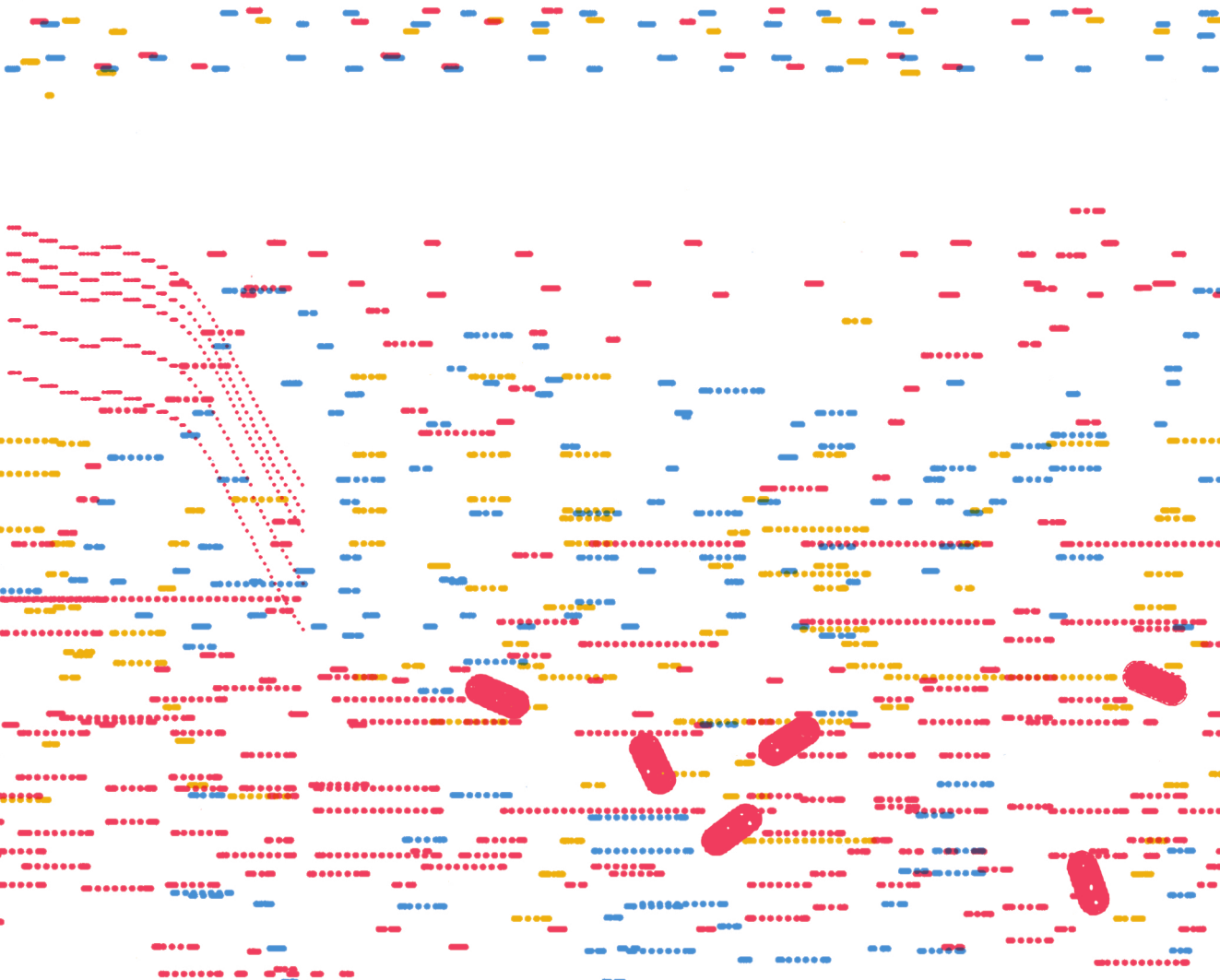
Towards a statistical physics of eco-evolutionary systems

Matteo Sireci

Supervisors : Miguel Ángel Muñoz Martínez, Jacopo Grilli
Doctorado en Física y Matemáticas

UNIVERSIDAD DE GRANADA



Editor: Universidad de Granada. Tesis Doctorales
Autor: Mateo Sirecci
ISBN: 978-84-1117-937-9
URI: https://hdl.handle.net/10481/83101

Towards a statistical physics of eco-evolutionary systems

Author:
Matteo Sireci

Supervisor:
Miguel Á. Muñoz
Jacopo Grilli

UNIVERSIDAD
DE GRANADA

Programa de doctorando en Física y Matemáticas

"Deus sive natura"

