \begin{cases} 
\mu_{N,J_n}(NG), & \text{supp}(G) \subset \text{supp}(J_n), \\
0, & \text{otherwise}.
\end{cases}
\]

(2.17)

The multinomial distribution (2.16) has mean \( NJ \). Hence, equation (2.17) implies \( \mathbb{E}[H_{n+1} \mid H_n, R_n] = \mathbb{E}[H_{n+1} \mid J_n] = J_n \).

### 2.6 Markov chain dynamics in the space of histograms

At the completion of the three phases during a daily cycle, the population size always returns to the large but fixed size \( N \). The cycle on day \( n \) thus induces a stochastic transition \( H_n \rightarrow H_{n+1} \) in the space of genetic histograms \( \mathcal{H} \subset \mathbb{R}^g \). The succession of daily cycles just described above generates
a time-homogeneous Markov chain \( \{H_n \to H_{n+1}\} \) on the state space \( \mathcal{H} \) of all histograms. However, for each fixed population size \( N \), the actual state space of the Markov chain \( \{H_n \to H_{n+1}\} \) is the finite set \( \mathcal{H}_N \) of \( N \)-rational histograms, which has size \( \text{card}(\mathcal{H}_N) \leq (N+1)^g \). The Markov chain’ transition kernel \( Q_N(H,G) = P(H_{n+1} = G \mid H_n = H) \) for \( H,G \in \mathcal{H}_N \) is given by the finite sum

\[
Q_N(H,G) = \sum_{r \in K_N(H)} P(R_n/N = r \mid H_n = H) P(H_{n+1} = G \mid R_n/N = r, H_n = H).
\]

(2.18)

Recall that \( \text{card}(K_N(H)) \leq [(1 + F_g)(N + 1)]^g \). This Markov chain generates random histogram trajectories \( H = [H_1, H_2, \ldots, H_T] \) of arbitrary duration \( T \). In the following sections, we develop an explicit large deviations theory for histogram trajectories. Since realistic experiments on bacterial evolution involve large populations with \( 5 \times 10^5 \leq N \leq 10^8 \), our rare events study naturally focuses on asymptotic results for \( N \to \infty \).

3 Large deviations for daily cycles

For the time-homogeneous Markov chain \( \{H_n \to H_{n+1}\} \), the distribution of histogram trajectories is essentially determined by its transition kernel \( Q_N \) given in (2.18). Rare events analysis for random histogram trajectories must hence start with a large deviations analysis of the daily transition kernel. In this section, we carry out a large deviations analysis for random mutations and for random selection, concluding with an explicit large deviations framework for daily transitions. The distribution of our daily random mutations matrices \( R_n \) is well approximated by products of Poisson distributions, and daily random selections follow multinomial distributions. For large \( N \) we combine precise large deviations results for Poisson and multinomial distributions to derive accurate approximations of the daily transition kernel. Technically however, to obtain smooth formulas one has to quantify accurately how to relax the \( N \)-rationality conditions for the random histograms \( H_n, J_n \) introduced above.

**Definition 3.1.** For any histogram \( H \), define the ball \( V_N(H) \) and the \( N \)-rational ball \( B_N(H) \) by \( V_N(H) = \{ H' \in \mathcal{H} : \| H' - H \| \leq \frac{2}{3N} \} \) and \( B_N(H) = \mathcal{H}_N \cap V_N(H) \).

Note that \( \text{card}(B_N(H)) \leq 2^g \). The following lemma characterizes "uniformity" within \( B_N(H) \) and will help to relax \( N \)-rationality.

**Lemma 3.2.** Fix any histogram \( H \). For \( N \) large enough, all histograms \( H' \) in the \( N \)-rational ball \( B_N(H) \) have the same support as \( H \), and an essential minimum \( b(H') \) larger than \( b(H)/2 \). More precisely these results hold as soon as \( N > \frac{2}{b(H)} \).

**Proof.** For \( H' \in B_N(H) \) and any \( k \in \text{supp}(H') \), one has \( H'(k) \geq 1/N \) since \( H' \) is \( N \)-rational. This implies

\[
H(k) \geq H'(k) - 2/(3N) > 1/(3N) > 0.
\]

Hence \( k \in \text{supp}(H) \) and therefore \( \text{supp}(H') \subset \text{supp}(H) \). Since \( \| H' - H \| \leq \frac{2}{3N} \), one has for \( N > \frac{2}{b(H)} \) and any \( j \in \text{supp}(H) \),

\[
H'(j) \geq H(j) - \frac{2}{3N} \geq b(H) - \frac{2}{3N} > \frac{2b(H)}{3}.
\]

This inequality yields \( b(H') \geq \frac{2b(H)}{3} \), which proves the lemma.

\( \square \)
3.1 Random mutations matrices and products of Poisson distributions

Given the initial histogram \( H_n = H \) for day \( n \), the conditional distribution of the random mutations matrix \( R_n \) was defined in two steps (see Section 2.4.3). First, one defines a companion random matrix \( Z_n \) of independent random variables \( Z_n(j,k) \), where each \( Z_n(j,k) \) is Poisson distributed with mean \( m q_{j,k} \) \( F_j H_n(j) \). Then for \( r \in K_n(H) \) one defines \( P(R_n/N = r \mid H_n = H) \) by the ratio

\[
\frac{P(Z_n/N = r \mid H_n = H)}{P(Z_n/N \in K(H) \mid H_n = H)}.
\]

In this section we will prove Theorem 3.4 which states that as \( N \to +\infty \), the random mutations matrix \( R_n \) and its Poisson companion \( Z_n \) (see Section 2.4.3) have nearly identical conditional distributions given \( H_n = H \). A key point will be to prove that the conditional probability \( P(Z_n/N \in K(H) \mid H_n = H) \) tends to 1 as exponential speed \( \simeq d(H)^N \) where \( 0 < d(H) < 1 \) is given by the following definition.

**Definition 3.3.** The essential minimum \( b(H) > 0 \) of any histogram \( H \) is given by the equation (2.1). Define the decay coefficient \( 0 < d(H) < 1 \) of any histogram \( H \) by \( d(H) = \exp(-\delta F_1 b(H)) \) where \( \delta = \log(1/m) - 1 > 0 \).

For \( H \in \mathcal{H}_N \), one has \( b(H) \geq 1/N \) and hence \( d(H)^N \leq e^{-\delta F_1} < 1 \) for all \( H \in \mathcal{H}_N \) and all \( N \). In actual laboratory experiments \([16,22,36,39]\), one typically has \( F_1 > 100 \) and mutation rates \( m \leq 10^{-6} \), so that \( \delta F_1 > 12800 \) and hence \( d(H)^N < 10^{-5500} \) which is practically zero.

**Theorem 3.4.** For \( N \) large, the random mutations matrix \( R_n \) on day \( n \) and its Poissonian companion \( Z_n \) have nearly identical conditional distributions given \( H_n = H \). Fix any histogram \( H^* \) and the \( N \)-rational ball \( B_N(H^*) \). Let \( c = 2 + \log(g)/F_1 \). Then, provided \( N > \frac{c}{b(H^* - 1)} \), one has for all \( n \), all \( H \in B_N(H^*) \), and all matrices \( z \in K(H) \),

\[
1 \leq \frac{P(R_n/N = z \mid H_n = H)}{P(Z_n/N = z \mid H_n = H)} \leq 1 + 2g d(H^*)^{N/2}.
\]

This result can be reformulated in terms of the conditional density \( f_N(z,H) = \frac{P(R_n/N = z \mid H_n = H)}{P(Z_n/N = z \mid H_n = H)} \) of \( R_n/N \) with respect to \( Z_n/N \).

Fix any \( 0 < a < 1 \). Then, \( f_N(z,H) \to 1 \) at exponential speed as \( N \to \infty \) and the convergence rate is uniform for \( b(H^*) \geq a \), \( H \in B_N(H^*) \), and \( z \in K(H) \). Therefore, the conditional joint distribution of the mutations matrix \( R_n(j,k) \) given \( H_n \in B_N(H^*) \) becomes, for large \( N \), practically equal to the product of Poisson distributions with respective means \( m q_{j,k} \) \( F_j H_n(j) \).

To prove Theorem 3.4, we first need two lemmas

**Lemma 3.5** (Poisson Large Deviations). Fix any \( u > 0 \), and let \( X \) be a random variable having a Poisson distribution with mean \( Nu \). For any \( v > u > w > 0 \) and for any integer \( N \), the following inequalities hold

\[
P(X \geq Nv) \leq \exp(-N \left[ u + v \log(v/u) - v \right]), \quad (3.2)
\]

\[
P(X \leq Nw) \leq \exp(-N \left[ u + w \log(w/u) - w \right]). \quad (3.3)
\]

Moreover, for any \( v \geq 0 \) such that \( Nu \) is an integer, one has

\[
\frac{1}{N} \log P(X = Nv) = -(u + v \log(v/u) - v) + o(N) \quad (3.4)
\]

with \( |o(N)| \leq 2 \log(N)/N \).
Proof. Express the Poisson random variable $X$ as $X = \sum_{s=1}^{N} X_s$ where $X_1, \ldots, X_N$ are i.i.d Poisson random variables with the same mean $u > 0$. Then for all $N$ and $v > u$ (see e.g. [215]), the empirical mean $X/N$ of the $X_i$ must verify the large deviations inequality $P(X/N \geq v) \leq \exp(-N\lambda(v))$ where $\lambda(v)$ is the large deviations rate for the Poisson distribution, and is given by the well-known formula $\lambda(v) = \frac{u + v \log(v/u) - v}{N}$ for all $v > u > 0$. This proves (3.2)  and a similar argument proves (3.3).

The classical proof of Stirling formula for factorials can be easily modified to yield the following uniform inequality

$$|\log(N!) - N \log(N/e)| \leq 2 \log(N) \quad \text{for all } N \geq 1. \quad (3.5)$$

For any $v \geq 0$ such that $Nv$ is an integer, one has $P(X = Nv) = e^{-Nv}(Nu)^{Nv}/(Nv)!$ so that

$$\frac{1}{N} \log P(X = Nv) = -u + v \log(Nu) - \frac{1}{N} \log((Nv)!)$$

For $v > 0$, (3.5) implies

$$\frac{1}{N} \log((Nv)!) = v \log(N) + v \log(v) - v + o(N)$$

with $|o(N)| \leq 2 \log(N)/N$. The last two equations prove (3.4) for $v > 0$. Finally, (3.4) is trivially true for $v = 0$ with the convention $0 \log(0) = 0$. This concludes the lemma.

Lemma 3.6. Fix a positive sequence $\epsilon(N)$ such that $\epsilon(N) \to 0$ as $N \to \infty$. Fix any set $E$ and any function $\lambda(s) \geq 0$ defined for all $s \in E$. Fix $c > 0, \beta > 0$, and let $E_N$ be a finite subset of $E$ with $\text{card}(E_N) \leq cN^\beta$. Consider fast-vanishing exponentials $p_N(s) > 0$ indexed by $s \in E_N$ such that $\frac{1}{N} \log p_N(s) = -\lambda(s) + o(s) \lambda(N)$ where $|o(s)| \leq \epsilon(N)$. Define $\Lambda(E_N) = \inf_{s \in E_N} \lambda(s)$. The sum $p(E_N) = \sum_{s \in E_N} p_N(s)$ satisfies

$$\frac{1}{N} \log p(E_N) = -\Lambda(E_N) + o(N)$$

for all $N$ with $|o(N)| \leq \epsilon(N) + \beta \log(N)/N + \log(c)/N$.

Proof. Select $s(N) \in E$ such that $\Lambda(E_N) = \lambda(s(N))$. This yields the lower bound

$$\frac{1}{N} \log p(E_N) \geq \frac{1}{N} \log p(s(N)) = -\lambda(s(N)) + o_{s(N)}(N) \geq -\Lambda(E_N) - \epsilon(N). \quad (3.6)$$

We have $p_N(s) \leq \exp(-N[\Lambda(E_N) + \epsilon(N)])$ for all $s \in E_N$ by definition of $\Lambda(E_N)$. This gives $p(E_N) \leq \exp(-N[\Lambda(E_N) + \epsilon(N)])$, which implies

$$\frac{1}{N} \log p(E_N) \leq -\Lambda(E_N) + \epsilon(N) + \log(\text{card}(E_N))/N.$$

Combining this upper bound with (3.6) concludes the proof.

Proof of Theorem 3.4

Proof. Consider any $N$-rational histogram $H$. Given $H_n = H$, let $Z_n$ be a matrix of independent Poisson random variables $Z_n(j,k)$ with means $mNq_jkF_jH(j)$. The sums $S(j) = \sum_{k=1}^{q} Z_n(j,k)$ then have Poisson distributions with respective means $\mathbb{E}[S(j)] = NmF_jH(j)$. First, we show that for each $j$,

$$P(Z_n \in NK(j,H) \mid H_n = H) \geq 1 - d(H)^N \quad (3.7)$$
holds for all $N$ and $H$. To do so, we separately consider the two cases $H(j) > 0$ and $H(j) = 0$.

**Case 1:** Suppose $H(j) > 0$. For any $s > m$, apply (3.2) to $X = S(j)$ with $v = sF_j H(j)$ and $u = mq(j,k)F_j H(j)$ to obtain

$$P(S(j) \geq sNF_j H(j) \mid H_n = H) \leq \exp\left(-sNF_j H(j)[\log(1/m) - 1]\right). \tag{3.8}$$

Since $\log(1/m) - 1 = \delta > 0$, the inequality (3.8) implies

$$P(S(j) \geq sNF_j H(j) \mid H_n = H) \leq \exp\left(-\delta sNF_j H(j)\right) \leq d(H)^sN \tag{3.9}$$

for all $N, H$ and $s > m$. By definition (2.4) of $K(j, H)$,

$$P(Z_n/N \in K(j, H) \mid H_n = H) = 1 - P(S(j) \geq NF_j H(j) \mid H_n = H).$$

Equation (3.9) with $s = 1$ implies (3.7) for all $N$ and $H$.

**Case 2:** Suppose $H(j) = 0$. This implies $P(Z_n(j,k) = 0 \mid H_n = H) = 1$ for all $k$. Therefore, $S(j) = 0$ and $P(Z_n \in NK(j, H_n) \mid H_n = H) = 1$, which trivially satisfies (3.7).

Now, since $K(H) = \bigcap_{j=1}^\delta K(j, H)$ and (3.7) holds for all $N$ and $H$, then

$$P(Z_n/N \in K(H) \mid H_n = H) \geq 1 - g(d(H))N \tag{3.10}$$

also holds for all $N$ and $H$. The constraint $N > N(H) = \log(2g)/(12F_1 b(H))$ forces $g(d(H)) < 1/2$ so that $1/(1 - g(d(H))N) \leq 1 + 2g(d(H))N$. For $N > N(H)$, equation (3.10) then implies

$$1 \leq \frac{1}{P(Z_n/N \in K(H) \mid H_n = H)} \leq 1 + 2g(d(H))N. \tag{3.11}$$

For all matrices $z \in K(H)$, we have

$$P(R_n/N = z \mid H_n = H) = \frac{P(Z_n/N = z \mid H_n = H)}{P(Z_n/N \in K(H) \mid H_n = H)}.$$

For $H_n = H$, $z \in K(H)$, and $N > N(H)$, equation (3.11) yields

$$P(Z_n/N = z \mid H_n = H) \leq P(R_n/N = z \mid H_n = H) \leq (1 + 2g(d(H))N) P(Z_n/N = z \mid H_n = H). \tag{3.12}$$

Consider any fixed $H^* \in \mathcal{H}$ and any $H_n = H \in V_N(H^*)$, which is equivalent to $H \in B_N(H^*)$ since $H_n$ is $N$-rational. By Lemma 3.2, for $N > \frac{\delta}{b(H^*)}$, $\text{supp}(H) = \text{supp}(H^*)$ and $b(H) \geq b(H^*)/2$ so that $d(H) < d(H^*)^{1/2}$. Let $N_1 = \frac{c}{b(H^*)}$ with $c = 2 + \frac{\log(\delta)}{F_1}$. Then $N > N_1$ will force $N > N(H)$ provided $H \in B_N(H^*)$. Equation (3.12) then implies, for all $N$-rational $H \in B_N(H^*)$, $z \in K_N(H)$, and $N > N_1$,

$$1 \leq \frac{P(R_n/N = z \mid H_n = H)}{P(Z_n/N = z \mid H_n = H)} \leq 1 + 2g(d(H^*))^{N/2}, \tag{3.13}$$

which proves (3.1). The uniform convergence in the statement of the theorem is an easy consequence of (3.1). \qed
3.2 Large deviations for random mutations matrices $R_n$

In this section, we develop explicit large deviations formulas for the daily matrices $R_n$ of random mutations. We have shown that with probability tending to 1 at exponential speed for large $N$, the random variables $R_n(j,k)$ are conditionally independent given $H_n = H$, and that their conditional distributions $\pi_{H,i,j}$ are Poissonian with means depending on $H$ and $j,k$. This indicates that given $H_n = H$ the conditional large deviations rate function of $R_n$ should be computed by summing over all $j,k$ the Poisson rate functionals of the $\pi_{H,i,j}$. However, proving this requires a few meticulous technical steps.

Since the non-zero coefficients of $R_n$ are all of order $N$, the key matrices to obtain large deviations results are the matrices $r_n = R_n/N$, which belong to the following convex cone $\mathcal{R}$.

**Definition 3.7.** The random matrices $r_n = R_n/N$ always belong to the convex cone $\mathcal{R}$ of all $g \times g$ matrices $r$ such that $r_{j,k} \geq 0$ and $r_{j,j} = 0$ for all $j,k$. For $r \in \mathcal{R}$, denote $\|r\| = \max_{j,k} |r_{j,k}|$ the $L_\infty$ norm of $r$.

In $\mathcal{R}$ determining relevant "bounded" sets $S$ of matrices $r$ involves controlling the *smallest positive* $r_{j,k}$ over $r \in S$. So for $r \in \mathcal{R}$, we define the support $\text{supp}(r)$ and the essential minimum $b(r) > 0$ of $r$, by

$$\text{supp}(r) = \{(j,k) \mid r_{j,k} > 0\} \quad \text{and} \quad b(r) = \min_{(j,k) \in \text{supp}(r)} r_{j,k}.$$  

As was done above for histograms, denote $V_N(r)$ the $L_\infty$ ball with center $r \in \mathcal{R}$ and radius $\frac{2}{N}$. Let $B_N(r) \subset V_N(r)$ be the set of $N$-rational matrices in $V_N(r)$. Note that $\text{card}(B_N(r)) \leq 2^r$.

Given $H_n = H$, we showed above that the random matrices $R_n/N$ always belongs to the set $K_N(H)$ of all $N$-rational matrices within the convex set $K(H) \subset \mathcal{R}$ defined by $2^9$. Note that for $r \in K(H)$ and $(j,k) \in \text{supp}(r)$, one must have $H(j) > 0$ and $q_{j,k} > 0$.

We can now define the adequate large deviations rate function controlling the probabilities of rare random mutations matrices.

**Definition 3.8** (Rate function for random mutations matrices). For any histogram $H \in \mathcal{H}$, any matrix $r \in K(H)$, and any $j,k$ such that $H(j)q_{j,k} > 0$, denote $\pi(H,j,k)$ the Poisson distribution with mean $m q_{j,k} F_j H(j)$. Note that for $N$ large $\pi(H,j,k)$ will be extremely close to the distribution of $R_n(j,k)$. Well known Large deviations formulas for Poisson distribution show that the rate function $L_{j,k}(H,r) \geq 0$ for $\pi(H,j,k)$ is given by

$$L_{j,k}(H,r) = m q_{j,k} F_j H(j) + r_{j,k} \log\left(\frac{r_{j,k}}{e m q_{j,k} F_j H(j)}\right) \quad \text{when} \quad H(j) q(j,k) > 0. \quad (3.14)$$

When $H(j) q(j,k) = 0$ we conventionally set $L_{j,k}(H,r) = 0$. For Poisson distributions such as $\pi(H,j,k)$, the rate function $L_{j,k}(H,r)$ is a strictly convex function of $r_{j,k}$ for $r_{j,k} > 0$, and one has $L_{j,k}(H,r) = 0$ if and only if $r_{j,k} = m q_{j,k} F_j H(j)$. Denote $L(r,H)$ the $g \times g$ matrix $L_{j,k}(H,r)$.

We will show further on that large deviations for the random mutations matrices $r_n = R_n/N$ are controlled by the rate function $\text{mut}(H,r)$ given by

$$\text{mut}(H,r) = \sum_{j,k} L_{j,k}(H,r) = \sum_{(j,k) \mid m q_{j,k} H(j) > 0} m q_{j,k} F_j H(j) + r_{j,k} \log r_{j,k} - r_{j,k} \log(e m q_{j,k} F_j H(j)) \quad (3.15)$$
Proposition 3.9. For all histograms $H$, the mutations rate function $\text{mut}(H, r)$ is a finite, non-negative, continuous, strictly convex function of $r \in K(H)$. Moreover, $\text{mut}(H, r)$ is a continuous, convex function of $H$ for fixed $r$. For any fixed $0 < a < 1$, there is an explicit constant $c(a)$ depending only on $a$ and on the process parameters $g, F, Q, m$, such that

$$\text{mut}(H, r) \leq c(a) \text{ for all } H \text{ with } b(H) \geq a \text{ and all } r \in K(H),$$

(3.16)

Note also that $\text{mut}(H, r) = 0$ if and only if $r_{j,k} = m q_{j,k} F_j(j)$ for all $j, k$.

