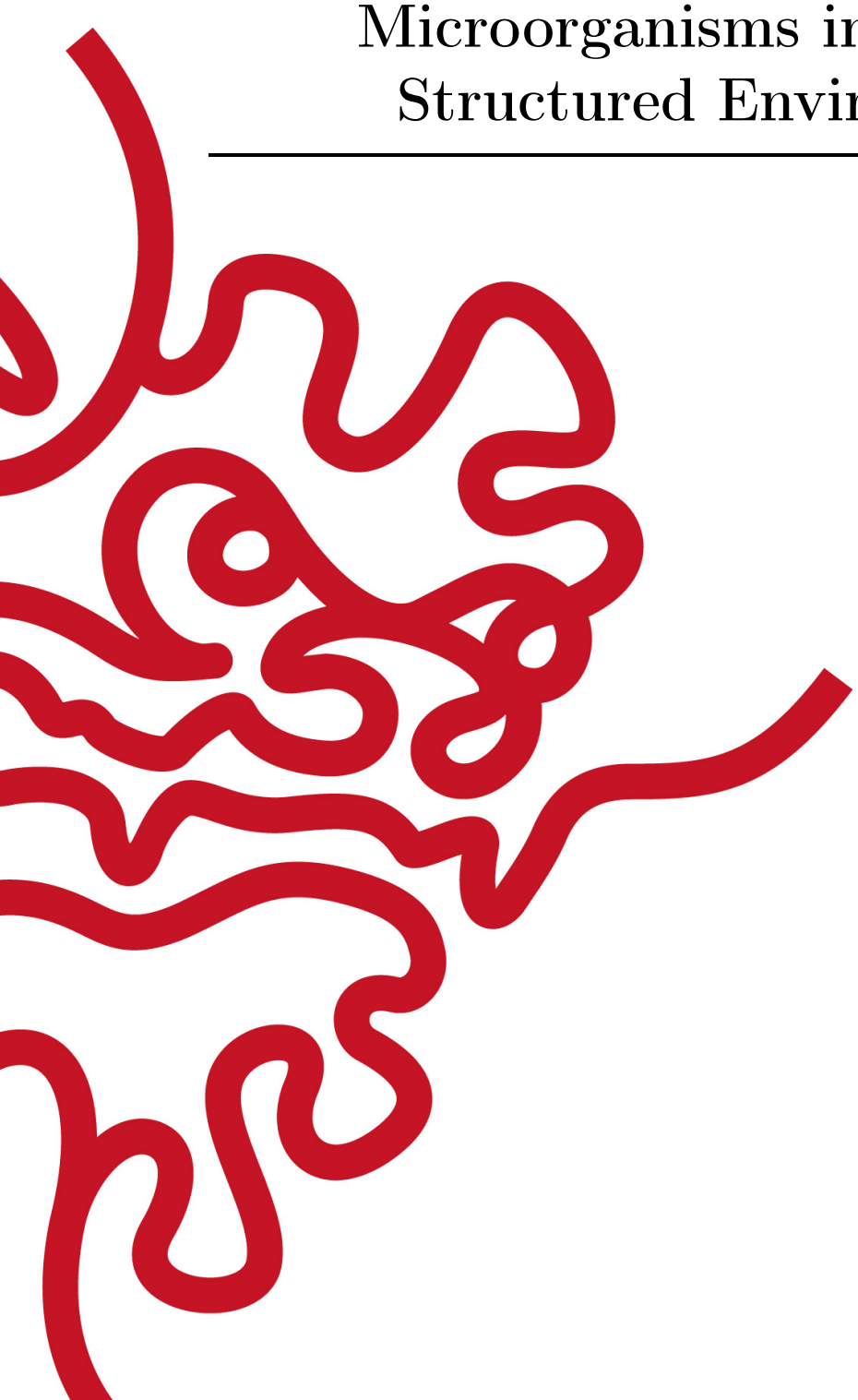
Okinawa Institute of Science and Technology
Graduate University

Thesis submitted for the degree

Doctor of Philosophy

Population Dynamics of Microorganisms in Spatially Structured Environments



by

Anzhelika Koldaeva

Supervisor: **S. Pigolotti**

September 2023

Declaration of Original and Sole Authorship

I, Anzhelika Koldaeva, declare that this thesis entitled *Population Dynamics of Microorganisms in Spatially Structured Environments* and the data presented in it are original and my own work.

I confirm that:

- No part of this work has previously been submitted for a degree at this or any other university.
- References to the work of others have been clearly acknowledged. Quotations from the work of others have been clearly indicated, and attributed to them.
- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.
- None of this work has been previously published elsewhere, with the exception of the following:

  1. **Koldaeva, A.**, Tsai, H.F., Shen, A.Q. and Pigolotti, S. "Population genetics in microchannels", Proceedings of the National Academy of Sciences, 2022.
  2. Tsai, H.F., Carlson, D.W., **Koldaeva, A.**, Pigolotti, S. and Shen, A.Q. "Optimization and Fabrication of Multi-Level Microchannels for Long-Term Imaging of Bacterial Growth and Expansion", Micromachines, 2022.
  3. Martín, P.V., **Koldaeva, A.** and Pigolotti, S. "Coalescent dynamics of planktonic communities", Physical Review E, 2022.

Date: September 2023
Signature:

Abstract

Microbial populations live and grow in spatially structured environments. These structures lead to spatial patterns in populations and alter the course of their natural evolution. Such phenomena are theoretically studied using spatially explicit models. However, these models are still poorly understood due to their analytical and numerical complexity. In this thesis, we study two systems of microorganisms living and proliferating in different spatially structured environments. The first system consists of populations of *Escherichia coli* growing in rectangular microchannels with two open ends. We study such populations with a lattice model in which cells shift each other along lanes as they reproduce. The model predicts rapid diversity loss along the lanes, with neutral mutations appearing in the middle of the channel being the most likely to fixate. These theoretical predictions are in agreement with our experimental observations. The second system is constituted by planktonic microorganisms that are transported by chaotic oceanic currents. To replicate their dynamics, we employ an individual-based coalescence model. The model predicts the effect of oceanic currents on the biodiversity of planktonic communities, as observed in metabarcoding data sampled from oceans and lakes around the world.

Acknowledgment

First, I would like to express my deepest gratitude to my supervisor, Professor Simone Pigolotti, without whom my PhD journey would not be possible. Also, I'm grateful to all my colleagues from the Biological Complexity Unit, for being encouraging and responsive. In particular, I would like to thank Paula Villa Martín for her support and assistance. Special thanks to the Micro/Bio/Nanofluidic Unit, in particular, Professor Amy Shen, Paul Hsieh Fu Tsai, and Daniel Carlson for working on the experimental part of my project. Also, I would like to acknowledge the Scientific Computing and Data Analysis Section (SCDA) at OIST for providing computing resources. Finally, I would like to thank all my friends and family for being caring and supportive throughout my PhD.

Abbreviations

AmpR	ampicillin-resistant gene
DNA	deoxyribonucleic acid
E. coli	bacteria *Escherichia coli*
IBD	isolation by distance
MRCA	most recent common ancestor
SADs	species abundance distributions

Contents

List of Figures

Chapter 1

Introduction

1.1 Outline of the Thesis

In this thesis, we study two systems of microorganisms evolving in spatially structured environments. First, we study populations of bacteria *E. coli* growing in rectangular microchannels with two open ends. Second, we explore the evolution of planktonic communities living in aquatic environments with and without advection.

In Chapter 1, we introduce the main concepts of the field. In Section 1.2, we introduce the biodiversity of a population from the ecological and genetic point of view. In Section 1.3, we motivate why we focus on populations of microorganisms. In particular, we discuss dense populations of bacteria and freely moving plankton microorganisms in aquatic environments. In Section 1.4, we provide a brief historical overview and the main theoretical concepts of population genetics.

Chapter 2 is devoted to the mathematical modeling of population dynamics and genetics. In Section 2.1, we introduce the main terminology and notations. We start with introducing the models for well-mixed populations in Section 2.2. Then, we introduce the lattice models in Section 2.3. In Section 2.4, we briefly discuss the concept of the coalescent theory and introduce the basic coalescent model. Finally, we introduce off-lattice forward and backward models in Section 2.5.

In Chapter 3, we briefly introduce neutral populations of *E. coli* growing in rectangular microchannels with two open ends and provide references to our papers.

In Chapter 4, we theoretically study the competition of two non-neutral strains of *E. coli* in rectangular microchannels with two open ends. We find their fixation times and probabilities for different initial distributions of mutants. In Section 4.1, we study one mutant at a random location. In Section 4.2, we explore a cluster of mutants located at an open end. In Section 4.3, we investigate the case of randomly distributed mutants in a channel.

In Chapter 5, we briefly discuss our results on the evolution of planktonic communities in aquatic environments with and without advection and provide a reference to our paper.

Finally, in Chapter 6, we provide concluding discussions and further perspectives on both systems studied in the thesis.

1.2 Biodiversity of populations

Life forms on our planet are astonishingly diverse. This diversity is a result of a complex interplay between individuals and the environment they live and reproduce in. A group of interacting individuals inhabiting a particular environment forms a population. The "diversity" of a population is a term with different meanings, depending on the level of biological organization one is interested in.

On the population level, individuals are subdivided into species. Individuals belonging to a species have common characteristics and are capable of interbreeding. The number of different species in a given population is a main measure of species diversity. Species emerge and become extinct as a consequence of ecological factors acting on them, such as food, water, shelter, predation and so on. This interplay between individuals and their environment shapes the species diversity and is studied by population ecology, see Fig. 1.1. Within one species, individuals may differ by variants of the same gene called *alleles*. The number of distinct alleles within a population of species constitutes its genetic diversity. New alleles emerge due to random mistakes in the DNA replication called *mutations*. Some of these mutations are more likely to survive in the next generation than others, due to the action of *natural selection*. Moreover, due to random events, such as natural disasters, some individuals can die and not pass their alleles to the next generations. This phenomenon results in a loss of alleles and is called *genetic drift*. Finally, due to *gene migration*, populations of species can exchange alleles between each other, thereby increasing their genetic diversity. Mutations, genetic drift, natural selection, and gene migration are usually thought to be the four main evolutionary forces. The way in which these forces shape the genetic diversity of a population of species is studied by population genetics, see Fig. 1.1.

Both population ecology and population genetics are conceptually similar, i.e. they study how a group of biological forms change under external factors acting on them. From this point of view, there is a natural correspondence between the two theories: individuals and species in population ecology can be associated with genes and alleles in population genetics [5]. Because of this correspondence, the mathematical formalism of population ecology and population genetics are similar as well. For this reason, we use the term "biodiversity" in a broad sense, meaning either the ecological or genetic diversity. By "evolution" of a population we understand a change of its biodiversity in time, under the influence of ecological and evolutionary factors. In addition to ecological and evolutionary factors, spatial structures affect the biodiversity of a population [9, 92]. Such structures are formed by spatial inhomogeneities in the environment that the population inhabits. Due to the variety of different spatial environments observed in nature, and the difficulty in their study, this aspect still remains poorly understood. Understanding of the role of spatially complex environments in the biodiversity of populations is the main focus of this thesis.

1.3 Populations of microorganisms

In this thesis, we restrict our attention to microorganisms. There are three main reasons for this.

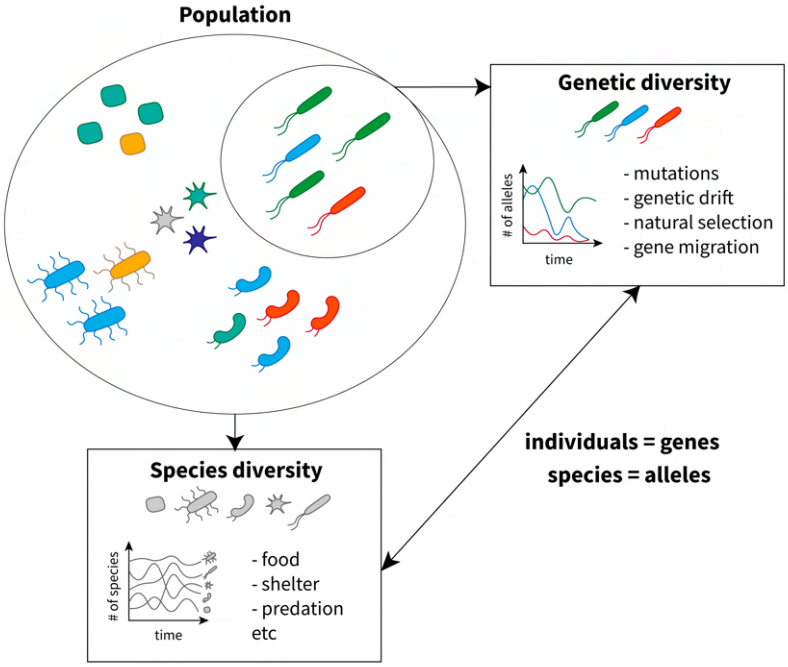


Figure 1.1: Ecological and genetic concepts of biodiversity. The biodiversity of a population can be understood as either species or genetic diversity. The correspondence "individuals ↔ genes", and "species ↔ alleles" makes the species and genetic diversity conceptually similar.

First, microorganisms are found everywhere in nature and play a leading role in numerous biogeochemical processes [35]. For example, bacteria living in soil provide plants with the chemical elements that are critical for their survival, and are responsible for about 90% of all biomass decomposition [57, 62]. Marine microorganisms, such as phytoplankton, are responsible for virtually all the photosynthesis that occurs in the ocean, and form the basis of aquatic food webs [39]. Despite the central role of microorganisms in various biological processes, we know very little about their diversity and underlying evolutionary dynamics [83]. Understanding the evolution of microorganisms has countless applications and can shed light on some possible solutions to global issues, such as climate change [15] and antibiotic resistance [75].

Second, microorganisms are simple and convenient model systems for studying evolution in action [23, 26, 61]. Specifically, they enable us to test predictions of mathematical models of large populations evolving over many generations. In particular, microorganisms have a short lifetime, which enables us to study their dynamics in real time for many generations [34]. Also, the small size of microorganisms allows an analysis of large samples with millions of individuals. Moreover, microorganism genomes are often simple, which makes them easier to sequence, classify and modify if necessary [65]. The findings of such studies can be extended to populations of different sizes and nature.

Finally, microorganisms have a small dispersal distance, i.e. the distance traveled by an individual in one generation. This helps to explore how the populations expand in their habitat, and how the spatial structure of the habitat shapes the evolution of

such populations.

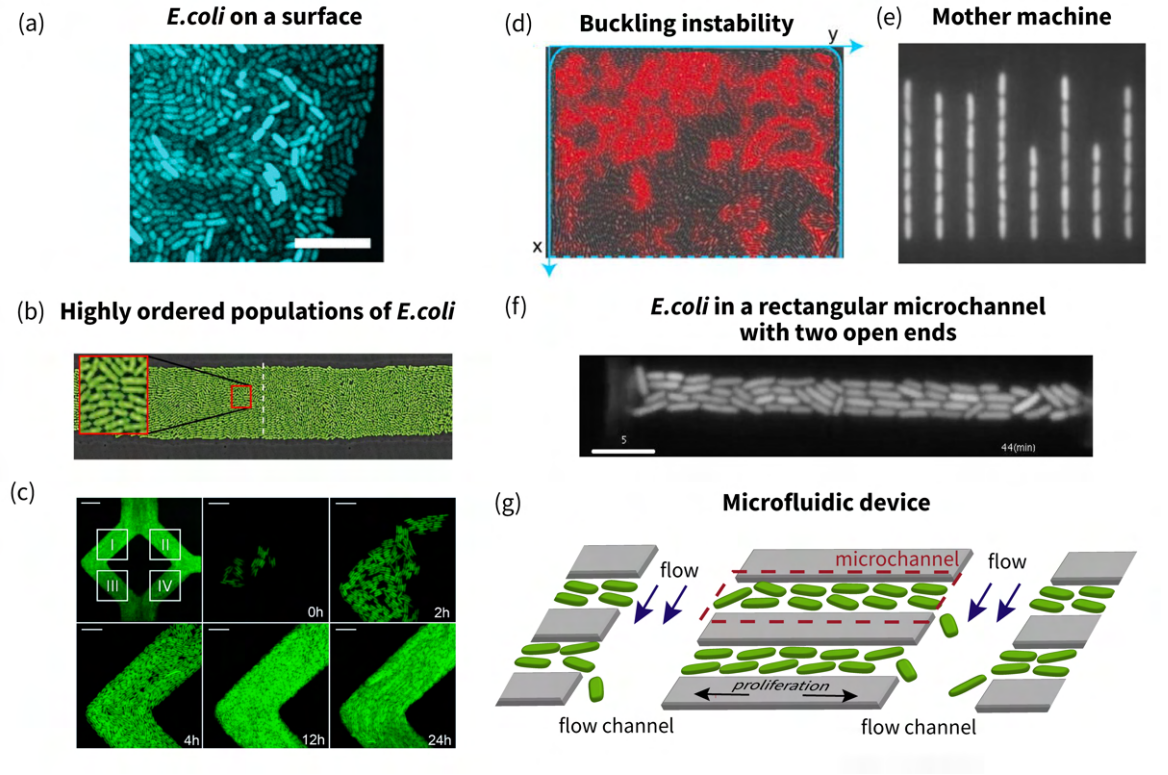


Figure 1.2: *E. coli* in spatially structured environments. (a) *E. coli* proliferating on a surface, taken from [97]. (b, c) Highly ordered populations of *E. coli* proliferating in narrow microchannels, originally found in [17] and [89], respectively. (d) Disordered population of *E. coli* growing in a wide rectangular microchannel with one open end, taken from [7]. (e) *E. coli* in a narrow microchannel with one open end, called "the mother machine", [91]. (f) *E. coli* in a microchannel with two open ends, taken from [37]. (g) Schematic representation of a microfluidic device.

In this thesis, we study two different spatially structured populations of microorganisms living at radically different spatial scales: dense populations of tightly packed bacteria in microchannels, and planktonic microorganisms that freely move in an aquatic environment. We aim at understanding how the environment shapes the biodiversity of such communities, and what effect it has on the course of their evolution. We introduce these systems in details in the following sections.