Spinoza

Abstract

In this thesis, we employ the principles of statistical physics to investigate the ecology and evolution of bacterial communities. By studying the collective behavior of large ensembles of molecules or components, statistical physics uncovers new phenomena and transitions between different phases . This approach, originally applied to physical systems, has been extended to complex systems like biology and social systems, leading to the emergence of the discipline of complex systems. Ecosystems, exemplifying complex systems, consist of numerous interacting species that generate emergent properties such as diversity, stability, and functionality. Given the continuous evolution of species, constructing a statistical physics framework for ecological systems is a challenging task. To tackle this, the thesis focuses on bacterial populations as a relatively accessible system. Recent technological advancements allow for easy sampling, analysis, and sequencing of bacterial communities, making them ideal for studying emergent phenomena. The thesis is divided into three parts, corresponding to ecology, evolution and non-equilibrium physics.

In Chapter 1, we provide an overview of the motivations and content of this thesis, along with a general introduction to complex systems and the ecology and evolution of bacteria. Part I centers on microbial macroecology, with Chapter 2 specifically investigating interactions within bacterial ecosystems. Through extensive data analysis, we discover a universal macroecological law that relates pairwise correlations between species to their phylogenetic distance. Utilizing a statistical physics approach, we develop a stochastic model that reproduces this empirical pattern, attributing it to coupled environmental fluctuations, also known as environmental filtering.

Part II focuses on bacterial eco-evolution, particularly the formulation of a new theoretical framework and its application to antibiotic tolerance evolution. In Chapter 3, we establish a general framework for trait distributions using statistical physics tools, investigating various evolutionary phenomena such as evolutionary branching.

In Chapter 4, we employ this framework to study the evolution of antibiotic tolerance in bacteria through lag-time adaptation. By presenting a stochastic individual-based model that replicates experimental results, we derive analytical predictions using our framework. Finally, in Part III, we delve into the concept of irreversibility in non-equilibrium statistical physics.

Chapter 5 examines the geometric properties of non-equilibrium currents in stochastic thermodynamics, extracting theoretical insights. This geometric information is then utilized to comprehend the relationship between irreversibility, dissipation, and current symmetry breaking in non-equilibrium stationary states.

In Chapter 6, we analyze the irreversible properties of evolution using the general framework introduced in Chapter 3. Evolution is found to be constantly out of equilibrium due to the simultaneous presence of selection and mutations, and we explore its irreversibility in various examples, including evolutionary branching. Ultimately, in Chapter 7, we present general conclusions drawn from our findings and suggest potential avenues for future research. Also, in appendix E we include a resume and the concluions of the thesis in english.

La física estadística proporciona un marco teórico para estudiar las propiedades colectivas de sistemas compuestos por muchos componentes que interactúan entre sí. Aunque inicialmente se aplicó a sistemas físicos, se ha extendido a otros sistemas complejos como la biología y los sistemas sociales. En esta tesis,empleamos la física estadística para investigar la ecología y evolución de comunidades bacterianas, las cuales son sistemas ecológicos ideales debido a la enorme cantidad de datos disponibles.
La tesis se divide en tres partes, correspondientes a los argumentos investigados: ecología, evolución y física estadística fuera del equilibrio. En el Capítulo 1, ofrecemos una visión general de las motivaciones y el contenido de esta tesis, junto con una introducción general a los sistemas complejos y la ecología y evolución de las bacterias.
En la Parte I, nos enfocamos en la macroecología microbiana y exploramos las interacciones dentro de los ecosistemas bacterianos. A través del análisis de datos, descubrimos una ley macroecológica universal que relaciona la correlación entre especies con su distancia filogenética. Mediante el desarrollo de un modelo estocástico basado en la física estadística, atribuimos este patrón a fluctuaciones ambientales acopladas, conocidas como filtro ambiental.
La parte II se adentra en la eco-evolución bacteriana. En el Capítulo 3 establecemos un nuevo marco teórico utilizando herramientas de física estadística para estudiar la distribución de fenotipos y diversos fenómenos evolutivos, como la especiación simpátrica. Además, en el Capítulo 4 empleamos este marco para investigar la evolución de la tolerancia a los antibióticos en bacterias mediante la adaptación del tiempo de lag. Presentamos un modelo estocástico que reproduce los resultados experimentales y obtenemos predicciones analíticas utilizando nuestro marco teórico.
En la Parte III, examinamos el concepto de irreversibilidad en la física estadística de sistemas fuera de equilibrio. El Capítulo 5 se centra en las propiedades geométricas de las corrientes en termodinámica estocástica y sus implicaciones. Analizamos la relación entre irreversibilidad, disipación y rupturas de simetría de las corrientes en estados estacionarios fuera de equilibrio. En particular, conseguimos generalizar el principio de Prigogine utilizando el exceso de entropía.
En el Capítulo 6, exploramos las propiedades irreversibles de la evolución Darwiniana utilizando el marco teórico general presentado anteriormente. Descubrimos que la evolución se mantiene constantemente fuera de equilibrio debido a la presencia de selección y mutaciones, y estudiamos la irreversibilidad en ejemplos como la especiación.
Finalmente, en el Capítulo 7, presentamos conclusiones generales y sugerimos posibles direcciones para futuras investigaciones. Al emplear la física estadística, esta tesis contribuye a la comprensión de los sistemas complejos, proporcionando conocimientos sobre el comportamiento colectivo, la ecología y la evolución de las comunidades bacterianas. Un resumen y las conclusiones de la tesis en castellano están incluidos en la apéndice E.