Proof. As all Poisson rate functions, $L_{j,k}$ must be strictly convex in $r_{j,k}$. This proves strict convexity in $r$ for $\text{mut}(H, r)$. Each $L_{j,k}$ is also clearly convex in $H$. This implies convexity in $H$ for $\text{mut}(H, r)$. To prove (3.16), note the elementary inequalities

$$x \log x \leq D(v) \text{ for all } 0 \leq x \leq v;$$

$$\log x \leq E(u, v) \text{ for all } 0 < u \leq x \leq v.$$  

where $D(v) = 1/e + \log(1 + v)$, and $E(u, v) = \log(1 + 1/u) + \log(1 + v)$.

For $H(j) q_{j,k} > 0$, $r \in K(H)$, and $b(H) \geq a$, one has

$$r_{j,k} \leq F_j H(j) \leq F_g$$

and

$$0 < a \cdot m b(Q) \leq m q_{j,k} F_j H(j) \leq F_g m \|Q\|.$$  

Combining the last three inequalities yields

$$|r_{j,k} \log(r_{j,k})| \leq D(F_g),$$

$$|r_{j,k} \log(e m q_{j,k} F_j H(j))| \leq F_g |\log(e m q_{j,k})| + F_j H(j) |\log(F_j H(j))| \leq E(a \cdot m b(Q), m F_g \|Q\|) + D(F_g).$$

This concludes the proof due to formula (3.15).

To obtain uniform Hölder continuity bounds for $\text{mut}(H, r)$, the histogram $H$ must belong to one of the compact sets $\mathcal{H}(a)$ defined next.

Definition 3.10. For any fixed $0 < a < 1$, define the compact set of histograms $\mathcal{H}(a) \subset \mathcal{H}$ by

$$\mathcal{H}(a) = \{ H \in \mathcal{H} \mid b(H) \geq a \}.$$  

(3.17)

Proposition 3.11. Fix $0 < a < 1$ and any Hölder exponent $0 < \alpha < 1$. For all histograms $H', H \in \mathcal{H}(a)$ with $\text{supp}(H') = \text{supp}(H)$, and all matrices $r \in K(H)$, $r' \in K(H')$ the mutations rate function $\text{mut}(H, r)$ verifies

$$|\text{mut}(r', H') - \text{mut}(r, H)| \leq c \left( \|r' - r\|^\alpha + \|H' - H\| \right) \quad \text{with} \quad c = \frac{14g^2 F_g \log(F_g)}{a(1 - \alpha)}.$$  

(3.18)

Proof. The proof requires several meticulous inequalities and is presented in the Appendix.

We now conclude the analysis of random mutations with a key asymptotic large deviations result for random mutations matrices.

Proposition 3.12. Let $R_n$ be the random matrix of mutations on day $n$. Let $\text{mut}(H, r)$ be the mutations rate function defined by (3.15). Fix $a > 0$ and the parameters $P$ of the bacterial evolution model (definition 2.2). Then there is a constant $N_0 = N_0(a, P)$ such that for $N > N_0$, the large deviations formula

$$\frac{1}{N} \log P(R_n/N = r \mid H_n = H) = -\text{mut}(H, r) + o(N),$$

(3.19)

with $|o(N)| \leq 4g^2 \log N/N$ will hold uniformly for all $H \in \mathcal{H}(a) \cap \mathcal{H}_N$ and $r \in K_N(H)$.

Proof. The proof of this proposition is presented in Appendix.
3.3 Large deviations for random selection

After daily growth and mutations, the current population \( P O P_n \) has a random size much larger than \( N \), and the day \( n \) cycle ends with the random selection of a sample of size \( N \). This sample becomes \( pop_{n+1} \) and has a histogram \( H_{n+1} \). Given \( H_n = H \) and \( R_n = R \), \( P O P_n \) has large size \( \eta(H)N \), and the conditional distribution \( \text{mult}_{H,R}(N) \) of \( NH_{n+1} \) is multinomial. Since both population and sample have very large sizes of order \( N \), one cannot apply directly the classical large deviations for "standard" multinomial distributions. Moreover the histogram of \( P O P_n \) given \( H_n, R_n \) is itself random. Therefore, in this section we develop an explicit rate function for large deviations of the conditional multinomial \( \text{mult}_{H,R}(N) \), and demonstrate that this functional involves the well-known Kullback-Leibler divergence. These results have then to be transposed into precise rate functions for conditional large deviations of \( H_{n+1} \) given \( H_n = H \) and \( R_n = R \). We also analyze the regularity of these conditional rate functions when \( H \) and \( R/N \) vary that is an important technical point to prove large deviations results for the daily transition from \( H_n \) to \( H_{n+1} \).

**Definition 3.13** (Kullback-Leibler divergence). The classical *Kullback-Leibler divergence* between two histograms \( G \) and \( J \) is defined by

\[
KL(G, J) = \sum_{j \in \text{supp}(G)} G(j) \log \frac{G(j)}{J(j)} \quad \text{when} \quad \text{supp}(G) \subset \text{supp}(J)
\]

and by \( KL(G, J) = +\infty \) otherwise.

Note that \( KL(G, J) \geq 0 \), and that \( KL(G, J) = 0 \) if and only if \( G = J \). When \( \text{supp}(G) \subset \text{supp}(J) \), the function \( KL(G, J) \) is finite, continuous, and is strictly convex in \( J \) for \( G \) fixed, as well as strictly convex in \( G \) for \( J \) fixed, since \( KL(G, J) \) has the following partial derivatives for all \( i, j \in \text{supp}(G) \subset \text{supp}(J) \)

\[
\partial_{G(i)}KL(G, J) = 1 + \log\left(\frac{G(i)}{J(i)}\right) \quad \text{and} \quad \partial_{J(i)}KL(G, J) = -\frac{G(i)}{J(i)};
\]

\[
\partial_{G(i)}\partial_{G(j)}KL(G, J) = \left(\frac{1_{\{i=j\}}}{G(j)}\right) \partial_{J(i)}\partial_{J(j)}KL(G, J) = 1_{\{i=j\}} \frac{G(i)}{J(i)^2}.
\]

On day \( n \), we denote the random population histogram right after growth and mutations (phases 1 and 2) as \( J_n \). For any two histograms \( G \) and \( J \) such that \( \text{supp}(G) \subset \text{supp}(J) \), equation (2.17) shows that

\[
P(H_{n+1} = G \mid J_n = J) = \mu_{N,J}(NG)
\]

where \( \mu_{N,J}(V) \) is the multinomial distribution defined by (2.16). For \( N \) large we now derive an asymptotic expression of \( \frac{1}{N} \log(\mu_{N,J}(NG)) \).

**Proposition 3.14.** For any histogram \( J \in \mathcal{H} \), and any \( N \)-rational histogram \( G \in \mathcal{H} \) such that \( \text{supp}(G) \subset \text{supp}(J) \), the multinomial distribution \( \mu_{N,J} \) defined by (2.16) verifies

\[
\frac{1}{N} \log(\mu_{N,J}(NG)) = -KL(G, J) + o(N)
\]

with uniform remainder \( |o(N)| \leq 2(g + 1) \log N/N \).

**Proof.** The proof is presented in Appendix \( \square \)
We can now prove the main large deviations evaluation for random selections. Combined with Proposition 3.12 this result will be used to obtain the large deviation estimate for the full daily transition kernel \( P(H_{n+1} = G \mid H_n = H) \).

**Proposition 3.15.** Fix \( 0 < a < 1 \), and set \( N^* = 2/a + \log(F_g) \). On day \( n \), let \( J_n \) be the population histogram at the end of the mutations phase (Phase 2). For any \( G \) with \( b(G) > a \), any \( G' \) in the \( N \)-rational ball \( B_N(G) \), and any \( N > N^* \), one has the large deviations estimate

\[
\frac{1}{N} \log P(H_{n+1} = G' \mid H_n, R_n) = -KL(G, J_n) + o(N)
\]

with uniform remainder \(|o(N)| \leq (5g + 1) \log N/N\).

**Proof.** The proof is presented in Appendix C. \( \square \)

### 3.4 Large deviations estimates for the one-step transition kernel

For each random phase of any day \( n \) cycle, namely mutations and selection, we have so far proved precise conditional large deviations estimates. We can now combine these two results to develop a large deviations analysis for the one-step transition kernel \( Q_N(H, G) = P(H_{n+1} = G \mid H_n = H) \). In this section, we prove that, in the state space \( \mathcal{H} \) of population histograms, the Markov transition kernel \( Q_N(H, G) \) converges to a deterministic kernel at exponentially fast speed as \( N \to \infty \), and we compute explicitly the corresponding rate function \( C(H, G) \geq 0 \) as the limit \( \lim_{N \to \infty} \frac{1}{N} \log(Q_N(H, G)) \).

Equations (3.15) and (3.25) provide two explicitly computed “partial” rate functions, namely, the rate function \( \text{mut}(H, r) \geq 0 \) to approximate \(-\frac{1}{N} \log(P(R_n/N = r \mid H_n = H)) \) and the Kullback-Leibler rate function \( KL(G, J) \geq 0 \) to approximate \(-\frac{1}{N} \log(P(H_{n+1} = G \mid J_n = J)) \) (where \( J_n \) is the histogram after growth and mutation phases). The deterministic function \( \Psi(H, r) \) given by (28) provides the approximation \( J_n \approx \Psi(H_n, R_n/N) \) with accuracy \( \approx 1/N \). We thus expect \( KL(G, \Psi(H, r)) \) to be the rate function approximating \(-\frac{1}{N} \log(P(H_{n+1} = G \mid H_n = H, R_n/N = r)) \).

We will define an explicit composite transition rate \( \tau(H, r, G) \geq 0 \) which for large \( N \) provides an estimate of the form

\[
\tau(H, r, G) \approx -\frac{1}{N} \log(P(\{H_{n+1} = G\} \& \{R_n/N = r\} \mid H_n = H)).
\]

Hence for \( H, G \) fixed, the probability \( P(\{H_{n+1} = G\} \& \{R_n/N = r\} \mid H_n = H) \) will be nearly maximized over \( r \) when the random mutations matrix \( R_n \) is close to \( N r^* \), where \( C(H, G) = \tau(H, r^*, G) \) is the minimum of \( \tau(H, r, G) \) over \( r \). Later on we show that \( C(H, G) \) is the rate function approximating \(-\frac{1}{N} \log(P(H_{n+1} = G \mid H_n = H)) \).

We now present a rigorous development of this large deviations analysis.

**Definition 3.16 (Composite Transition Rate).** To control large deviations for the composite daily transition \( (H_n = H) \to (R_n/N = r) \to (H_{n+1} = G) \), we introduce the composite transition rate \( \tau(H, r, G) \geq 0 \) defined by

\[
\tau(H, r, G) = \text{mut}(H, r) + KL(G, \Psi(H, r))
\]

for \( H, G \in \mathcal{H} \) and \( r \in K(H) \).
Recall that \( \text{mut}(H, r) \) is a finite continuous convex function of \( r \in K(H) \) (see prop. \ref{prop:mut-def}), and \( KL(G, J) \) is a continuous convex function of \( J \in \mathcal{H} \), with strict convexity whenever \( KL(G, J) \) is finite (see def. \ref{def:KL}); moreover due to \((2.3)\), \( J = \Psi(H, r) \) is an affine function of \( r \), so that \( KL(G, \Psi(H, r)) \) is continuous convex in \( r \) on the compact convex closure of \( K(H) \).

Hence \( \tau(H, r, G) \geq 0 \) is continuous and convex in \( r \) on the compact convex closure of \( K(H) \), with strict convexity in \( r \) whenever \( \tau(H, r, G) \) is finite. Similarly one verifies that whenever the function \( \tau(H, r, G) \) is finite, it is convex in \( H \) for \((r, G)\)-fixed, and convex in \( G \) for fixed \((H, r)\).

Since \( \text{mut}(H, r) \) is finite, one can have \( \tau(H, r, G) = +\infty \) if and only if \( Kl(G, \Psi(H, r)) = +\infty \), which is equivalent to the existence of a genotype \( j \) such that \( G(j) > 0 \) and \( \Psi_j(H, r) = 0 \). Due to \((2.12)\), we see that for \( r \in K(H) \), one has \( \tau(H, r, G) = +\infty \) if and only if there is a genotype \( j \) such that \( G(j) > 0 \) and \( H(j) = r_{k,j} = r_{j,k} = 0 \) for all \( k \).

\textbf{Definition 3.17} (Feasible Transitions and One-Step Cost). By construction, for any \( H, G \in \mathcal{H} \), the transition kernel \( P(H_{n+1} = G \mid H_n = H) \) is strictly positive if and only if for any \( j \) such that \( G(j) > 0 \) and \( H(j) = 0 \), one can find a \( k \) such that \( H(k) q_{k,j} > 0 \). We then say that \((H \rightarrow G)\) is a feasible transition. Let \( \tau(H, r, G) \geq 0 \) be the composite transition rate defined by \((3.26)\). For any \( H, G \in (H) \), define the one-step cost function \( C(H, G) \geq 0 \) by

\[
C(H, G) = \min_{r \in K(H)} \tau(H, r, G) = \min_{r \in K(H)} \left[ \text{mut}(H, r) + KL(G, \Psi(H, r)) \right].
\]

An explicit expression of the one-step cost \( C(H, G) \) will be developed later in this section in Theorem \ref{thm:one-step-cost}.

\textbf{Lemma 3.18. Finiteness and convexity of \( C(H, G) \):} For all \( H, G \in \mathcal{H} \), the one-step cost function \( C(H, G) \) is finite if and only if \((H \rightarrow G)\) is a feasible transition.

When \( C(H, G) \) is finite, there is a unique \( \rho \) in the closure of \( K(H) \) such that \( C(H, G) = \tau(H, \rho, G) \); moreover \( C(H, G) \) is then convex in \( H \) for \( G \) fixed and convex in \( G \) for \( H \) fixed.

\textbf{Proof.} By definition, \( C(H, G) \) is finite if and only if there is at least one \( r \in K(H) \) with \( \tau(H, r, G) \) finite. As seen above, this occurs if and only if for each \( j \) such that \( G(j) > 0 \) and \( H(j) = 0 \), there is a \( k \) with \( r_{k,j} > 0 \). For any \((j, k)\) verifying such a condition, one must also have \( m q_{k,j} > 0 \) due to \((2.4)\) so that \((H \rightarrow G)\) is a feasible transition.

Since \( \tau(H, \rho, G) \) is continuous in \( \rho \) on \( K(H) \), there is at least one \( \rho \in K(H) \) such that \( C(H, G) = \tau(H, \rho, G) \). When \( C(H, G) \) is finite, this \( \rho \) is unique in the compact, convex set \( K(H) \) since \( \tau(H, \rho, G) \) is strictly convex in \( \rho \) whenever \( \tau(H, \rho, G) \) is finite; the asserted convexity properties of \( C(H, G) \) follow then from the separate convexity in \( H \) and \( G \) of \( \tau(H, \rho, G) \).

A key large deviations result (see theorem \ref{thm:ldr} below) will relate \( C(H, G) \) to the one-step Markov transition kernel \( Q(H, G) \) by a large deviations estimate of the form \( 1/N \log(Q(H, G)) \approx -C(H, G) \), valid for large \( N \). The proof of theorem \ref{thm:ldr} will require precise Holder continuity estimates for \( KL(H, G) \) and \( \tau(H, r, G) \), which we now state in the next two technical results.

\textbf{Lemma 3.19.} (Continuity estimates for \( KL(G, J) \)) For \( G, J \in \mathcal{H} \), then \( KL(G, J) \) is finite if and only if \( \text{supp}(G) \subset \text{supp}(J) \), and one then has

\[
KL(G, J) \leq \log \beta(G, J) \leq \frac{KL(G, J) + \log(g)}{b(G)} \text{ where } \beta(G, J) = \max_{k \in \text{supp}(G)} \frac{1}{J(k)}.
\]

(3.28)
Proof. The proof of this technical lemma is given in the appendix (see Appendix C). \qed 

Proposition 3.20. (Holder continuity for the composite transition cost $\tau_{au}(H, r, G)$) Fix $A > 0$, $a > 0$, and the process parameters $\mathcal{P}$ (definition 2.2). For all $H', H, G \in \mathcal{H}$ with $\min(b(H), b(H'), b(G)) > a$, and matrices $r' \in K(H')$ and $r \in K(H)$ verifying

$$supp(H') = supp(H), \quad \|r' - r\| \leq 1, \quad \tau(H, r, G) \leq A, \quad \tau(H', r', G) < \infty$$

one then has the following bound

$$|\tau(H', r', G) - \tau(H, r, G)| \leq \eta \left[\|r' - r\|^{1/2} + \|H' - H\|\right].$$

(3.33)

where the constant $\eta = \eta(A, a, \mathcal{P})$ is given by

$$\eta = 3 g^{2+1/a} e^{A/a} F_g / F_1 + 20 F_g \log(F_g) / a.$$

(3.34)

Proof. Proof of this proposition is given in Appendix C \qed

We first state a key large deviations result for the one-step transition kernel controlling daily transitions. The following theorem will be essential to develop the large deviations formalism for random histogram trajectories in Section 4.

Theorem 3.21. Fix any $0 < a < 1 < d$ and the process parameters $\mathcal{P}$ (definition 2.2). One-step large deviations for the Markov chain $H_n$ are then controlled by two constants $c = c(d, a, \mathcal{P})$ and $N_0 = N_0(d, a, \mathcal{P})$ as follows. Consider any $N$-rational histograms $H, G \in \mathcal{H}$ with $\min(b(H), b(G)) > a$ and transition cost $C(H, G) \leq d$. Then, the one-step transition kernel $Q(H, G)$ has a uniform large deviations approximation valid for all $N > N_0$

$$\frac{1}{N} \log Q(H, G) = -C(H, G) + o(N)$$

(3.35)

with $|o(N)| \leq c / \sqrt{N}$. 

18
Proof. We sequentially construct constants $c_i > 0, N_i > 0$, which will only depend on the fixed positive $(d,a)$ and the process parameters $\mathcal{P}$.

On day $n$, given $H_n = H$ and $R_n/N = r \in K_N(H)$, the population histogram $J_n$ after growth and mutations is a deterministic function $J_n = J(H, r)$ defined by (2.7). Recall that $J(H, r)$ can be precisely estimated by $I = \Psi(H, r)$ (see equation (2.10)). To transition from histograms $H_n = H$ to $H_{n+1} = G$, we define a set $A_N$ of “acceptable” mutations matrices $r_n = R_n/N$ by

$$A_N = A_N(H, G) = \{r \in K_N(H) \mid KL(G, J) < \infty \text{ and } KL(G, I) < \infty\}. \quad (3.36)$$

Introduce the approximate transition cost $\hat{\tau}(H, r, G) = \text{mut}(r, H) + KL(G, J(H, r))$, which we expect to be very close to the explicit composite transition cost $\tau(H, r, G)$ (defined in def 3.26). We now give a technical but more tractable expression for the transition kernel.

Indeed there are constants $c_0, N_0$ such that, for $N > N_0$ and for any $N$-rational triple $(H, r, G)$ verifying $H, G \in \mathcal{H}, r \in K_N(H)$, $\min(b(H), b(G)) > a$, then the Markov transition kernel $Q(H, G)$ (see definition 2.18) verifies

$$Q(H, G) = \sum_{r \in A_N} \exp \left(-N \left[\hat{\tau}(H, r, G) + o(N)\right]\right) \quad (3.37)$$

with $|o(N)| \leq c_0 \log N/N$.

To prove (3.37), apply (2.9) and the discussion in Section 2.4.4 to $J = J(H, r)$ and $I = \Psi(H, r)$ to conclude that for $N > N_1 = 20 + 2F_g/a$ and $c = 13F_g/F_1$, one has

$$\text{supp}(J) = \text{supp}(I) \quad \text{and} \quad \|J - I\| \leq c/N. \quad (3.38)$$

This forces either supp($G$) $\subset$ supp($J$) so that $KL(G, I)$ and $KL(G, J)$ are both finite, or supp($G$) $\not\subset$ supp($J$) so that $KL(G, I) = KL(G, J) = \infty$. Define $\kappa = \kappa(H, r, G) = \min\{KL(G, J), KL(G, I)\}$. Set $c_1 = g^{1+1/a}$ and $c_2 = c_1 c$. Combine (3.29) with (3.38) to obtain, for $r \in A_N$

$$|KL(G, J) - KL(G, I)| \leq c_1 \epsilon^{\kappa/a} \|J - I\| \leq c_2/N \quad (3.39)$$

which implies

$$|\hat{\tau}(H, r, G) - \tau(H, r, G)| \leq c_2/N.$$

The transition kernel $Q$ verifies by construction

$$Q(H, G) = \sum_{r \in K_N(H)} P(R_n/N = r \mid H_n = H) P(H_{n+1} = G \mid H_n = H, R_n/N = r). \quad (3.40)$$

Since $\|r\| \leq F_g$, for all matrices $r \in K_N(H)$, one must have card($K_N(H)$) $\leq [(N + 1)F_g]^2$. From (3.26) we get, for $N > N_2 = N_1 + 2/a + \log F_g$,

$$P(H_{n+1} = G \mid H_n = H, R_n/N = r) = \exp \left(-N[KL(G, J(H, r)) + o_1(N)]\right) \quad (3.41)$$

with $|o_1(N)| \leq (5g + 1) \log N/N$. In (3.31), the right-hand side is 0 unless $r \in A_N$. Hence the sum in (3.40) can be restricted to $r \in A_N$. Then, (3.19) yields, for $N > N_3 = N_2 + [4 + 2 \log(g)]/a$,

$$P(R_n/N = r \mid H_n = H) = \exp \left(-N[\text{mut}(r, H) + o_2(N)]\right) \quad (3.42)$$

19
with $|o_2(N)| \leq 4g^2 \log N/N$. For $r \in A_N$ and $N > N_3$, substitute \textcolor{red}{(3.31)} and \textcolor{red}{(3.32)} into \textcolor{red}{(3.30)} to finally prove \textcolor{red}{(3.37)} with $o(N) = o_1(N) + o_2(N)$, provided one sets $N_0 = N_3$ and $c_0 = 5g + 1 + 4g^2$.