1.3.1 Dense populations of bacteria

Bacteria tend to live in dense populations of closely interacting individuals. Due to the lack of free space between cells, each newborn individual shifts the others in order to make space for itself. In some cases, this shifting dynamics is the only mechanism that makes individuals move in such populations. Moreover, the freedom of movement of each individual is restricted by the surface the population resides on. These two factors result in formation of complex spatial structures in populations. For instance,

rod-shaped bacteria *E. coli* proliferating on a surface form patches of co-aligned cells [97], see Fig.1.2, (a). The formation of such dense and sessile populations is important to withstand environmental stresses and maintain quorum sensing [70].

In nature, bacteria often live in environments characterized by spatial barriers. Often, bacteria actively seek shelter in small cavities and channels to survive in a harsh environment. For example, the soil consists of micropores and channels that are inhabited by microbes [69]. Micropores retain fluid which supply microbes with nutrients [20] and act as a shelter against larger predators [98].

The boundaries of such pores and channels create extra limitations for bacteria movement and affect the spatial organization of the entire population. For instance, *E. coli* proliferating in narrow channels reach a high global ordering of cells co-aligned with the walls of the channels [17, 89], see Fig.1.2, (b) and (c). In contrast, bacteria inhabiting a wide channel are persistently disordered due to a buckling instability [7], see Fig.1.2, (d). Shifting dynamics in combination with the size and shape of a channel are the key elements that determine the spatial structure of such populations, and, therefore, their evolution.

The simplest example of a population with shifting dynamics is a single lane of cells proliferating in a rectangular channel with one open end. In this case, cells reproduce and shift each other towards the open end. As the result, a cell at the open end of the channel is pushed out from the population. In applications, such channels are called "the mother machine" and the cell at the close end is called "the mother cell" [91], see Fig.1.2, (e). This is due to the fact that the mother cell is never eliminated from the population, and at some point, it becomes the ancestor of all individuals in the population [68].

A more complex case is that of microorganisms growing in rectangular channels with two open ends [37], see Fig.1.2, (f). In such populations, bacteria shift each other towards either the right or the left open end as they reproduce. The direction of the shift depends on the cell's position along the channel and the ordering of the other cells. Similarly to the mother machine, cells at the open ends are pushed out from the channel.

Such populations of bacteria are experimentally studied with microfluidic devices [81]. These devices consist of multiple microchannels that harbor populations of cells. Microchannels are connected with flow channels through which the nutrients are delivered to the growing populations. The flows in the flow channels remove the expelled cells and cell waste from the microchannels, preventing clogging, see Fig. 1.2, (g). The main advantage of microfluidic devices is their design flexibility [100]. In particular, microchannels can be designed in different shapes and sizes of interest, to mimic spatially complex habitats of bacteria observed in nature.

Microfluidic devices are also used in studying population dynamics and properties at a single-cell level [58, 60]. For example, the mother machine has been widely used for studying the variability of properties of cells born from the same ancestor [91]. A device with microchannels with two open ends has been employed for studying the interplay between the population growth rate and the single-cell doubling time [37].

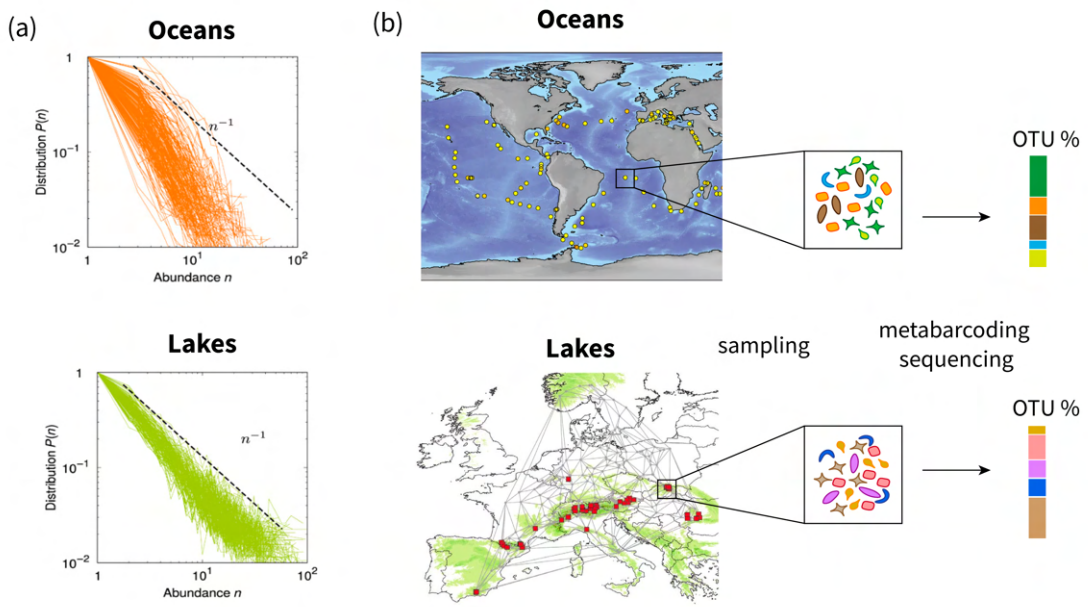


Figure 1.3: Biodiversity of planktonic communities in different environments. (a) SADs (species abundance distributions) of planktonic populations sampled in oceans (top) and lakes (bottom). The figure is taken from [88]. (b) Planktonic communities sampled in oceans and lakes (left) are sequenced with metabarcoding technique (right). The leftmost figures are originally found in [85] and [6].

1.3.2 Freely moving plankton microorganisms

Plankton microorganisms can be transported across very large distances by currents and can be found everywhere in the global ocean [28]. The biodiversity of planktonic communities is extremely rich [22].

The cause of plankton biodiversity remains unclear. In relatively-homogeneous aquatic environments, plankton microorganisms compete for a few limited resources such as nutrients and light. The niche theory of biodiversity predicts that such an environment can be inhabited only by a limited number of species, each of which would occupy its own niche. This prediction is not consistent with the astronomical number of planktonic species observed in nature. This conflict is known as the "paradox of the plankton", as formulated by Hutchinson in 1961 [42].

The paradox suggests that additional mechanisms, besides niches, promote planktonic biodiversity. Several possible mechanisms have been proposed, such as the seasonal cycle of the environment, intrinsically chaotic dynamics, incomplete mixing, and so on [82]. Notably, it has been shown that spatial heterogeneity of aquatic environments is an important reason for coexistence of species [8, 45].

Chaotic oceanic flows act like barriers and limit dispersal of plankton microorganisms. This reduces competition among microorganisms at opposite sides of barriers. As a result, more competitive species can coexist with each other, and rare species have more chances to survive in such an environment. In contrast, the absence of flows increases competition between species and positively affects the diversity of very

abundant species. This phenomenon can be quantified via the species abundance distributions (SADs), defined as the relationship between the frequency of species $P(n)$ observed in a sample as a function of their observed abundance n. SADs of plankton sampled in oceans are steeper than the ones of plankton sampled in lakes, see Fig. 1.3, (a). This phenomenon has been extensively studied experimentally and theoretically [85, 88].

Studying such diverse populations as planktonic communities in their natural habitats requires sampling and sequencing a large number of individuals. In the TARA ocean expedition, millions of protist microorganisms were sampled from 121 different locations in the global ocean [85]. Another example is a study in which plankton microorganisms were sampled in 218 European freshwater lakes [6], see Fig. 1.3, (b) left.

Such biological samples are sequenced with DNA metabarcoding techniques. DNA metabarcoding relies on high-throughput DNA sequencing technology and allows identification of millions of species within one sample [38]. The technique is based on targeted amplification and sequencing of specific highly conserved DNA regions, called DNA barcodes. Since those regions are highly conserved, sequences in the sample with a high degree of genetic similarity are likely to originate from individuals within the same taxonomic group. These groups, identified by genetic similarity, are called operational taxonomic units (OTUs). The biodiversity of a population can be estimated from the number and abundance of OTUs found in a sample, see Fig. 1.3, (b) right.

1.4 Theoretical frameworks in population genetics

1.4.1 Emergence of evolutionary biology

The origin and cause of the great biodiversity of species we observe around us have been puzzling the greatest minds for centuries. The first attempts at understanding the diversity of animals and plants have roots in ancient Greek philosophy, and the controversial theory of *essentialism* (around 600 BC) [13]. According to this theory, the species are determined by a series of attributes, that had always been present and that can not be changed or modified over time. However, in 1859, Charles Darwin proposed a new revolutionary idea in his monumental work "The Origin of Species" [21]. He stated that all species evolved from a common ancestor due to natural selection through a branching pattern of evolution. Despite the initial criticism, Darwin's ideas became the backbone of current *evolutionary biology*.

Only a few decades later, in 1900, Mendel's Principles of Heredity have been accepted in the scientific community [4]. Mendel proposed the idea of discrete traits determined by a combination of recessive and dominant "factors" inherited from two parents. Whereas, Darwin explained inheritance through *pangenesis* - a mechanism that continuously blends the traits of each parent [41]. At that time, these two theories were incompatible and led to a split in the scientific society. Moreover, neither Mendel's, nor Darwin's theories were individually sufficient to fully explain all the phenomena observed in the natural world.

1.4.2 Population genetics

Mendel's and Darwin's ideas were reconciled in the 1920s by a statistician, R. Fisher. He proposed the idea of cumulative Mendelian factors that explain the distribution of human traits [63]. The next significant steps were taken at the beginning of 1930s by Fisher, Haldane and Wright. They developed the first mathematical model, called the *Wright-Fisher model*, to explore the effect of population size on random fluctuations of gene frequency under the influence of evolutionary forces (natural selection, mutation and random drift) [94]. Almost 20 years later, another model was formulated, called the *Moran model* as a modification of the Wright-Fisher model for overlapping populations [64]. These two models started the new field called *population genetics*. They remain the dominant methodology of modern population genetics and find numerous applications beyond this field.

The Wright-Fisher and Moran models describe evolution of *well-mixed populations*, i.e. populations with no spatial structure where each individual can interact with any other, see Fig. 1.4, (a). In part, this is because mathematical predictions are possible for such simplified models. Well-mixed population models are useful reference points to which spatial models can be compared to understand the impact of the spatial structure.

1.4.3 Spatial population genetics

Real biological populations are often subject to constraints and are far from being well-mixed. Gravity, physical barriers, availability of food, and many other factors result in the formation of complex spatial structures in such populations. The effect of spatial structure on evolution might be so severe that the theory obtained for well-mixed populations cannot be directly applied to such populations. As a result, a new sub-branch arose called *spatial population genetic* [9]. It studies the interplay between the evolutionary forces and spatial structures of populations.

Originally, spatial models considered populations that are subdivided into "islands" or "demes" of well-mixed populations, with individuals migrating between the demes. This concept is based on the "isolation by distance" (IBD) principle proposed by Wright in 1940s [95, 96]. He formulated the idea that distance between populations can act as a barrier and restrict gene flows. As a consequence, populations that live in close proximity to each other are more genetically linked, whereas distant populations are genetically more different.

This idea was further extended by Kimura in the so-called stepping-stone model [47, 93]. Kimura formalized the IBD principle by placing populations on a lattice and restricting migration of individuals in each generation to nearby demes only, see Fig. 1.4, (b), (c). The model was solved by Kimura and Weiss in one-, two- and three-dimensions [50]. They demonstrated that the decrease of genetic relatedness with distance highly depends on dimensionality. The genetic relatedness is stronger in one dimension and becomes weaker as the dimensionality increases.

A stepping stone model with a single individual ("voter") in each deme reduces to the voter model [40]. The individuals can assume one of N states ("opinions"). The connections between individuals indicate that they can interact and adopt the state

of each other, see Fig. 1.4, (c). The voter model represents an idealized description for the evolution of opinion formation [14]. However, the model has been thoroughly investigated also in population genetics for studying competition between species [25, 72, 99].

Models with population structure have been further generalized in evolutionary graph theory, introduced by E. Lieberman et al. [2, 56]. In this framework, a population is represented as a graph with individuals occupying vertices, and edges representing spatial structure, see Fig. 1.4, (d). In contrast to the previous models, this approach allows the study of complex irregular hierarchical organization with heterogeneous connectivity.

In the last half of the 20th century, technologies that enable the study of populations at a single-cell level emerged and started a new era in spatial population genetics. With more insights into how populations of cells actually grow, targeted models of their dynamics were proposed [27]. One class of such models implements shifting dynamics, which can be observed in densely packed populations of cells discussed in Section 1.3.1. The second class is individually-based models of plankton microorganisms living in chaotic aquatic environments discussed in Section 1.3.2. We discuss these models in the following sections.

1.4.4 Populations with shifting dynamics

Models with shifting dynamics constitute a new paradigm for spatial population genetics. In the spatial models introduced in Section 1.4.3, the location of each individual remains the same from its birth to death. In models with shifting dynamics, each new reproduction changes the position of existing individuals, shifting them in some direction. If a population grows in a confined area, individuals can be pushed out of the area and eliminated from the population. This creates competition among individuals to remain within the growth area.

Due to the complex interplay between the individuals, models with shifting dynamics are challenging to solve analytically. There are exact results only for a few simple models. An important example is the linear array process, which replicates the architecture of epithelial compartments and cells proliferating in the mother machine [68]. In this model, a population is arranged into an array of cells that inhabit a space-limited area. Each cellular reproduction shifts existing individuals to the right in order to accommodate a newborn individual, see Fig. 1.4, (e). As a result, the rightmost cell is expelled from the population. A similar model with a population arranged in a cycle has been studied in [1].

Models with shifting dynamics have been mostly investigated with computational approaches. For example, Hashimoto et al. proposed a two-dimensional lattice model with shifting dynamics for studying reproduction rates of cells proliferating in rectangular microchannels with two open ends [37]. In this model, a newborn cell can take any neighboring position to its mother cell in any direction, see Fig. 1.4, (f). Each reproduction is followed by a shift of the other cells either to the left or to the right, causing the expulsion of a cell at the corresponding end of a microchannel. A similar model for three-dimensional populations has been employed for studying competition between two strains of bacteria [86].

Lattice models offer a simplified representation of population dynamics. They can be a good approximation of small and highly-ordered populations but might not be suitable for populations of disordered cells. For such cases, computational tools that model off-lattice populations of bacteria at a single-cell level are used. In an off-lattice model, individuals move continuously in space according to force laws governing the mechanical interactions between cells, see Fig. 1.4, (g). Such models were employed to simulate dense populations of cells growing on a surface [80] and in wide microchannels [60].

1.4.5 Populations in velocity fields

The transport of marine microorganisms under the influence of oceanic currents can be described with off-lattice models. In such models, each plankton microorganism is assigned to a tracer, i.e. a spatial trajectory moving in time with currents, see Fig. 1.4, (h), top. Such tracers are modeled with a system of ordinary differential equations, which explicitly incorporate the velocity field of the aquatic environment. In the absence of any currents, individuals freely diffuse, mimicking movements in small-scale turbulence, see Fig. 1.4, (h), bottom.

The standard way of modeling such populations is forward in time. In "forward" models, individuals displace, reproduce and die. Since such forward models are challenging to solve analytically due to their complexity, their analysis is often limited to numerical simulations. However, planktonic communities are composed of many rare species, which requires simulations of a huge number of individuals in order to reconstruct their biodiversity. This is not always feasible even with state-of-the-art computational approaches.

Coalescence theory offers a solution to this problem [67, 90]. The main idea of coalescent theory is to consider the evolution of a sample of a population backward in time [78]. The models of such samples are called "backward" models. With such an approach, the ancestry of a sample is reconstructed. Each individual is represented as a particle that can coalesce with any other one if they happen to be close to each other. Such coalescence events indicate that these two individuals have a common ancestor, and correspond to reproduction events in the forward model. Therefore, there is a mapping between the forward and backward models. In other words, we say that these two models are *dual*.

Backward models have advantages over forward models. First, they are easier to solve analytically. For example, many analytical results for the voter model have been obtained using its dual backward representation - coalescing random walkers [10, 19, 72]. Second, since backward models simulate a sample of individuals rather than the entire population, and do not consider individuals that do not contribute to the final biodiversity, they are more computationally efficient [16, 77].

A backward model was employed to reconstruct the biodiversity of planktonic communities living in aquatic environments with and without currents [88]. The model predicts a positive effect of oceanic currents on rare planktonic species.