Contents

List of Abbreviations

AFD	**A**bundance **F**luctuation **D**istribution
AD	**A**daptive **D**ynamics
CSLM	**C**orrelated **S**tochastic **L**ogistic **M**odel
EGT	**E**volutionary **G**ame **T**heory
GLVM	**G**eneralized **L**otka **V**olterra **M**odel
LGT	**L**ateral **G**enetic **T**ransfer
NT	**N**eutral **T**heory
NESS	**N**onequilibrium **S**tationary **S**tate
SAD	**S**pecies **A**bundance **D**istribution
SLM	**S**tochastic **L**ogistic **M**odel

Chapter 1

Introduction

1.1 Prelude: life out of equilibrium

Depending on the scale at which it is observed, life manifests itself in totally different phenomena and characteristics. Nevertheless, a common aspect pervades it: it is out of equilbrium, at physical, chemical or purely dynamical level.
Inside the cell, chemical reactions work at an unstoppable pace in transforming molecules or in producing energy. Yet, the cell persists. In multi-cellular organisms, single cells are constantly destroyed and created to maintain the individual alive. Energy flows from one being to another in the act of eating. Species are persistently reshaped by evolution and adaptation, creating a temporal lineage that goes beyond the survival of the individual. Ecosystems are an entangled web of relations emerging from species interactions. Independently of the scale at which life is observed, the apparent stasis is just the result of a hidden movement. This dynamic motion relates all the individuals with others, and is the common nature of all life on earth. As the philosopher Emanuele Coccia points out, life is is always an inheritance from somebody else, a *Metamorphoses* where the form changes but the substance persists [1].
Ecology, is the study of how populations, species and ecosystems are subjected to this flow of matter, biomass and energy. The process by which individuals and their relations change in time is *evolution*.
Along its history, ecology has been an attractive subject for physicists. The American physical chemist Alfred J. Lotka, in his classic book "Elements of physical biology" [2], proposed an interpretation of ecology in terms of flow of biological matter (*biomass*) and energy through organisms [3]. Lotka was probably one of the first physicists interested in ecology, and, in his book, he distinguished between biophysics, the study of morphological and physical phenomena at the organism level, and his physical biology, whose aim is to apply physical principles to biological systems. Being very influenced by his education, he proposed to study aggregates of individuals, i.e. populations, in the same way a chemist analyzes aggregates of molecules.
Following his physical inclination, Lotka proposed the idea that ecological relations could derive from energetic principles, and introduced words as "system", "dynamics" and "kinetics" together with differential equations in the field of theoretical ecology. In his view, individuals are just a metastable accumulation of biomass in constant exchange of energy with the environment.
It is inspiring to contemplate the distribution of biomass on earth by agreeing momentarily with Lotka's view. Life is divided into five kingdoms, each of them with characteristic metabolic and ecological properties. How much biomass does each kingdom store?
The biomass composition of the entire biosphere consists of a census of 550 (approx) gigatons of carbon (Gt C) distributed among all of the kingdoms of life [4]. Plants accumulate ≈450 Gt C and are the dominant kingdom, primarily present in terrestrial areas. Bacteria

(≈70 Gt C) and archaea (≈7 Gt C) are the second and third dominant kingdom and are predominantly located in deep subsurface environments. Finally, fungal biomass consists of ≈ 15 Gt C while animals one is ≈ 2 Gt C (mainly in marine environments.) [4]. Figure 1.1 illustrates graphically the biomass distribution across kingdoms, and in particular in the animal one. Indeed, organisms can be divided into two classes depending on their energetic strategy.
As an inspiring exercise let us assume this purely energetic perspective and imagine the flow of biomass in the whole earth. Consider a sphere in an empty space: the inside is organic matter, while the outside is inorganic matter. On the surface of the sphere lie the autotroph, like plants algae and many bacteria, that are organism able to grow by absorbing energy from inorganic matter. Photosyntesis, i.e. that capacity of using light, water and carbon dioxide to produce oxygen and energy in form of sugar, is a typical behavior of this class of organisms. Hence, they create a net change by converting inorganic matter into organic one. On the other hand, the interior of the sphere is the residence of heterotroph, organisms that need to draw energy from other life forms, like animals, fungi and some bacteria. Many of this organisms sustain themselves by predating on other individuals, being them autotroph or heterotroph, and hence transforming inorganic matter into new forms.
These two classes of organisms constitute the basis of a *food web*, i.e. a graphical representation of what-its-what in a ecosystem and a central tool in theoretical ecology.
The continuous flux of energy through the food web produces several universal laws relating the biomass of its different components [5], such as predator-prey [6], plant energetics [7] and mammals body-mass [8].
Even if this thesis is not about ecosystems energetics, we consider the Lotka's view very inspiring and motivating.
The next section is devoted to present the contents of this thesis.