For $H, G, r$ as above, and $N > N_0$, we have now shown that

$$\text{supp}(J) = \text{supp}(I) \quad \text{and} \quad \|J - I\| \leq c/N \text{ for } r \in K_N(H),$$

$$|KL(G, J) - KL(G, I)| \leq c_1 e^{c/a} \|J - I\| \text{ for } r \in A_N,$$

and we have proved equation \textcolor{red}{(3.37)}, namely

$$Q(H, G) = \sum_{r \in A_N} \exp(-N \left[ \tau(H, r, G) + o(N) \right]),$$

with $|o(N)| \leq c_0 \log N/N$.

Assume that $C(H, G) < d$. Let $U_N = \exp(-N[T(H, r, G) + o(N)])$. Partition $A_N$ into two subsets $M$ and $R$ defined by

$$M = \{ r \in A_N \mid KL(G, J) > 4d \} \quad \text{and} \quad R = \{ r \in A_N \mid KL(G, J) \leq 4d \}$$

so that $Q(H, G) = S(R) + S(M)$ with $S(R) = \sum_{r \in R} U_N$ and $S(M) = \sum_{r \in M} U_N$. For $N > N_4 = N_0 + (c_0/2d)^2$ and $r \in M$, the definitions of $U_N$ and $T(H, r, G)$ yield

$$-\frac{1}{N} \log(U_N) \geq KL(G, J) - C_0 \log N/N \geq 4d - C_0 \log N/N \geq 2d.$$

Since $\text{card}(M) \leq (N + 1)^g$, this entails

$$S(M) \leq (N + 1)^g e^{-2N^d} \leq e^{-N^d} \quad (3.46)$$

for $N > N_5 = N_4 + (2g^2/d)^2$. By definition of $\kappa(H, r, G)$, when $r \in R$, one must have $\kappa < KL(G, J) \leq 4d$, and \textcolor{red}{(3.44)} yields for $N > N_5$

$$|KL(G, J) - KL(G, I)| \leq e^{\kappa/a} (c_0^2/N) \leq c_3/N$$

with $c_3 = c_0^2 e^{4d/a}$. This forces $-\frac{1}{N} \log(U_N) = \tau(H, r, G) + o_1(N)$ with $|o_1(N)| \leq |o(N)| + c_3/N \leq c_4 \log(N)/N$ with $c_4 = c_0 + c_3$. By definition of $S(R)$, this implies, for $N > N_5$

$$S(R) = \sum_{r \in R} \exp(-N[\tau(H, r, G) + o_1(N)]). \quad (3.47)$$

By definition of $C(H, G)$, one has $\tau(H, r, G) \geq C(H, G)$ so that \textcolor{red}{(3.37)} yields

$$S(R) \leq (N + 1)^g \exp(-N \ C(H, G) + c_4 \ log(N)) \leq \exp(-N \ C(H, G) + c_5 \ log(N)) \quad (3.48)$$

for $N > N_5$ and $c_5 = c_4 + g^2 + 1$. Equation \textcolor{red}{(3.46)} then yields $S(M) \leq e^{-Nd} \leq e^{-NC(H, G)}$, since $C(H, G) \leq d$. Combining this with \textcolor{red}{(3.48)} we obtain for $N > N_5$ and $c_6 = 3c_5$,

$$Q(H, G) = S(M) + S(R) \leq (1 + Nc_6) e^{-NC(H, G)} \leq \exp(-N \ C(H, G) + c_6 \ log(N)).$$

This yields the large deviations upper bound

$$\frac{1}{N} \log Q(H, G) \leq -C(H, G) + c_6 \ log N/N. \quad (3.49)$$
By definition of $C(H, G)$, there exists a matrix $w = w(H, G, N) \in K(H)$ such that

$$C(H, G) \leq \tau(H, w, G) \leq C(H, G) + 1/\sqrt{N}. \quad (3.50)$$

Lemma 2.10 shows that for $N > N_0 = N_5 + \frac{2^2}{\alpha R_1}$, there is an $N$-rational matrix $s \in K_N(H)$ such that $\text{supp}(s) = \text{supp}(w)$ and $\|s - w\| \leq g/N$. Equation (3.50) implies $\tau(H, w, G) \leq d + 1$ so that $\tau(H, s, G)$ must be finite since $\text{supp}(s) = \text{supp}(w)$. A fortiori $KL(G, \Psi(H, s))$ is finite, implying $\text{supp}(\Psi(H, s)) \subset \text{supp}(s)$. However, since $s \in K_N(H)$, equation (3.43) forces $\text{supp}(J(H, s)) = \text{supp}(\Psi(H, s))$, so that $KL(G, J(H, s))$ must also be finite. Therefore, one has $s \in A_N$. The Holder continuity in $r$ for the composite cost function $\tau(H, r, G)$ was established above by (3.33), which we now apply to get $c_7 = \eta(d + 1, a, P)$ such that

$$|\tau(H, s, G) - \tau(H, w, G)| \leq c_7 \|s - w\|^{1/2} \leq c_7 \sqrt{g}/\sqrt{N}. \quad (3.51)$$

Set $c_8 = 1 + c_7 \sqrt{g}$ and $N_7 = N_0 + c_8^2$. Using (3.50), we get for $N > N_7$

$$\tau(H, s, G) \leq \tau(H, w, G) + c_7 \sqrt{g}/\sqrt{N} \leq C(H, G) + c_8/\sqrt{N} \leq d + 1. \quad (3.52)$$

Then, $\kappa(H, s, G) \leq KL(G, \Psi(H, s)) \leq \tau(H, s, G)) \leq d + 1$. For $N > N_7$, since $s \in A_N$, apply (3.44) to the triple $(H, s, G)$ to get

$$|KL(G, J(H, s)) - KL(G, \Psi(H, s))| \leq c_8/N$$

with $c_8 = c_0^2 e^{(d+1)/a}$. Since $d \geq 1$, this implies

$$KL(G, J(H, s)) \leq KL(G, \Psi(H, s)) + c_8/N \leq d + 1 + c_8/N < d + 2 < 4d$$

for $N > N_8 = N_7 + 1/c_8$. Hence for $N > N_8$, the matrix $s \in A_N$ must belong to $\mathcal{R}$, and (3.47) provides the lower bound

$$S(\mathcal{R}) = \sum_{r \in \mathcal{R}} \exp(-N[\tau(H, r, G) + o_1(N)]) \geq \exp(-N[\tau(H, s, G) + o_1(N)]). \quad (3.53)$$

Due to the bound on $o_1(N)$ and (3.52),

$$\frac{1}{N} \log S(\mathcal{R}) \geq -\tau(H, s, G) + o_1(N) \geq -C(H, G) - c_8(1/\sqrt{N} + \log N/N)$$

for $N > N_8$. Finally, since $Q(H, G) \geq S(\mathcal{R})$, setting $c_9 = 2c_8$ yields

$$\frac{1}{N} \log Q(H, G) \geq -C(H, G) - c_9/\sqrt{N}.$$\

In view of (3.49), we now obtain for $N > N_8$,

$$-C(H, G) - c_9/\sqrt{N} \leq \frac{1}{N} \log Q(H, G) \leq -C(H, G) + c_6/\sqrt{N}. \quad (3.54)$$

This concludes the proof of theorem 3.21. \qed

The Holder continuity property of the composite cost $\tau(H, r, G)$ proved above can be used to derive uniform holder continuity for the one-step cost function $C(H, G)$, as outlined in the next theorem.
Theorem 3.22. Fix $0 < a < 1, d > 0$, and the parameter set $\mathcal{P}$ (definition 2.2). Consider any histograms $H, G, H', G'$ in the compact set $\mathcal{H}(a)$ defined by (3.17), and verifying

$$\text{supp}(G') = \text{supp}(G), \quad \text{supp}(H) = \text{supp}(H'), \quad \|H' - H\| \leq a/F_g, \quad C(H, G) \leq d$$

Then there exists a constant $c \equiv c(d, a, \mathcal{P})$ such that

$$|C(H', G') - C(H, G)| \leq c \left(\|H' - H\|^{1/2} + \|G' - G\|\right).$$

Proof. this proof is quite technical and is displayed in the Appendix C. □

3.5 Computation of the one-step transition cost

One of the main goals of this paper was to develop a computationally implementable set of rigorous large deviations formulas in order to compute numerically the most likely evolutionary paths connecting any known initial histogram $H$ to any histogram $G$ observed at some later time $T$. In fact, for many experimental observations of $\{H, G\}$ the time $T$ may not even be known precisely. To this end, we will provide an explicitly computable expression for the one-step cost $C(H, G)$ defined by the theoretical formula (3.27). This formula requires minimizing the explicit convex function $\tau(H, r, G)$ over all $g \times g$ matrices $r \in K(H)$. We discovered that for small mutation rates $m \leq 10^{-6}$ and large population sizes $N \geq 10^6$, the minimization of $\tau(H, r, G)$ can be done explicitly using an asymptotic expansion, and this minimization procedure provides a very accurate approximation of $C(H, G)$ with error terms of order $m^2 < 10^{-12}$.

The explicit computation of precise approximations for $C(H, G)$ is here deliberately restricted to the simpler case when both $H$ and $G$ are interior histograms, i.e. verify $H(j) > 0$ and $G(j) > 0$ for all $j$. This enables a first order Taylor expansion of $C(H, G)$ in the mutation rate $m$. For more generic $H, G$ having some coordinates equal to 0, good approximations of $C(H, G)$ are feasible but are more complicated because Taylor formulas in $m$ are no longer valid (see remarks below).

Definition 3.23. A histogram $H \in \mathcal{H}$ is called interior if $H(j) > 0$ for all $1 \leq j \leq g$. We denote the open set of interior histograms as $\hat{\mathcal{H}}$.

We will prove below that $C(H, G)$ is smooth in $H, G$ whenever $H, G \in \hat{\mathcal{H}}$. Recall (see eq. 2.24) that matrices $r$ in the convex set $K(H)$ satisfy

$$r_{j,k} \geq 0 \text{ for all } (j,k), \quad \sum_k r_{j,k} < F_j H(j) \text{ for all } j, \quad r_{j,k} = 0 \text{ when } q_{j,k} = 0. \quad (3.57, 3.58, 3.59)$$

The interior $\hat{K}(H)$ of $K(H)$ is the set of all $r$ verifying (3.57), (3.58), and $r_{j,k} > 0$ whenever $q_{j,k} > 0$. The following theorem gives an explicit computation of the one-step cost function $C(H, G)$.

Theorem 3.24. Fix the vector $F = \{F_1, \ldots, F_g\}$ of growth factors, the transition matrix $Q = \{q_{j,k}\}$ for mutations, and any $0 < a < 1$. Let $\Gamma(a) \subset \hat{\mathcal{H}}$ be the set of all histograms $H$ such that $H(j) > a$ for all $j = 1, \ldots, g$. There is then a constant $c \equiv c(a, F, Q) > 0$ such that for all
Let \( H, G \in \Gamma(a) \), and all mutation rates \( 0 \leq m < c \), the transition cost \( C(H, G) \) is a finite \( C^\infty \) function of \( (m, H, G) \). Moreover, \( C(H, G) \) has an explicit first-order expansion in \( m \) given by

\[
C(H, G) = KL(G, \Phi(H)) + m \sum_{j,k} F_j H(j) q_{j,k} [1 - U_k / U_j] + O(m^2),
\]

where

\[
\Phi_j(H) = F_j H(j) / \langle F, H \rangle, \quad KL(G, \Phi(H)) = \sum_i G(i) \log G(i) / \Phi_j(H), \quad U_j = \exp \left( \frac{G(j)}{F_j H(j)} \right)
\]

Proof. Denote \( f_{j,k} = q_{j,k} F_j H(j) \). For \( r \in K(H) \) recall that \( \tau(H, r, G) = \text{mut}(H, r) + KL(G, \Psi(H, r)) \)

\[
\Psi_i = \Psi_i(H, r) = \frac{1}{\langle F, H \rangle} \left( F_i H(i) - \sum_k r_{i,k} + \sum_k r_{k,i} \right),
\]

\[
KL(G, \Psi(H, r)) = \sum_i G(i) \log G(i) / \Psi_i,
\]

\[
\text{mut}(H, r) = \sum_{(j,k) \in \text{supp}(Q)} [m f_{j,k} + r_{j,k} \log r_{j,k} - r_{j,k} - r_{j,k} \log(m f_{j,k})],
\]

To minimize \( \tau(H, r, G) \), we only need to consider \( r \in K(H) \) with \( \tau(H, r, G) \) finite, which holds iff \( \Psi_i(H, r) > 0 \) for all \( i \). Due to convexity in \( r \) of \( \tau(H, r, G) \) we first seek to minimize \( \tau(H, r, G) \) over \( r \in K(H) \). Such a minimizing \( r \) must verify

\[
\partial_{r_{j,k}} \tau(H, r, G) = 0 \quad \text{whenever} \quad q_{j,k} > 0.
\]

The set \( \{(j, k) | q(j, k) > 0\} \) is by definition the support \( \text{supp}(Q) \) of the fixed transition matrix \( Q = (q_{j,k}) \). For \( (j, k) \in \text{supp}(Q) \), basic derivations and algebra reduce (3.61) to the system

\[
0 = \log \frac{r_{j,k}}{m f_{j,k}} - \sum_i \frac{G(i)}{\Psi_i} \partial_{r_{j,k}} \Psi_i
\]

The coefficients \( p(i, j, k) = \partial_{r_{j,k}} \Psi_i \) are constants given by

\[
\langle F, H \rangle p(i, j, k) = \begin{cases} -1, & i = j \neq k, \\ 1, & j \neq k = i, \\ 0, & \text{otherwise}. \end{cases}
\]

Let \( d = \text{card}(\text{supp}(Q)) \). Define the vector \( x \in (\mathbb{R}^+)^d \) by \( x_{j,k} = r_{j,k} / (m f_{j,k}) \) for \( q_{j,k} > 0 \). Note the double-index notation for the coordinates \( x_{j,k} \) of the vector \( x \). Substitute (3.66) into (3.65) to obtain

\[
\log x_{j,k} = -G(j) / A_j + G(k) / A_k,
\]

where the \( A_i = A_i(x) \) are given by

\[
A_i(x) = F_i H(i) + \sum_k m [-f_{i,k} x_{i,k} + f_{k,i} x_{k,i}].
\]
To enforce (3.58), the vector \( x \) must verify, for all \( 1 \leq i \leq g \), the set of strict linear constraints

\[
F_i H(i) - m \sum_k f_{i,k} x_{i,k} > 0.
\]

(3.68)

Call \( \mathcal{O} \) the open set of all \( x \in (\mathbb{R}^+)^g \) verifying (3.68). Rewrite (3.67) as an implicit equation \( W(m, x) = 0 \) for \( x \in \mathcal{O} \) where the function \( W(m, x) \) is given by

\[
W_{j,k}(m, x) = x_{j,k} - \exp\left(-G(j)/A_j + G(k)/A_k\right)
\]

for all \((j, k) \in \text{supp}(Q), x \in \mathcal{O}, \) and \( m \geq 0 \). We can treat \( W(m, x) \) as a vector-valued function \( W(m, x) : (\mathbb{R}^+)^d \to \mathbb{R}^d \). Then \( W(m, x) \) is of class \( C^\infty \) in \((m, x), \) as well as in \( H, G \in \mathcal{H}^c \). For \( m = 0 \) and \((j, k) \in \text{supp}(Q), \) the system \( W(m, x) = 0 \) has a unique solution \( \bar{x} \in \tilde{K} \) given by

\[
\bar{x}_{j,k} = \exp\left(-G(j)/[F_j H(j)] + G(k)/[F_k H(k)]\right).
\]

(3.69)

For all \((s, t) \in \text{supp}(Q)\) and all \( i \), the derivatives \( a(i, s, t) = \partial_{x_s,t} A_i(m, x) \) verify \( a(i, s, t) = mb(i, s, t) \) where the only nonzero terms of \( b(i, s, t) \) are \( b(i, i, t) = -f_{i,t} \) and \( b(i, s, i) = f_{s,i} \). The Jacobian \( JAC(m, x) \) of \( W(m, x) \) has coefficients \( \partial_{x_s,t} W_{j,k} \) given for all \((j, k) \) and \((s, t) \) in \( \text{supp}(Q) \) by

\[
\partial_{x_s,t} W_{j,k} = \delta_{(j,k),(s,t)} - m Y_{(j,k),(s,t)} \exp\left(-G(j)/A_j + G(k)/A_k\right)
\]

where \( Y_{(j,k),(s,t)} = b(j, s, t) G(j)/A_j^2 - b(k, s, t) G(k)/A_k^2 \), and \( \delta_{(j,k),(s,t)} \equiv \delta_{j,s} \delta_{k,t} \) is the Kronecker delta function. At the point \((0, \bar{x})\), the Jacobian \( JAC(0, \bar{x}) \) is hence equal to the identity matrix and thus invertible. The classical implicit function theorem then applies to \( W(m, x) = 0 \) and provides \( c = c(a, F, Q) > 0 \) such that the equation \( W(m, x) = 0 \) has a unique solution \( x^*(m) \in \mathcal{O} \) for \( m \leq c \) and \( H, G \in \Gamma(a) \). The same classical theorem implies that \( x^*(m) = x(m, H, G) \) is of class \( C^\infty \) in \((m, H, G)\). Define \( r^*(m) \) by

\[
r_{j,k}^*(m) = \begin{cases} m x_{j,k}^*(m), & \text{when } q_{j,k} > 0, \\ 0, & \text{otherwise.} \end{cases}
\]

Then, \( r^*(m) \) is a solution of (3.64) and inherits from \( x^*(m) \) the \( C^\infty \) smoothness in \((m, H, G)\). Moreover, \( r^*(m) \) verifies the constraints (3.57)–(3.59) since \( x^*(m) \in \mathcal{O} \). The positivity of all coordinates of \( x^*(m) \) implies \( r_{j,k}^*(m) > 0 \) for \( q_{j,k} > 0 \). Hence, \( r^*(m) \in \tilde{K}(H) \) for \( m < c \) and solves (3.64). The strict convexity of \( \tau(H, r, G) \) on the open convex set \( \tilde{K}(H) \) forces \( r^*(m) \) to be the unique minimizer of \( \tau(H, r, G) \) for \( r \in K(H) \) so that \( C(H, G) = \tau(H, r(m), G) \). The function \( (H, r, G) \to \tau(H, r, G) \) is \( C^\infty \) for \( H, G \in \mathcal{H}^c \) and \( r \in \tilde{K}(H) \). Hence the function \((m, H, G) \to C(H, G) \) is also \( C^\infty \) for \( m < c \) and \( H, G \in \Gamma(a) \).

To obtain the first-order expansion for the solution \( r^*(m) \) and the cost function \( C(H, G) \), define

\[
U_j = \exp\left(G(j)/[F_j H(j)]\right) \quad \text{and} \quad E_{j,k} = U_k/U_j \quad \text{for all} \quad 1 \leq j, k \leq g,
\]

(3.70)

so that \( \bar{x}_{j,k} = f_{j,k} E_{j,k} = F_j H(j) q_{j,k} U_k/U_j \). To simplify notation, write \( u \simeq v \) whenever \( u(m) = v(m) + O(m^2) \) as \( m \to 0 \). As \( m \to 0 \), the differentiability of \( x^*(m) \) gives

\[
r_{j,k}^*(m) = m x_{j,k}^*(m) \simeq m \bar{x}_{j,k} = m f_{j,k} E_{j,k}.
\]

(3.71)
Inserting (3.71) into (3.63) yields \( \log(H, r^*(m)) \simeq m \mu \) with \( \mu = \sum_{j,k} f_{j,k} (1 - E_{j,k} + E_{j,k} \log E_{j,k}) \).

Since \( \log(E_{j,k}) = -G(j)/[F_j H(j)] + G(k)/[F_k H(k)] \), we have

\[
\mu = \sum_{j,k} f_{j,k} (1 - E_{j,k}) + \sum_{j,k} q_{j,k} E_{j,k} \left( -G(j) + G(k) \frac{F_j H(j)}{F_k H(k)} \right). \tag{3.72}
\]

Substituting (3.71) into (3.61) we obtain

\[
\Psi_j = \Psi_j (H, r^*(m)) \simeq (F_j H(j) - m t_j) / \langle F, H \rangle;
\]

\[
t_j = \sum_k F_j H(j) q_{j,k} E_{j,k} - \sum_k F_k H(k) q_{k,j} E_{k,j}.
\]

This yields \( \log(\Psi_j) \simeq \log(F_j H(j)/\langle F, H \rangle) - m \frac{t_j}{F_j H(j)} \). Recall that \( \Phi_j(H) = F_j H(j)/\langle F, H \rangle \) is the population histogram at the end of the daily deterministic growth starting with histogram \( H \).