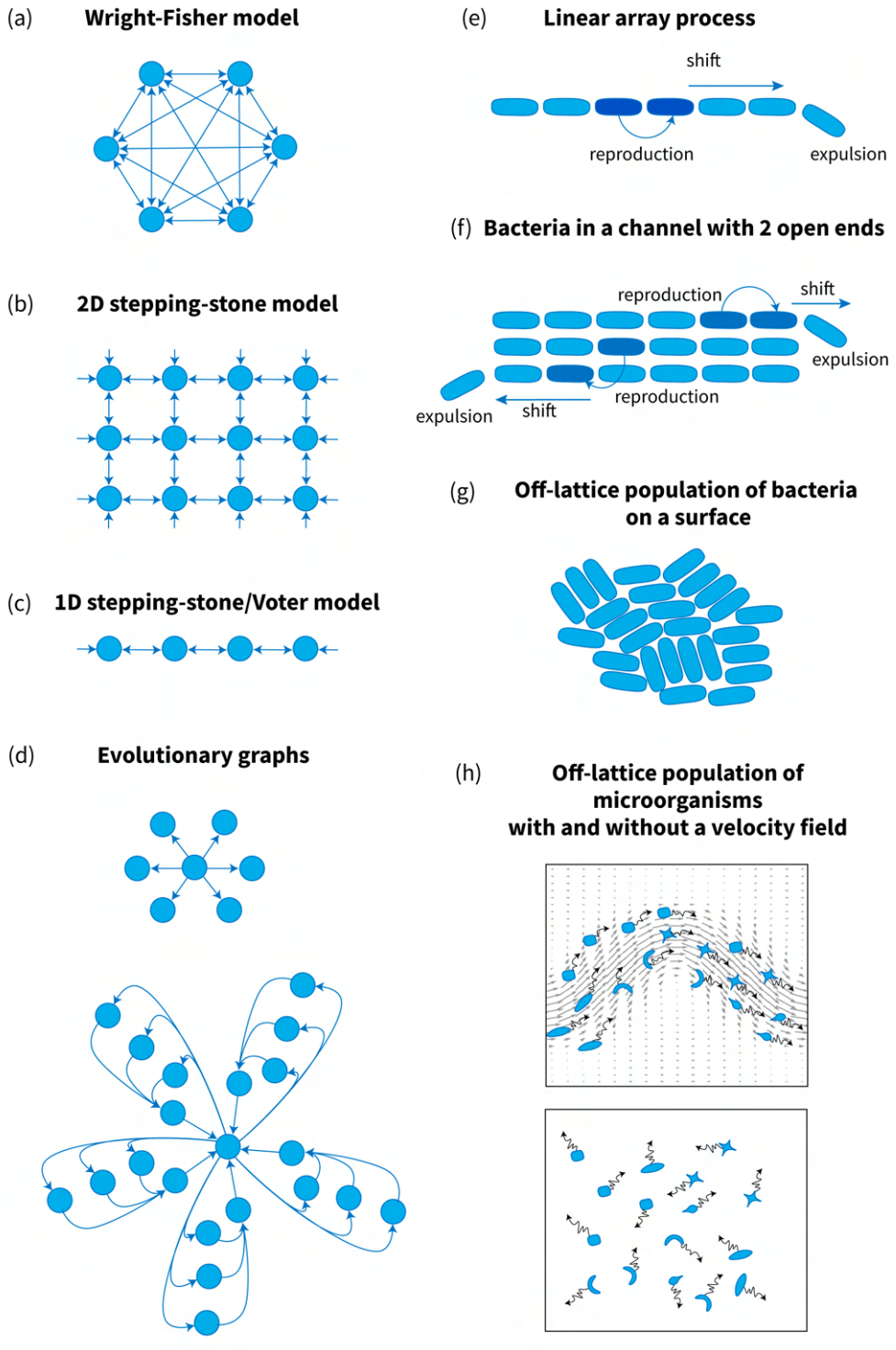


Figure 1.4: Models of population genetics. (a) Wright-Fisher model. (b) 2D stepping-stone model. (c) 1D stepping-stone model or Voter model. (d) Evolutionary graphs: one-root graph (top), super-star graph (bottom). (e) Linear array process. (f) Model of a population of cells in a channel with two open ends. (g) Off-lattice population of cells on a surface. (h) Off-lattice population of microorganisms in a velocity field (top) and without a velocity field, i.e. freely diffusing (bottom).

Chapter 2

Mathematical formalism

This chapter is devoted to mathematical modelling of population genetics. We introduce the main concepts and notations that we use in this thesis. We start with introducing simple models of well-mixed populations. Then, we discuss lattice models of spatially organized populations, such as stepping-stone models, evolutionary graphs, and models with shifting dynamics. Finally, we introduce the coalescent theory and off-lattice models of populations evolving in a force field.

2.1 Terminology

In this section, we introduce basic terminology and notations for this chapter.

We consider a population that consists of two types of individuals: A and B. These types can represent either different species or gene alleles, as discussed in Section 1.2. Individuals die and reproduce in such a way that the total population size remains always constant and equal to N. We focus on *haploid* populations, i.e. each individual carries one copy of a gene (type) and passes it to its offspring after a reproduction. We denote by X_t the total number of individuals of type A at time t. Consequently, the total number of individuals of type B at time t is $(N - X_t)$. The frequencies of individuals of type A and B at time t are then equal to $f_t = X_t/N$ and $(1 - f_t)$, correspondingly.

We focus on models in which the number of individuals of a given type in the next generation conditionally depends on the number of individuals of that type in the present generation, not on the previous generations. This property is called the *Markov property* and, therefore, X_t is called a *Markov chain*.

At some point, the population will reach a *fixation state*, i.e. a state in which there is only one remaining type in the population. There are two possible fixation state: $X_t = 0$ and $X_t = N$. They are also called *absorbing states* of the Markov chain X_t. The time at which the fixation state is reached is called the *fixation time* and is defined as:

$$\tau = \inf\{t \geq 0 : X_t = 0 \text{ or } X_t = N\}.$$

The probabilities of reaching either of the absorbing states $P(X_\tau = N)$ and $P(X_\tau = 0)$ are called the *fixation probabilities*.

If both types of individuals A and B reproduce at the same rate b, the population

is called *neutral*. If the two types reproduce at different rates, the population is called *non-neutral*. In such a case, we denote the reproduction rate of individuals A as $b(s+1)$, where $s > 0$ is a *selective advantage coefficient*. We call individuals of type A and B *mutants* and *wild-types*, correspondingly.

The fixation probability of mutants depends on the coefficient s. The higher this dependence, the stronger the effect of natural selection in the evolution of a population. If fixation probabilities are largely independent of s, this means that random drift dominates. We explore this selection-drift balance in the mathematical models considered in this chapter.

Also, we consider diffusion approximations of the process $\{X_t\}$ in the limit $N \gg 1$. In this case, the dynamics of the system is described with a partial differential equation that contains a standard Brownian motion W_t term. Such equations are not well defined because W_t is discontinuous at every point. For this reason, we need to specify a discretization rule for the stochastic integration. In this thesis, we use the Ito convention [32, 54], which defines a stochastic integral as

$$I = \int_0^T f(t')dW = \lim_{dt\to 0} \sum_{t_i=0}^{T/dt} f(t_i)[W(t_i + dt) - W(t_i)],$$

where $f(t)$ is an arbitrary continuously differentiable function and $t_i = idt$.

2.2 Models for well-mixed populations

We start with introducing the simplest models for well-mixed populations: Wright-Fisher and Moran models.

2.2.1 Wright-Fisher model

In the Wright-Fisher model, a population of N individuals evolves in discrete generations. At each generation, an individual samples a parent uniformly and independently from the previous generation and inherits its type, see Fig. 2.1, (a). The individuals sample their parent with replacement, i.e. the same parent can be chosen by multiple offsprings. As a result, the entire generation is renewed with no overlapping with the previous one. Therefore, the number X_n of individuals of type A in the nth generation is binomial:

$$P(X_{n+1}|X_n = i) \sim \text{Binom}\Big(n = N, p = \frac{i}{N}\Big).$$

In a neutral population, the fixation probability of individuals of type A is simply equal to their initial frequency:

$$P(X_\tau = N) = \frac{i_0}{N} = f_0. \tag{2.1}$$

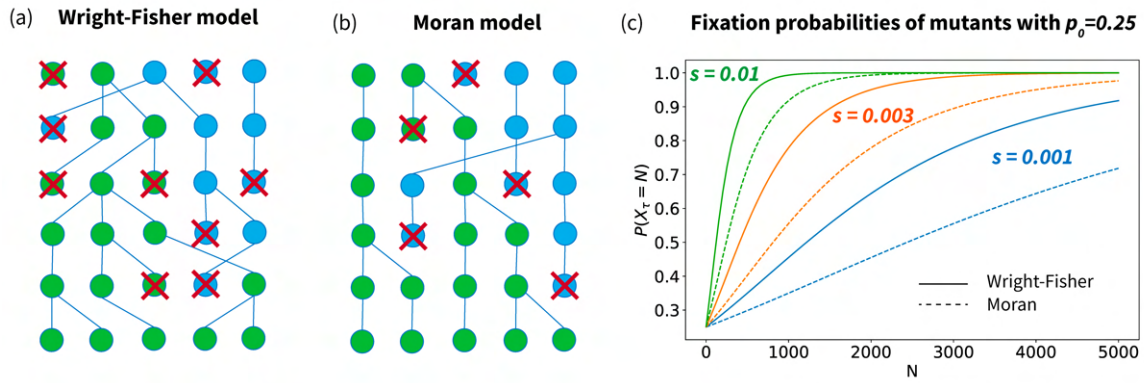


Figure 2.1: Models of well-mixed populations. (a) Wright-Fisher model. At each generation, an individual randomly chooses a parent from the previous generation and inherits its type. (b) Moran model. At each generation, one random individual dies, and one random individual reproduces. (c) Fixation probabilities of mutants with the selective advantage coefficients $s = 0.001, 0.003$, and 0.01 in the Wright-Fisher model (solid lines, Eq. (2.3)) and Moran model (dashed lines, Eq. (2.7)). The initial frequency of the mutants is $p_0 = 0.25$.

The average fixation time of the entire population linearly depends on the population size N and is expressed by

$$E(\tau) = -2N\Big(f_0 \log(f_0) + (1 - f_0)\log(1 - f_0)\Big), \tag{2.2}$$

where the time τ is measured in generations, [49].

In the non-neutral case, the fixation probability of mutants reproducing at rate $b(s+1)$ with initial frequency f_0 has the form

$$P(X_\tau = N) = \frac{1 - e^{-2Nsf_0}}{1 - e^{-2Ns}} \approx \frac{1 - (s+1)^{-2Nf_0}}{1 - (s+1)^{-2N}}, \tag{2.3}$$

[48]. In this case, the fixation probability depends on the compound parameter Ns. Therefore, even small selective differences can have a decisive effect on fixation probability, provided the population is large enough. This indicates a strong effect of natural selection in populations evolving according to the Wright-Fisher model.

For large population sizes (in the limit $N \gg 1$), the Wright-Fisher process converges to a limiting unbiased process which traces out a continuous path as time evolve. This limiting process is called *Wright-Fisher diffusion* and has the form

$$df_t = \sqrt{f_t(1 - f_t)}dW_t \quad \text{(Ito)}, \tag{2.4}$$

where W_t is a standard Brownian motion.

The Wright-Fisher process for non-neutral mutants reproducing at rate $b(s+1)$, $s > 0$ converges to a diffusion with a non-zero drift term for $N \gg 1$, which has the form

$$df_t = Nsf_t(1 - f_t)dt + \sqrt{f_t(1 - f_t)}dW_t \quad \text{(Ito)}. \tag{2.5}$$

2.2.2 Moran model

In contrast to the Wright-Fisher model, the Moran model represents populations with overlapping generations. At each time step, two individuals are sampled at random with replacement, i.e. the same individual can be chosen twice. One of the two individuals is chosen to reproduce and another one is chosen to die, see Fig. 2.1, (b). The reproducing individuals passes its type to the offspring. Thus, the Moran process is a simple birth and death process which makes it more tractable analytically.

Similarly to the Wright-Fisher model, the fixation probability of a neutral allele is simply equal to its initial frequency f_0 in a population, see Eq. (2.1). The fixation time of neutral individuals having the initial frequency f_0 is equal to

$$E(\tau) = -2\Big(f_0 \log(f_0) + (1 - f_0)\log(1 - f_0)\Big), \tag{2.6}$$

where the time τ is measured in units of N generations. Converting the time to the units of generations in Eq. (2.6), the Moran model is characterized by exactly the same fixation time as the Wright-Fisher model, see Eq. (2.2).

In the non-neutral case, the individuals reproducing at rate $b(1 + s), s > 0$ fixates with a probability given by

$$P(X_\tau = N) = \frac{1 - e^{-Nsf_0}}{1 - e^{-Ns}} \approx \frac{1 - (s + 1)^{-Nf_0}}{1 - (s + 1)^{-N}}. \tag{2.7}$$

Notably, in the fixation probability given by Eq. (2.7) there is a factor s rather than $2s$ comparing to the fixation probability formula for the Wright-Fisher model, see Eq. (2.3). This is due to different samplings in these two models and, accordingly, different rates of random genetic drift, see Fig. 2.1, (c).

2.3 Lattice models

In the simplest scenario, a spatially structured population can be represented by a lattice model. Here, we consider lattice models, in which each site is occupied either by a well-mixed population or one individual. We consider lattice models in one-, two-, and three-dimensional Euclidean space.

2.3.1 Stepping-Stone models

The stepping-stone model is represented as an infinite one-dimensional array of demes indexed by an integer i. Each deme is inhabited by a well-mixed population with M individuals of type A and B. We denote the frequency of individuals of type A within deme i by f_i. Individuals can migrate between adjacent colonies at the migration rate per generation m. Therefore, from each deme, $(mM)/2$ individuals migrate to the deme on the right and $(mM)/2$ individuals migrate to the deme on the left, see Fig. 2.2, (a). We assume that individuals of both types have the same migration rate. Therefore, the stepping-stone model is a system of interacting Wright-Fisher diffusions.

In the general case, in the limit $M \gg 1$, if individuals of type A reproduce at rate

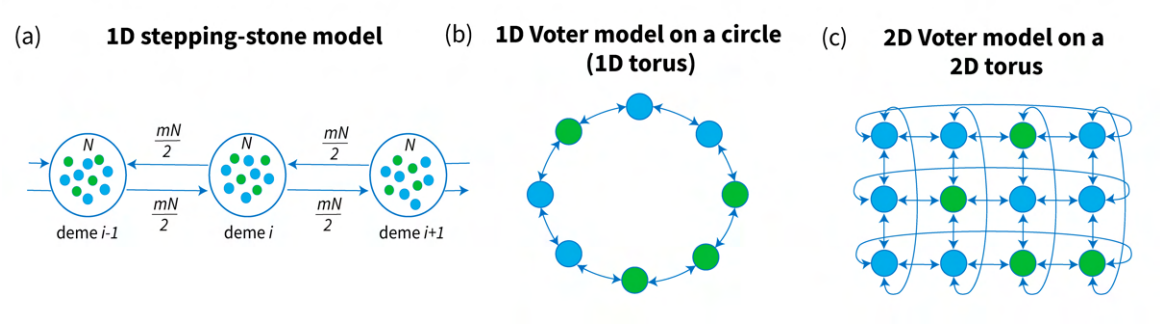


Figure 2.2: Stepping-stone models. (a) One-dimensional stepping stone model of an infinite number of interacting demes with M individuals within each deme. The individuals migrate at rate $mM/2$ between the neighboring demes, where m is the migration rate per generation. (b) One-dimensional voter model on a circle (one-dimensional torus). (c) Two-dimensional voter model on a two-dimensional torus.

$b(s+1)$, their frequency obeys the stochastic differential equation

$$df_i(t) = \frac{m}{2}(f_{i-1} + f_{i+1} - 2f_i)dt + Nsf_i(1-f_i)dt + \sqrt{f_i(1-f_i)}dW_t \quad \text{(Ito)}. \tag{2.8}$$

The voter model is equivalent to a stepping-stone model with $M = 1$. In such a case, one randomly chosen individual is replaced by an offspring of one of the neighboring individuals at each step. A population in a voter model can have different geometric configurations with different connectivity between individuals. We consider the simplest cases of finite voter models with N individuals on a torus in $Z^d, d = 1, 2$ and 3 with periodic boundary conditions, see Fig. 2.2, (b), (c). For such models, the average fixation time heavily depends on their dimensionality [19]:

$$E(\tau) \sim \begin{cases} N^2, & d = 1 \\ N\log(N), & d = 2 \\ N, & d = 3 \end{cases} \tag{2.9}$$

The fixation probabilities of mutants are equal to those for the Moran model. In the neutral case, they are equal to the initial frequency f_0 of the mutants. In the non-neutral case, the fixation probabilities are given by Eq. (2.7). This fact is explained in the next Section 2.3.2.

2.3.2 Evolutionary graphs

Evolutionary graphs represent populations as directed, weighted networks. Individuals occupy the vertices of a graph. The edges of the graph are specified by a matrix $W = \{w_{ij}\}$ satisfying $\sum_i w_{ij} = 1$ for all j, where the quantity w_{ij} defines the weight of the directed edge from a vertex v_i to a vertex v_j, see Fig 2.3, (a). A spatial organization of a population is determined via the structure of the graph: the connectivity of the vertices and the distribution of the weights of edges.

We call a graph isothermal if $\sum_j w_{ij} = 1$, i.e. all the vertices are equally likely

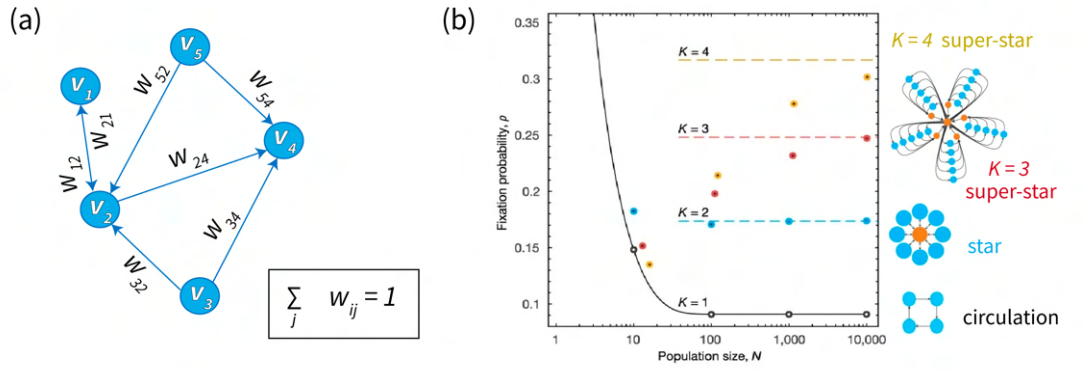


Figure 2.3: Evolutionary graphs. (a) An example of an evolutionary graph with vertices $\{v_i\}$ and weights $\{w_{ij}\}$. (b) Fixation probabilities of a mutant with $s = 0.1$ in on graphs depicted on the left. Points represent simulations. Dashed lines represent the theoretical predictions (dashed lines) obtained using Eq. (2.10). The figure is adopted from [56].

to be replaced. In this terminology, the Moran model corresponds to an isothermal, fully connected graph with evenly weighted edges. For such a structure, a fixation probability of a mutant with a reproduction rate $b(s + 1)$ is given by Eq. (2.3) for $p_0 = 1/N$. Lieberman et. al. demonstrated that the fixation probability of a mutant in any isothermal graph is equal to that in the Moran model given by Eq. (2.7) [56]. The voter models on torus in $Z^d, d = 1, 2, 3$ introduced in Section 2.3.1 are also isothermal graphs, which makes their fixation probabilities equal to the ones of the Moran model as well.