1.2 Motivations and contents of this thesis

Statistical physics studies the macroscopic properties of aggregates of many molecules, or components, ranging from gases to ferromagnetic systems. Instead of following the single trajectory of each components, statistical physics takes the ensamble approach, and studies the probability of a certain collective configuration. By moving to the macroscopic level, new collective and cooperative phenomena are revealed, such as different "phases" and particular transition between them [9]. By applying this paradigm to non-physical systems, like biological or social one, physicists contributed to create the discipline of *complex system*, which aim is to study natural collective phenomena.
Ecosystems are a paradigmatic example of complex systems, given that they are composed by a large number of species, which interactions generate collective emergent properties, such as diversity, stability and functions. Nevertheless, ecological communities are far more complex than gases, because in the first one, individual differences have an effect, while in the latter they are averaged away by the enormous size of 10^{23} molecules per mole. Furthermore, while in statistical physics the system is in thermodynamic equilibrium, ecological communities are away from equilibrium, due the nontrivial species interactions and the constant fluctuations of the environment. Finally, species are always in a process of *evolution*, that, even if on long timescales, causes a continuous transformation of the system components and interactions. Hence, constructing a statistical physics of ecological systems is far more difficult objective than the one faced by Boltzmann and Gibbs.
To make the venture smoother, we will try to construct such an ambitious theory in one

of the easiest available system: bacterial populations. Indeed, thanks to recent technological advances, bacterial (or more in general microbial) communities, such as the one human gut microbiome, can be easily sampled, analyzed and sequenced, obtaining the abundances of typically 10^3 species for a total order of 10^{13} individuals ! By looking at the number, it is not as big as Avogadro's number, but it is more than the total number of trees on the planet. Hence, microbial communities are the ideal ecological system to study emergent phenomena.
Furthermore, thanks to their typical short division time and their large mutation rate (for E.Coli in the lab the first is of 20 mins, while the second is circa 10^{-10} per base per generation), bacterial populations are suitable to study evolutionary phenomena, like the emergence of tolerance to antibiotics [10], or rise of different species [11]. Hence, the development of a predictive physics of microbial communities would have deep consequences in medicine, given the highly non trivial rate of the gut microbiome in chronic diseases, and the rapid rise of resistance to antibiotics.
Hence, we claim that non-equilibrium statistical physics is an important tool to understand biological phenomena.
Part I of this thesis , and in particular Chapter 2, is devoted to the study of collective interactions in microbial communities. Using a macroecological approach, we reveal the presence of a universal mocroecological law relating species correlations and phylogenetic similarity. Using the tools of statistical physics, we formulate a model able to reproduce such a empirical pattern. In particular, shared environmental flctuations, also known as environmental filtering, are individuated as the cause of the macrecological law.
Part II of this thesis is devoted to the bacterial phenotypic evolution. In Chapter 3 we formulate an a original eco-evolutionary framework for bacterial population, using the tools of non-equilibrium statistical physics such as stochastic processes and kinetic theory such as stochastic processes and kinetic theory. Thanks to this framework, we are able to generalize the existent theory of Adaptive Dynamics (AD) to bacterial populations. In Chapter 4 we use the aforementioned framework to study a relevant example of bacterial evolution: the emergence of tolerance to antibiotics. In particular, we consider a recent experiments where bacteria extend their lag phase to survive to antibiotic exposition and formulate a compuational model able to reproduce the empirical results. Furthermor, our eco-evolutionary framework is used to derive an analytical form of the model and to make quantitative predictions.
In part III we delve deep into the non-equilibrium statistical physics of irreversible phenomena. In particular, Chapter 5 is dedicate to study the general problems of non-equilibrium systems using the tools of stochastic thermodynamics, and in particular to study the relation between dissipation, irreversibility and the geometrical properties of currents. We show that non-equilibrium conditions induce a chiral symmetry breaking in the currents and derive from it important thermodynamics consequences. In particular, when approaching a non-equilibrium steady state the systems tends to minimize the entropy production but to minimize the housekeeping heat emerging from the aforementioned symmetry breaking.
To conclude this part, in Chapter 6 we study the eco-evolutionary framework introduced in 3 under the lens of non-equilibrium statistical physics. In particular, we derive the entropy production, i.e. a measure of irreversibilty , of adaptive evolution, concluding that the contemporary presence of selection and mutations drives the population away from equilibrium. The entropy production is studied in many different cases, included the diversification phenomenon of evolutionary branching.
To close this thesis, in Chapter 7 we derive some general conclusion across and in each of three parts and sketch the way to future research.