Substituting the expansion of \( \log(\Psi_j) \) in (3.20) we obtain

\[
KL(G, \Psi(H, r^*(m)) \simeq \kappa + m \eta;
\]

\[
\kappa = KL(G, \Phi) = \sum_j G(j) \log \left( \frac{G(j)}{\Phi_j} \right) > 0; \quad \eta = \sum_j \frac{G(j) t_j}{F_j H(j)}.
\]

This implies \( \eta = \sum_{j,k} \left[ G(j) q_{j,k} E_{j,k} - \frac{G(j)}{\langle F_j H(j) \rangle} F_k H(k) q_{k,j} E_{k,j} \right] \). Exchange \( j \) and \( k \) in the second term of the previous sum to get

\[
\eta = \sum_{j,k} q_{j,k} E_{j,k} \left[ G(j) - G(k) \frac{F_j H(j)}{F_k H(k)} \right]. \tag{3.73}
\]

Hence, \( C(H, G) = \tau(H, r^*(m), G) = \kappa + m (\eta + \mu) + O(m^2) \) is the first-order expansion of \( C(H, G) \).

Combine (3.72) and (3.73) to obtain \( \eta + \mu = \sum_{j,k} F_j H(j) q_{j,k} (1 - U_k/U_j) \), which concludes the proof of (3.60) and of the theorem.

In a future paper, we will analyze how approximations of \( C(H, G) \) for small \( m \) extend to histograms \( H \) and \( G \) which are not interior histograms (i.e. which have some coordinates equal to 0). The computation of one-step cost \( C(H, G) \) for such "boundary" histograms is still feasible but considerably more complex than for the interior histograms. For instance, when \( H \in H \) is not an interior histogram, the first order expansion of \( C(H, G) \) contains computable terms of order \( \log m \) as well as terms of order \( m \), and differentiability in \( H \) for \( C(H, G) \) involves restrictive constraints on \( H \).

## 4 Large Deviations for Evolutionary Trajectories

For fixed large population size \( N \), and fixed small mutation rate \( m \), the stochastic genetic evolution of the daily populations \( \text{pop}_n \) across discrete time \( n \) is modeled here by the random histogram path \( \{H_1, \ldots, H_T\} \) where \( H_n \) is the histogram of \( \text{pop}_n \). The discrete Markov chain \( \{H_n, n = 1, 2, \ldots\} \) belongs to the compact convex state space \( H \), and has Markov transition kernel \( Q_N \). The large
deviations analysis of daily transitions discussed in section \( \text{[3]} \) provides an explicit formula for the one-step rate function \( C(H, G) \geq 0 \), and for \( N \) large, the precise large deviations approximation

\[
\frac{1}{N} \log Q_N(H, G) \approx -C(H, G)
\]

that roughly says that \( Q_N(H, G) \approx \exp(-NC(H, G)) \). In this section, we rigorously extend this one-step large deviations approximation to the "path space" of random histogram trajectories for the Markov chain describing the time-evolution of histograms.

### 4.1 Large deviations for a single trajectory

**Definition 4.1.** For any fixed time horizon \( T \geq 2 \), denote \( \Omega_T \equiv \mathcal{H}^T \) the path space of all possible histogram trajectories \( H = [H_1 \ldots H_T] \) with all \( H_n \in \mathcal{H} \).

**Definition 4.2.** The distance between trajectories \( H \) and \( H' \) is defined by

\[
\|H - H'\| = \max_{n=1 \ldots T} \|H_n - H'_n\|.
\]

Intuitively a ball \( B(H) \subseteq \Omega_T \) of small radius can then be viewed as a "thin tube" of trajectories around the tube axis \( H \). Since the random population histograms \( H_n \) are \( N \)-rational, our Markov chain actual trajectories will always belong to the set of \textit{N-rational} trajectories \([H_1 H_2 \ldots H_T]\) such that \( H_n \in \mathcal{H}_N \) for \( 1 \leq n \leq T \).

Next, we construct an explicit rate functional \( \lambda(H) \geq 0 \) defined for all \( H \in \Omega_T \), which extends the one-step rate function \( C(H, G) \) to a \( T \)-steps rate function defined on the path space. Furthermore, will prove that \( \exp(-N \lambda(H)) \) roughly approximates the probability that the trajectory of the underlying Markov chain remains within any given very thin tube of trajectories centered around \( H \).

**Definition 4.3.** For any trajectory \( H \in \Omega_T \), we define the large deviations rate function \( \lambda : \Omega_T \to [0, \infty) \) by

\[
\lambda(H) = \sum_{n=1}^{T-1} C(H_n, H_{n+1}). \tag{4.1}
\]

We then define the large deviations set functional \( \Lambda(\Gamma) \) for any \( \Gamma \subseteq \Omega_T \) by

\[
\Lambda(\Gamma) = \inf_{H \in \Gamma} \lambda(H). \tag{4.2}
\]

The next theorem justifies calling \( \lambda \) the rate function on path space. This is a uniform large deviations result for single \( N \)-rational trajectories, and will be a key point to obtain large deviations estimates for subsets of the path space (see Theorem \( \text{[4.9]} \)).

**Definition 4.4.** For any path \( H \in \Omega_T \), define its essential minimum \( b(H) = \min_{n=1 \ldots T} b(H_n) \) where \( b(H_n) \) is the essential minimum of \( H_n \) defined in def \( \text{[2.6]} \).

**Theorem 4.5.** Fix the path length \( T \geq 2 \), two positive constants \( a, d > 0 \), and parameters \( \mathcal{P} \). Denote \( H = [H_1, \ldots, H_T] \in \Omega_T \) the random path of the Markov chain describing daily transitions. Then \( (d, a, \mathcal{P}) \) determine positive constants \( c \) and \( N_0 \) such that the following holds. For any given fixed \( N \)-rational path \( h = [h_1 h_2 \ldots h_T] \in \Omega_T \) such that \( \lambda(h) \leq d \) and \( b(h) \geq a \), one has

\[
\frac{1}{N} \log P(H = h \mid H_1 = h_1) = -\lambda(h) + o(N) \text{ for all } N > N_0 \tag{4.3}
\]

with \( |o(N)| \leq cT/\sqrt{N} \).
Proof. For all $n \leq T - 1$, we must have $C(h_n, h_{n+1}) \leq d$ since $\lambda(h) \leq d$. Hence, Theorem 3.21 provides constants $N_0$ and $c$, determined by $(d, a, \mathcal{P})$, such that the Markov transition kernel $Q$ verifies $\frac{1}{N} \log Q(h_n, h_{n+1}) = -C(h_n, h_{n+1}) + \eta_n(N)$ with $|\eta_n(N)| \leq c/\sqrt{N}$ for all $h$ as above, $n \leq T - 1$, and $N > N_0$. The Markov property yields $\log P(H = h | H_1 = h_1) = \sum_{n=1}^{T-1} \log Q(h_n, h_{n+1})$. Since $\lambda(h) = \sum_{n=1}^{T-1} C(h_n, h_{n+1})$, we obtain

$$\frac{1}{N} \log P(H = h | H_1 = h_1) = -\lambda(h) + o(N)$$

(4.4)

with $o(N) \leq \sum_{n=1}^{T-1} |\eta_n(N)| \leq cT/\sqrt{N}$ for $N > N_1$. \hfill $\square$

Next, we prove that the rate function $\lambda(H)$ has a uniform Hölder property.

Theorem 4.6. Fix two constants $d > 0$, $a > 0$ and the process parameters $\mathcal{P}$. Consider any (deterministic and fixed) path $h \in \Omega_T$ such that $b(h) \geq a$ and $\lambda(h) \leq d$. Let $H \in \Omega_T$ be any path such that $\text{supp}(H_n) = \text{supp}(h_n)$ for all $n = 1, \ldots, T$ and verifying

$$b(H) \geq a \quad \text{and} \quad \|H - h\| \leq a/F_g.$$  

(4.5)

Then there is a constant $c \equiv c(d, a, \mathcal{P})$ such that for all $T$, one has

$$|\lambda(H) - \lambda(h)| \leq cT \|H - h\|^{1/2}.$$  

(4.6)

Proof. By Theorem 3.22 there is a constant $c_1(d, a, \mathcal{P})$ such that, for all $n = 1, \ldots, T - 1$.

$$|C(H_n, h_{n+1}) - C(h_n, h_{n+1})| \leq c_1 \left(\|H_n - h_n\|^{1/2} + \|H_{n+1} - h_{n+1}\|\right)$$

and hence $|C(H_n, h_{n+1}) - C(h_n, h_{n+1})| \leq 2c_1 \|H - h\|^{1/2}$. By definition of $\lambda$, this yields

$$|\lambda(H) - \lambda(h)| \leq \sum_{n=1}^{T-1} |C(H_n, h_{n+1}) - C(h_n, h_{n+1})| \leq cT \|H - h\|^{1/2},$$

which proves (4.6) with $c = 2c_1$. \hfill $\square$

4.2 Large deviations for sets of trajectories

Theorem 4.7 now sets the stage for a large deviations result for sets of trajectories. In particular, we will show that for large $A$, the probability of observing random paths $H$ of population histograms such that $\lambda(H) > A$ is bounded above by $e^{-cAN}$ for some constant $c > 0$. This will naturally lead to the main result of this section given by Theorem 4.9. We will first need to define an open neighborhood of paths that is analogous to Definition 3.21 for histograms.

Definition 4.7. For any $\Gamma \subset \Omega_T$, define the open neighborhood $V_N(\Gamma)$ as the union of all balls $V_N(H)$ with radius $\frac{2}{3N}$ and center $H \in \Gamma$. Denote $B_N(\Gamma)$ the (finite) set of all $N$-rational paths in $V_N(\Gamma)$. Define also $b(\Gamma) = \inf_{H \in \Gamma} b(H)$.

Theorem 4.8. Fix $0 < a < 1$ and the model parameters $\mathcal{P}$ (definition 2.22). Then, there is a constant $c \equiv c(a, \mathcal{P}) > 0$ such that for all $A > cT$ and for any initial histogram $Z$ such that $b(Z) \geq a$, the random paths $H$ starting at $H_1 = Z$ will verify

$$P(\lambda(H) > A \quad \text{and} \quad b(H) > a \mid H_1 = Z) \leq e^{-\frac{c}{2T}NA}$$

(4.7)

for all $N > cA/T$. 27
Proof. Recall the notations used to prove Theorem 3.21. Consider any N-rational $H,G \in \mathcal{H}(a)$ and $r \in K_N(H)$. Given $(H_n = H, R_n/N = r)$, the population histogram $J_n$ after growth and mutations is a function $J_n = J(H,r)$ of $H,r$ (see (2.27)), with $J(H,r)$ very close to $I = \Psi(H,r)$ (see equation (2.10)), and we defined the set $A_N$ of acceptable mutations matrices $r = R_n/N$ by

$$A_N = A_N(H,G) = \{ r \in K_N(H) \mid KL(G,J) < \infty \text{ and } KL(G,I) < \infty \}. \quad (4.8)$$

We also defined the approximate transition cost $\bar{\tau}(H,r,G) = \text{mut}(r,H) + KL(G,J(H,r))$ in Theorem 3.21. The proof of Theorem 3.21 yielded equations (3.37) and (3.44), which provided constants $c_0(o,P)$ and $N_0(o,P)$ such that for $N > N_0$,

$$\text{supp}(J(H,r)) = \text{supp}(I(H,r)) \quad \text{with} \quad \|J - I\| \leq c_0/N, \quad \text{when} \quad r \in A_N, \quad (4.9)$$

and

$$Q(H,G) = \sum_{r \in A_N} \exp(-N[\bar{\tau}(H,r,G) + o(N)]), \quad (4.10)$$

where $|o(N)| \leq c_0 \log N/N$.

Suppose that $C(H,G) > A$ so that $\tau(H,r,G) \geq C(H,G) > A$ for $r \in K(H)$. From (3.16), we have a constant $c \equiv c(a,P)$ such that $\text{mut}(H,r) \leq c$ whenever $b(H) \geq a$ and $r \in K(H)$. We then have $KL(G,I) = \tau(H,r,G) - \text{mut}(H,r) > A - c$. Hence (3.28) yields $\beta(G,I) = \max_{k \in \text{supp}(G)} 1/J_k(H,r) \geq KL(G,I) > A - c$. Therefore, for some $j \in \text{supp}(G)$, one has $I_j(H,r) < 1/(A - c)$ so that $J_j(H,r) < 1/(A - c) + c_0/N$ by (4.9). This yields $1/J_j(H,r) > 3A/(4A - c)$ provided $N > N_0 + 8c_0A$. A fortiori, we get $\beta(G,J) = \max_{k \in \text{supp}(G)} 1/J_k(H,r) > \frac{3}{4}(A - c)$. For $N > N_0 + 8c_0A$ and $r \in A_N$,

$$\bar{\tau}(H,r,G) \geq KL(G,J) > b(G)\beta(G,J) - \log g > \frac{3a}{4}(A - c) - \log(g) = \frac{3aA}{4} - c_1 \quad (4.11)$$

by (3.28) with $c_1 = \log(g) + 3ac/4$. Since $\text{card}(E_N) \leq [(N + 1)F_0]^g$, combining (4.11) and (4.10) yields

$$Q(H,G) \leq [(N + 1)F_0]^g \exp \left( -N \left[ \frac{3a}{4} - c_1 \right] + c_0 \log N \right) \leq \exp \left( N \left[ -\frac{3a}{4} + c_2 \right] \right) \quad (4.12)$$

for $N > N_0 + 8c_0A$ with $c_2 = c_1 + c_0 + g^2(\log F_g + 1)$. Let $\Gamma(A) = \{ H \in \Omega_T \mid H_1 = Z, \lambda(H) > A, b(H) > a \}$. Let $\Gamma_N(A)$ be the set of $N$-rational paths in $\Gamma(A)$. Then, $P(H \in \Gamma_N(A) \mid H_1 = Z) = P(H \in \Gamma(A) \mid H_1 = Z)$ since all random paths $H$ are $N$-rational. For $H \in \Gamma_N(A)$, the relation $\lambda(H) > A$ provides at least one time step $\nu = \nu(H) \leq T$ such that $C(H_\nu, H_{\nu+1}) > A/T$. Apply (4.12) to $(H_\nu, H_{\nu+1})$ to get

$$P(H = h \mid H_1 = Z) \leq P(H_{\nu+1} = h_{\nu+1} \mid H_\nu = h_\nu) \leq \exp \left( N \left[ -\frac{3a}{4T}A + c_2 \right] \right)$$

for $N > N_0 + 8c_0A/T$ and $H \in \Gamma_N(A)$ and any fixed path $h \in \Gamma_N(A)$. Since $\text{card}(\Gamma_N(A)) \leq (N + 1)^Tg$, this yields

$$P(H \in \Gamma(A) \mid H_1 = Z) = P(H \in \Gamma_N(A) \mid H_1 = Z)$$

$$\leq (N + 1)^Tg \exp \left( N \left[ -\frac{3a}{4T}A + c_2 \right] \right)$$

$$\leq \exp \left( N \left[ -\frac{3a}{4T}A + c_3 \right] \right)$$
for $N > N_0 + 8c_0 A/T$ with $c_3 = c_2 + Tg$. Impose now $A/T > 4c_3$ to get

$$P(H \in \Gamma(A) \mid H_1 = Z) \leq \exp \left( -\frac{a}{2T} N A \right)$$

for $N > N_0 + 8c_0 A/T$. Set $c = 4c_3 + N_0/8c_0 + 16c_0$, so that the simpler constraints $A/T > c$ and $N > cA/T$ force $A/T > c_3$ and $N > N_0 + 8c_0 A/T$. This completes the proof. \qed

**Theorem 4.9.** Let $\Lambda$ be the large deviations set functional of the Markov chain of population histograms defined by (1.1) and (1.2). Fix the process parameters $P$, the path length $T$, and $0 < a < 1$. Denote $H$ a random histogram trajectory of our Markov chain, starting at some deterministic histogram $Z$ such that $b(Z) \geq a$. Let $T = 0$ be any set of histogram trajectories starting at $Z$. Assume that $\Lambda(\Gamma) \leq L$ for some fixed finite $L > 0$, and that $b(\Gamma) \geq a$ (see def 1.7).

Let $V_N(\Gamma)$ be the open neighborhood of $\Gamma$ with (small) radius $\frac{2}{\sqrt{N}}$ (see def 4.4). Then $(T,a,P,L)$ determine constants $c$ and $N_0$ such that for all $N > N_0$, one has the uniform large deviations result

$$\frac{1}{N} \log P(H \in V_N(\Gamma) \mid H_1 = Z) = -\Lambda(\Gamma) + o(N)$$

(4.13)

holds with $|o(N)| \leq c/\sqrt{N}$. This yields the asymptotic large deviations limit

$$\lim_{N \to \infty} \frac{1}{N} \log P(H \in V_N(\Gamma) \mid H_1 = Z) = -\Lambda(\Gamma).$$

**Proof.** Let $\Gamma$ be any subset of $\Omega_T$ of paths starting at $Z$ with $b(\Gamma) \geq a$ and finite $\Lambda(\Gamma) = L > 0$.

For any $A > 0$, let $E(A)$ be the set of all $N$-rational paths $h \in \Omega_T$ with $h_1 = Z$ such that $\lambda(h) > A$ and $b(h) > a/2$. Theorem 4.3 provides $c = c(a,P)$ such that $P(H \in E(A) \mid H_1 = Z) \leq e^{-\frac{a}{2T} N A}$ for all $A > cT$ and $N > cA/T$. Set $d = T(c + 4L/a) + L$. Set $A = d$, which forces $A > cT$ and $\frac{a}{2T} A > 2L$ so that

$$P(\omega \in E(d) \mid H_1 = Z) \leq e^{-2NL}$$

(4.14)

for all $N > N_1 = cd/T$. Let $W_N(\Gamma) \subset B_N(\Gamma) \subset V_N(\Gamma)$ be the set of $N$-rational paths $h \in V_N(\Gamma)$ such that $\lambda(h) \leq d$ and $h_1 = Z$. Then, $B_N(\Gamma) \subset W_N(\Gamma) \cup E(d)$ so that (4.14) yields

$$P(H \in B_N(\Gamma) \mid H_1 = Z) \leq P(H \in W_N(\Gamma) \mid H_1 = Z) + e^{-2NL}$$

(4.15)

for $N > N_1$. Due to Theorem 4.6 there is a constant $c_1 > 0$ determined by $(d,a,P)$, and thus by $(T,L,a,P)$, such that for any paths $h,h' \in \Omega_T$, the inequality

$$\lambda(h') - \lambda(h) \leq c_1 \|h' - h\|^{1/2}$$

(4.16)

must hold whenever the following holds for all integers $1 \leq n \leq T$:

$$\min \left\{ \lambda(h), \lambda(h') \right\} \leq d, \quad b(h') \geq a, \quad \|h' - h\| \leq a/F_g, \quad \text{supp}(h'_n) = \text{supp}(h_n).$$

Since $b(\Gamma) \geq a$, Lemma 3.2 implies $b(V_N(\Gamma)) \geq a/2$ so that all $h \in W_N(\Gamma)$ verify $\lambda(h) \leq d$ and $b(h) > a/2$. Therefore, Theorem 4.5 provides $c_2$ and $N_2 > N_1$ determined by $(d,a,P)$, and thus by $(T,L,a,P)$, such that

$$\frac{1}{N} \log P(H = h \mid H_1 = Z) = -\lambda(h) + o_1(N)$$

(4.19)
Combining this with (4.22) yields the lower bound verify both (4.17) and (4.18) so that (4.16) applies and yields \( N > N \) for simulation starting with a known initial histogram \( H \) for provided \( N > N \) or \( N > N \). Theorem 4.9 indicates that the likelihood of observing the for \( N > N \) after an arbitrary number of "days". Theorem 4.9 is crucial for computing the most likely random evolutionary path \( h^* \) of a population starting with a known initial histogram \( H \) and ending with another known histogram \( G \) after an arbitrary number of "days". Theorem 4.9 indicates that the likelihood of observing the random histogram trajectory \( H \) within any fixed set of paths \( \Gamma \subset \Omega_T \) is roughly of the order of \( \exp(-N\Lambda(\Gamma)) \), which is an extremely small exponential for \( N \) large whenever \( \Lambda(\Gamma) > 0 \). Therefore,
if $\Lambda(\Gamma) > 0$, then observing a random trajectory $H \in \Gamma$ is a rare event. The actual calculation of $\Lambda(\Gamma)$ involves minimizing the rate function $\lambda(h)$ over all paths $h \in \Gamma$. Let $\Gamma^* \subset \Gamma$ be the set of all such minimizing paths for the rate function $\lambda(h)$. Whenever the rare event $H \in \Gamma$ is realized by some observed random histogram trajectory $H$, then with a conditional probability extremely close to 1, the trajectory $H$ must follow very closely one of the minimizing path $h \in \Gamma^*$. We explicitly quantify this argument in Section 5.