For non-isothermal graphs, this drift-selection balance tilts to either of the sides depending on the structure of the graph. For example, in graphs with one *root*, i.e. one vertex with no incoming edges, a mutation fixates with probability 1 irrespective of its selective advantage if and only if it arises in the root vertex, see Fig. 1.4, (d), top. Therefore, selection is completely suppressed in such a graph and random drift dominates. In contrast, the star and funnel graph structures amplify selection and suppress drift. An example of a star graph is shown in Fig. 2.3, (b), left. In such graphs, the fixation probability of a mutant is given by

$$P(X_\tau = N) = \frac{1 - \exp^{-2Ks}}{1 - \exp^{-2NKs}}, \tag{2.10}$$

where the parameter K represents the *complexity* of a graph structure. For instance, the super-star graph shown in Fig. 2.3, (b), left has $K = 3$. The fixation probabilities of any advantageous mutant in Eq. (2.10) converge to one as $N \to \infty$, see Fig. 2.3, (b).

2.3.3 Models with shifting dynamics

The simplest example of a model with shifting dynamics describes a one-dimensional linear array of N cells [68]. Cells reproduce at a constant rate b. After the reproduction of the ith cell in the array, a newborn cell is placed on the $(i + 1)$th position with

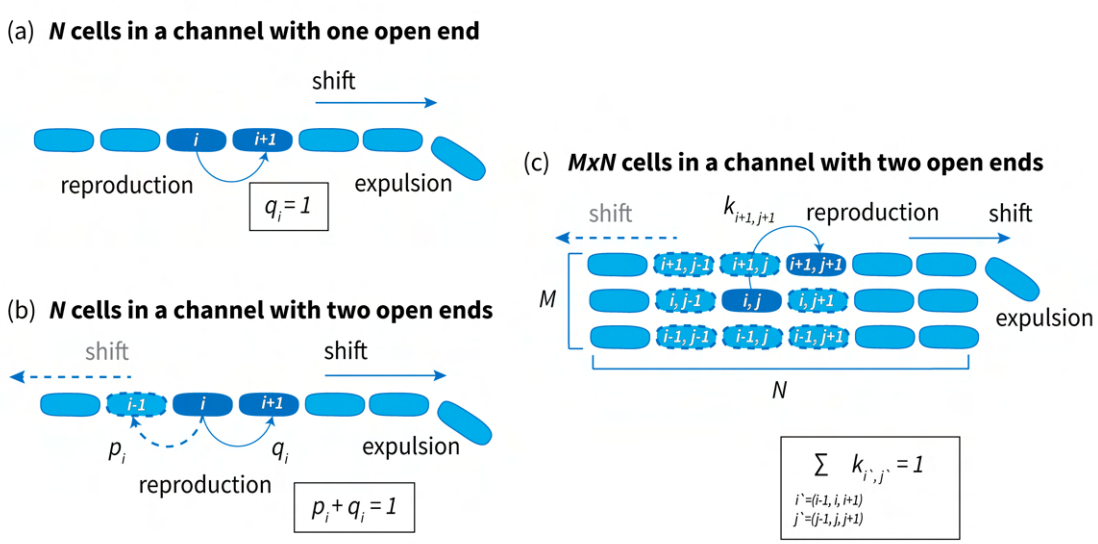


Figure 2.4: Models with shifting dynamics. (a) Array of N cells proliferating in a channel with one open end. After each reproduction, cells are shifted to the right with probability $q_i = 1$. (b) Array of N cells proliferating in a channel with two open ends. After each reproduction, cells are shifted either to the left (dashed arrow) or to the right (solid arrow) with probabilities p_i and q_i, correspondingly. (c) A two-dimensional $M \times N$ array of cells in a channel with two open ends. A newborn cell can take any of the neighboring positions (shown with dashed outlines) with probabilities $k_{i',j'}$, where $i' = (i-1, i, i+1)$ and $j' = (j-1, j, j+1)$.

probability $q_i = 1$, see Fig. 2.4, (a). Reproduction is followed by a shift of the existing cells to the right by one position. As a result, the Nth cell is eliminated from the population. A mutant that reproduces at rate $b(s+1)$, $s > 0$ can reach fixation if and only if it arises at the first position in the array, regardless of the value of s. Therefore, similarly to the one-root graph discussed in Section 2.3.2, the strength of natural selection is completely suppressed.

Now, we consider a one-dimensional linear array of N cells reproducing in a channel with two open ends. A newborn of the ith cell in the array can take either the $(i-1)$th or $(i+1)$th positions with probabilities $p(i)$ and $q(i)$ respectively, such that $p(i)+q(i) = 1$ for any i, see Fig. 2.4, (b). Depending on the chosen direction, the existing cells are shifted by one position either to the right or to the left. Consequently, either the 1st or the Nth cell is expelled from the population.

A more complex case is a two-dimensional $M \times N$ array of cells proliferating in a rectangular microchannel with two open ends [37]. After a reproduction of a cell at (i, j)th position, a newborn individual can take one of the neighboring positions either within the ith lane or in the neighboring $(i+1)$ and $(i-1)$ lanes, see Fig. 2.4, (c). If the mother cell is located at the boundary of a channel, i.e. $i = 1$ or M, a newborn cell has 6 positions to take. Otherwise, there are 8 possible positions to take. Each position can be taken with a probability $k_{i',j'}$, with $i' = (i-1, i, i+1)$ and $j' = (j-1, j, j+1)$. In the simplest case, the reproduction probabilities are uniform: $p(i) = q(i) = 1/2$ for the one-dimensional model, and $k_{i',j'} = 1/6$ or $1/8$ for the two-dimensional model.

We employ this model for studying *E. coli* proliferating in rectangular microchannels with two open ends. Based on experimental observations, we generalize the reproduction probability distributions and represent them as functions of the positions i of cells in a one-dimensional population, and (i, j) in two-dimensional populations, see Chapter 3.

2.4 Coalescent theory

A coalescent process is a model of a population that traces genetic ancestors of individuals in a population. Such series of ancestors are called *genetic ancestral lineages*. The coalescent process on a sample of n individuals consists of $(n-1)$ coalescent events. After each coalescence, the number of genetic ancestral lineages decreases by one until there is the last single remaining lineage. This last lineage is called *the most common recent common ancestor (MRCA)* of a population. The process can be represented as a tree structure, with the main parameters being the coalescence times $0 < T_i < \infty$, $2 \leq i \leq n$, i.e. the times during which there were exactly i lineages.

The duality between a retrospective coalescent model and a standard model means that there is a mapping between the trajectories predicted by the two models [40, 72].

Kingman coalescent

The Kingman coalescent is the simplest coalescent model for well-mixed populations [52]. It describes the ancestral genetic process for a sample of fixed size n in the limit of the population size as $N \to \infty$ in the Wright-Fisher and Moran models [51]. The coalescent process is an approximation to the behavior of a relatively small sample from a large population, i.e we assume $n \ll N$.

Kingman showed that, as $N \to \infty$, the coalescent times T_i are independent and exponentially distributed as

$$f_{T_i}(t_i) = \binom{i}{2} e^{-\binom{i}{2} t_i}, \quad t_i \geq 0, \quad i = 2, \ldots, n, \tag{2.11}$$

where time is measured in units of N generations. Eq. (2.11) implies that the coalescent times T_i has mean $2/(i(i-1))$. Therefore,

$$E(T_{MRCA}) = 2 \sum_{i=2}^{n} \frac{1}{i(i-1)}. \tag{2.12}$$

This time requires a scaling factor of N to convert time units to generations so that T_{MRCA} linearly depends on the population size N, similarly to the Wright-Fisher model, see Eq. (2.3). Also, Eq. (2.12) predicts that the last coalescence time in which the remaining two lineages coalesce into the MRCA is expected to be the longest among the others, see Fig. 2.5, (a).

A sample of size n from a population evolving according to the Wright-Fisher model with time measured in units of N generations can be approximated by the Kingman coalescent with $(n-1)$ coalescent events. Notably, we do not assign types to individuals

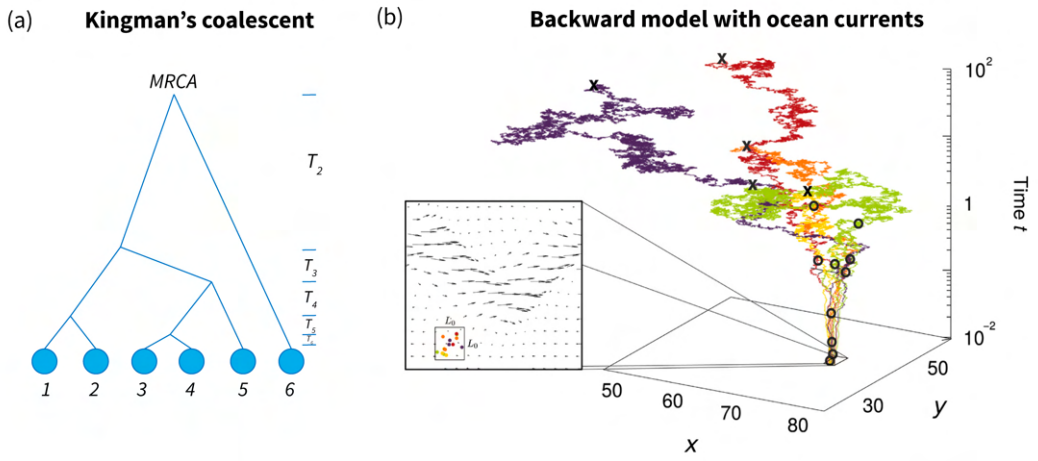


Figure 2.5: Coalescence models. (a) Kingman coalescence model. (b) Coalescence model of planktonic microorganisms in oceanic currents, adapted from [88].

in the initial sample in the Kingman coalescent. Instead, we assign the types to the ancestral lineages at some time point in the past and then trace genealogy to the present.

Coalescing random walks

The voter model discussed in the Section 2.3.1 is dual to a system of coalescing random walkers [11, 19]. Suppose there is one particle located at each lattice site $[1, N]$ at $t = 0$. At each discrete time step, a randomly chosen walker is moved to a one of the neighboring sites that are chosen randomly with equal probabilities. If the site is already occupied, the two walkers coalesce into one.

The particular example of a voter model with speciation (i.e. the model in which individuals can mutate a become a new species) is dual to a system of coalescing random walkers with annihilation [72].

2.5 Off-lattice populations in a force field

Plankton microorganisms moving in aquatic environments are modeled as passive particles driven by flows. Each individual is associated with a Lagrangian tracer with spatial coordinates (x, y). We briefly introduce forward and backward (coalescent) models for such populations in the following sections. Both of the models can predict biodiversity of a population in the presence and absence of oceanic currents.

The detailed discussion of the models, their predictions, and duality can be found in Chapter 5.

2.5.1 Forward approach

Initially, individuals are homogeneously distributed in a square $L \times L$. Each time step of the forward model consists of three events.

First, individuals reproduce at a constant rate λ. A newborn cell is placed in a $l \times l$ neighborhood of its mother cell. Second, the tracers move in space according to the following system of advection-diffusion Langevin equations:

$$\begin{aligned} \frac{d}{dt}x &= v_x(x,y,t) + \sqrt{2D}\xi_x(t), \\ \frac{d}{dt}y &= v_y(x,y,t) + \sqrt{2D}\xi_y(t), \end{aligned} \tag{2.13}$$

where the terms proportional to $\sqrt{2D}$ represent effective diffusion, the functions $\xi_x(t), \xi_y(t)$ are independent white noise sources such that $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_i(t') \rangle = \delta_{ij}\delta(t-t')$. The functions v_x and v_y represent an advecting fluid flow. Finally, individuals die at rate $\lambda\hat{n}$, where $\hat{n}$ is the number of other individuals in the $l \times l$ neighborhood. Such density-dependent death events represent competition between individuals for common resources.

The two-dimensional velocity field can be defined in terms of a stream function $\phi(x,y,t)$. The components of the field are related to the stream function by the following equations:

$$\begin{aligned} v_x(x,y,t) &= -\frac{d}{dy}\phi(x,y,t), \\ v_y(x,y,t) &= \frac{d}{dx}\phi(x,y,t), \end{aligned} \tag{2.14}$$

which automatically imply the incompressibility condition $\vec{\nabla} \cdot \vec{v} = 0$, where $\vec{v} = (v_x, v_y)$ and $\vec{\nabla} = (\partial/\partial x, \partial/\partial y)$. Details of the model are discussed in Chapter 5.

In order to model individual movement under diffusion only, one needs to impose $v_x(x,y,t) = v_x(x,y,t) = 0$ in Eqs. (2.13). Such a model is used as a reference to reveal the effect of oceanic currents on the predicted biodiversity.

2.5.2 Backward approach

With the backward approach, we model a sample of individuals rather than the entire population. Initially, individuals are homogeneously distributed in a sample square $L_s \times L_s$, where $L_s \ll L$. In contrast to the forward model, the backward model has only two events at each time step. First, individuals displace with respect to the backward advection-diffusion equations

$$\begin{aligned} \frac{d}{dt}x &= -v_x(x,y,t) + \sqrt{2D}\xi_x(t), \\ \frac{d}{dt}y &= -v_y(x,y,t) + \sqrt{2D}\xi_y(t), \end{aligned} \tag{2.15}$$

where all the terms have the same definitions as in Eqs. (2.13). Second, tracers are selected one by one and removed from a population with a probability μdt, which represents immigration events. Then, a pair of tracers coalesce at a rate λdt if they are in the same $l \times l$ neighborhood at time t.

The population is evolved until all individuals are assigned to species via either coalescence or immigration event, see Fig. 2.5, (b).

Chapter 3

Population genetics in microchannels

The main results of this chapter have been published as:

- **Koldaeva, A.**, Tsai, H.F., Shen, A.Q. and Pigolotti, S. "Population genetics in microchannels", Proceedings of the National Academy of Sciences, 2022,

see Appendix A. The experimental protocol has been published as:

- Tsai, H.F., Carlson, D.W., **Koldaeva, A.**, Pigolotti, S. and Shen, A.Q. "Optimization and Fabrication of Multi-Level Microchannels for Long-Term Imaging of Bacterial Growth and Expansion", Micromachines, 2022,

see Appendix B.

Outline of the project and results

In this project, we study populations of rod-shaped bacteria *E. coli* growing in rectangular microchannels with two open ends. Our experiments with two strains of *E. coli* tagged with green and red fluorescent protein reveal a striking effect: the strains segregate into lanes along the channel, see Fig. 3.1, (a). This observation demonstrates a strong influence of the boundaries of a channel in the underlying dynamics in a population and, therefore, its evolution. We aim at understanding the dynamics of such populations by combining theory, numerical simulations, and experimental observations.

Our analysis is based on a lattice model of a population with shifting dynamics, see Fig. 3.1, (b). The population consists of M lanes with N cells within each lane. We discussed the simplest version of this model with uniform reproduction probability distributions in Section 2.3.3. However, our experimental observations demonstrate that such distributions are not uniform and have two main sources of bias. First, within each lane, cells tend to reproduce towards the closest open ends due to the smaller mass of cells to be shifted. Second, due to the rod shape, cells tend to reproduce within their original lane rather than in the neighboring lane. We incorporate these two effects in the lattice model and evaluate all its parameters from experimental data with *E. coli* proliferating in microchannels.

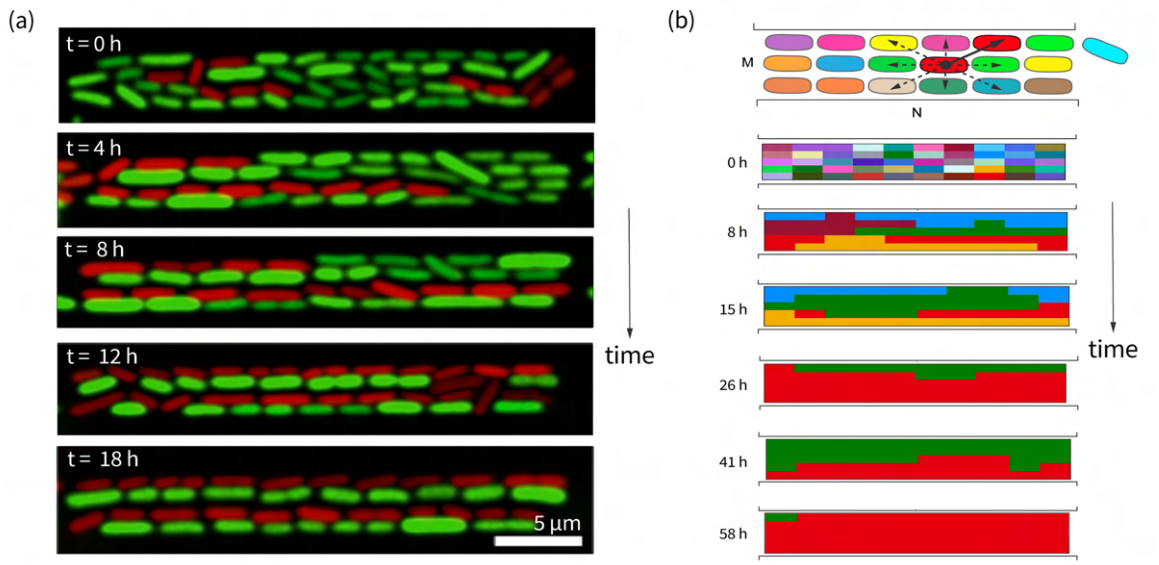


Figure 3.1: *E. coli* in microchannels with two open ends. (a) An experiment with two competing strains of *E. coli*. The strains segregate into four stripes in a channel of width 3 μm harboring four lanes of cells. (b) $M \times N$ lattice model of a population of *E. coli* in a microchannel with two open ends and its numerical simulation. Different colors represent different neutral mutants.

Our model predicts two regimes of biodiversity loss. The first regime is characterized by the fast fixation of one of the mutant strains within each lane. The second regime is characterized by slow competition between the lanes. Moreover, our model predicts that the mutants located in the middle of a channel and next to the boundaries have higher fixation probabilities. We find analytical forms of the biodiversity loss in time and fixation probabilities and run extensive numerical simulations of the model.