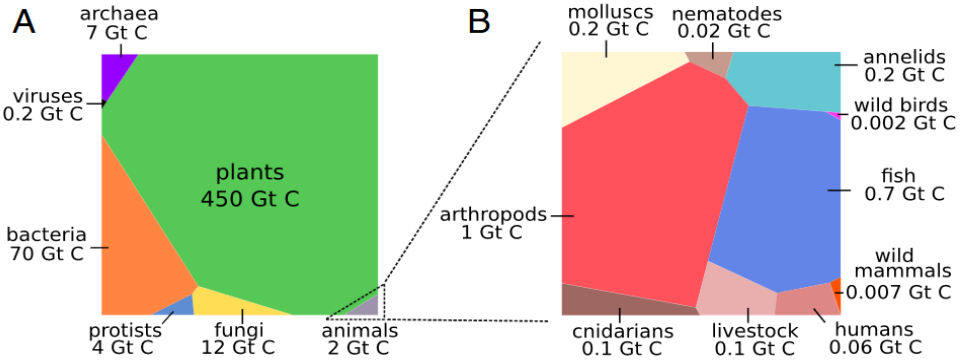


FIGURE 1.1: Distribution of biomass on earth
. Figure from Ref. [4]. (A) Graphical representation of the biomass distribution across principle taxa (colors): animals, protists, fungi, plants, bacteria, archea and viruses. The are of the polygons is proportional to the taxa biomass. Clearly, the majority of earth biomass is accumulated in plants. (B) Zoom in in the biomass distribution in animals. The major contribution of biomass are from arthropods and fishes.

To complement the results exposed in the thesis, we present also four appendices, containing the more technical material of data analysis, numerical methods and analytic derivations.
In this introduction, we both present the biological system of interest, i.e. the ecology and evolution of microbial world, and the physical tools we use to describe it.
In Section 1.3, we introduce briefly both microbial evolution and ecology. We begin by giving an overview of bacterial evolution and by showing how its peculiar characteristics hinder the univocal definition of species. The following section 1.3.2 is devoted to microbial ecology, illustrating briefly the nature and the interaction of microbial communities. In particular in Sec.1.3.2, *Macroecology* is introduced as the framework to study emergent and general statistical patterns in ecosystems.
In Sec.1.4 we discuss how theoretical physics can describe and study biological collective processes as complex systems. In particular, Section 1.5 is devoted to the presentation of the key aspects of non-equilibrium statistical physics, such as stochastic dynamics and multiplicative fluctuations.
After having discussed the basic concepts and formalism, Sections 1.6, 1.7 and 1.8 are devoted to a detailed introductions to the investigations reported in the three parts of the thesis. Section 1.3.2, is and introduction to part I where we discuss some recent results on microbial macroecology, that are the basis of the research reported in this thesis.
On the other hand, in section 1.7 we discuss evolutionary dynamics and the theory of Adaptive dynamics. These theories are the basis of part II, were will generalize and apply them to bacterial populations. Finally, section 1.8 is a brief introduction to irreversibility and dissipation in non-equilibrium statistical physics and complex systems.