4.3 Mean evolution and zero-cost trajectories

When $\Gamma$ is a thin open tube around a zero cost path $h$, i.e a path such that $\lambda(h) = 0$, then $\Lambda(\Gamma) = 0$, and our next theorem implies that $P(H \in \Gamma)$ will approach 1 at exponential speed for large $N$. This motivates our explicit analysis and construction of all zero-cost paths. In particular, we now show that given an initial histogram $h_1$, the unique zero-cost trajectory starting at $h_1$ is recursively computable by an explicit formula.

**Theorem 4.10.** Fix a path length $T$. A histogram path $h = [h_1, \ldots, h_T] \in \Omega_T$ satisfies $\lambda(h) = 0$ if and only if $h_{n+1} = \zeta(h_n)$ for $1 \leq n \leq T - 1$, where the histogram valued function $H : \zeta(\mathcal{H})$ is defined for all $H \in \mathcal{H}$ by

$$
\zeta_j(h) = \frac{1}{\langle F, H \rangle} \left( F_j H(j) - m \sum_k q_{j,k} F_j H(k) + m \sum_k q_{k,j} F_k H(k) \right), \quad j = 1, \ldots, g. \tag{4.24}
$$

Hence a zero cost path $h$ is uniquely determined by its starting point $h_1$. Consider any zero cost path $h$. Let $V_N(h)$ be the ball with a center at $h$ and radius $\frac{2}{N} \sqrt{N}$ in $\Omega_T$. Then, the initial histogram $h_1$ and the process parameters $P$ determine constants $c$ and $N_0$ such that the random path $H \in \Omega_T$ verifies

$$
1 \geq P(H \in V_N(h) \mid H_1 = h_1) \geq 1 - e^{-c\sqrt{N}} \tag{4.25}
$$

for all $N > N_0$.

**Proof.** Given any two histograms $H, G \in \mathcal{H}$ with finite cost $C(H, G)$, Lemma 3.18 proves the existence of a matrix $r \in K(\mathcal{H})$ such that $C(H, G) = \tau(H, r, G)$. Hence, $C(H, G) = 0$ if and only if $\tau(H, r, G) = 0$, which is equivalent to $\text{mut}(H, r) = KL(G, \Psi(H, r)) = 0$. From proposition 3.9, one has $\text{mut}(H, r) = 0$ iff

$$
r_{j,k} = m q_{j,k} F_j H(j) = m q_{j,k} F_j H(j) \tag{4.26}
$$

for all $j, k$. By definition of the Kullback-Leibler divergence, one has $KL(G, \Psi(H, r)) = 0$ if and only if $\Psi(H, r) = G$. Combine this relation with (2.8) and (4.26) to conclude that $C(H, G) = 0$ if and only if one has

$$
G(j) = \frac{1}{\langle F, H \rangle} \left( F_j H(j) - \sum_k m q_{j,k} F_j H(k) + \sum_k m q_{k,j} F_k H(k) \right)
$$

for all $j$. Hence $C(H, G) = 0$ if and only if $G = \zeta(H)$. Now, for any path $h \in \Omega_T$, the relation $\lambda(h) = 0$ holds if and only if $C(h_n, h_{n+1}) = 0$ for all $1 \leq n \leq T - 1$, which is equivalent to $h_{n+1} = \zeta(h_n)$ for all $1 \leq n \leq T - 1$. Finally, the bound in (4.25) is an immediate consequence of (4.13) applied to the set $F = \{h \mid \lambda(h) = 0\}$.  

31
Next, we demonstrate that as \( n \to \infty \), any infinite zero-cost path \( h = [h_n]_{n=1}^{\infty} \) achieves near fixation of some explicitly determined genotype.

**Definition 4.11.** For any non-empty set \( S \) of genotypes, define the set \( R(S) \) of genotypes reachable from \( S \) as the set of all genotypes such that there is a genotype sequence \( k_1, \ldots, k_T \) of arbitrary length \( T \) such that \( k_1 \in S, k_T = j \), and all \( q_{k_t,k_t+1} \) > 0.

**Theorem 4.12.** Let \( h \) be any zero-cost histogram path of infinite length starting at \( h_1 \in \mathcal{H} \). Denote \( S_n = \text{supp}(h_n) \). Then, the set \( S_n \) increases with \( n \) and stabilizes after a finite time \( n_0 \) so that \( S_n = S_\infty \) for all \( n > n_0 \). Moreover, \( S_\infty = R(S_1) \) where \( R(S_1) \) is the set of all genotypes reachable from \( S_1 \) as in the definition 4.11. Let \( s \in S_\infty \) be the fittest genotype within \( S_\infty \), i.e. such that \( F_s = \max_{j \in S_\infty} F_j \). Then, the initial histogram \( h_1 \) and the parameters \( \mathcal{P} \) determine \( m_0 > 0 \) such that for all mutation rates \( m \leq m_0 \),

\[
\lim_{n \to \infty} h_n = h_\infty
\]

where \( h_\infty \) is the unique solution of \( h_\infty = \zeta(h_\infty), \) has support \( \text{supp}(h_\infty) = S_\infty \), and is a \( C^\infty \) function of \( m \) with the first-order expansion

\[
h_\infty(i) \simeq mq_{s,i}/(F_s - F_i) \quad \text{for} \quad i \in (S_\infty \setminus s),
\]

\[
h_\infty(s) \simeq 1 - m \sum_{i \in (S_\infty \setminus s)} q_{s,i}/(F_s - F_i).
\]

**Proof.** The detailed proof is displayed in the Appendix [D].

## 5 Most Likely Evolution from Initial to Terminal Histograms

Theorem 4.11 indicates that for any set \( \Gamma \subset \Omega_T \) of histograms paths, the probability \( P(H \in \Gamma) \) that the random histogram trajectory \( H \) belongs to \( \Gamma \) is roughly estimated by \( \exp(-N\lambda(\Gamma)) \). We now apply this estimate to the set \( \Gamma(H, G) \) of all histograms trajectories \( H = \{H_1, \ldots, H_T\} \) such that \( H_1 = H \) and \( H_T = G \), where the initial and terminal histograms \( H \) and \( G \) are known and fixed. For large \( N \) and any trajectory \( h \in \Gamma(H, G) \), we have shown that \( \frac{1}{N} \log(P(H = h)) \) tends to \( -\lambda(h) \). This indicates that the most likely evolutionary path \( h^* \) "followed" by a random histograms trajectory \( H^* \) connecting \( H \) to \( G \) in \( T \) steps should be a minimizer of \( \lambda(h) \) over all paths \( h \in \Gamma(H, G) \). This is a natural question of interest for laboratory experiments on bacterial genetic evolution, since actual detailed observations of \( H_n \) may be costly and hence may occur only every \( T \) steps. In long experiments, the time \( T \) itself may not even be known precisely, so the reconstruction of the whole most likely evolutionary path connecting \( H \) to \( G \) given only \( H \) and \( G \) should provide insights into key genotype changes necessary to realize a rare evolution from \( H \) to \( G \). We shall also outline how to compute the most likely number of days \( T \) enabling a rare transition from \( H \) to \( G \) in \( T \) steps.

### 5.1 Interior Paths in the space of histograms trajectories

Note that for interior histograms \( H, G \in \mathcal{H} \), the transition cost \( C(H, G) \) is always finite. A path \( H = [H_1 \ldots H_T] \) will be called an interior path if \( H_n \in \mathcal{H} \) are interior histograms for all \( 1 \leq n \leq T \). To develop explicit computational schemes, we focus on interior paths and histograms for the remainder of this paper.
5.2 Sets of Thin Tubes realizing rare events

Definition 5.1. For any small $\eta > 0$ and any set of paths $\Gamma \subset \Omega_T$, we define the $\eta$-neighborhood $U^\eta(\Gamma)$ of $\Gamma$ as the union of all open balls of radius $\eta$ centered at arbitrary $h$ belonging to $\Gamma$. For small $\eta$, the set $U^\eta(\Gamma)$ is a set of thin tubes of paths with “axes” $h \in \Gamma$.

Let $H \in \Omega_T$ be the random trajectory of population histograms. For any closed set of interior paths $E \subset \Omega_T$, one can easily prove that $a = b(E) > 0$. Assume moreover that $\Lambda(E) > 0$. Then for any $0 < A < \Lambda(E)$ and large enough $N > N_0(a, A, P)$, Theorem 4.9 provides the fast-vanishing bound

$$P(H \in E) \leq P(H \in V_N(E)) \leq 2e^{-NA}$$

so that the random events $\{H \in E\}$ and $\{H \in V_N(E)\}$ are both rare events.

We now show (Theorem 5.2) that for large population size $N$, rare evolutionary events $\{H \in E\}$ with finite $\Lambda(E) > 0$ can only be realized by random population evolutions following a very thin tube of radius $1/N$ around one of the (possibly several) paths $h^* = \arg \min_{h \in E} \lambda(h)$.

More precisely, let $T = B_N(H)$ be the thin tube of trajectories centered around the random trajectory $H$ and having vanishingly small radius $1/N$. Fix a set of paths $E$ such that $\Lambda(E) > 0$, so that the event $H \in E$ is a rare event. The next theorem proves that whenever we know that the thin random tube $T$ contains a path belonging to $E$, then with conditional probability extremely close to 1, the random trajectory $H$ itself must be arbitrarily close to a trajectory $h^*$ minimizing the rate function over all paths in $E$. Of course, this result is of higher interest when there is only one such minimizer $h^*$, but the uniqueness of $h^*$ cannot be guaranteed in general. Computing such minimizing paths requires efficient numerical strategies discussed in Section 6.

Theorem 5.2. Fix the path length $T$ and an initial interior histogram $H$. Denote $P_H$ the probability distribution of random histogram paths $H \in \Omega_T$ starting at $H$. Let $E \subset \Omega_T$ be any closed set of interior paths starting at $H$ and satisfying $0 < \Lambda(E) < \infty$. Let $E^* \subset E$ be the set of all paths $h^*$ minimizing the rate function $\lambda(h)$ over all $h \in E$. Then $E^*$ is a closed subset of $E$. For any fixed $\eta > 0$, the $\eta$-neighborhood $U^\eta(E^*)$ of $E^*$ verifies

$$\lim_{N \to \infty} P_H(H \in U^\eta(E^*) \mid H \in V_N(E)) = 1$$

with exponential speed of convergence.

Proof. Set $a = b(E) > 0$ and $L = \Lambda(E) > 0$. Let $E(c) = \{h \in E \mid \lambda(h) \leq c\}$. Then the function $\lambda(h)$ is continuous on $E(2L)$ due to Theorem 4.6 applied to interior paths. Hence, $E(2L)$ is closed and must contain any path minimizing $\lambda(h)$ over $E(2L)$. However, the two sets of minimizers of $\lambda$ over $E$ and over $E(2L)$ are obviously identical. This proves $E^* \subset E(2L) \subset E$.

Theorem 4.6 applied to interior paths provides a constant $c = c(T, L, a, P)$ such that for all $h \in E^*$ and $h' \in E$ with $|h' - h| < a/F_g$, one has $|\lambda(h') - \lambda(h)| \leq c||h' - h||^{1/2}$. Fix any $0 < \eta < a/F_g$. Let $K \subset E$ be the open $\eta$-neighborhood of $E^*$ within $E$, and set $W = E - K$. For each $h' \in K$, there is one $h \in E^*$ with $||h' - h|| < \eta$. Therefore, Theorem 4.6 implies $\lambda(h') < L + c\sqrt{\eta}$. This forces $\lambda(h) \geq L + c\sqrt{\eta}$ for all $h \in W$ so that $\Lambda(W) \geq L + c\sqrt{\eta}$. Apply Theorem 4.9 to get $N_0$ and $c_0$ determined by $(T, L, a, P, \eta)$ such that for $N > N_0$,

$$\frac{1}{N} \log P_H(H \in V_N(W)) = -L - c_0 + o_1(N),$$

$$\frac{1}{N} \log P_H(H \in V_N(E)) = -L + o_2(N),$$

33
with \(|o_1(N)| < c_0/\sqrt{N}\) and \(|o_2(N)| < c_0/\sqrt{N}\). These results yield

\[
P_H(H \in V_N(W)) \leq \exp(-NL - Nc/\sqrt{\eta} + c_0\sqrt{N}),
\]

\[
P_H(H \in V_N(E)) \geq \exp(-NL - c_0\sqrt{N}),
\]

so that

\[
\frac{P_H(H \in V_N(W))}{P_H(H \in V_N(E))} \leq \exp(-Nc/\sqrt{\eta} + 2c_0\sqrt{N}).
\]

Impose \(N > N_0 + 16c_0^2c^2/\eta\) to force \(N(c/2)/\sqrt{\eta} > 2c_0\sqrt{N}\) to give

\[
\frac{P_H(H \in V_N(W))}{P_H(H \in V_N(E))} \leq \exp(-N(c/2)/\sqrt{\eta}).
\]

Since \(E = W \cup K\), one has \(V_N(E) \subseteq V_N(W) \cup V_N(K)\). Therefore,

\[
P_H(H \in V_N(K)) \geq P_H(H \in V_N(E)) - P_H(H \in V_N(W)).
\]

From \(P_H(H \in V_N(K) \mid H \in V_N(E)) = \frac{P_H(H \in V_N(K))}{P_H(H \in V_N(E))}\), we now obtain

\[
1 \geq P_H(H \in V_N(k) \mid H \in V_N(E)) \geq 1 - \frac{P_H(H \in V_N(K))}{P_H(H \in V_N(E))} \geq 1 - \exp(-N(c/2)/\sqrt{\eta})
\]

for \(N > N_1\). For \(N > N_1 + 1/\eta\), one has \(V_N(K) \subseteq U^{2\eta}(E^*) \cap V_N(E)\). Therefore,

\[
1 \geq P_H(H \in U^{2\eta} \mid H \in V_N(E)) \geq P_H(H \in V_N(K) \mid H \in V_N(E)) \geq 1 - \exp(-N(c/2)/\sqrt{\eta})
\]

which concludes the proof. \(\square\)

### 5.3 Most likely path connecting two histograms

**Definition 5.3.** Let \(Q = (q_{i,j})\) be the mutations transfer matrix and \(m\) the mutations rate. As in definition 5.1, for any subset \(S\) of the set of genotypes \(\{1, \ldots, g\}\) we denote \(R(S)\) the set of genotypes reachable from \(S\) in some finite number of steps. For any given histograms \(H\) and \(G\), we say that \(G\) is reachable from \(H\) if \(\text{supp}(G) \subseteq R(\text{supp}(H))\).

Note that if there exists some power \(p\) such that all coefficients of \(Q^p\) are positive, one has \(R(S) = \{1, \ldots, g\}\) for any non-empty set of genotypes \(S\), and hence, any \(G\) is reachable from any \(H\). The next theorem answers an important question for bacterial genetic evolution: how can one reconstruct the most likely evolutionary path starting at a known initial histogram \(H\) and reaching a known terminal histogram \(G\) after a fixed number \(T\) of daily cycles.

**Theorem 5.4.** Fix a path length \(T\) and a small \(0 < a < 1\). Fix any interior histograms \(H\) and \(G\) such that \(G\) is reachable from \(H\), and verifying \(b(H) > a\), \(b(G) > a\). Let \(B = B(G, \eta)\) be the open ball of center \(G\) and a small radius \(\eta < a/2\). Define the following compact set \(E_T\) of paths starting at \(H\) and reaching \(B\) for the first time at time \(T\), by

\[
E_T = \{h \in \Omega_T \mid b(h) \geq a; h_1 = H; h_T \in \bar{B}; h_n \neq B(G) \text{ for } 1 \leq n < T\}.
\]

Then for \(\eta\) small enough, \(\Lambda(E_T) = \inf_{h \in E_T} \lambda(h)\) must be finite. The set \(E_T^*\) of all paths \(h \in E_T\) such that \(\lambda(h) = \Lambda(E_T)\) is then non-empty and contains only interior paths.
When $\Lambda(E_T) > 0$, then the event $\mathcal{H} \in E_T$ is a rare event, and for any open neighborhood $U$ of $E_T^*$, one has

$$ \lim_{N \to \infty} \frac{1}{N} \log P(H \in U \mid H \in E_T) = 1 $$

with convergence at exponential speed.

When $\Lambda(E_T) = 0$, then there is a unique zero-cost trajectory $h^* \in E$ determined by equation \eqref{eq:4.3}, with $h_1 = H$ and $h_T \in B$. Then for any open neighborhood $U$ of $h^*$, one has

$$ \lim_{N \to \infty} P(H \in U) = \lim_{N \to \infty} P(H \in E_T) = 1 $$

with convergence at an exponential speed.

**Proof.** We omit the detailed proof which is a direct application of the preceding theorem \ref{thm:5.2} and of our results on zero-cost trajectories.

**Intuitive interpretation of theorem \ref{thm:5.4}.** For a large $N$ and a very small $\eta$, the preceding theorem indicates that a stochastic population evolution connecting interior histograms $H$ and $G$ in exactly $T$ days must follow a very thin tube of paths centered around an "optimal" path $h^*$ that minimizes $\lambda(h)$ over all $h$ such that $h_1 = H$, $h_T = G$, and $h_j \neq G$ for $j < T$. We discuss this minimization problem in the next section.

For laboratory experiments observing bacterial populations over a very long time (e.g. several years), a plausible situation may involve two histograms $H$ and $G$ where $H$ is observed on day $n = 1$, but the exact observation time $T$ of $G$ is only known to be smaller than some given $T_{\text{max}}$. The following theorem indicates how one can estimate $T$.

**Theorem 5.5.** We use here the same notations $a, H, G, \eta, E_T, B = B(G, \eta)$ and hypotheses as in the preceding theorem \ref{thm:5.4}, but $T$ can now vary in a fixed interval $[0, T^+]$. For any path $h \in \Omega_T$ denote $\theta(h)$ the first hitting time of the small ball $B = B(G, \eta)$.

Denote $E = \bigcup_{T \leq T^+} E_T$ is the set of all paths starting at $H$ with the hitting time $\theta \leq T^+$. Then $\Lambda(E) = \inf_{T \leq T^+} \Lambda(E_T)$ is finite and there is at least one time $T^* < T^+$ such that $\Lambda(E) = \Lambda(E_{T^*})$. If $T^*$ is unique and $\Lambda(E_{T^*}) > 0$, then one has

$$ \lim_{N \to \infty} \frac{1}{N} \log P(\theta(H) = T^* \mid H \in E) = 1 $$

with convergence at an exponential speed.

**Proof.** This result is a direct corollary of theorem \ref{thm:5.2} since a finite union of rare events $E_T$ can essentially ”only be realized” if the $E_T$ having the smallest $\Lambda(E_T)$ is realized.

6 Recursive Computation of Cost-Minimizing Trajectories connecting two histograms

The previous section on large deviations demonstrates that given two interior histograms $H$ and $G$, the most likely bacterial population evolution from $H$ to $G$ in a given number of days, $T$, must be a very thin tube of histograms trajectories centered around a deterministic path $h^* = [h_1^* h_2^* \ldots h_T^*]$ minimizing the rate function $\lambda(h)$ over all $h \in \Omega_T$ such that $h_1 = H$ and $h_T = G$. By analogy with
control problems and the definition of energy minimizing geodesics $\gamma(t)$ on Riemannian manifolds, we call any such $h^*$ a \textit{geodesic} from $H$ to $G$ whenever $\lambda(h^*)$ is finite. Indeed, in this discrete version of geodesics the one-step cost $C(h_n, h_{n+1})$ at the discrete time $n$ is analogous to the Riemannian energy $\|\frac{dh^*}{dt}\|^2$ at a continuous time $t$ (see our companion paper [34] where this point of view is explored further). When $h^* \in \mathcal{H}^0$ for all $1 \leq n \leq T$, we call $h^*$ an \textit{interior geodesic}. The numerical computation of geodesics presents multiple numerical and mathematical challenges in our discrete context. We now develop a key recursive equation in reverse time that must be verified by all geodesics. This equation is an essential practical tool for the efficient numerical computation of geodesics connecting two histograms $H$ and $G$.

\textbf{Theorem 6.1.} Fix any small $a > 0$ and the process parameters $\mathcal{P}$. If the mutation rate $m$ is small enough, then for any pair $y, z \in \mathcal{H}^0$ of interior histograms verifying $b(y) \geq a$, and $b(z) \geq a$, there is a unique interior histogram $x = \chi(m, y, z)$ such that $[x, y, z]$ is an interior geodesic connecting $x$ to $z$ in two steps. There is a positive constant $m_0 \equiv m_0(a, \mathcal{P})$ such that, for all $0 < m < m_0$, the following hold

1. the histogram valued function $x = \chi(m, y, z)$ is $C^\infty$ in $m, y, z$,

2. the one-step costs $C(x, y)$ (from $x$ to $y$) and $C(y, z)$ (from $y$ to $z$) are both differentiable in $y \in \mathcal{H}^0$,

3. $x = \chi(m, y, z)$ is the unique histogram solving $\nabla_y [C(x, y) + C(y, z)] = 0$ where the gradient $\nabla_y$ is computed in the convex set of interior histograms $x, y \in \mathcal{H}^0$,

4. for any $T \geq 2$, any interior geodesic $h = [h_1 h_2 \ldots h_T]$ such that $b(h) > a$ must verify the reverse recurrence relation

\begin{equation}
    h_n = \chi(m, h_{n+1}, h_{n+2}), \quad \text{for } 1 \leq n \leq T - 2,
\end{equation}

5. the geodesic $h$ is completely determined by its last two histograms $h_T$ and $h_{T-1}$,

6. the 1st order Taylor expansion in $m$ of $x = \chi(m, y, z)$ is given by $x = \hat{x} + m (\hat{x}, w) + O(m^2)$ where the interior histogram $\hat{x}$ and the vector $w \in \mathbb{R}^g$ both depend only on the histograms $(y, z)$, and are explicitly computed below by the succession of five formulas (6.11), (6.12), (6.13), (6.14), (6.15).