We run experiments with *E. coli* proliferating in microchannels that fit one, two, three, and four lanes of cells. We process the experimental recording with a custom-developed single-cell tracking algorithm. The theoretical predictions are in quantitative agreement with experimental observations.

Chapter 4

Selective advantage in microchannels

In this chapter, we study the competition of two non-neutral strains of cells in microchannels with two open ends. In particular, we aim at understanding the effect of the spatial structure on the strength of selection in such populations.

As we showed in Chapter 3, the fixation probability of a single neutral mutation heavily depends on its initial position in a microchannel. For this reason, we consider three different choices of the initial distributions: one mutant at a random position, a cluster of mutants concentrated at an open end, and randomly distributed mutants in a channel, see Fig. 4.1. As previously, the wild-type cells reproduce at rate b and mutants reproduce at rate $b(s+1), s > 0$. We focus on the values of the selection coefficients in the range from $s = 0.01$ to $s = 0.05$ comparable with typical beneficial mutations in *E. coli* observed in the lab [43].

4.1 One mutant

In the neutral case, the fixation probabilities of a single neutral mutant can be approximated by Gaussian distributions with mean $\mu = (N-1)/2$ and variance $\sigma^2 = (N-1)/4$, see Chapter 3 and Appendix A. Therefore, the fraction of mutants that accounts for more than 90% of the fixation probability mass is concentrated in the middle of a

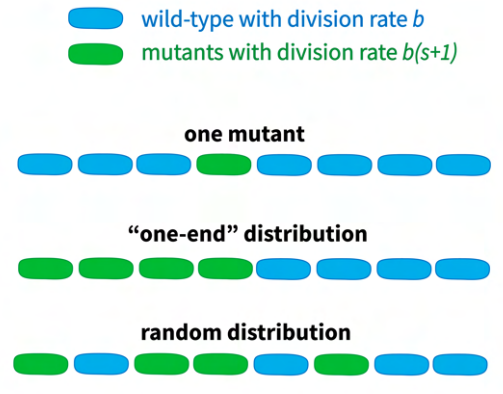


Figure 4.1: Initial distributions of two non-neutral strains.

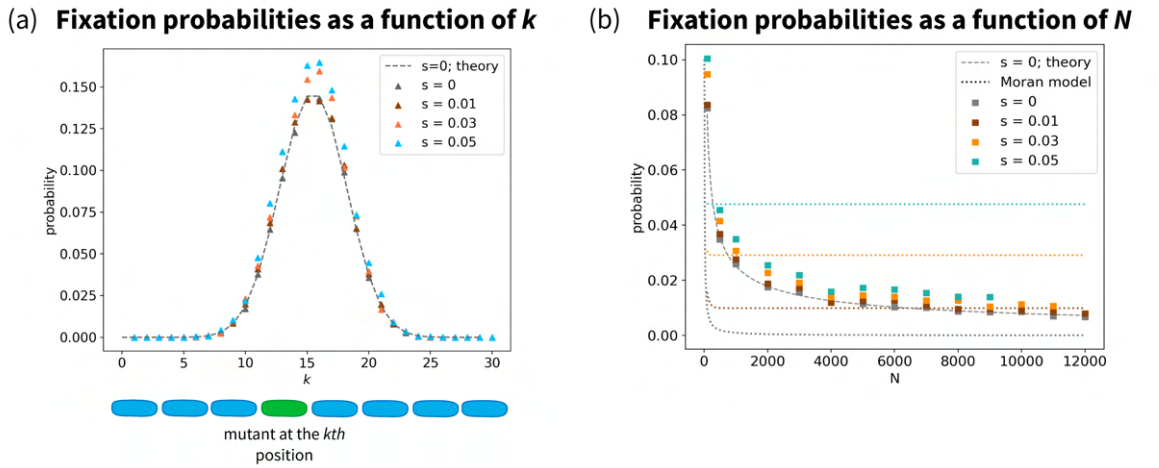


Figure 4.2: One mutant in a channel. (a) Fixation probability as a function on the initial position k of a mutant and different value of the selective advantage coefficient s in a population of $N = 30$ cells. (b) Fixation probability of a mutant located in the middle of a population as a function of N. Dotted curves represent fixation probabilities given by Eq. (2.7) for one mutant in the Moran model for different values of coefficient s.

channel and decreases as $1/\sqrt{N}$ as $N \to \infty$. Due to this, only a small portion of mutants have a chance to fixate, whereas the majority of them have almost zero fixation probability. For this reason, selection has an effect only on the mutants in the narrow region in the middle, see Fig. 4.2, (a). It has the highest effect on the mutant with the highest fixation probability, i.e. the mutant located in the middle of a channel. We fix $k = N/2$ and study this case in more detail in the rest of the section.

The fixation probability of a neutral mutant at the position $k = N/2$ is equal to $2/\sqrt{2\pi(N-1)}$. This value is larger than the fixation probability of a neutral mutant in a well-mixed population, which is equal to $1/N$. However, this situation drastically changes in the non-neutral case. In a well-mixed population, a mutant with reproduction rate $b(s+1)$ fixates at the probability

$$P = \frac{1-(s+1)^{-1}}{1-(s+1)^{-N}} \to 1-(s+1)^{-1} \text{ as } N \to \infty,$$

see Eq. (2.7). Whereas, our numerical simulations of populations with one non-neutral mutant in the middle of a channel demonstrate a much weaker effect of $s > 0$ on the fixation probability, see Fig. 4.2, (b). For example, in a well-mixed population of $N = 2000$ cells, a selective advantage coefficient $s = 0.05$ increases the fixation probability by almost 95 times. For the same mutant located in the middle of a population growing in a channel, the fraction probability increases by about 1.5 times.

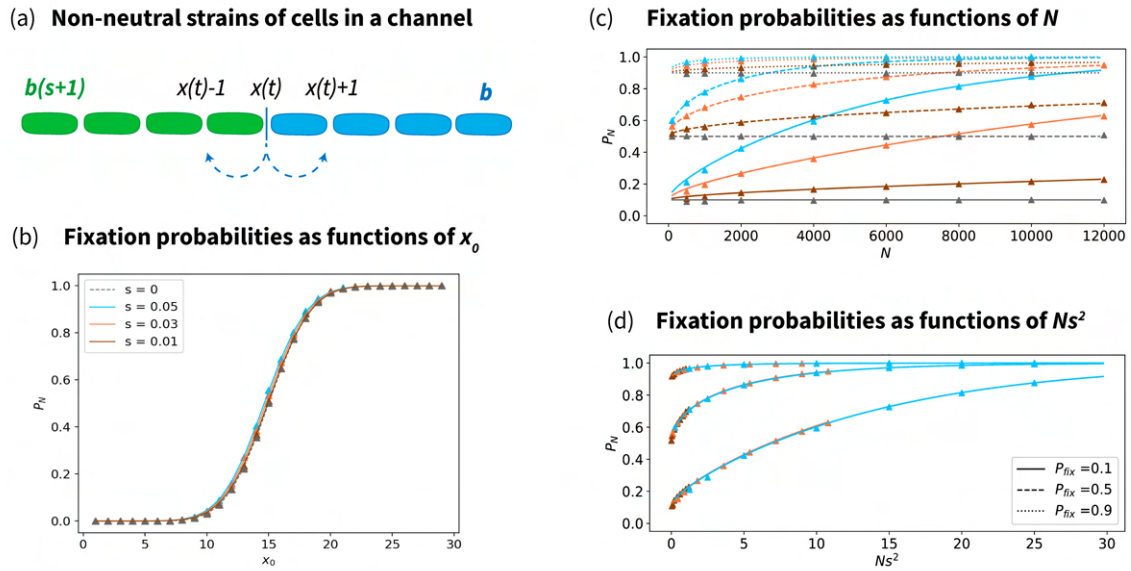


Figure 4.3: Population with the "one-end" initial condition. (a) Schematic representation of a population and the interface between the strains. (b) The fixation probabilities Eq. (4.4) as a function of x_0 for$N = 30$ and different values of the selection coefficient s. Solid lines represent the theoretical solution, the triangle-shape points represent numerical simulations. (c) Fixation probabilities Eq. (4.4) as a function of the number of cells N for different numbers of mutants that result in the fixation probabilities $P_{fix} = 0.1, 0.5$ and 0.9 in the neutral case and different values of the selection coefficient s. Solid lines represent the theoretical solution; triangle-shape points represent numerical simulations. (d) Fixation probabilities from (c) plotted as a function of Ns^2.

4.2 One-end initial condition

Eventually, a mutant population will reach one of the open ends and compete for fixation with the wild-type cells. Therefore, we consider a scenario in which the two populations occupy a portion of the channel adjacent to each open end. Without loss of generality, we assume that the population on the right is the wild-type population, with reproduction rate b. The population on the left is constituted by mutants with reproduction rate $b(1 + s)$.

Since the two strains can not mix, the state of the population can be described by the position of the boundary between the strains, called the *interface*. We denote the position of the interface at time t by $x(t)$, $0 \leq x(t) \leq N$, see Fig. 4.3, (a). The interface $x(t)$ moves to the left or right as cells reproduce. Therefore,

$$x \to x + 1 \quad \text{with rate} \quad \frac{xb(s+1)}{2},$$
$$x \to x - 1 \quad \text{with rate} \quad \frac{(N-x)b}{2}.$$

The process $x(t)$ has two absorbing states: 0 and N. If the process $x(t)$ reaches the

absorbing state N, it means that the mutant fixates. The absorbing state 0 corresponds to the fixation of the wild-type.

4.2.1 Fixation probabilities

In order to find the fixation probabilities, we use the Langevin equation associated with the process $x(t)$

$$\frac{dx}{dt} = v(x,t) + \sigma(x,t)\xi(t), \tag{4.1}$$

where

$$v(x,t) = \frac{(x(2+s) - N)b}{2}, \quad \sigma^2(x,t) = \frac{(N+xs)b}{2},$$

and $\xi(t)$ is a Gaussian white noise term. The Langevin equation (4.1) is interpreted in the Ito sense.

We calculate the fixation probabilities by finding a transformation $y = F(x)$ such as $y(t)$ is a martingale [76]. By substituting y to (4.1), we obtain the expression

$$\frac{dy}{dt} = \frac{dy}{dx}[v(x,t) + \sigma(x,t)\xi(t)] + \frac{d^2y}{dx^2}D(x). \tag{4.2}$$

Since y is a martingale, we need to assume that the drift term in Eq. (4.2) is zero. Therefore,

$$\frac{dy}{dx}v(x) + \frac{d^2y}{dx^2}D(x) = 0. \tag{4.3}$$

We solve Eq. (4.3) and use the Doob's optional stopping theorem to find fixation probabilities of the process $x(t)$. We find that for large N the probability that the mutants take over the channel is expressed by

$$\pi_N(f_0) = \frac{\gamma\left(z, w(f_0 s + 1)\right) - \gamma\left(z, w\right)}{\gamma\left(z, w(s+1)\right) - \gamma\left(z, w\right)}, \tag{4.4}$$

where $f_0 = x_0/N$ is the initial frequency of mutants, $\gamma(x,y) = \int_0^y t^{x-1}e^{-t}dt$ is the lower incomplete gamma function, $w = 2N(2+s)/s^2$, and $z = 1+4N(1+s)/s^2$, see Appendix C.

Equation (4.4) predicts that the fixation probability as a function of the fraction f_0 of the channel occupied by mutants is a steep function, see Fig. 4.3, (b). If this fraction is much smaller than 0.5, mutants are very likely to be expelled, whereas if it is much larger than 0.5 they are very likely to take over. This conclusion is rather insensitive to the value of s and is related to the fact that cells at the center of the channel have a high positional advantage.

If s is assumed to be small such that $(1+s) \approx 1$, then Eq. (4.4) predicts that the fixation probability depends on the compound parameter Ns^2. In order to visualize it, we need to understand the connection between the initial frequency f_0 of mutants and their fixation probability P_{fix} in the neutral model, i.e. when $s = 0$. In contrast to the neutral Moran, Voter and stepping-stone models, P_{fix} is not equal to f_0 in this case, see Sections 2.2.2, 2.3.1. According to our results for the neutral model of N cells proliferating in a microchannel, the fixation probabilities of one neutral mutation have a Gaussian distribution with mean $\mu = (N-1)/2$ and variance $\sigma^2 = (N-1)/4$,

see Chapter 3. Due to the symmetry of this Gaussian distribution function, we have that $f_0 = P_{\text{fix}} = 1/2$. For $f_0 \neq 1/2$, we use the quantile function of the Gaussian distribution. We find that the frequency of mutants needed to achieve the fixation probabilities P_{fix} is equal to

$$f_0 = \frac{\mu + \sigma\sqrt{2}\text{erf}^{-1}(2P_{\text{fix}} - 1)}{N}, \tag{4.5}$$

where $\text{erf}(z) = 2\pi^{-1}\int_0^z e^{-t^2}dt$ is the error function. We find the frequencies f_0 given by Eq. (4.5) for different N and a fixed value of P_{fix}. Then, we calculate the fixation probabilities given by Eq. (4.4) for the found frequencies f_0. For the fixation probabilities found in such a way, the rescaling Ns^2 holds, see Fig. 4.3, (c), (d).

The dependence on the parameter Ns^2 means that the fixation probability is largely insensitive to the value of the selective advantage. For comparison, in well-mixed populations, the classic expression for the fixation probability derived by Kimura depends on the parameter Ns, see Eq. (2.3) in Section 2.2.1. In this case, if $Ns \gg 1$, the selective advantage significantly biases the fixation process. In the opposite regime $Ns \ll 1$, fluctuations due to finiteness of the population dominate. In other words, mutations conferring selective advantages smaller than $1/N$ are not likely to be fixated and contribute to the evolution of the population. In microchannels instead, the same line of thought leads to conclude that selective advantages need to be larger than $1/\sqrt{N}$ in order to reach fixation. In practice, our result implies that evolution in a microchannel should be much slower than in a well-mixed system.

4.2.2 Fixation time

To find the average fixation time of the process $x(t)$, we use the Fokker-Plank equation

$$\partial_t P(x,t) = -\partial_x[A(x,t)P(x,t)] + \partial_x^2[B(x,t)P(x,t)], \tag{4.6}$$

where

$$A(x) = \frac{(x(2+s) - N)b}{2}, \quad B(x,t) = \frac{1}{2}\sigma^2(x,t) = \frac{(N+xs)b}{4}.$$

The average fixation time of the process $x(t)$ with absorbing states 0 and N can be found by integration the Fokker-Plank equation (4.6). The resulting formula for the fixation time is [31]

$$T(x_0) = \frac{\left(\int_0^{x_0}\psi(y)dy\right)\int_{x_0}^N\psi(y')dy'\int_0^{y'}\frac{dz}{B(z)\psi(z)} - \left(\int_{x_0}^N\psi(y)dy\right)\int_0^{x_0}\psi(y')dy'\int_0^{y'}\frac{dz}{B(z)\psi(z)}}{\int_0^N\psi(y)dy}, \tag{4.7}$$

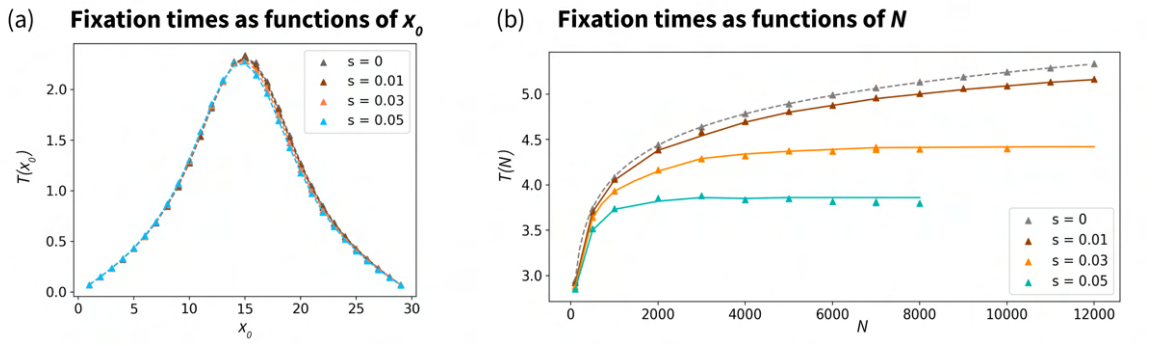


Figure 4.4: Fixation times for populations with the "one end" initial condition. (a) Fixation time given by Eq. (4.7) as a function of the initial number of mutants x_0 for different values of the selection coefficient s. (b) Fixation time given by Eq. (4.7) as a function of N for different values of the selection coefficient s. The value of x_0 is fixed and equal to $N/2$. In both figures, triangle-shaped points represent simulations and curves represent numerical simulations of the analytical solutions.

where $\psi(x) = e^{-\int_0^x \frac{A(x')}{B(x')}dx'}$. Substituting $A(x)$ and $B(x)$ from Eq. (4.6) and performing integration, we obtain

$$\psi(x) = \begin{cases} e^{-\frac{(x-N/2)^2}{N/2}+N/2}, & \text{if } s = 0 \\ \left(\frac{N}{N+sx}\right)^{\frac{-4N(s+1)}{s^2}} e^{-\frac{2(2+s)}{s}x}, & \text{if } s \neq 0. \end{cases} \tag{4.8}$$

We solve the integrals in Eq. (4.7) numerically, using the QUADPACK integration [71].

First, we explore how the average fixation time varies depending on the initial number of mutants x_0 in a population. Eq. (4.7) demonstrates that the average fixation time is maximal for populations with the initial number of mutants and wild-types close to $N/2$, see Fig. 4.4, (a). These observations agree with the results obtained in [79].

In order to understand how the fixation time given by Eq. (4.7) depends on N, we fix $x_0 = N/2$ further on in the section. For neutral mutants with $s = 0$ and the initial number $x_0 = N/2$, we find that Eq. (4.7) transforms to

$$T\left(\frac{N}{2}\right) = \frac{\sqrt{\pi}}{b}\int_0^{\sqrt{N/2}} e^{-t^2}\mathrm{erfi}(t)dt = \Gamma(3/2)\sum_{n=1}^{\infty}\frac{(-N/2)^n}{\Gamma(1/2+n)n}, \tag{4.9}$$

see Appendix D for more details. The fixation time given by Eq. (4.9) has a slow logarithmic growth as $N \to \infty$, see Fig. 4.4, (b). For the non-neutral mutants with $s > 0$, we find $T\left(\frac{N}{2}\right)$ given by Eq. (4.4) numerically. The solution agrees with numerical simulations, see Fig. 4.4, (b).