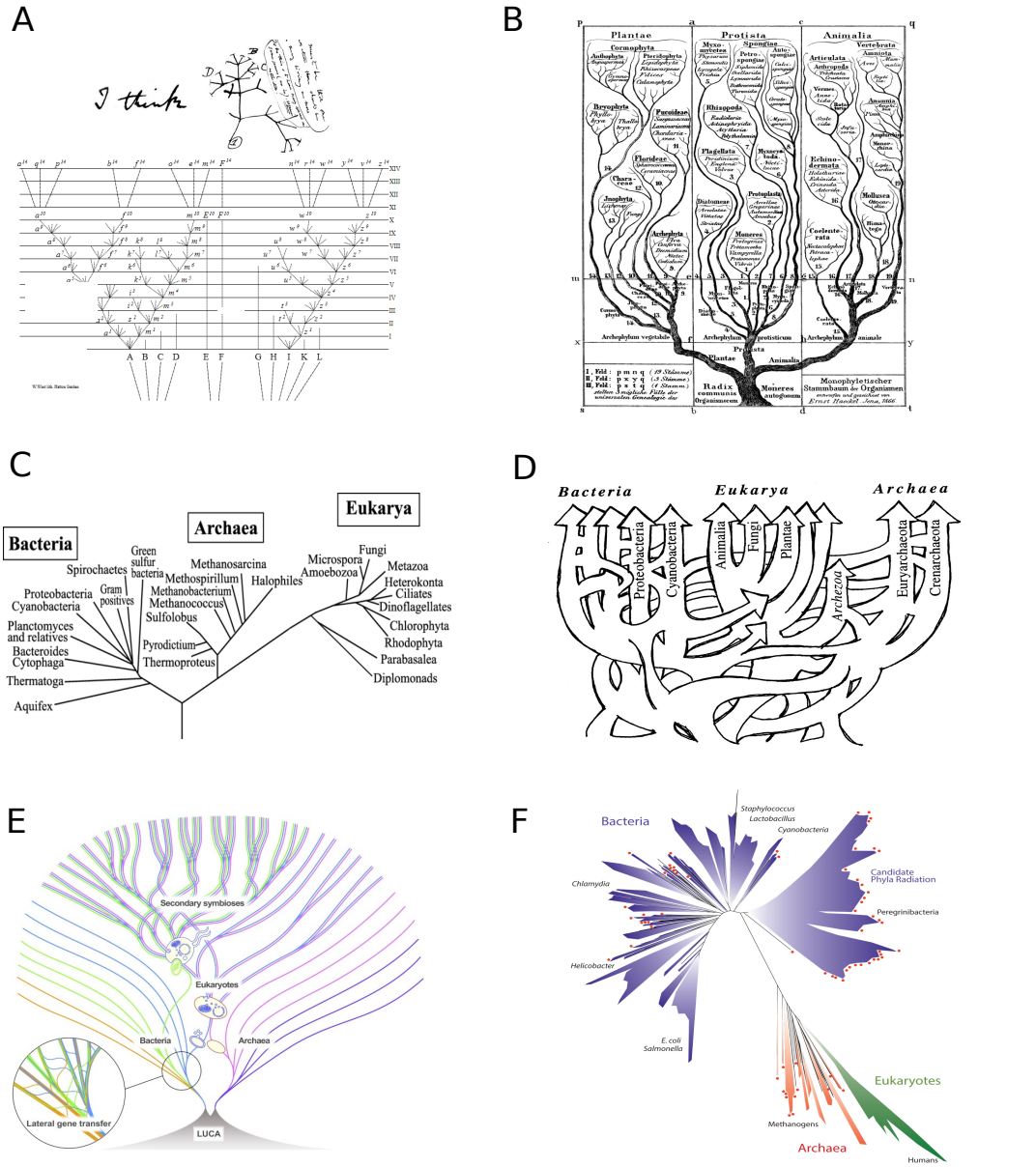


FIGURE 1.2: **The history of the evolutionary tree metaphor** A) Darwin's tree. (Top) In his travel diary, Darwin drew a tree like shape together with the words "I think" [12]. This is the first proof the intuition of representing evolution in diagram with the form of a tree. (Bottom)A more detailed representation of the evolutionary tree appeared in the Origins[13]. B) the graphically interpretation of Darwin's divergence diagram as a tree by E. Haeckel C) Using the 16s rRNA method, Carl Woese constructed a phylogenetic tree that included also bacteria, and eukaryotes, as life's domain. The research results revealed the existence of the third domain of archea (figure from [14]). D) After the discoveries of lateral genetic transfer and endosymbiosis, W. Ford Doolittle defined the new notion of "reticulated tree" to represent natural evolution (figure from [15]). E) William Martin realized a graphical illustration of the reticulated tree that included clearly the effect of LGT, the first symbiosis producing eukaryotes and further symbiotic events (figure from [16]). F) The tree of life obtained by J. Banfield and collaborators obtained sequencing thousands of *wild* microbes with a new method involving sixteen ribosomial proteins.