\textbf{Proof.} Any sub-segment $[h_n = x, h_{n+1} = y, h_{n+2} = z]$ of an interior geodesic $h$ is also a geodesic connecting $x = h_n$ to $z = h_{n+2}$ in two steps. For the 2-steps interior geodesic $(x, y, z)$, the two-step cost function $u(x, y, z) = C(x, y) + C(y, z)$ is minimized in $y$ by $y = h_{n+1}$. For any three histograms $x, y, z \in \mathcal{H}^0$, both $C(x, y)$ and $C(y, z)$ are finite and differentiable in $y$ by Theorem 3.24 and convex in $y$ by Lemma 3.18. Hence, for fixed $x, z \in \mathcal{H}^0$, the function $f(y) = C(x, y) + C(y, z)$ is finite, convex, and differentiable for all $y$ in the open convex set $\mathcal{H}^0$. If $y \in \mathcal{H}^0$ is a minimizer of $f(y)$ over all $\mathcal{H}^0$, any vanishingly small $\Delta y$ ($\Delta y$ is a histogram in this context) so that $y + \Delta y \in \mathcal{H}^0$ must verify $\sum_j \Delta y(j) = 0$ and $f(y + \Delta y) \geq f(y)$. Hence the gradient of $\left[ f(y) + \mu \sum_j y(j) \right]$ must be 0 for some Lagrange multiplier $\mu$. For each $s \in \{1, \ldots, g\}$, this yields the system

\begin{equation}
    \frac{\partial}{\partial y_s} f(y) + \mu = 0.
\end{equation}

36
Denote \( u(x, y, z) = f(y) = C(x, y) + C(y, z) \). For given \( y, z \in \mathcal{H}^c \), extend (6.2) into the following system of \((1 + g)\) equations to be solved for a histogram \( x \) and a Lagrange multiplier \( \mu \):

\[
\frac{\partial}{\partial y_s} u(x, y, z) + \mu = 0, \quad 1 \leq s \leq g, \tag{6.3}
\]

\[
\sum_k x_k = 1. \tag{6.4}
\]

This provides \((g + 1)\) equations for \((g + 1)\) unknowns \((x, \mu)\). We now show that for \( m = 0 \), this system has a unique explicit solution \((\hat{x}, \nu)\). By Theorem 3.24, for \( m < m_0 \) with \( m_0 > 0 \) small enough, the function \( u(x, y, z) = C^\infty \) in \((m, x, y, z)\), with explicit first-order Taylor expansion in \( m \) given by (3.60). Recall our earlier notations

\[
E_{s,k}(x, y) = \exp \left( -\frac{y_s}{F_s x_s} + \frac{y_k x_k}{F_k x_k} \right), \quad E_{s,k}(y, z) = \exp \left( -\frac{z_s}{F_s y_s} + \frac{z_k}{F_k y_k} \right).
\]

Taking derivatives in \( y \) of the first-order expansions (3.60) of \( C(x, y) \) and \( C(y, z) \) readily yield the following first-order Taylor expansions in \( m \) with remainders of order \( m^2 \leq 10^{-12} \)

\[
\frac{\partial}{\partial y_s} C(x, y) \simeq A_s(x, y) + m \hat{A}_s(x, y), \tag{6.5}
\]

\[
A_s(x, y) = 1 + \log \frac{y_s}{F_s} - \log x_s, \tag{6.6}
\]

\[
\hat{A}_s(x, y) = \sum_k \left( Q_{s,k} E_{s,k}(x, y) - \frac{F_k x_k}{F_s x_s} Q_{k,s} E_{k,s} \right), \tag{6.7}
\]

\[
\frac{\partial}{\partial y_s} C(y, z) \simeq D_s(y, z) + m \hat{D}_s(y, z), \tag{6.8}
\]

\[
D_s(y, z) = \frac{F_s}{(F, y)} - \frac{z_s}{y_s}, \tag{6.9}
\]

\[
\hat{D}_s(y, z) = F_s \sum_k Q_{s,k} - \left( \frac{F_s}{(F, y)} + \frac{z_s}{y_s} \right) \sum_k E_{s,k}(y, z) Q_{s,k} - \frac{z_s}{F_s y_s} \sum_k F_k y_k Q_{k,s} E_{k,s}(y, z). \tag{6.10}
\]

For \( m = 0 \), the system in (6.3) becomes \( A_s(x, y) + D_s(y, z) + \mu = 0 \) for each \( s \), which yields

\[
1 + \log \frac{y_s}{F_s x_s} + \frac{F_s}{(F, y)} - \frac{z_s}{y_s} + \mu = 0.
\]

Hence, we have \( x = e^{1+\mu} X \) where the vector \( X \) is given by

\[
X_s = \frac{y_s}{F_s} \exp \left( \frac{F_s}{(F, y)} - \frac{z_s}{y_s} \right) > 0. \tag{6.11}
\]

The constraint given by (6.4) gives \( e^{1+\mu} = \frac{1}{\sum_t X_t} \). Therefore, for \( m = 0 \) and all \( s \), the unique solution \( \hat{x} = x(0, y, z) \) and \( \hat{\mu} = \mu(0, y, z) \) of the system given by (6.3) and (6.4) is

\[
\hat{x}_s = \frac{X_s}{\sum_t X_t}, \quad X_s = \frac{y_s}{F_s} \exp \left( \frac{F_s}{(F, y)} - \frac{z_s}{y_s} \right), \quad 1 + \hat{\mu} = -\log \left( \sum_t X_t \right). \tag{6.12}
\]
The expressions for some fixed \( \mu \) Therefore, the implicit function theorem applies to the system given by (6.3) and (6.4). Hence, for \( x, \mu \)

\[ U(s, t) = \frac{\partial}{\partial x_t} \frac{\partial}{\partial y_s} [A_s(x, y) + D_s(y, z)] = -1_{\{s=t\}} \frac{1}{x_s}. \]

The \( g \times g \) matrix \( U \) is thus diagonal with non-zero diagonal terms, and, therefore, \( U \) is invertible. Therefore, the implicit function theorem applies to the system given by (6.3) and (6.4). Hence, for some fixed \( m_1 > 0 \), there is a unique solution \((x(m, y, z), \mu(m, y, z))\) to the system (6.3)–(6.4), and the functions \((x, \mu)\) are \( C^\infty \) in \((m, y, z)\).

Let \( x \simeq \bar{x} + mV \) and \( \mu \simeq \bar{\mu} + m\mu_1 \) be the first-order Taylor expansions of \((x, \mu)\) in \( m \). Denote \( V_s = \bar{x}_s w_s \) so that \( x_s \simeq \bar{x}_s(1 + mw_s) \). The constraint (6.4) then implies \( \langle \bar{x}, w \rangle = 0 \). The first-order expansion of (6.3) becomes

\[ A_s(\bar{x} + mV, y) + D_s(y, z) + \bar{\mu} + m(\bar{A}_s(\bar{x}, y) + \bar{D}_s(y, z) + \mu_1) \simeq 0. \]

(6.13)

Since \( A_s(\bar{x} + mV, y) \simeq A_s(\bar{x}, y) - mw_s \), the zero-order term in (6.13) vanishes due to the values of \( \bar{x} \) and \( \bar{\mu} \). The first-order term must vanish as well, which gives for all \( s \)

\[ w_s = \bar{A}_s(\bar{x}, y) + \bar{D}_s(y, z) + \mu_1. \]

Since \( \langle \bar{x}, w \rangle = 0 \), this yields

\[ \mu_1 = -\sum_t \bar{x}_t[A_t(\bar{x}, y) + D_t(y, z)]. \]

We then have

\[ w_s = \bar{A}_s(\bar{x}, y) + \bar{D}_s(y, z) - \sum_t \bar{x}_t[A_t(\bar{x}, y) + D_t(y, z)]. \]

Define vectors \( \alpha \) and \( \beta \) by \( \alpha_s = \bar{A}_s(\bar{x}, y) \) and \( \beta_s = \bar{D}_s(y, z) \) so that

\[ w = \alpha + \beta - \langle \bar{x}, \alpha + \beta \rangle. \]

(6.14)

The expressions for \( \bar{A}_s, \bar{D}_s, \) and \( E_{s,k} \) given above yield directly

\[ \alpha_s = \sum_k \left( Q_{s,k} e_{s,k} - \frac{F_k X_k}{F_s X_s} Q_{k,s} e_{k,s} \right) \]

(6.15)

with \( e_{s,k} = 1/e_{k,s} = \exp[-y_s/(F_s \bar{x}_s) + y_k/(F_k \bar{x}_k)] \) and

\[ \beta_s = F_s \sum_k Q_{s,k} - \left( F_s + \frac{z_s}{y_s} \right) \sum_k f_{s,k} Q_{s,k} - \frac{z_s}{F_s y_s} \sum_k F_k y_k Q_{k,s} f_{k,s} \]

(6.16)

with \( f_{s,k} = 1/f_{k,s} = \exp[-z_s/(F_s y_s) + z_k/(F_k y_k)] \). Note that \( \bar{x}, \alpha, \beta, \) and \( w \) depend only on \( y \) and \( z \). The preceding formulas provides the explicit first-order expansion \( x_s = \bar{x}_s + m\bar{x}_s w_s \), concluding the proof.

The reverse recurrence relation in Theorem 6.1 introduces several challenges regarding numerical implementation. Once a target histogram \( h^*_T = G \) is fixed, each geodesic is completely determined by its penultimate histogram \( h^*_T = Z \). However, the penultimate histogram \( Z \) is not known

38
Performing a straightforward exhaustive numerical search for the optimal penultimate histogram, even for a very small number \( g \) of genotypes, can be inefficient. We performed a preliminary numerical investigation of geodesics computations for a small number of genotypes \( g = 3 \) in [21]. Our empirical preliminary exploration indicates that geodesics connecting two interior histograms \( H \) and \( G \) seem to contain only interior histograms and are interior geodesics. In addition, we discuss the numerical implementation of the optimization algorithm in more detail in a companion paper [34] where we extensively explore and test efficient strategies to numerically compute geodesics for a larger number of genotypes, up to \( g = 20 \).

Another potential research direction is to extend the mathematical formalism and the numerical computation of geodesics starting at a boundary histogram \( H \). For a boundary histogram, the number of \( j \)-cells is zero for some \( j \) (same as \( H(j) = 0 \)). Our analysis can be extended to this case, but this requires substantial research effort.

7 Conclusions

We developed an innovative large deviations analysis of rare genetic events for locked-box type stochastic models of the genetic evolution of large bacterial populations. We consider discrete time-homogeneous Markov chains that model daily cycles of growth, mutations, and selection, where each daily selection randomly reduces population size to a fixed large size \( N > 10^5 \). The daily population \( \text{pop}_n \) has then fixed size \( N \) and is characterized by its histogram \( H_n \in \mathcal{H} \subset \mathbb{R}^9 \), where \( H_n(j) \) is the frequency of genotype \( j \) in \( \text{pop}_n \). The state space of the Markov chain \( H_n \) is the compact convex set \( \mathcal{H} \) of all histograms in \( \mathbb{R}^9 \). Mutations occur at a fixed small rate \( m < 10^{-6} \). Other parameters (e.g. the mutation matrix \( q_{i,j} \)) in the Markov chain model can be fitted to emulate typical laboratory experiments on bacterial genetic evolution.

For the one step Markov transition kernel \( Q(H,G) \), we first proved the key uniform large deviation estimate \( \frac{1}{N} \log[Q(H,G)] \approx -C(H,G) \) valid for large \( N \) and arbitrary interior histograms \( H,G \), where the one-step cost function \( C(H,G) \geq 0 \) is computed explicitly up to terms of order \( m^2 \). This led us to obtain a uniform large deviations estimate for the random histogram trajectories \( H = [H_1, \ldots, H_T] \) of the Markov chain. For a fixed trajectory \( h = [h_1, \ldots, h_T] \) we demonstrate that \( \text{Prob}\{||H - h|| < 1/N\} \) is roughly \( \approx \exp(-N \lambda(h)) \) with a rate function \( \lambda(h) \) explicitly given by the sum \( C(h_1, h_2) + \cdots + C(h_{T-1}, h_T) \). For any set \( \Gamma \) of histogram trajectories of fixed duration \( T \), let \( \Gamma(N) \) be the \( \Gamma \)-neighborhood of radius \( 1/N \). We demonstrate that \( \text{Prob}\{H \in \Gamma(N)\} \) is of order \( \exp(-N \Lambda(\Gamma)) \), with rate functional is defined as \( \Lambda(\Gamma) = \inf_{h \in \Gamma} \lambda(h) \).

For a fixed \( T \), we applied our large deviations theoretical results to develop a computational optimization algorithm for computing numerically the most likely histogram trajectory \( h^* \) connecting a given initial histogram \( h_1^* = H \in \mathcal{H}^9 \) to a given terminal histogram \( h_T^* = G \in \mathcal{H}^9 \). This problem is of particular interest when \( G \) is not completely concentrated on the fittest genotype \( g \) since in that case the conditional probability \( P(H_T = G \mid H_1 = H) \) vanishes at an exponential speed when \( N \) becomes large. Indeed, when \( \Gamma(H,G) \) is the set of all histograms paths \( h \) such that \( h_1 = H \) and \( h_T = G \), and if \( \Lambda(\Gamma(H,G)) \neq 0 \), the random event \( \mathcal{E} = \{H \in \Gamma(H,G)\} \) must have exponentially vanishing probability for large \( N \). Our results then demonstrate that whenever \( \mathcal{E} \) actually occurs, any random histogram trajectory \( H \) realizing \( \mathcal{E} \) must lie in a very thin tube around the path \( h^* \) minimizing \( \lambda(h) \) over all \( h \in \Gamma(H,G) \). In laboratory experiments, knowing only \( H_1 = H \) and the observed terminal histogram \( H_T = G \), there is a practical interest to reconstruct the whole
unobserved random genetic evolution \( H \approx h^* \) between times 1 and \( T \).

In this paper we developed (theorem 6.1) a concrete theoretical and numerical approach for computing a minimizer \( h^* \) of \( \lambda(h) \) over all \( h \in \Gamma(H,G) \). By analogy with the geodesics minimizing Riemannian energy on a Riemannian manifold, we call any such minimizer \( h^* \) a large deviations "geodesic" connecting \( H \) to \( G \) in \( T \) steps. We demonstrate that any geodesic \( h^* \) is completely determined by its last two histograms, the known \( h^*_T = G \) and the unknown "penultimate histogram" \( h^*_{T-1} = Z \). This is due to the recursive relation in reverse time which expresses \( h^*_{n+2} \) as a function \( \chi(m, h^*_{n+1}, h^*_n) \). We compute explicitly the histogram-valued function \( \chi \) for small mutation rate \( m \). This reduces the search for \( h^* \) (given \( H \) and \( G \)) to a search for the optimal penultimate histogram \( Z \). In our companion paper [34], we explored this strategy to develop an efficient numerical algorithm for computing numerically the large deviations geodesics for a relatively large number of genotypes \( g \leq 20 \).

Our results open the way to interesting further applied work on the biological mechanisms that can trigger concrete rare genetic events, such as the emergence or fixation of non-dominant genotypes. Our explicit cost functions formulas can help quantify how the occurrence of specific rare genetic events is impacted by changes in the mutation rate \( m \), or by the emergence of deleterious mutations. Moreover, large deviations results developed in this paper offer a mathematical framework that, after adequate extensions, could be adapted to rare events analysis for the genetic evolution and transmission of viruses.

**Acknowledgment**

This research has been partially supported by the NSF grant DMS-1412927. I.T. was also partially supported by the NSF grant DMS-1903270.

**A  Summary of Notation**

Below, we summarize the notation used in this manuscript.

- \( g \in \mathbb{N} \): number of genotypes in the bacterial population
- \( N \in \mathbb{N} \): number of cells in the bacterial population
- \( [F_1, F_2, \ldots, F_g] \): ordered genotype growth factors with \( F_1 < F_2 < \ldots < F_g \)
- \( m \in \mathbb{R} \): mutation rate
- \( Q \in \mathbb{R}^g \times \mathbb{R}^g \) with entries \( q_{j,k} \geq 0 \) - conditional probabilities of mutation (see def 2.1)
- \( M = mQ \) mutation matrix (see def 2.8)
- Model parameters \( \mathcal{P} = \{N, g, F_1, \ldots, F_g, m, Q\} \) (see def 2.2)
- \( R_n(j,k) \): number of \( j \)-cells mutating into \( k \)-cells on day \( n \) (see def 2.7)
- \( H \in \mathbb{R}^g \): population histogram of bacterial frequencies; \( H = [H(1), H(2), \ldots, H(g)] \) with \( H(j) \geq 0 \) for \( j = 1, \ldots, g \) and \( \sum_{j=1}^{g} H(j) = 1 \)
- \( h_n(j) \in [0,1] \): frequency of genotype \( j \) in the population on day \( n \)
- \( h_n(j) \in [0,1] \): frequency of genotype \( j \) in the population on day \( n \)
$\mathcal{H} \subset \mathbb{R}^g$: space of all histograms (see def 2.4)

$\mathcal{H}^o = \{ H : H \in \mathcal{H} \text{ and } H(j) > 0 \text{ for all } j = 1, \ldots, g \}$ is the set of all interior histograms

$\mathcal{H} = \{ H_1, \ldots, H_T \}$: time-dependent path of length $T$ (days) in the space $\mathcal{H}$ connecting the initial histogram $H_1$ and the final histogram $H_T$

$\Omega_T$: space of all paths of length $T$ (see def 4.1)

$\Omega_T^+$: space of all interior paths of length $T$; if $\mathbf{H} \in \Omega_T^+$ then $H_n(j) \geq \varepsilon$ for all $1 \leq t \leq T$ and $1 \leq j \leq g$ and for some $\varepsilon \ll 1$

$supp(H) - \text{Support of a histogram } supp(H) = \{ j \mid H(j) > 0 \}$ for any $H \in \mathcal{H}$ (see def 2.6)

$b(H)$ - The essential minimum for a histogram $b(H) = \min_{j \in supp(H)} H(j)$ (see def 2.6)

$b(\mathbf{H})$ - The essential minimum for a path of histograms $b(\mathbf{H}) = \min_{n=1 \ldots T} b(H_n)$ (see def 4.4)

$\lceil u \rceil$ is the smallest integer greater than or equal to $u$ for $u \geq 0$

$\Phi_j(H) = F_jH(j)/\langle F, H \rangle$ approximation of the genetic histogram after the growth phase (see eq. (2.2))

$\Psi_j(H, r) = (F_jH(j) - \sum_k r_{j,k} + \sum_k r_{k,j}) / \langle F, H \rangle$ approximation of the genetic histogram after the mutation phase (see eq. (2.8))

A matrix $A$ is $N$-rational if $NA$ has non-negative integer coefficients

The set $\mathcal{Z}$ contains all $N$-rational $g \times g$ matrices

$K(j, H)$ is the set of $g \times g$ matrices with non-negative coefficients with constrains (2.4) (see def 2.9)

$K(H) = \bigcap_{j=1}^g K(j, H)$ and $K_N(H) = \mathcal{Z} \cap K(H)$

$N_{r_{j,r}} = R_{j,k}$ total number of $j$-cells mutating into $k$-cells

$Q_N(H, G) = P(H_{n+1} = G \mid H_n = H)$ Markov transition kernel (see def 2.18)

$\tau(H, r, G) = mut(r, H) + KL(G, \Psi(H, r))$ - Composite transition rate

$KL(G, J)$ - Kullback-Leibler divergence (see def 3.13)

$mut(H, r)$ - rate function for mutations (see 3.13)

$C(H, G) = \min_{r \in K(H)} \tau(H, r, G)$ - one-step cost function (see def 3.17)

$\Omega_T(a) \subset \Omega_T$ - the compact set of histogram trajectories such that $\mathbf{H} \in \Omega_T$ and $b(\mathbf{H}) \geq a$

The distance between trajectories $\| \mathbf{H} - \mathbf{H}' \| = \max_{n=1 \ldots T} \| H_n - H_n' \|

$\lambda(\mathbf{H})$ - large deviations rate function for the path $\mathbf{H}$; $\lambda(\mathbf{H}) = \sum_{n=1}^{T-1} C(H_n, H_{n+1})$ (eq. 4.1)

$\Lambda(F)$ - large deviations set functional; $\Lambda(F) = \inf_{\mathbf{H} \in F} \lambda(\mathbf{H})$ for any $F \subset \Omega_T$ (eq. 4.2)

$V_N(\Gamma)$ - open neighborhood of paths for any $\Gamma \subset \Omega_T$ (def 4.7)

$b(\Gamma)$ - the essential minimum for a set of paths; $b(\Gamma) = \inf_{\mathbf{H} \in \Gamma} b(\mathbf{H})$ for any $\Gamma \subset \Omega_T$ (def 4.7)

$B_N(\Gamma)$ - the (finite) set of all $N$-rational paths in $V_N(\Gamma)$
B Proofs of Theorems in Section 2

Proof of Lemma 2.10

In this proof only, denote \(|z|\) as the largest integer less than or equal to \(z\). Fix any \(j \in \text{supp}(H)\), and select any \(k^* = k^*(j)\) such that \(r_{j,k^*} = \max_k r_{j,k}\). For any \(k\), if \(r_{j,k^*} < g/N\), define \(s_{j,k}\) by

\[
s_{j,k} = \begin{cases} 
0, & r_{j,k} = 0; \\
1/N, & 0 < r_{j,k} < 1/N; \\
[Nr_{j,k}]/N, & 1/N \leq r_{j,k}.
\end{cases} \tag{B.1}
\]

If \(g/N \leq r_{j,k^*}\), define

\[
s_{j,k^*} = [N r_{j,k^*}]/N - (g - 1)/N. \tag{B.2}
\]

Equations (B.1) and (B.2) imply that \(\|s - r\| \leq g/N\) and \(\text{supp}(s) = \text{supp}(r)\). Let \(S(j) = \sum_k s_{j,k}\). To prove \(s \in K(j, H)\), we only need to show that \(S(j) < F_j H(j)\) for all \(j \in \text{supp}(H)\), which involves two cases. Thus, let \(j \in \text{supp}(H)\).