4.3 Random initial condition

In this section, we consider a population of randomly distributed wild-type cells and mutants. Each cell can be either a wild-type or a mutant with equal probabilities 0.5.

Our analysis is based on a theoretical guess. Instead of finite populations of size N, we consider populations of randomly distributed wild-type cells and mutants proliferating in an infinitely long channel. In such a case, the domains of wild-type and mutant cells grow exponentially, as no cells can be removed from the channel. We focus on the growth of two adjacent domains of mutants and wild-type cells. We study the competition between the two strains as the events when one of the strains grows up to N individuals before the other one. Our theoretical results for such populations are in good agreement with our numerical simulations of finite populations with N cells. We perform a convergence test to demonstrate that the theoretical solutions are exact and provide a possible explanation of the equivalence of these two systems.

4.3.1 Growing domains in infinite populations

The initial distribution of the domain size of both wild-type cells and mutants is

$$p_n^{WT}(t_0) = p_n^{mut}(t_0) = \frac{1}{2^n}. \tag{4.10}$$

The growth of each domain of either of the strains is a pure birth process since cells cannot be eliminated from a channel [73]. In this case, the evolution equation for the domain size of the wild type is

$$\frac{d}{dt}p_n = b\Big[(n-1)p_{n-1} - np_n\Big]. \tag{4.11}$$

The equation for the mutant is the same but with a factor $(1+s)$ on the right-hand side. We suppose that the distributions remain exponential at all times, i.e.

$$p_n(t) = (1 - e^{-k(t)})e^{-k(t)(n-1)}. \tag{4.12}$$

Substituting the solution given by Eq. (4.12) to Eq. (4.11) and using the initial condition $k(0) = \ln(2)$, we obtain that $k_{WT}(t) = -\ln\left(1 - \frac{e^{-bt}}{2}\right)$. Therefore, the distribution of the number of wild-type cells is

$$p_n(t) = \frac{e^{-bt}}{2}\Big(1 - \frac{e^{-bt}}{2}\Big)^{n-1}, \tag{4.13}$$

and the distribution of the number of mutants is

$$p_n(t) = \frac{e^{-b(s+1)t}}{2}\Big(1 - \frac{e^{-b(s+1)t}}{2}\Big)^{n-1}. \tag{4.14}$$

We are interested in the events when either the size of the mutant domain or the size of the wild-type domain becomes N first, see Fig. 4.5.

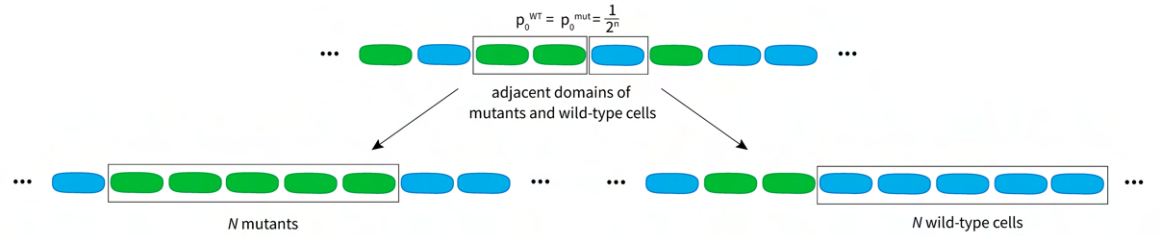


Figure 4.5: Wild-type cells and mutants in an infinite population. Two adjacent domains of mutants and wild-type cells start from n cells with probability $1/2^n$. We are interested in the probabilities and the average times at which either of the domains grows up to N cells first.

4.3.2 Fixation probabilities

We numerically simulate the finite populations of the size N with the random initial condition. Our simulations reveal a weak strength of selection in such populations. In fact, given the same initial frequency $f_0 = 0.5$, the effect of the value s on the fixation probabilities is much smaller than in populations with the "one-end" initial condition discussed in Section 4.2.1, see Fig. 4.6, (a).

For growing domains of two strains in an infinite channel introduced in Section 4.3.1, we calculate the probability that in two adjacent domains of different strains, the wild-type cells grow up to N cells before mutants. We denote by $X_{WT}(t)$ and $X_{mut}(t)$ the number of wild-types and mutants at time t. Also, we denote by T_{WT} and T_{mut} the time at which $X_{WT}(T_{WT}) = N$ and $X_{mut}(T_{mut}) = N$. Therefore, the probability that the mutant population grows to the size N before the wild-type one can be represented as

$$P(T_{mut} < T_{WT}) = \int_0^\infty \int_0^y f_{T_{mut},T_{WT}}(x,y)dxdy = \int_0^\infty \int_0^y f_{T_{mut}}(x) f_{T_{WT}}(y)dxdy, \tag{4.15}$$

where the density functions of T_{WT} and T_{mut} have the form

$$f_{T_{WT}} = N\frac{2^N}{2^N-1}\frac{e^{-bt}}{2}\left(1-\frac{e^{-bt}}{2}\right)^{N-1}, \tag{4.16}$$

$$f_{T_{mut}} = N(s+1)\frac{2^N}{2^N-1}\frac{e^{-b(s+1)t}}{2}\left(1-\frac{e^{-b(s+1)t}}{2}\right)^{(N-1)}. \tag{4.17}$$

Substituting the density functions $f_{T_{WT}}$ and $f_{T_{mut}}$ into Eq. (4.15) and performing the integration, we obtain

$$P(T_{mut} < T_{WT}) = -\frac{1}{2^N-1} + \frac{2^{2N}}{(2^N-1)^2}\frac{N}{2}\int_0^1 \left(1-\frac{u}{2}\right)^{N-1}\left(1-\frac{u^{s+1}}{2}\right)^N du, \tag{4.18}$$

see Appendix E for calculations. For the neutral case $s = 0$, Eq. (4.18) results in $P(T_{mut} < T_{WT}) = (T_{WT} < T_{mut}) = 1/2$, see Appendix E. For the non-neutral case $s > 0$, we solve Eq. (4.18) numerically with the QUADPACK integration [71]. This probability is in good agreement with our numerical simulations, see Fig.4.6, (a).

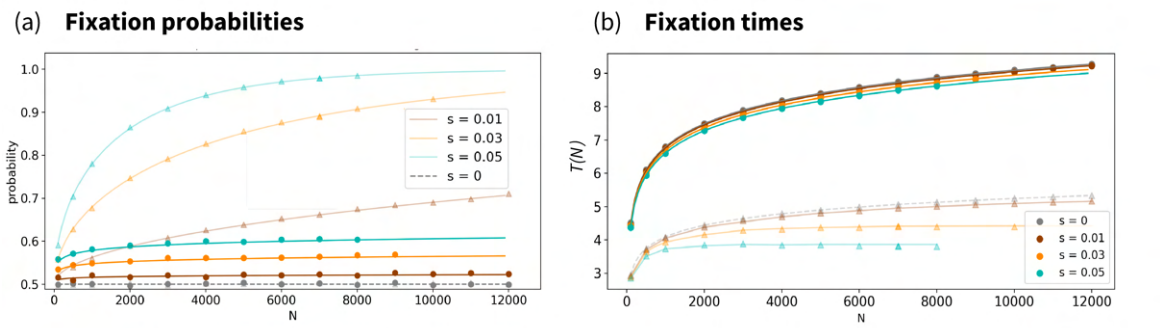


Figure 4.6: Fixation probabilities of populations with the random initial condition. (a) Fixation probability as a function of N for different values of the selection coefficient s. (b) Fixation time as a function of N for different values of the selection coefficient s. The light-colored curves with triangle-shaped points represent the "one-end" case in both figures.

As a convergence test, we demonstrate that the discrepancy between numerical simulations and the theoretical prediction given by Eq. (4.18) decreases as the number of iterations increases, see Appendix E for more details. Therefore, we conjecture that the solution given by Eq. (4.18) is exact.

4.3.3 Fixation time

Our numerical simulations of finite populations of size N demonstrate the low influence of selection on fixation times as well, see Fig.4.6, (b).

For the theoretical result, we find the average times $\langle T_{WT}\rangle$ and $\langle T_{mut}\rangle$ at which one of the adjacent domains of wild-types and mutants reach the size N. These average times have the forms

$$\langle T_{WT}\rangle = \int_0^\infty t f_{T_{WT}} dt = \frac{2^N}{b(2^N-1)}\Big[\log\Big(\frac{N}{2}\Big)+\gamma\Big], \tag{4.19}$$

$$\langle T_{mut}\rangle = \int_0^\infty t f_{T_{WT}} dt = \frac{2^N}{b(s+1)(2^N-1)}\Big[\log\Big(\frac{N}{2}\Big)+\gamma\Big], \tag{4.20}$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant, see Appendix F.

Using Eq. (4.18) and Eqs. (4.19), (4.20), we find the average time at which either of the wild-type and mutant domains reaches the size N. This average time can be found as

$$\langle T\rangle = P(T_{mut} < T_{WT})\langle T_{mut}\rangle + (1 - P(T_{mut} < T_{WT}))\langle T_{WT}\rangle, \tag{4.21}$$

see Appendix F for the full expression of Eq. (4.21).

For $s = 0$ and large N, Eq. (4.21) results in

$$\langle T\rangle = \frac{1}{b}\Big[\ln\Big(\frac{N-1}{2}\Big)+\gamma\Big], \tag{4.22}$$

see Appendix F. For $s \neq 0$, we solve Eq. (4.21) numerically. The results are in good

agreement with the simulations of finite populations, see Fig.4.6, (b).

Chapter 5

Coalescent dynamics of planktonic communities

This chapter has been published as:

- Martín, P.V., **Koldaeva, A.** and Pigolotti, S. "Coalescent dynamics of planktonic communities", Physical Review E, 2022,

see Appendix G.

Outline of the project and results

In this project, we explore the diversity of planktonic communities in aquatic environments with and without currents using spatial models that propagate forward and backward in time.

Both of the models simulate spatial trajectories of each individual, which we discuss in detail in Sections 2.5.2 and 2.5.2. As a result, the forward model reconstructs an ancestry tree of the entire population, whereas the backward (or coalescent) model predicts an ancestry of a sample from the population, see Fig. 5.1, (a) and (b). Therefore, as we discussed in Section 1.4.5, the backward model is preferable for simulating large planktonic communities, due it its simplicity over the forward model.

We explore the conditions upon which the backward model is dual to the forward model. In particular, we define two noise regimes called the *weak* and *strong* noise. In the weak noise regime, the stochastic fluctuations induced by birth and death processes in the forward model do not affect homogeneity of individuals' distribution in the forward model. Consequently, the total number of individuals does not deviate much from its original value. We claim that the duality between the forward and backward rigorously models holds in the weak-noise regime.

To demonstrate it, we run extensive numerical simulations of both models in the two noise regimes. Moreover, we simulate these models in the presence and in the absence of chaotic advection. We explore the predicted biodiversities using such biodiversity measures as $\alpha-$diversity, $\beta-$diversity, species-area relation, and species-abundance distribution (SAD). In the weak-noise regime, the forward and backward models yield nearly identical predictions for all these quantities.

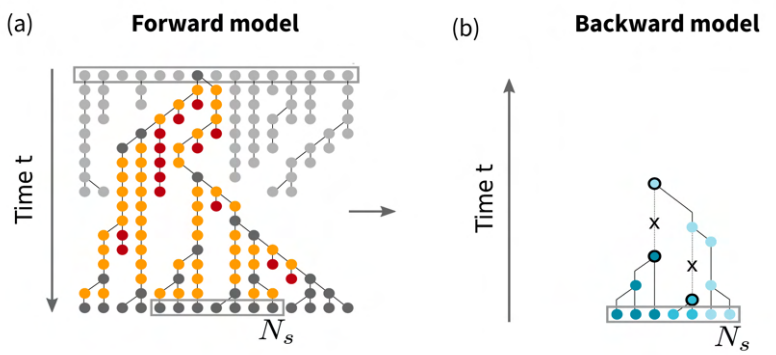


Figure 5.1: Two models of planktonic communities. (a) Forward model reconstructs ancestry tree of the entire population. (b) Backward model reconstructs ancestry tree of a sample from the population.

Finally, we employ the backward model to predict the impact of chaotic advection by oceanic currents on biodiversity of planktonic communities. We analyse the metagenomic data of plankton microorganisms sampled in oceans [85] and lakes [6] around the world. Both the experimental observations and numerical simulations of the model shows that SAD in the oceans are characterized by a steeper slope comparing to the ones in lakes.

Chapter 6

Conclusion

In this Thesis, we studied the population dynamics of two spatially organized populations: bacteria *E. coli* proliferating in rectangular microchannels with two open ends and planktonic communities evolving in a chaotic advection field. We analyzed both systems with individual-based mathematical models, and employed experimental data to validate theoretical predictions of the models. In this chapter, we provide detailed discussions of both systems and present further perspectives of the study.

6.1 *E. coli* in microchannels

The main focus of this thesis is the population dynamics of dense microbial communities growing in narrow rectangular microchannels. In such populations, a new reproduction is not necessarily followed by a death of a neighbor cell. In fact, reproduction creates a shift of an entire lane of cells toward an open end. Our experiments demonstrated that such dynamics, combined with the geometry of microchannels, result in the segregation of bacterial strains into stripes along microchannels. This is in contrast to populations growing on a surface, in which strains segregate into sector-like regions [36].

Our analysis is based on a lattice-based model [37]. We generalized the model and incorporated two effects observed in the experiments of proliferating *E. coli* in microchannels: the tendency of cells to reproduce within their original lane and towards the closest end. Our model predicts fast fixation within one lane of N cells with the time that scales as $\log(N)$. This time is much shorter than the average fixation times obtained for the Moran and linear Voter models, that scale as N and N^2, respectively. This demonstrates that shifting dynamics facilitates the fast accumulation of cells of the same strain that eventually takes over. Due to this, the dynamics of a population that consists of multiple lanes have two regimes: fast exponential diversity loss within lanes and slow competition between the lanes. Moreover, our model predicts a narrow Gaussian fixation probability distribution with the mean in the middle of a channel. In other words, the mutants arising in the middle of a channel have a high positional advantage, whereas the majority of other mutants have almost zero survival probability.

We discovered that the shifting dynamics in combination with the geometry of microchannels suppresses the strength of natural selection. Also, the strength of selection highly depends on the initial distribution of mutants in a channel. For a single

mutant with the highest positional advantage, the effect of selection advantage on fixation probability is lower by several magnitudes compared to the well-mixed case. The fixation probabilities of mutants with low positional advantage remain almost unaffected. We also explored two scenarios of mutants with the initial frequency greater than $1/N$: the mutants concentrated in one cluster next to an open end and random initial distribution. In the first case, the mutant's fixation probability depends on the compound parameter Ns^2, which demonstrates a weak strength of selection. Randomly distributed mutants result in an even weaker effect of selection on fixation probabilities. These findings demonstrate a remarkable characteristic of population dynamics in microchannels: the initial position of mutants is a dominating factor in their fate. Therefore, the balance between genetic drift and natural selection tilts toward the latter one. Moreover, this disbalance increases with the growth of the population size N.

Our findings are in contrast to predictions of other models of populations with shifting dynamics. For example, if such a population evolves on a cycle, the effect of selection is the same as in well-mixed populations [1]. Another example is the linear array process, in which cells propagate in a similar fashion as in the mother machine [68]. Since only the cell at the closed end has a non-zero fixation probability, selection is completely suppressed in such populations.

Our results are applicable to microbial ecology in soil. In order to find shelter from large predators and access to water, most bacteria inhabit micropores [98]. The diameter of such micropores is about three times the bacteria's body size [46]. This limits the number of individuals a bacteria interacts with within its population. It has been shown that this number is about 120 individuals on average [74]. Therefore, the size and organization of bacterial populations in soil are comparable with microchannels studied in this thesis.

Renewing epithelium tissue has a similar spatial organization with cells arranged into thin compartments [33, 44]. The cells in the compartments originate at a stem cell layer and push each other toward the top of the compartment in the same way as the cells propagate in the mother machine [3]. Understanding the dynamic of such populations can shed light on the emergence and propagation of deleterious mutations in epithelial tissue that leads to cancer [12, 29].

We aim to extend our study of competing non-neutral strains of bacteria *E. coli* to test the selective advantage in microchannels experimentally. To achieve this setup, we use two strains of *E. coli* with different antibiotic resistance levels [30]. One strain is sensitive to ampicillin and has been used in previous experimental work; the other has a modified plasmid, expressing a different fluorescent protein, with an additionally incorporated ampicillin-resistant gene (AmpR). Now, we achieve altered reproduction rates by adding ampicillin to the media with the two strains [55]. As a result, the resistant strain reproduces at a higher rate compared to the sensitive strain. Therefore, the value of the selection advantage coefficient s can be adjusted by changing the concentration of the antibiotic in the media. This experimental data will enable us to test the theoretical predictions of our model with a random initial distribution of the two strains in a channel.

Our theoretical findings on selection advantage in microchannels can be extended to populations consisting of multiple lanes of cells. However, in such a case, the sepa-

ration of the diversity loss into two regimes with the fast fixation within each lane and slow competition between lanes might not hold anymore. This might be due to the frequent invasion of mutants to their neighboring lanes. In order to understand this phenomenon, a detailed exploration of experimental data with two non-neutral strains along with extensive numerical simulations of the model is left for future work.

Also, our results can be extended to large populations growing in rectangular microchannels. The lattice model that we study in this thesis is a good approximation of the dynamics of highly ordered populations of bacteria growing in narrow microchannels. In larger populations, the destruction of the global cell ordering occurs due to the buckling instability [7]. In such a case, off-lattice models might give more realistic predictions [60, 80].