1.3 A tour in the microbial world

1.3.1 The species problem and bacterial evolution

Ecology and Evolution are sciences based on the concept of species. The utility of this concept comes from establishing a criterion to determine when two individuals are similar enough in their genomic and phenotypic properties to give them the same name. Yet, its definition is far from obvious and it underpin our conception of natural history. Even Darwin was not giving a clear definition:
"No one definition (of species) has as yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species. Generally the term includes the unknown element of a distinct act of creation. The term "variety" is almost equally difficult to define; but here community of descent is almost universally implied, though it can rarely be proved."
In this section we briefly give an overview of bacterial evolution and show how it hinders univocal definition of species.
Historically, species were defined way before the formulation of evolutionary theory as a consequence of the empirical observation that species, when reproducing, produce always an individual of the same species [17]. Linneus formalized the classification of species on the bases of *morphological* characteristics, such as shape, size, colour and internal structure like organs organization, skeleton etc. In this view, species are immutable entities.
In the Origins of species, Darwin, criticized these conceptions by claiming that species were, instead, the result of an evolutionary *process*. The central idea beyond Darwin's revolution is that species are not static entities, but the results of a transformation in time, that will continue to mutate them in the future [13]. The evolutionary tree, or diagram of divergence, is a graphical tool invented by Darwin to represent evolution in time. In metaphor, the top "leafs" represent species in the present, while the root is a common ancestor in the deep past. In Fig. 1.2 A we report the first sketch of the evolutionary tree in Darwin's diary, together with the illustration appeared in the Origins [13]. In Fig. 1.2 B we report also one of the earliest images representing Darwin's diagram as tree by E. Haeckel. As a consequence of Darwin's intuition, all the species have a *phylogenetic* relation, given that they share a common ancestor. In the present, that ancestor is called LUCA: latest universal common ancestor.
Notwithstanding Darwin's revolution, the modern concept of species was defined almost a century after him by Ernest Mayr, on the basis of breeding possibilities [18]:
"Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups" [19].
Even if this definition is far more clear than a comparison of morphological traits, it presents some problems. For example, how can researchers understand if individuals are capable of interbreeding? Does it refer to a wild or captivity condition? How can this definition be used to compare species that lived in different epochs? Furthermore, many ecologists remarked that even if the number of species is an important measure of the diversity of an ecosystem, it is not the whole story. Biological diversity is about the variety of living organisms at all levels, from individual genomic variants, to species, families and up to ecosystems.
The most important criticisism to the species concept come from bacteria. Indeed, prokaryotes are the most abundant form of life on earth, but they refuse to respect any species definition. First of all, bacteria reproduce asexually by duplication, while the majority of macroscopic organism reproduce sexually.
An early attempt of bacterial classification by morphological characteristics was put forward by the German biologist F. Cohn in 1868. He divided bacteria in four groups based

on the cell's shape: sphericals, short rods, threads, and spirals. Notwithstanding Cohn's proposal, cell's shape turned out to be a not reliable guide to classify bacteria. Shape is an adaptive feature, but as all results of adaptation it can be ancestral as convergent. For example, an elongated shape seems to be good for swimming, while the round one can be a good strategy against desiccation [17].
The smartest intent to identify bacterial species and to construct a universal evolutionary tree came from the molecular biologist Carl Woese in 1977 [20].
In 1965, Pauling and Zuckerkandl had the idea to use "molecules as documents of evolutionary history" to construct a general philogeny able to scanner the deep evolutionary past [21]. Woese had the intuition that a universal phylogentic tree could be constructed by comparing the sequences of Rna that compose the ribosome in different organisms. This claim comes from the idea of Pauling and Zuckerkandl to use "Molecules as documents of evolutionary history"[21] a simply but extremely deep observation: all life forms share the fact of using ribosomes to translate genetic information into proteins. Woese found out that the most conserved part of the ribosome, and hence the best indicator for phylogenetic relations, is the 16s unit. From this observation,the 16s rRNA method, the most used phylogenetic tool in microbiology, takes his name.
By sequencing thousands of microbes, Woese made an incredible discovery: apart from prokaryotes and eukaryotes, there is a third domain of life that he called *archea*. The names come from the fact that first individuals sequenced by Woese and Fox where methanogens (microorgansims that produce methanon as by-produc), and they assumed that their metabolism reflected the earth primordial environment. Until then, archea were classified as extremophile, bacteria that can live in environment with extreme levels of temperature, salinity, pH or radioactivity. Nevertheless, bacteria are, approximately, as different from us as they are from archea. The reason lays in arhcea genes and metabolic pathways, that are much more similar to the eukaryotic ones. Hence, a new evolutionary tree with three domains, prokaryotes, eukaryotes and archea, was proposed by Woese, see in Fig. 1.2 C a contemporary version of the illustration proposed by Woese in 1990 [22].
Nevertheless, the dream of drawing a universal tree of life and hence attaining a clear definition of bacterial species crashed due to bacteria's singular evolutionary history.
In animals, evolutionary novelties emerge as mutants in the population and can spread thanks to sex and recombination. Using Mayr's classification, different species cannot interbreed and hence they are "genetic island". Even if bacteria have long been though to reproduction asexually without recombination, the discovery of lateral genetic tansfer (LGT) revolutionized this conception [23]. Indeed, bacterial cells can exchange genetic material using various mechanisms without reproduction but "horizontally" [24]. For example, a bacterial cell can uptake and incorporate exogenous genetic material in the environment, for example a plasmid released by another dead bacterial cell (*transformation*). The equivalent of sex in bacterial cell is called *conjugation*, that consist of two cells building bridge-like structures between them through witch they can exchange genetic material (*conjugation*). Finally, bacteria can also receive new genetic material by injection by a virus (*transduction*). For example, bacteria can become pathogens by transformation, or acquire antibiotic resistance by conjugation [24, 23]. As an additional example, note that part of the enormous genetic diversity in strains of the cyanobacterium Prochlorococcus is due to gene acquisition by transduction [25].
Lateral genetic transfer has demolished the metaphor of the evolutionary tree, given that evolutionary branches can collide, something that is not possible in a natural tree. Nevertheless, the image was replace by the new metaphor of the " reticulated tree" [15], see Fig. 1.2 D.
Another reason for which bacteria defy the species classification is the phenomenon