**Case 1:** Suppose \(r_{j,k^*} < g/N\). Since \(F_j H(j) > F_1 b(H) > F_1 a\), impose \(N > g^2/aF_1\) to force \(F_j H(j) > g^2/N\). Definition (B.1) ensures that \(s_{j,k} < g/N\) for all \(j, k\), and hence \(S(j) < g^2/N < F_j H(j)\).

**Case 2:** Suppose \(g/N \leq r_{j,k^*}\). Let \(u(j) \leq (g - 1)\) be the number of indices \(k\) such that \(0 < r_{j,k} < 1/N\). Definitions (B.2) and (B.1) imply \(s_{j,k^*} \leq r_{j,k^*} - (g - 1)/N\) and \(\sum_{k \neq k^*} s_{j,k} \leq u(j)/N + \sum_{k \neq k^*} r_{j,k}\). This yields \(S(j) \leq \sum_k r_{j,k} < F_j H(j)\) since \(r \in K(H)\).

This concludes the proof.

C Proofs of Theorems in Section 3

First, we need the following technical lemma.

**Lemma C.1.** For any fixed \(0 < \alpha < 1\) and for all \(x, y \in [0, 1/2]\), the following inequality holds

\[
|y \log y - x \log x| \leq \frac{4}{1 - \alpha} |x - y|^\alpha. \tag{C.1}
\]

Moreover, for \(x, y \in [0, A]\) with \(A > 1\) the following inequality holds

\[
|x \log x - y \log y| \leq \left( \frac{5(1 + A)\log(A)}{1 - \alpha} \right) |x - y|^\alpha. \tag{C.2}
\]

**Proof.** Consider \(0 \leq x \leq 1\). Then

\[
|x \log(x)| = x \log(1/x) \leq (x/(1 - \alpha)) \log(1/x^{1-\alpha}) \leq x^\alpha/(1 - \alpha). \tag{C.3}
\]

Consider \(|y \log y - x \log x|\) with \(|y - x| < w \leq 1\). Then, either \(x, y \leq 2w\) or \(x, y \geq w\). When \(x, y \leq 2w\), equation (C.3) gives

\[
|y \log y - x \log x| \leq |x \log x| + |y \log y| \leq \frac{4}{(1 - \alpha)w^\alpha}.
\]
When both \( x, y \geq w \), Taylor’s formula and (C.3) yield
\[
|y \log y - x \log x| \leq w(1 + \log(1/w)) \leq \frac{4}{1-\alpha} w^\alpha.
\]
This proves equation (C.1). For \( x, y \in [0, A] \) with \( A > 1 \), apply equation (C.1) to \( x/A \) and \( y/A \) to obtain equation (C.2).

**Proof of Proposition 3.11**

Consider \( a \) and \( \alpha \) in proposition 3.11. Take \( H, H' \in H(a) \) with \( \text{supp}(H') = \text{supp}(H), \ r \in K(H), \ r' \in K(H') \). Fix \((j,k)\) with \( H(j)q_{j,k} > 0 \), which implies \( H'(j)q_{j,k} > 0 \). Then, \( x := r'_j \) and \( y := r_{j,k} \) verify \( |x - y| \leq \|r' - r\| \) and are bounded by \( \max \{ F_j H'(j), F_j H(j) \} \leq F_g \). By definition,
\[
\begin{align*}
L_{j,k}(H, r) &= y \log y - y \log(e F_j H(j)) + F_j H(j), \\
L_{j,k}(H', r') &= x \log x - x \log(e F_j H'(j)) + F_j H'(j).
\end{align*}
\]
Set \( \hat{L} := |L_{j,k}(H, r) - L_{j,k}(H', r')| \), which satisfies the inequality
\[
\hat{L} \leq |y \log y - x \log x| + |y - x| |\log(e F_j H(j))| + |x| |\log(e F_j H'(j))| + F_j |H'(j) - H(j)|. \tag{C.4}
\]
Reformulate (C.3) by writing \( \hat{L} \leq U_1 + U_2 + U_3 + U_4 \) where \( U_i \) is the \( i \)-th term in the right-hand side of (C.3). Clearly \( U_3 + U_4 \leq 2(F_g/a) \|H' - H\| \). Then, (C.2) gives \( U_1 \leq \frac{c_1}{a} \|r' - r\|^\alpha \) with \( c_1 = 5(1 + F_g)^2 \log(F_g) \). Since \( F_g \geq F_j H(j) \geq a F_1 \), one has \( |\log(e F_j H(j))| \leq c_2/a \) with \( c_2 = 2 + \log(F_g/F_1) \), and hence, \( U_2 \leq c_2 \|r' - r\| \). These bounds yield, for \( H(j)q_{j,k} > 0 \),
\[
\hat{L} \leq \frac{\hat{c}}{a(1-\alpha)} (\|r' - r\|^\alpha + \|H' - H\|) \tag{C.5}
\]
with \( \hat{c} = c_1 + c_2 + 2 F_g \leq 14 F_g \log(F_g) \). This result still holds when \( H(j)q_{j,k} = 0 \) since then \( \hat{L} = 0 \). Summing (C.5) over all \( j, k \), we get
\[
|\text{mut}(H', r') - \text{mut}(H, r)| \leq \frac{\hat{c} g^2}{a(1-\alpha)} (\|r' - r\|^\alpha + \|H' - H\|) \leq c (\|r' - r\|^\alpha + \|H' - H\|),
\]
where \( c \) is the constant stated in this proposition. This concludes the proof.

**Proof of Proposition 3.12**

Take \( H \) and \( r \) as stated in proposition 3.12. Given \( H_n = H \), the coefficients \( Z_n(j,k) \) of the companion matrix \( Z_n \) are independent and have Poisson distributions \( \pi(H,j,k) \) with respective means \( N m q_{j,k} F_j H(j) \). For \( j \in \text{supp}(H) \) and any \( k \), set \( u = m q_{j,k} F_j H(j) \) and \( v = r_{j,k} \). Since \( N v \) is an integer, apply (5.4) to \( X = Z_n(j,k) \) to obtain
\[
\frac{1}{N} \log P(Z_n(j,k)/N = r_{j,k} \mid H_n = H) = -L_{j,k}(H, r) + o_1(N) \tag{C.6}
\]
with \( |o_1(N)| \leq 2 \log N/N \). This equation remains true for \( j \notin \text{supp}(H) \) and all \( k \) since if \( j \notin \text{supp}(H) \) then \( H(j) = 0 \) and \( Z_n(j,k) = 0 \).
Conditional independence of the \( Z_n \) coefficients then yields

\[
P(Z_n/N = r \mid H_n = H) = \prod_{j,k} P(Z_n(j,k)/N = r_{j,k} \mid H_n = H).
\]

Equation (C.6) implies

\[
\frac{1}{N} \log P(Z_n/N = r \mid H_n = H) = - \sum_{j,k} L_{j,k}(H, r) + o_2(N) = -\text{mut}(H, r) + o_2(N) \quad (C.7)
\]

with \( |o_2(N)| \leq g^2 |o_1(N)| \leq 2g^2 \log N/N. \)

Set \( c = 2 + \log(g) \), and impose \( N > c/a \geq c b(H) \). Applying (3.1) yields

\[
\left| \frac{1}{N} \log P(R_n/N = r \mid H_n = H) - \frac{1}{N} \log P(Z_n/N = r \mid H_n = H) \right| \leq \frac{1}{N} \log \left(1 + 2g d(H)^{N/2}\right). \quad (C.8)
\]

The right-hand side of (C.8) is bounded above by \( \frac{1}{N} \log(1 + 2g) \) so that equation (C.7) implies

\[
\frac{1}{N} \log P(R_n/N = r \mid H_n = H) = -\text{mut}(r, H) + o(N) \quad (C.9)
\]

with \( |o(N)| \leq \frac{1}{N} \left(\log(1 + 2g) + 2g^2 \log N\right) \leq 4g^2 \log(N)/N. \)

This concludes the proof.

**Proof of Proposition 3.14**

The coordinates of \( V = NG \) are non-negative integers that sum up to \( N \). Therefore, (2.17) gives

\[
\frac{1}{N} \log(\mu_{N,J}(NG)) = \frac{1}{N} \log N! - \sum_{j \in \text{supp}(G)} \frac{1}{N} \log V(j)! + \sum_{j \in \text{supp}(G)} \frac{V(j)}{N} \log J(j). \quad (C.10)
\]

For \( j \in \text{supp}(G) \subseteq \text{supp}(J) \), we apply Stirling’s formula (3.5) to \( V(j)! = [NG(j)]! \) and obtain

\[
\frac{1}{N} \log V(j)! = G(j) \log N + G(j) \log G(j) - G(j) + o_j(N) \quad (C.11)
\]

with \( |o_j(N)| \leq 2|\log(NG(j))/N| \). Since \( G \) is \( N \)-rational, then for each \( j \in \text{supp}(G) \) one has \( NG(j) > 0 \), and hence \( 1 \leq NG(j) \leq N \). This yields \( 0 \leq \log(NG(j)) \leq \log(N) \) and \( |o_j(N)| \leq 2 \log N/N. \)

Since \( \text{supp}(G) \subseteq \text{supp}(J) \), we apply (3.5) to \( \log(N)! \) along with (C.11) and equation (C.10) to get

\[
\frac{1}{N} \log(\mu_{N,J}(NG)) = \sum_{j \in \text{supp}(G)} [-G(j) \log G(j) + G(j) \log J(j)] + o(N) \quad (C.12)
\]

with uniform remainder \( |o(N)| \leq 2(g + 1) \log N/N. \) Notice that the sum in (C.12) is equal to the Kullback-Leibler divergence \( -KL(G, J) < 0 \) given by equation (3.20).

This concludes the proof.
Proof of Proposition 3.15

Consider \(G, G'\), as stated in proposition 3.15. Then \(N > N^*\) forces \(N > 2/b(G)\) so that \(\text{supp}(G') = \text{supp}(G)\) by Lemma 3.2. Let \(J_n\) be the population histogram at the end of the mutations phase (Phase 2). Suppose first that \(\text{KL}(G, J_n)\) is finite so that \(\text{supp}(G') = \text{supp}(G) \subset \text{supp}(J_n)\). By construction of the Markov chain \(H_n\), one has \(P(H_{n+1} = G' \mid H_n, R_n) = P(H_{n+1} = G' \mid J_n)\). Given \(J_n\), the conditional distribution of \(NH_{n+1}\) is the multinomial \(\mu_{N, J_n}\) given by (2.16). Since \(G'\) is \(N\)-rational, \(P(H_{n+1} = G' \mid J_n) = \mu_{N, J_n}(NG')\). Then, (3.24) yields

\[
\frac{1}{N} \log P(H_{n+1} = G' \mid J_n) = -\text{KL}(G', J_n) + o(N)
\]

with \(|o(N)| \leq 2(g + 1) \log N/N\). From the KL-derivative (3.21), we also deduce

\[
|\text{KL}(G', J_n) - \text{KL}(G, J_n)| \leq \frac{2g}{3N}(1 + |\log b(G)| + |\log b(J_n)|).
\]

Condition \(b(J_n) \geq 1/(N F_0)\) holds due to (2.11). Since \(b(G) \geq 2/N\), the right-hand side of (C.14) is bounded above by \(2g \log N/N\) provided \(N > N^*\). Equation (C.13) then implies

\[
\frac{1}{N} \log P(H_{n+1} = G' \mid J_n) = -\text{KL}(G, J_n) + o_1(N)
\]

with \(|o_1(N)| \leq 5g \log N/N\). Hence, (3.25) is proved when \(\text{KL}(G, J_n)\) is finite.

When \(\text{KL}(G, J_n) = +\infty\), we have \(\text{supp}(G) \notin \text{supp}(J_n)\). For \(G' \in B_N(G)\), one has \(\text{supp}(G') = \text{supp}(G) \notin \text{supp}(J_n)\) so that the transition from \(J_n\) to \(H_{n+1} = G'\) is impossible during Phase 3. Thus, both sides of (3.25) are equal to \(-\infty\).

This concludes the proof.

Proof of Lemma 3.19

Define the entropy \(\mathcal{E}(G)\) as \(0 \leq \mathcal{E}(G) = -\sum_{j \in \text{supp}(G)} G(j) \log(G(j)) \leq \log(g)\). By definition of \(\text{KL}(G, J)\), one has

\[
\text{KL}(G, J) + \mathcal{E}(G) = \sum_{j \in \text{supp}(G)} G(j) \log(1/J(j)).
\]

For each \(k \in \text{supp}(G)\), this yields \(b(G) \log (1/J(k)) \leq \text{KL}(G, J) + \mathcal{E}(G)\). Hence the term \(\beta(G, J)\) in (3.28) verifies \(\log \beta(G, J) \leq |\text{KL}(G, J) + \log(g)/b(G)|\). Since \(G\) is a histogram, equation (C.16) implies

\[
\text{KL}(G, J) \leq \text{KL}(G, J) + \mathcal{E}(G) \leq \max_{j \in \text{supp}(G)} \log(1/J(j)) = \log(\beta(G, J)).
\]

This proves equation (3.28). Take now \(G, J, J' \in \mathcal{H}\) with \(\text{KL}(G, J) \leq \text{KL}(G, J') < \infty\). Formula (3.21) implies \(|\partial_{J(k)} \text{KL}(G, J)| \leq \beta(G, J)\) for all \(k \in \text{supp}(G)\), with \(\beta(G, J)\) as above. By Taylor’s formula and (3.28), we obtain

\[
|\text{KL}(G, J) - \text{KL}(G, J')| \leq g \beta(G, J) \|J - J'\| \leq c \|J - J'\|
\]

with \(c = g \exp([\text{KL}(G, J) + \log(g)/b(G)])\), proving equation (3.29). Now take \(a, G, G'\), and \(J\) as stated in (ii). For \(j \in \text{supp}(G) \cup \text{supp}(G')\), let \(u(j) = [G'(j) - G(j)] \log J(j)\) and \(v(j) = G'(j) \log G'(j) - G(j) \log G(j)\). By definition of Kullback-Leibler divergence, we have

\[
|\text{KL}(G', J) - \text{KL}(G, J)| \leq \sum_{j \in \text{supp}(G), \text{supp}(G')} (|u(j)| + |v(j)|).
\]
Equation (C.4) with \( \alpha = 1/2 \) implies
\[
|v(j)| \leq 8 \| G' - G \|^{1/2}, \quad j \in \text{supp}(G) \cup \text{supp}(G').
\]
(C.19)

For \( j \in \text{supp}(G) \cup \text{supp}(G') \) one has \( j \in \text{supp}(J) \), and the bound (3.28) gives
\[
|u(j)| \leq \| G' - G \| \max (\log \beta(G, J), \log \beta(G, J))
\]
\[
\leq \| G' - G \| \left( \frac{1}{a} \left[ \log g + \max \left( KL(G, J), KL(G', J) \right) \right] \right).
\]
(C.20)

Combine (C.20), (C.19), and (C.18) to get
\[
|KL(G', J) - KL(G, J)| \leq c_1 \| G' - G \|^2
\]
with \( c_1 = \frac{a}{g} (8 + \log g + \max \{ KL(G, J), KL(G', J) \}) \), proving equation (3.30). For \( G, G', J \) as above, consider the special case \( \text{supp}(G) = \text{supp}(G') \). From (3.28), one gets for \( j \in \text{supp}(G) = \text{supp}(G') \)
\[
|u(j)| \leq \frac{1}{a} \| G' - G \| (\log g + \min \{ KL(G, J), KL(G', J) \}).
\]
and Taylor’s formula implies,
\[
|v(j)| \leq \| G' - G \| \left[ 1 + \log(1/a) \right].
\]
(C.22)

These last two bounds along with (C.18) prove equation (3.31).

This concludes the proof.

**Proof of Proposition 3.20**

Consider \( H', H, G, r' \), and \( r \) verifying (i)-(iv) in proposition 3.20. Apply (3.18) with Hölder coefficient \( \alpha = 1/2 \) to obtain
\[
|\text{mut}(r', H') - \text{mut}(r, H)| \leq c_0 \left[ \| r' - r \|^{1/2} + \| H' - H \| \right],
\]
(C.23)

where \( c_0 = \frac{20}{a} F_g \log(F_g) \). From (2.15), we also obtain
\[
\| \Psi(H', r') - \Psi(H, r) \| \leq \frac{3gF_g}{F_1} \left( \| r' - r \| + \| H' - H \| \right).
\]
(C.24)

By definition of \( \tau(H, r, G) \) in (3.26) and item (iii) above, one has
\[
KL(G, \Psi(H, r)) \leq \tau(H, r, G) \leq A.
\]
(C.25)

Since \( \tau(H', r', G) \) and \( \tau(H, r, G) \) are finite, both \( \text{supp}(\Psi(H, r)) \) and \( \text{supp}(\Psi(H', r')) \) contain \( \text{supp}(G) \). Therefore, we can apply (C.25) and (3.29) to obtain
\[
|KL(G, \Psi(H', r')) - KL(G, \Psi(H, r))| \leq c_1 \| \Psi(H', r') - \Psi(H, r) \|
\]
with \( c_1 = g^{1+1/\alpha} e^{A/\alpha} \). In view of (C.24), this yields
\[
|KL(G, \Psi(H', r')) - KL(G, \Psi(H, r))| \leq c_2 \left[ \| r' - r \| + \| H' - H \| \right]
\]
with \( c_2 = 3c_1 g F_g / F_1 \). Combining this last result with (C.23) and setting \( \eta = c_0 + c_2 \) yields
\[
|\tau(H', r', G) - \tau(H, r, G)| \leq \eta \left[ \| r' - r \|^{1/2} + \| H' - H \| \right].
\]

Noting that \( \eta \) here is precisely (3.34) concludes the proof.
Proof of Theorem 3.22

Consider $G, G', H \in \mathcal{H}(a)$ verifying (5.55). One can then select $r \in K(H)$ with $\tau(H, r, G) \leq d + 1$ so that $I = \Psi(H, r)$ must verify $\text{supp}(G') = \text{supp}(G) \subset \text{supp}(I)$ and $KL(G, I) \leq d + 1$, due to the definitions of $C(H, G)$ and $\tau(H, r, G)$ by equations (3.27), (3.26). Apply (3.30) to get the Lipschitz bound

$$|KL(G', I) - KL(G, I)| \leq c_1 \|G' - G\|$$

(C.26)

with $c_1 = \frac{2}{a}(3 + \log g + d)$. This implies, by definition of $\tau(H, r, G)$,

$$|\tau(H, r, G') - \tau(H, r, G)| = |KL(G', I) - KL(G, I)| \leq c_1 \|G' - G\|.$$

and then, since $C(H, G') \leq \tau(H, r, G')$,

$$C(H, G') \leq \tau(H, r, G) + c_1 \|G' - G\| \leq d_1$$

(C.27)

with $d_1 = d + 1 + c_1$. In (C.27), take the infimum of the middle term over $r \in K(H)$ to obtain

$$C(H, G') \leq C(H, G) + c_1 \|G' - G\|.$$  

(C.28)

Let $c_2 = \frac{2}{a}(3 + \log g + d_1)$. Since $C(H, G') \leq d_1$, the generic result (C.28) can now also be rewritten by switching the roles of $G$ and $G'$ provided one also replaces $d$ with $d_1$ and $c_1$ with $c_2$. This yields $C(H, G) \leq C(H, G') + c_2 \|G' - G\|$. Therefore, the bound in (C.28) yields

$$|C(H, G) - C(H, G')| \leq c_2 \|G' - G\|.$$  

(C.29)

Define a linear mapping $\mathcal{L}$ for all $\rho \in K(H)$ denoted $\hat{\rho} := \mathcal{L}(\rho)$ given by

$$\hat{\rho}_{j,k} = \rho_{j,k} H(j)/H'(j) \text{ for } j \in \text{supp}(H) \text{ and all } k$$  

(C.30)

$$\hat{\rho}_{j,k} = 0 \text{ for } j \notin \text{supp}(H) \text{ and all } k$$  

(C.31)