6.2 Plankton in velocity field

The second system we studied in this thesis consists of planktonic communities evolving in a chaotic aquatic environment. Our model is based on a coalescence principle, which enables reconstructing the evolution of a population backward in time. With the backward model, we reconstructed the biodiversity of a sample instead of the entire community. Moreover, with the backward approach, we modeled only the individuals that contributed to the biodiversity of the final population. This makes the backward model more computationally efficient compared to the standard forward models.

We demonstrated that there are two noise regimes. The weak noise regime is characterized by small stochastic fluctuations of the average number of individuals induced by birth and death processes in the forward model. Whereas, in the strong noise regime, fluctuations dominate altering the average size of a population. We showed that the backward and forward models are equivalent in the weak noise regime. Moreover, we demonstrated that the predictions of the models are in good agreement with the OTU composition of microbial communities sampled in aquatic environments with and without currents.

Our model can be extended to a non-neutral case, considering distinct planktonic species reproducing at different rates, potentially influenced by environmental changes [18]. The incorporation of selection into our model allows us to examine how differential reproductive rates and environmental factors impact the abundance and diversity of planktonic species, providing valuable insights into the effect of selection on biodiversity in changing environments [66, 84].

Bibliography

[1] B. Allen and M. A. Nowak. Evolutionary shift dynamics on a cycle. *Journal of theoretical biology*, 311:28–39, 2012.

[2] B. Allen and M. A. Nowak. Games on graphs. *EMS surveys in mathematical sciences*, 1(1):113–151, 2014.

[3] S. P. Bach, A. G. Renehan, and C. S. Potten. Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis*, 21(3):469–476, 2000.

[4] W. Bateson. *Mendel's principles of heredity*. At the University Press, 1913.

[5] R. A. Blythe and A. J. McKane. Stochastic models of evolution in genetics, ecology and linguistics. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07018, 2007.

[6] J. Boenigk, S. Wodniok, C. Bock, D. Beisser, C. Hempel, L. Grossmann, A. Lange, and M. Jensen. Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *Metabarcoding and Metagenomics*, 2:e21519, 2018.

[7] D. Boyer, W. Mather, O. Mondragón-Palomino, S. Orozco-Fuentes, T. Danino, J. Hasty, and L. S. Tsimring. Buckling instability in ordered bacterial colonies. *Physical biology*, 8(2):026008, 2011.

[8] A. Bracco, A. Provenzale, and I. Scheuring. Mesoscale vortices and the paradox of the plankton. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1454):1795–1800, 2000.

[9] G. S. Bradburd and P. L. Ralph. Spatial population genetics: it's about time. *Annual Review of Ecology, Evolution, and Systematics*, 50:427–449, 2019.

[10] M. Bramson and J. L. Lebowitz. Asymptotic behavior of densities for two-particle annihilating random walks. *Journal of statistical physics*, 62:297–372, 1991.

[11] M. Bramson, J. T. Cox, and R. Durrett. Spatial models for species area curves. *The Annals of Probability*, 24(4):1727–1751, 1996.

[12] J. Cairns. Mutation selection and the natural history of cancer. *Nature*, 255 (5505):197–200, 1975.

[13] R. L. Cartwright. Some remarks on essentialism. *The Journal of Philosophy*, 65 (20):615–626, 1968.

[14] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009.

[15] R. Cavicchioli, W. J. Ripple, K. N. Timmis, F. Azam, L. R. Bakken, M. Baylis, M. J. Behrenfeld, A. Boetius, P. W. Boyd, A. T. Classen, et al. Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*, 17(9):569–586, 2019.

[16] M. Cencini, S. Pigolotti, and M. A. Munoz. What ecological factors shape species-area curves in neutral models? *PloS one*, 7(6):e38232, 2012.

[17] H. Cho, H. Jönsson, K. Campbell, P. Melke, J. W. Williams, B. Jedynak, A. M. Stevens, A. Groisman, and A. Levchenko. Self-organization in high-density bacterial colonies: efficient crowd control. *PLoS biology*, 5(11):e302, 2007.

[18] S. Collins, B. Rost, and T. A. Rynearson. Evolutionary potential of marine phytoplankton under ocean acidification. *Evolutionary applications*, 7(1):140–155, 2014.

[19] J. T. Cox. Coalescing random walks and voter model consensus times on the torus in Z^d. *The Annals of Probability*, pages 1333–1366, 1989.

[20] K. Z. Coyte, H. Tabuteau, E. A. Gaffney, K. R. Foster, and W. M. Durham. Microbial competition in porous environments can select against rapid biofilm growth. *Proceedings of the National Academy of Sciences*, 114(2):E161–E170, 2017.

[21] C. Darwin. *The origin of species by means of natural selection*, volume 247. EA Weeks, 1859.

[22] C. De Vargas, S. Audic, N. Henry, J. Decelle, F. Mahé, R. Logares, E. Lara, C. Berney, N. Le Bescot, I. Probert, et al. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348(6237):1261605, 2015.

[23] S. Del Duca, A. Vassallo, A. Mengoni, and R. Fani. Microbial genetics and evolution, 2022.

[24] J. L. Doob. *Stochastic processes*. John Wiley & Sons, 1953.

[25] R. Durrett and S. Levin. Spatial models for species-area curves. *Journal of Theoretical Biology*, 179(2):119–127, 1996.

[26] S. F. Elena and R. E. Lenski. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, 4(6):457–469, 2003.

[27] J. Ferrer, C. Prats, and D. López. Individual-based modelling: an essential tool for microbiology. *Journal of biological physics*, 34:19–37, 2008.

[28] B. J. Finlay. Global dispersal of free-living microbial eukaryote species. *science*, 296(5570):1061–1063, 2002.

[29] S. A. Frank. *Dynamics of cancer: incidence, inheritance, and evolution*. Princeton university press, 2007.

[30] L. Galera-Laporta and J. Garcia-Ojalvo. Antithetic population response to antibiotics in a polybacterial community. *Science advances*, 6(10):eaaz5108, 2020.

[31] C. Gardiner. *Stochastic methods*, volume 4. Springer Berlin, 2009.

[32] C. W. Gardiner et al. *Handbook of stochastic methods*, volume 3. Springer Berlin, 1985.

[33] R. L. Gardner, D. I. Gottlieb, and D. R. Marshak. *Stem cell biology*. Cold Spring Harbor Laboratory Press, 2001.

[34] B. Gibson, D. J. Wilson, E. Feil, and A. Eyre-Walker. The distribution of bacterial doubling times in the wild. *Proceedings of the Royal Society B*, 285(1880): 20180789, 2018.

[35] J. A. Gilbert and J. D. Neufeld. Life in a world without microbes. *PLoS biology*, 12(12):e1002020, 2014.

[36] O. Hallatschek, P. Hersen, S. Ramanathan, and D. R. Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930, 2007.

[37] M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto. Noise-driven growth rate gain in clonal cellular populations. *Proceedings of the National Academy of Sciences*, 113(12):3251–3256, 2016.

[38] P. D. Hebert, A. Cywinska, S. L. Ball, and J. R. DeWaard. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512):313–321, 2003.

[39] M. E. Heinrichs, C. Mori, and L. Dlugosch. Complex interactions between aquatic organisms and their chemical environment elucidated from different perspectives. *Youmares*, pages 279–297, 2020.

[40] R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975.

[41] K. Holterhoff. The history and reception of Charles Darwin's hypothesis of pangenesis. *Journal of the History of Biology*, 47:661–695, 2014.

[42] G. E. Hutchinson. The paradox of the plankton. *The American Naturalist*, 95 (882):137–145, 1961.

[43] M. Imhof and C. Schlötterer. Fitness effects of advantageous mutations in evolving Escherichia coli populations. *Proceedings of the National Academy of Sciences*, 98(3):1113–1117, 2001.

[44] S. M. Janes, S. Lowell, and C. Hutter. Epidermal stem cells. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 197(4):479–491, 2002.

[45] G. Károlyi, Á. Péntek, I. Scheuring, T. Tél, and Z. Toroczkai. Chaotic flow: the physics of species coexistence. *Proceedings of the National Academy of Sciences*, 97(25):13661–13665, 2000.

[46] G. Kilbertus et al. Microhabitats in soil aggregates. Their relationship with bacterial biomass and the size of the procaryotes present. *Revue d'Ecologie et de Biologie du Sol*, 17(4):543–557, 1980.

[47] M. Kimura. Stepping-stone model of population. *Ann. Rept. Nat. Inst. Genetics*, 1953.

[48] M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713, 1962.

[49] M. Kimura and T. Ohta. The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61(3):763, 1969.

[50] M. Kimura and G. H. Weiss. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561, 1964.

[51] J. F. Kingman. On the genealogy of large populations. *Journal of applied probability*, 19(A):27–43, 1982.

[52] J. F. C. Kingman. The coalescent. *Stochastic processes and their applications*, 13(3):235–248, 1982.

[53] A. Koldaeva, H.-F. Tsai, A. Q. Shen, and S. Pigolotti. Population genetics in microchannels. *Proceedings of the National Academy of Sciences*, 119(12): e2120821119, 2022.

[54] K. S. Korolev, M. Avlund, O. Hallatschek, and D. R. Nelson. Genetic demixing and evolution in linear stepping stone models. *Reviews of modern physics*, 82(2): 1691, 2010.

[55] M. Li, Q. Liu, Y. Teng, L. Ou, Y. Xi, S. Chen, and G. Duan. The resistance mechanism of Escherichia coli induced by ampicillin in laboratory. *Infection and drug resistance*, pages 2853–2863, 2019.

[56] E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, 2005.

[57] K. Maheep et al. Bacteria involving in nitrogen fixation and their evolutionary correlation. *International Journal of Current Microbiology and Applied Sciences*, 3(3):824–830, 2014.

[58] J. Männik, R. Driessen, P. Galajda, J. E. Keymer, and C. Dekker. Bacterial growth and motility in sub-micron constrictions. *Proceedings of the National Academy of Sciences*, 106(35):14861–14866, 2009.

[59] P. V. Martín, A. Koldaeva, and S. Pigolotti. Coalescent dynamics of planktonic communities. *Physical Review E*, 106(4):044408, 2022.

[60] W. Mather, O. Mondragón-Palomino, T. Danino, J. Hasty, and L. S. Tsimring. Streaming instability in growing cell populations. *Physical review letters*, 104 (20):208101, 2010.

[61] M. J. McDonald. Microbial experimental evolution–a proving ground for evolutionary theory and a tool for discovery. *EMBO reports*, 20(8):e46992, 2019.

[62] K. L. McGuire and K. K. Treseder. Microbial communities and their relevance for ecosystem models: decomposition as a case study. *Soil Biology and Biochemistry*, 42(4):529–535, 2010.

[63] P. Moran and C. Smith. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52: 899–438, 1918.

[64] P. A. P. Moran. Random processes in genetics. In *Mathematical proceedings of the cambridge philosophical society*, volume 54, pages 60–71. Cambridge University Press, 1958.

[65] H. Mori, M. Kataoka, and X. Yang. Past, present, and future of genome modification in Escherichia coli. *Microorganisms*, 10(9):1835, 2022.

[66] J. Norberg, D. P. Swaney, J. Dushoff, J. Lin, R. Casagrandi, and S. A. Levin. Phenotypic diversity and ecosystem functioning in changing environments: a theoretical framework. *Proceedings of the National Academy of Sciences*, 98(20): 11376–11381, 2001.

[67] M. Nordborg. Coalescent theory. *Handbook of Statistical Genomics: Two Volume Set*, pages 145–30, 2019.

[68] M. A. Nowak, F. Michor, and Y. Iwasa. The linear process of somatic evolution. *Proceedings of the national academy of sciences*, 100(25):14966–14969, 2003.

[69] A. G. O'Donnell, I. M. Young, S. P. Rushton, M. D. Shirley, and J. W. Crawford. Visualization, modelling and prediction in soil microbiology. *Nature Reviews Microbiology*, 5(9):689–699, 2007.

[70] S. Park, P. M. Wolanin, E. A. Yuzbashyan, H. Lin, N. C. Darnton, J. B. Stock, P. Silberzan, and R. Austin. Influence of topology on bacterial social interaction. *Proceedings of the National Academy of Sciences*, 100(24):13910–13915, 2003.

[71] R. Piessens, E. de Doncker-Kapenga, C. W. Überhuber, and D. K. Kahaner. *Quadpack: a subroutine package for automatic integration*, volume 1. Springer Science & Business Media, 2012.

[72] S. Pigolotti, M. Cencini, D. Molina, and M. A. Muñoz. Stochastic spatial models in ecology: a statistical physics approach. *Journal of Statistical Physics*, 172: 44–73, 2018.

[73] M. Pinsky and S. Karlin. *An introduction to stochastic modeling*. Academic press, 2010.

[74] X. Raynaud and N. Nunan. Spatial ecology of bacteria at the microscale in soil. *PloS one*, 9(1):e87217, 2014.

[75] W. C. Reygaert. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS microbiology*, 4(3):482, 2018.

[76] É. Roldán, I. Neri, R. Chetrite, S. Gupta, S. Pigolotti, F. Jülicher, and K. Sekimoto. Martingales for physicists. *arXiv preprint arXiv:2210.09983*, 2022.

[77] J. Rosindell and S. J. Cornell. Species–area relationships from a spatially explicit neutral model in an infinite landscape. *Ecology letters*, 10(7):586–595, 2007.

[78] J. Rosindell, Y. Wong, and R. S. Etienne. A coalescence approach to spatial neutral ecology. *Ecological Informatics*, 3(3):259–271, 2008.

[79] J. Rothschild, T. Ma, J. N. Milstein, and A. Zilman. Spatial exclusion leads to tug-of-war ecological dynamics between competing species within microchannels. *bioRxiv*, pages 2023–01, 2023.

[80] T. J. Rudge, P. J. Steiner, A. Phillips, and J. Haseloff. Computational modeling of synthetic microbial biofilms. *ACS synthetic biology*, 1(8):345–352, 2012.

[81] R. Rusconi, M. Garren, and R. Stocker. Microfluidics expanding the frontiers of microbial ecology. *Annual review of biophysics*, 43:65–91, 2014.

[82] M. Scheffer, S. Rinaldi, J. Huisman, and F. J. Weissing. Why plankton communities have no equilibrium: solutions to the paradox. *Hydrobiologia*, 491:9–18, 2003.

[83] K.-H. Schleifer. Microbial diversity: facts, problems and prospects. *Systematic and applied microbiology*, 27(1):3, 2004.

[84] T. W. Schoener. The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *science*, 331(6016):426–429, 2011.

[85] E. Ser-Giacomi, L. Zinger, S. Malviya, C. De Vargas, E. Karsenti, C. Bowler, and S. De Monte. Ubiquitous abundance distribution of non-dominant plankton across the global ocean. *Nature ecology & evolution*, 2(8):1243–1249, 2018.

[86] T. Shimaya and K. A. Takeuchi. Lane formation and critical coarsening in a model of bacterial competition. *Physical Review E*, 99(4):042403, 2019.

[87] H.-F. Tsai, D. W. Carlson, A. Koldaeva, S. Pigolotti, and A. Q. Shen. Optimization and fabrication of multi-level microchannels for long-term imaging of bacterial growth and expansion. *Micromachines*, 13(4):576, 2022.

[88] P. Villa Martín, A. Buček, T. Bourguignon, and S. Pigolotti. Ocean currents promote rare species diversity in protists. *Science advances*, 6(29):eaaz9037, 2020.

[89] D. Volfson, S. Cookson, J. Hasty, and L. S. Tsimring. Biomechanical ordering of dense cell populations. *Proceedings of the National Academy of Sciences*, 105 (40):15346–15351, 2008.

[90] J. Wakeley. *Coalescent Theory, an Introduction*. Roberts and Company, 2005.

[91] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun. Robust growth of Escherichia coli. *Current biology*, 20(12):1099–1103, 2010.

[92] B. G. Weiner, A. Posfai, and N. S. Wingreen. Spatial ecology of territorial populations. *Proceedings of the National Academy of Sciences*, 116(36):17874–17879, 2019.

[93] G. H. Weiss and M. Kimura. A mathematical analysis of the stepping stone model of genetic correlation. *Journal of Applied Probability*, 2(1):129–149, 1965.

[94] S. Wright. *Statistical Genetics in Relation to Evolution*. Actualités scientifiques et industrielles. Hermann, 1939.

[95] S. Wright. Breeding structure of populations in relation to speciation. *The American Naturalist*, 74(752):232–248, 1940.

[96] S. Wright. Isolation by distance. *Genetics*, 28(2):114, 1943.

[97] Z. You, D. J. Pearce, A. Sengupta, and L. Giomi. Geometry and mechanics of microdomains in growing bacterial colonies. *Physical Review X*, 8(3):031065, 2018.

[98] I. Young and K. Ritz. Tillage, habitat space and function of soil microbes. *Soil and Tillage Research*, 53(3-4):201–213, 2000.

[99] T. Zillio, I. Volkov, J. R. Banavar, S. P. Hubbell, and A. Maritan. Spatial scaling in model plant communities. *Physical review letters*, 95(9):098101, 2005.

[100] K. Ziółkowska, R. Kwapiszewski, and Z. Brzozka. Microfluidic devices as tools for mimicking the in vivo environment. *New Journal of Chemistry*, 35(5):979–990, 2011.

Appendix A

Population genetics in microchannels

Koldaeva, A., Tsai, H.F., Shen, A.Q. and Pigolotti, S. "Population genetics in microchannels", Proceedings of the National Academy of Sciences, 2022, [53].

Contributions:

A.K. and S.P. designed the research. A.K. performed numerical simulations, analytical calculations, and analyzed the data. H.-F.T. and A.Q.S. designed the experiments. H.-F.T. performed experiments. A.K. and S.P. wrote the paper, with input from H.-F.T. and A.Q.S

Appendix B

Optimization and Fabrication of Multi-Level Microchannels for Long-Term Imaging of Bacterial Growth and Expansion

Tsai, H.F., Carlson, D.W., **Koldaeva, A.**, Pigolotti, S. and Shen, A.Q. "Optimization and Fabrication of Multi-Level Microchannels for Long-Term Imaging of Bacterial Growth and Expansion", Micromachines, 2022, [87] .

Contributions:

Conceptualization, H.-F.T., A.K., S.P. and A.Q.S.; Methodology, H.-F.T., A.K.; Numerical simulation, H.-F.T. and D.W.C.; Writing—original draft preparation, H.-F.T. and D.W.C.; Writing—review and editing, A.K., A.Q.S. and S.P.

Appendix C

Fixation probabilities for the "one-end" initial condition

In this appendix, we solve Eq. (4.3) and derive the fixation probabilities given by Eq. (4.4) of mutants with the initial frequency f_0 and selection coefficient s concentrated at one end in a microchannel.

First, we need to solve Eq. (4.3), which has the following form

$$\frac{dy}{dx}v(x) + \frac{d^2y}{dx^2}D(x) = 0.$$

We introduce a new notation $g(x) = \frac{dy}{dx}$. Thus, the equation takes the form

$$\frac{dg}{dx}D(x) + gv(x) = 0,$$

or equivalently

$$\frac{g'(x)}{g(x)} = -\frac{v(x)}{D(x)}.$$

We integrate the latest equation and get the following expressions

$$\int \frac{g'(x)}{g(x)}dx = -\int \frac{v(x)}{D(x)}dx,$$

$$\ln(g(x)) = -\int \frac{((2+s)x - N)b/2}{(N+xs)b/4}dx = -2\Big[(2+s)\int \frac{x}{N+xs}dx - N\int \frac{dx}{N+xs}\Big] =$$

$$= -2\Big[(2+s)\int \frac{sx+N-N}{s(N+xs)}dx - \frac{N}{s}\ln(xs+N)\Big] =$$

$$= -2\Big[\frac{2+s}{s}\int \Big(1 - \frac{N}{s(N+xs)}dx\Big) - \frac{N}{s}\ln(xs+N)\Big] =$$

$$= -2\Big[\frac{2+s}{s}\Big(x - \frac{N}{s}\ln(xs+N)\Big) - \frac{N}{s}\ln(xs+N)\Big].$$

Therefore,

$$\ln g(x) = -\frac{2(2+s)x}{s} + \frac{4N(s+1)\ln(xs+N)}{s^2}.$$

Taking exponential function of both sides on the latest equation, we find the function $g(x)$ in the form

$$g(x) = (xs+N)^{\frac{4N(s+1)}{s^2}} e^{-\frac{2(2+s)x}{s}}.$$

From the definition $g(x) = \frac{dy}{dx}$, we find the function $y(x)$ by integration:

$$y(x) = \int_0^x (zs+N)^{\frac{4N(s+1)}{s^2}} e^{-\frac{2(2+s)z}{s}} dz = \int_0^x \frac{\left(\frac{2(2+s)}{s^2}(zs+N)\right)^{\frac{4N(s+1)}{s^2}}}{\left(\frac{2(2+s)}{s^2}\right)^{\frac{4N(s+1)}{s^2}}} e^{-\frac{2(2+s)z}{s} \pm N\frac{2(2+s)}{s^2}} dz =$$

$$= const \int_0^x \left(\frac{2(2+s)}{s^2}(zs+N)\right)^{\frac{4N(s+1)}{s^2}} e^{-\frac{2(2+s)(zs+N)}{s^2}} d(zs+N).$$

Define a new variable $t = \frac{2(2+s)}{s^2}(zs+N)$. Thus,

$$y(x) = C \int_{\frac{2(2+s)}{s^2}N}^{\frac{2(2+s)}{s^2}(xs+N)} t^{\frac{4N(s+1)}{s^2}} e^{-t} dt =$$

$$= C\Big[\gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}(xs+N)) - \gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}N)\Big],$$

where C is a constant. By the definition of $x(t)$,

$$0 \le x(t) \le N,$$

thus

$$C\Big[\gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}N(s+1)) - \gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}N)\Big] \le y(t) \le$$

$$\le C\Big[\gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}N) - \gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}N)\Big].$$

By the Doob's optional stopping theorem [24], the expected value of the bounded martingale $y(t)$ is equal to $y(0) = y(x_0)$. Therefore

$$y(0) = p_1 a_0 + p_0 a_1, \quad \text{where} \quad p_0 + p_1 = 1,$$

and a_0 and a_1 are the left and right ends of the interval for y. Therefore,

$$p_1 = \frac{y_0 - a_1}{a_0 - a_1} = \frac{\gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}(x_0 s+N)) - \gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}N)}{\gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}N) - \gamma(\frac{4N(s+1)}{s^2}+1, \frac{2(2+s)}{s^2}N)}$$

which is equivalent to the fixation probabilities given by Eq. (4.4) .

Appendix D

Fixation time for the "one-end" initial condition

In this appendix, we derive the average fixation time given by Eq. (4.9) from the general formula given by Eq. (4.7). Since we focus on the neutral case with $x_0 = N/2$, the fixation time in Eq. (4.9) can be written in the following way

$$T(x_0) = \frac{1}{2}\int_{x_0}^{N} \psi(y')dy' \int_0^{y'} \frac{dz}{B(z)\psi(z)} - \frac{1}{2}\int_0^{x_0} \psi(y')dy' \int_0^{y'} \frac{dz}{B(z)\psi(z)}. \tag{D.1}$$

This is due to the fact that the sub-populations of mutants and wild-types have the same fixation probabilities equal to $1/2$. In order to calculate the fixation time with Eq. (D.1), we first find function $\psi(x)$.

$$\psi(x) = e^{-\int_0^x \frac{v(x')}{D(x')}dx'} = exp\Big(-2\int_0^x \frac{2x-N}{N}\Big) = exp\Big(-\frac{2}{N}\int_0^x (2x-N)dx\Big) =$$

$$= exp\Big(-\frac{2}{N}(x^2 - xN)\Big) = exp\Big(-\frac{2}{N}(x-\frac{N}{2})^2 + \frac{N}{2}\Big) = e^{-\frac{(x-\frac{N}{2})^2}{N/2}} e^{N/2}.$$

Next, we calculate the terms of Eq. (D.1) separately using the expression of the function $\psi(x)$. The first term have the following representation

$$\int_{N/2}^{N} \psi(y')dy' \int_0^{y'} \frac{dz}{B(z)\psi(z)} = \int_{N/2}^{N} e^{-\frac{(y'-\frac{N}{2})^2}{N/2}} e^{N/2} dy' \int_0^{y'} \frac{2}{Nb} e^{\frac{(z-\frac{N}{2})^2}{N/2}} e^{-N/2} dz =$$

$$= \frac{2}{Nb}\sqrt{\frac{N}{2}} \int_{N/2}^{N} e^{-\frac{(y'-\frac{N}{2})^2}{N/2}} \int_{-\sqrt{N/2}}^{\frac{y'-N/2}{\sqrt{(N/2)}}} e^{t^2} dt = \frac{\sqrt{\pi}}{2b}\int_0^{\sqrt{N/2}} e^{-t^2}[erfi(t) + erfi(\sqrt{N/2})]dt =$$

$$= \frac{\sqrt{\pi}}{b}\Big[\int_0^{\sqrt{N/2}} e^{-t^2} erfi(t)dt + \frac{\sqrt{\pi}}{2} erfi(\sqrt{N/2}) erf(\sqrt{N/2})\Big].$$

The second term differs from the first one only in the integration limits. Therefore, it can be written in the following form.

$$\int_0^{N/2}\psi(y')dy'\int_0^{y'}\frac{dz}{B(z)\psi(z)}=\int_0^{N/2}e^{-\frac{(y'-\frac{N}{2})^2}{N/2}}e^{N/2}dy'\int_0^{y'}\frac{2}{Nb}e^{\frac{(z-\frac{N}{2})^2}{N/2}}e^{-N/2}dz=$$

$$=\frac{2}{Nb}\sqrt{\frac{N}{2}}\int_0^{N/2}e^{-\frac{(y'-\frac{N}{2})^2}{N/2}}\int_{-\sqrt{N/2}}^{\frac{y'-N/2}{\sqrt{(N/2)}}}e^{t^2}dt=\frac{\sqrt{\pi}}{2b}\int_{-\sqrt{N/2}}^{0}e^{-t^2}[erfi(t)+erfi(\sqrt{N/2})]dt=$$

$$=\frac{\sqrt{\pi}}{b}\Big[\int_{-\sqrt{N/2}}^{0}e^{-t^2}erfi(t)dt+\frac{\sqrt{\pi}}{2}erfi(\sqrt{N/2})erf(\sqrt{N/2})\Big].$$

Substituting these two terms to Eq. (D.1), we get the following expression

$$T(x)=\frac{\sqrt{\pi}}{2b}\Big[\int_0^{\sqrt{N/2}}e^{-t^2}erfi(t)dt-\int_{-\sqrt{N/2}}^{0}e^{-t^2}erfi(t)dt\Big]=$$

$$=\frac{\sqrt{\pi}}{b}\int_0^{\sqrt{N/2}}e^{-t^2}erfi(t)dt=\frac{N}{2}{}_2F_2(1,1;\frac{3}{2},2;-\frac{N}{2}),$$

where ${}_2F_2$ is the generalized hypergeometric function and it's has the following representation

$${}_2F_2(1,1;\frac{3}{2},2;-\frac{N}{2})=\Gamma(3/2)\sum_{k=0}^{\infty}\frac{(-N/2)^k}{\Gamma(3/2+k)(k+1)}.$$

Therefore,

$$T(x)=\Gamma(3/2)\sum_{n=1}^{\infty}\frac{(-N/2)^n}{\Gamma(1/2+n)n},$$

which is equivalent to Eq. (4.9).

Appendix E

Fixation probabilities for the random initial condition

Derivation of the fixation probability

In this section, we derive the fixation probability given by Eq. (4.15).

The fixation probability of mutants can be found as the probability that the population of mutants grows up to N cells before the population of wild-type cells. We find this probability in the form

$$P(T_{mut} < T_{WT}) = \int_0^\infty \int_0^y f_{T_{mut},T_{WT}}(x,y)dxdy = \int_0^\infty \int_0^y f_{T_{mut}} f_{T_{mut}T_{WT}}(x,y)dxdy, \tag{E.1}$$

We substitute $f_{T_{mut}}$ and $f_{T_{mut}T_{WT}}$ given by Eqs. (4.16) and (4.17) and perform integration. As a result, we obtain

$$P(T_{mut} < T_{WT}) = \frac{2^{2N}N^2(s+1)}{(2^N-1)^2}\int_0^\infty \int_0^y \frac{e^{-b(s+1)x}}{2}\Big(1-\frac{e^{-b(s+1)x}}{2}\Big)^{N-1}\frac{e^{-bx}}{2}\Big(1-\frac{e^{-bx}}{2}\Big)^{N-1}dxdy =$$

$$= \frac{2^{2N}N^2}{(2^N-1)^2}\int_0^\infty \frac{e^{-by}}{2}\Big(1-\frac{e^{-by}}{2}\Big)^{N-1}\int_0^y \Big(1-\frac{e^{-b(s+1)x}}{2}\Big)^{N-1}d\Big(\frac{e^{-b(s+1)x}}{2}\Big) =$$

$$= \frac{2^{2N}N}{(2^N-1)^2}\int_0^\infty \frac{e^{-by}}{2}\Big(1-\frac{e^{-by}}{2}\Big)^{N-1}\Big(-\frac{1}{2^N}+\Big(1-\frac{e^{-b(s+1)y}}{2}\Big)^N\Big)dy =$$

$$= -\frac{2^N N}{(2^N-1)^2}\int_0^\infty \frac{e^{-by}}{2}\Big(1-\frac{e^{-by}}{2}\Big)^{N-1}dy + \frac{2^{2N}N}{2(2^N-1)^2}\int_0^1 \Big(1-\frac{u}{2}\Big)^{N-1}\Big(1-\frac{u^{s+1}}{2}\Big)^N du =$$

$$= -\frac{2^N}{(2^N-1)^2}\Big(-\frac{1}{2^N}+1\Big) + \frac{2^{2N}N}{2(2^N-1)^2}\int_0^1 \Big(1-\frac{u}{2}\Big)^{N-1}\Big(1-\frac{u^{s+1}}{2}\Big)^N du =$$

$$= -\frac{1}{2^N-1} + \frac{2^{2N}N}{2(2^N-1)^2}\int_0^1 \Big(1-\frac{u}{2}\Big)^{N-1}\Big(1-\frac{u^{s+1}}{2}\Big)^N du.$$

The latest expression is equivalent to Eq. (4.18).

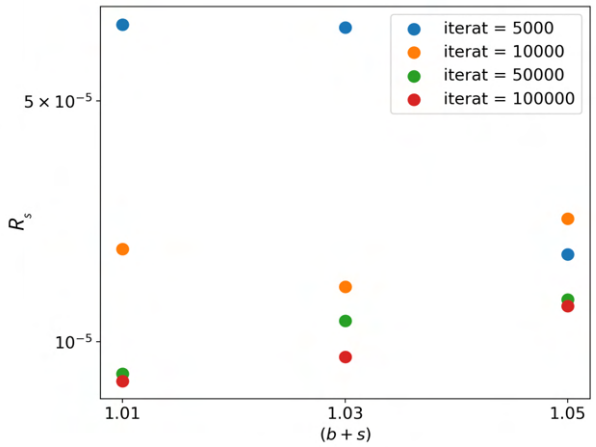


Figure E.1: Convergence test. The value of the mean squared error given by Eq. (E.2) decreases as the number of iterations grows.

For $s = 0$, Eq. (4.18) transforms to

$$P(T_{mut} < T_{WT}) = -\frac{1}{2^N - 1} + \frac{2^{2N} N}{2(2^N - 1)^2} \int_0^1 \left(1 - \frac{u}{2}\right)^{2N-1} du =$$

$$= -\frac{1}{2^N - 1} + \frac{2^{2N} N}{(2^N - 1)^2} \int_0^{1/2} \left(1 - t\right)^{2N-1} dt = -\frac{1}{2^N - 1} + \frac{2^{2N}}{2(2^N - 1)^2} \frac{2^{2N} - 1}{2^{2N}} =$$

$$= \frac{-2(2^N - 1) + 2^{2N} - 1}{2(2^N - 1)^2} = \frac{(2^N - 1)^2}{2(2^N - 1)^2} = \frac{1}{2}.$$

Convergence test

We demonstrate that the numerical fixation probabilities found for finite populations with the random initial condition converge to the theoretical fixation probabilities given by Eq. (4.18).

We run numerical simulations of populations of the size N for the same range of N and the selection advantage coefficient s as in Fig. 4.6. As the measure of discrepancy between numerical and theoretical predictions, we use the mean squared error in the form

$$R_s = \frac{1}{n} \sum_{i=1}^{n} \left(P_s^{num}(i) - P_s^{theor}(i)\right)^2, \tag{E.2}$$

where $P_s^{num}(i)$ and $P_s^{theor}(i)$ are numerical and theoretical predictions of the fixation probability, and n is the number of points (different values of the parameter N).

We find the value of R_s given by Eq. (E.2) for different values of s and different numbers of numerical simulations. Fig. E.1 shows that the value of R_s decreases as the

number of simulations grows.

Appendix F

Fixation time for the random initial condition

In this section, we find the average fixation time given by Eq. (4.21).

First, we find the average times at which populations of mutants and wild-type cells grow up to N cells.

$$\langle T_{mut}\rangle = \int_0^\infty t f_{T_{mut}} dt = \frac{2^N N(s+1)}{2^N-1}\int_0^\infty t\frac{e^{-b(s+1)t}}{2}\Big(1-\frac{e^{-b(s+1)t}}{2}\Big)^{N-1}dt =$$

$$= \frac{2^N N}{(2^N-1)(s+1)}\int_0^{1/2}\ln(2u)\Big(1-u\Big)^{N-1}du =$$

$$= \frac{2^N N}{(2^N-1)(s+1)}\frac{H_{[N/2]}}{[N/2]} \to \frac{1}{s+1}\frac{2^N}{2^N-1}\Big(\ln\Big[\frac{N}{2}\Big)+\gamma\Big],\ \text{as } N\to\infty,$$

where $[N/2]$ is the integer part of $N/2$, $H_{[N/2]} = \sum_{k=1}^{[N/2]}\frac{1}{k}$ is the $[N/2]$th harmonic number, and $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

The average time $\langle T_{WT}\rangle$ at which populations of wild-type cells grow up to N cells equals to $\langle T_{mut}\rangle$ for $s = 0$.

The total average time at which either of the wild-type and mutant domains reaches the size N in the form

$$\langle T\rangle = P(T_{mut} < T_{WT})\langle T_{mut}\rangle + (1 - P(T_{mut} < T_{WT}))\langle T_{WT}\rangle =$$

$$\Big[-\frac{1}{2^N-1}+\frac{2^{2N}}{(2^N-1)^2}\frac{N}{2}\int_0^1\Big(1-\frac{u}{2}\Big)^{N-1}\Big(1-\frac{u^{s+1}}{2}\Big)^N du\Big]\frac{1}{s+1}\frac{2^N}{2^N-1}\Big[\ln\Big(\frac{N}{2}\Big)+\gamma\Big]+$$

$$+\Big[1+\frac{1}{2^N-1}-\frac{2^{2N}}{(2^N-1)^2}\frac{N}{2}\int_0^1\Big(1-\frac{u}{2}\Big)^{N-1}\Big(1-\frac{u^{s+1}}{2}\Big)^N du\Big]\frac{2^N}{2^N-1}\Big[\ln\Big(\frac{N}{2}\Big)+\gamma\Big].$$

We solve the integrals in this expression numerically.

For $s = 0$ and large N, the average time $\langle T\rangle$ becomes

$$\langle T\rangle = \frac{1}{2}\Big[\ln\Big(\frac{N}{2}\Big)+\gamma\Big]+\frac{1}{2}\Big[\ln\Big(\frac{N}{2}\Big)+\gamma\Big] = \ln\Big(\frac{N}{2}\Big)+\gamma,$$

which is equivalent to Eq. (4.22).

Appendix G

Coalescent dynamics of planktonic communities

Martín, P.V., **Koldaeva, A.** and Pigolotti, S. "Coalescent dynamics of planktonic communities", Physical Review E, 2022, [59].

Contributions:

P.V.M. and S.P. designed the research. P.V.M. designed the model; A.K. performed numerical simulations and analyzed the results; P.V.M., A.K., and S.P. wrote the paper.