By definition of $K(H')$ and $K(H)$, one readily verifies that $\hat{\rho} \in K(H)$. Thus, $\mathcal{L}$ maps $K(H')$ into $K(H)$ bijectively with an inverse mapping $\hat{\rho} \mapsto \rho$ defined by exchanging $\rho$ and $\hat{\rho}$ as well as $H'$ and $H$ in (C.30). Definition (C.30) forces $\text{supp}(\rho) = \text{supp}(\hat{\rho})$. Since $\text{supp}(H') = \text{supp}(H)$, then (2.13) yields $\text{supp}(\Psi(H', \rho)) = \text{supp}(\Psi(H, \hat{\rho}))$. Hence, either these two supports contain $\text{supp}(G)$ or neither of them do. This implies the equivalence

$$\tau(H', \rho, G) < \infty \iff \tau(H, \hat{\rho}, G) < \infty.$$  

(C.32)

For $\rho \in K(H')$, the bound $\|\rho\| \leq F_g$ holds due to (2.31). Thus, we get $|1 - H(j)/H'(j)| \leq \|H' - H\|/a$ for $j \in \text{supp}(H') = \text{supp}(H)$. Hence, for all $\rho \in K(H')$,

$$\|\rho - \hat{\rho}\| \leq \|\rho\| \|H' - H\|/a \leq \frac{F_g}{a} \|H' - H\| \leq 1.$$  

(C.33)

Define $U = \{r \in K(H) | \tau(H, r, G) \leq d\}$. This implies $C(H, G) = \min_{r \in U} \tau(H, r, G)$ since $C(H, G) \leq d$. For $\hat{\rho} \in U$, the composite cost $\tau(H, \hat{\rho}, G)$ is finite so that $\tau(H', \rho, G)$ is also finite due to (C.32). Apply Proposition 3.20 to get a constant $\eta = \eta(D, a, \mathcal{P})$ given by (3.31), and such that

$$|\tau(H', \rho, G) - \tau(H, \hat{\rho}, G)| \leq \eta \left(\|\rho - \hat{\rho}\|^{1/2} + \|H' - H\|\right).$$  

47
For $\hat{\rho} \in U$, the bound in (C.33) yields
\[
|\tau(H', \rho, G) - \tau(H, \hat{\rho}, G)| \leq \eta_0 \|H' - H\|^{1/2}
\] (C.34)
with $\eta_0 = 2(F_g/a)^{1/2}$. Since $\mathcal{L}$ is a bijection and $U \subset K(H)$, any $r \in U$ is of the form $r = \hat{s}$ for some $s = s(r) \in K(H')$. Then, for any $r \in U$, applying (C.34) to $\rho = s$ and $r = \hat{s}$ implies
\[
C(H', G) \leq \tau(H', s, G) \leq \tau(H, r, G) + \eta_0 \|H' - H\|^{1/2}.
\]
Take the infimum of the right-hand side over all $r \in U$ to get
\[
C(H', G) \leq C(H, G) + \eta_0 \|H' - H\|^{1/2} \leq d + \eta_0.
\] (C.35)
Define $V = \{\rho \in K(H) \mid \tau(H', \rho, G) \leq d + \eta_0\}$. This implies $C(H', G) = \min_{\rho \in V} \tau(H', \rho, G)$. For $\rho \in V$, the cost $\tau(H', \rho, G)$ is finite so that $\tau(H, \hat{\rho}, G)$ is also finite. Apply again Proposition 3.20 with $A = d + \eta_0$ to get $\eta_1 = \eta(d + \eta_0, a, \mathcal{P})$ such that, with (C.33),
\[
|\tau(H', \rho, G) - \tau(H, \hat{\rho}, G)| \leq \eta_1 \left(\|\rho - \hat{\rho}\|^{1/2} + \|H' - H\|\right) \leq \eta_2 \|H' - H\|^{1/2}
\] (C.36)
with $\eta_2 = 2(F_g/a)^{1/2}$. This implies, for all $\rho \in V$,
\[
\tau(H', \rho, G) \geq \tau(H, \hat{\rho}, G) - \eta_2 \|H' - H\|^{1/2} \geq C(H, G) - \eta_2 \|H' - H\|^{1/2}.
\]
Take the infimum of the left-hand side over $\rho \in V$ to get $C(H', G) \geq C(H, G) - \eta_2 \|H' - H\|^{1/2}$. Combine this with (C.35), and set $\gamma = \max\{\eta_1, \eta_2\}$ to obtain
\[
|C(H', G) - C(H, G)| \leq \gamma \|H' - H\|^{1/2}.
\] (C.37)
This forces $C(H', G) \leq d + \gamma$. Apply then (C.29) to obtain $c_3 = c_2(d + \gamma, a, \mathcal{P})$ such that
\[
|C(H', G) - C(H', G')| \leq c_3 \|G' - G\|.
\]
Combining the above expression with (C.37) yields the equation (B.56) with the constant $c = c_3 + \gamma$ and concludes the proof.

D Proofs of Theorems in Section 4

Proof of Theorem 4.12

Define $c > 1$ and $0 < \eta < 1$ by $c = 1 + \max_j (\sum_k q_{j,k})$ and $1 - \eta = \max_{1 \leq k \leq g-1} F_k/F_{k+1}$. Fix temporarily a constant $0 < A < 1/2$ to be selected later on. Impose $mc \leq A$ so that all $m_j = m \sum_k Q_{k,j} \leq A$. For $H \in \mathcal{H}$, set $U_j(H) = F_j H(j)(1 - m_j) + m \sum_k Q_{k,j} F_k H(k)$. Setting $G(j) = \zeta_j(H) = U_j(H)/< F, H >$, we have that $G(j) = 0$ if and only if $H(j) = 0$ and $H(k) = 0$ for all $k$ such that $Q_{k,j} > 0$. This implies supp$(H) \subset$ supp$(\zeta(H))$. Consequently, since $h_{n+1} = \zeta(h_n)$, the set $S_n = \operatorname{supp}(h_n)$ must increase with $n$. Therefore, there exists a fixed (sub)set of genotypes $S_\infty$ and an $n_0$ such that $S_n = S_\infty$ for $n \geq n_0$. An easy recurrence based on the conditions above for $G(j) = \zeta_j(H) = 0$ shows that $S_\infty$ is the set of all $j$ reachable by some finite sequence $k_1, k_2, \ldots, k_T = j$ with $k_1 \in \operatorname{supp}(h_1)$ and all $q_{k_t, k_{t+1}} > 0$. 

48
Now, call a genotype \( q \) dominant in \( H \) if \( H(q) = \max_k H(k) \). Then, we have \( \langle F, H \rangle \leq gF_\mu H(q) \), \( 1/g \leq H(q) \), and
\[
(1 - A)F_j H(j) \leq U_j(H) \leq F_j H(j) + A\langle F, H \rangle \leq F_j H(j) + gAF_\mu H(q) \quad (D.1)
\]
for all \( j \). Since \( G(j)/G(q) = U_j(H)/U_q(H) \), this implies
\[
\left( \frac{(1 - A)F_j}{(F_q + gAF_\mu)} \right) \left( \frac{H(j)}{H(q)} \right) \leq \frac{G(j)}{G(q)} \leq \frac{F_j H(j) + gAF_\mu H(q)}{F_q H(q)(1 - A)}.
\]
(D.2)

For \( j < q \), one has \( H(j) \leq H(q) \) and \( F_j \leq (1 - \eta)F_q \) by definition of \( c \) and \( \eta \). Then, (D.2) yields
\[
\frac{G(j)}{G(q)} \leq \left( \frac{1 - \eta}{1 - A} \right) \frac{H(j)}{H(q)} + \frac{gAF_\mu}{1 - A} \leq \frac{1 - \eta + AF_\mu/F_q}{1 - A}.
\]
(D.3)

Impose \( A < A_1 = \frac{\eta F_1}{1 + gF_\mu F_q} \) to get \( G(j)/G(q) \leq 1 - \eta/2 < 1 \) for all \( j < q \). Hence, the dominant genotype \( q(G) \) in \( G = \zeta(H) \) verifies \( q(G) \geq q(H) \). Consequently, the dominant genotype \( q_n = q(h_n) \) of \( h_n \) verifies \( q_{n+1} \geq q_n \) for all \( n \) so that there exists a finite \( n_1 \geq n_0 \) and a genotype \( q^* \) such that \( q_n = q^* \) for all \( n > n_1 \). From (D.1), we get \( h_n(q_n) \geq 1/g \) and hence \( q^* \in S_\infty \). Let \( s = \max(S_\infty) \) be the fittest genotype in \( S_\infty \) so that \( q^* \leq s \). Assume there is a \( j \in S_\infty \) verifying \( q^* < j \leq s \). We will proceed by contradiction to show that such a \( j \) cannot exist. For \( n > n_1 \) one has \( h_n(j) \leq h_n(q^*) \) and, by (D.2),
\[
\theta \left( \frac{h_n(j)}{h_n(q^*)} \right) \leq \frac{h_{n+1}(j)}{h_{n+1}(q^*)} \leq 1
\]
with \( \theta = (1 - A)F_j/(F_q + gAF_\mu) \). Since \( j > q^* \), we get \( F_j \geq F_q/(1 - \eta) \) by definition of \( \eta \) so that
\[
\theta \geq \frac{1 - A}{(1 - \eta)(1 + gAF_\mu/F_q)} \geq \frac{1 - A}{(1 - \eta)(1 + AgF_\mu/F_1)}.
\]
(D.4)

Impose \( A < A_2 = \frac{\eta^2 F_1}{2gF_\mu} \) to force \( \theta > 1 + \eta \). By recurrence, (D.4) implies for \( s \geq j > q^* \) and \( n' > n > n_1 \) that \( (1 + \eta)^{n-n} \frac{h_n(i)}{h_n(q^*)} \leq 1 \). Fix \( n > n_1 \). Since \( j \in S_\infty \), one has \( h_n(j) > 0 \). Letting \( n' \to \infty \) yields a contradiction. Hence there is no \( j \in S_\infty \) with \( q^* < j \leq s \), so that \( q^* = s \).

For \( i \in S_\infty \) with \( i < s \) and \( n > n_1 \), the terms \( y_n(i) = h_n(i)/h_n(s) \) verify \( y_{n+1}(i) \leq \mu y_n(i) + \nu \) by (D.3) and \( q_n = s \) with \( \mu = 1 - \eta/2 \) and \( \nu = 2gAF_\mu/F_1 \). Iterating this inequality gives \( y_n(i) \leq \mu^{n-n_1} y_1(i) + \frac{\nu}{1-\mu} \). Select \( n_2 = n_2(A) > n_1 \) to force \( (1 - \eta/2)^{n-n_1} y_1(i) < A \) for all \( n > n_2 \) and \( i \in (S_\infty \setminus s) \). Then, \( y_n(i) \leq c_1 A \), with \( c_1 = 1 + \frac{AF_\mu}{\eta F_1} \). This yields \( h_n(i) \leq c_1 A \) for all \( n > n_2(A) \) and \( i \in (S_\infty \setminus s) \). Hence \( h_n(s) \geq 1 - c_1(g - 1)A \) for \( n > n_2(A) \), since \( \text{supp}(h_n) = S_\infty \). We now fix \( A = A_3 = \min \left\{ A_2, \frac{1}{2gct^2} \right\} \) and set \( n_3 = n_2(A_3) \) to get the fixed lower bound \( h_n(s) \geq 1/2 \) for all \( n > n_3 \). For \( n > n_3 \) let \( z_n = \text{vector of all the } z_n(i) = y_n(i) \) with \( i \in (S - s) \). Since \( \text{supp}(h_n) = S \), the nonzero \( U_i(h_n) \) only depend on \( z_n \) and can be denoted \( U_i(z_n) \). We have then \( z_{n+1} = f(z_n) \) with \( f_i(z) = U_i(z)/U_s(z) \). Set \( p = \text{card}(S) - 1 \). The rational fraction \( f(z) \) is well-defined on the set \( \Gamma \) of all \( z \in [0,1]^p \) such that \( U_s(z) \geq 1/2 \) for \( 0 \leq m \leq A_3/c \). For the limit case \( m = 0 \) of no mutations, the function \( f(z) \) has the form \( f_i(z) = F_i/F_s \) for all \( i \in (S - s) \). The Jacobian matrix \( D_z f \) obviously verifies
\[
\left\| D_z f \right\| \leq \max_{i \in (S - s)} \frac{F_i}{F_s} \leq 1 - \eta \quad (D.5)
\]
for all $z \in [0,1]^p$. For $z \in \Gamma$, the numerator and denominator of each rational fraction $f_i(z)$ are separately affine in $m \leq A_3/c$ and $z$, with denominator bounded below by $1/2$ and uniformly bounded coefficients. Elementary algebraic computations then prove that as $m \to 0$, the Jacobian matrix $D_z f(z)$ tends to $D_z f(0)$ uniformly over all $z \in \Gamma$. Due to (D.5), this provides a constant $0 < A_4 < A_3/c$ such that $\|D_z f(z)\| \leq (1-\eta/2)$ for all $m < A_4$ and all $z \in \Gamma$. This yields $\|z_n+1 - z_n\| \leq (1-\eta/2) \|z_n - z_{n-1}\|$ for $m < A_4$ and $n > 1 + n_3$. This contraction property classically shows that $z^* = \lim_{n \to \infty} z_n$ exists with $z^* \in \Gamma$. Furthermore, $z^*$ is the unique solution of $z^* - f(z^*) = 0$. Note that $z^*$ is an implicit function of $m$ for $0 \leq m < A_4$, with $z^*(0) = 0$. Due to (D.5), the Jacobian $Id - D_z f$ of $z - f(z)$ is invertible for $m = 0$. Since $f(z)$ is a $C^\infty$ function of $(m,z) \in [0,A_4] \times \Gamma$, the implicit function theorem applies to $z - f(z) = 0$ and proves the existence of a constant $0 < A_5 < A_4$ such that $z^*$ is a $C^\infty$ function of $m \in [0,A_5)$. Since $(1-h_n(s))/h_n(s) = \sum_{i \in (S-s)} z_n(i)$ and $h_n(i) = h_n(s)z_n(i)$ for $i \in (S-s)$, we see that $H = \lim_{n \to \infty} h_n$ exists with $\text{supp}(H) \subset S$ and verifies $H(s) = 1/[1 + \sum_{i \in (S-s)} z^*(i)]$ and $H(i) = H(s)z^*(i)$. Hence, $H$ is a $C^\infty$ function of $m \in [0,A_5)$ and $\text{supp}(H) = S$. For $m = 0$, the solution of $z - \hat{f}(z) = 0$ is clearly $z^* = 0$, and the associated histogram $H$ verifies $H(s) = 1$ and $\text{supp}(H) = \{s\}$. The first-order Taylor expansion of $H$ as a function of $m$ is of the form $H(i) \simeq mv_i$ for $i \in (S-s)$ and $H(s) \simeq 1 - um$. Substitute this into $H = \zeta(H)$ to get $v_i = Q_{s,i}/(F_s - F_i)$ and $u = \sum_i v_i$. This concludes the proof.

References

[1] R. Azencott. Large deviations theory and applications. In Hennequin P. L., editor, Probabilities at Saint-Flour VIII-1978, pages 1–176. Springer Lect. Notes Math. vol 774, 1980.

[2] R. Azencott, M. I. Freidlin, and S.R.S. Varhadan. Large Deviations Theory and Applications. In Hennequin P. L., editor, Probabilities at Saint-Flour VIII-1978, pages 1–176. Springer Lect. Notes Math. vol 774, 1980.

[3] R. Azencott and G.Ruget. Random mixing of differential equations: Large deviations theory. Zeitschrift Wahrschein.Th. und ver. Gebiete, 38:1–54, 1977.

[4] J. E. Barrick and R. E. Lenski. Genome dynamics during experimental evolution. Nature Reviews. Genetics, 14(12):827–839, 2013.

[5] J. E. Barrick, C. C. Strelioff, R. E. Lenski, and M. R. Kauth. Escherichia coli rpoB mutants have increased evolvability in proportion to their fitness defects. Molecular Biology and Evolution, 27(6):1338–1347, 2010.

[6] Fabio A. C. C. Chalub and José Francisco Rodrigues, editors. The Mathematics of Darwin’s Legacy (Mathematics and Biosciences in Interaction). Birkhäuser, 2011.

[7] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. Stoch. Proc. Appl., 116(8):1127–1160, 2006.

[8] N. Champagnat, R. Ferriere, and G. Ben Arous. The canonical equation of adaptive dynamics: A mathematical view. Selection, 2:71–81, 2001.

[9] N. Champagnat and A. Lambert. Evolution of discrete populations and the canonical diffusion of adaptive dynamics. Ann. Appl. Prob., 17:102–155, 2007.
[10] Nicolas Champagnat, Régis Ferrière, and Sylvie Méléard. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. *Theoretical Population Biology*, 69(3):297 – 321, 2006. ESS Theory Now.

[11] T. F. Cooper, D. E. Rozen, and R. E. Lenski. Parallel changes in gene expression after 20,000 generations of evolution in *E. coli*. *Proc. Nat. Acad. Sci. USA*, 100:1072–1077, 2003.

[12] V. S. Cooper, D. Schneider, M. Blot, and R. E. Lenski. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli*. *Journal of Bacteriology*, 183:2834–2841, 2001.

[13] J. A. Moura de Sousa, P. R. A. Campos, and I. Gordo. An abc method for estimating the rate and distribution of effects of beneficial mutations. *Genome Biology and Evolution*, 5(5):794–806, 2013.

[14] Daniel E. Deatherage, Kepner L. Jamie, Albert F. Bennett, Richard E. Lenski, and Jeffrey E. Barrick. Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. *Proc. Nat. Acad. Sci. USA*, 114(10):E1904–E1912, 2017.

[15] A. Dembo and O. Zeitouni. *Large deviations Techniques and applications*. Springer-verlag, 1998.

[16] M. Desai and D. Fisher. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–1798, 2007.

[17] Richard Durrett. *Probability Models for DNA Sequence Evolution (Probability and Its Applications)*. Springer, 2008.

[18] S. F. Levy et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181 – 186, 2015.

[19] J. W. Fox and R. E. Lenski. From here to eternity - the theory and practice of a really long experiment. *PLoS Biology*, 13(6):e1002185, 2015.

[20] M. Freidlin and A. Wentzell. *Random perturbations of dynamical systems*, volume 260. Springer, 1998.

[21] Brett Joseph Geiger. *Large Deviations for Dynamical Systems with Small Noise*. PhD thesis, University of Houston, https://hdl.handle.net/10657/4801, 2017.

[22] I. Gordo, L. Perfeito, and A. Sousa. Fitness effects of mutations in bacteria. *Journal of Molecular Microbiology and Biotechnology*, 21(1-2):20–35, 2012.

[23] Matthew Hegreness, Noam Shoresh, Daniel Hartl, and Roy Kishony. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science*, 311(5767):1615–1617, 2006.

[24] C. Illinworth and V. Mustonen. A method to infer positive selection from marker dynamics in asexual population. *Bioinformatics*, 28(6):831–837, 2012.
[25] R. Korona, C. H. Nakatsu, L. J. Forney, and R. E. Lenski. Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. *Proc. Nat. Acad. Sci. USA*, 91(19):9037–9041, 1994.

[26] Sylvie Meleard. *Modèles aléatoires en Ecologie et Evolution (Mathématiques et Applications, 77)* (French). Springer, 2016.

[27] Sylvie Meleard and Vincent Bansaye. *Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior (Mathematical Biosciences Institute Lecture Series, 1.4)*. Springer, 2015.

[28] F. Peng, S. Widmann, A. Wünsche, K. Duan, K. A. Donovan, R. C. J. Dobson, R. E. Lenski, and T. F. Cooper. Effects of beneficial mutations in *pykF* gene vary over time and across replicate populations in a long-term experiment with bacteria. *Molecular Biology and Evolution*, 10(6):msx279, 2017.

[29] L. D. Plank and J. D. Harvey. Generation time statistics of *Escherichia coli* B measured by synchronous culture techniques. *Journal of General Microbiology*, 115:69–77, 1979.

[30] S. H. Rice. *Evolutionary Theory*. Sinauer Associates, 2004.

[31] S. M. Ross. *Stochastic Processes*. John Wiley & Sons, New York, 1996.

[32] Justin Schoneman. *Stochastic Models for Genetic Evolution*. PhD thesis, University of Houston, [http://hdl.handle.net/10657/3540](http://hdl.handle.net/10657/3540), 2016.

[33] D. Simon. *Evolutionary Optimization algorithms*. John Wiley & Sons, New York, 2013.

[34] Yingxue Su, Brett Geiger, Andreas Mang Ilya Timofeyev, and Robert Azencott. Rare events analysis and computation for stochastic evolution of bacterial populations, https://arxiv.org/abs/2308.14135, 2023.

[35] S. R. S. Varadhan. *Large deviations and applications, CBMS-NSF*, volume 46. SIAM, 1984.

[36] F. Vasi, M. Travisano, and R. Lenski. Long-term experimental evolution in *Escherichia coli*. ii. changes in life-history traits during adaptation to a seasonal environment. *The American Naturalist*, 144(3):432–456, 1994.

[37] Dominik Wodarz and Natalia Komarova. *Computational Biology of Cancer: Lecture Notes and Mathematical Modeling*. World Scientific, 2005.

[38] R. J. Woods, J. E. Barrick, T. F. Cooper, U. Shrestha, M. R. Kauth, and R. E. Lenski. Second-order selection for evolvability in a large *Escherichia coli* population. *Science*, 331(6023):1433–1436, 2011.

[39] W. Zhang, V. Sehgal, D. Dinh, R. R. Azevedo, T. Cooper, and R. Azencott. Estimation of the rate and effect of new beneficial mutations in asexual populations. *Theoretical Population Biology*, 81(2):168–178, 2012.