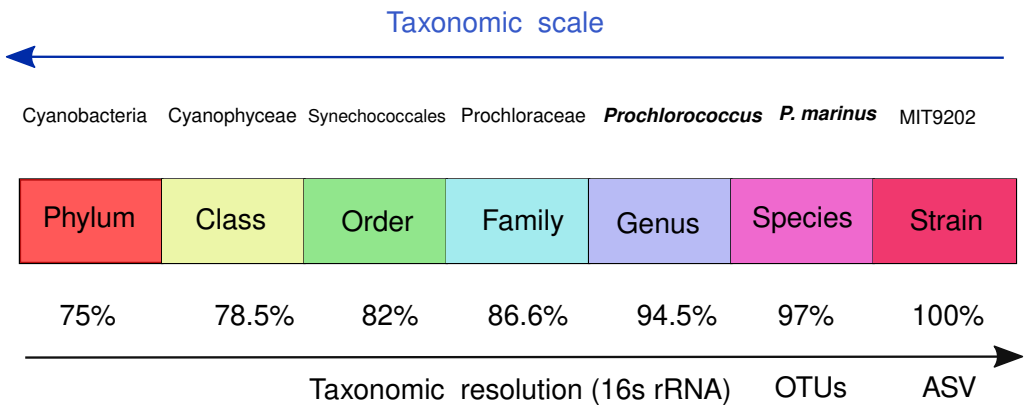


FIGURE 1.3: Graphical representation of bacterial taxonomy. Ranks in bacterial taxonomy goes from phylum (a large scale/small resolution), to strain (small scale/large resolution). OTUs constructed with the 16s rRNA method at 97% similarity correspond roughly to the rank of species, while higher ranks can be constructed with smaller similarity thresholds (values taken from Ref.[28]). Depending on the single case, OTUs can also be defined up to 99% similarity. To explore the finer scale of strains, ASV or metagenomics methods are instead needed. To give an example, we also represent the taxonomy of Prochlorococcus p. marinus, the most abundant photosynthetic organism, including also an example of strain sequence, "MIT9202" [29].

of *endosymbiosis*, i.e. the process of living in a mutualistic relations inside the body of another organisms [26, 17]. This relation can be more or less tight and irreversible. For example, eukaryotes probably originated by a symbiosis between an archea and a bacterium. Indeed, the mithocondria are the residuals of a proteobacteria entered in endosymbiosis with his host; similarly, chloroplasts in plants originated by ancestral cyanobacteria. To represent these concrete symbiotic events, spectacular images of the reticulated tree have been constructed, as the tree depicted in Fig. 1.2 E by W. Martin.
Finally, while Woese and collaborators used just the 16s unit of cultivated microbes, in 2016 J. Banfield and collaborators have constructed a new tree of life sequencing thousands of *wild* microbes with a new shotgun method involving sixteen ribosomial proteins [27]. The result is depicted in Fig. 1.2 F and expanded dramatically our perception of life diversity by discovering hundreds of new species. In summary, species are a scientific concept with many limits and a given region of validity. Nevertheless, an operational definition is possible, and microbial ecology is temporarily safe.

Bacterial taxonomy

Thanks to the 16s rRNA method, bacterial taxonomy started a path towards scientific categorization.
To identify bacteria, ribosomial sequences are amplified with the method PCR, and the compared one with the other. In contemporary research, the 16s method is usually used as a comparative classification, i.e. each researcher compares his particular sequences with genetic references stored in a bioinformatic database. OTUs (Operational Taxonomic Units) are created by clustering sequences with at least a minimum similarity (generally at 97 %) with the reference one. This approach is computationally economic but is not able to identify unknown sequences.
Even if not in a unique way, it is possible to construct a taxonomy with this method. Generally, 97% similarity OTUs correspond to bacterial species, and by lowering the similarity threshold, one can group OTUS and define more coarsened ranks, as phyla. For

example, bacterial phyla can be defined at 75% similarity, obtaining a total of 89 of them, such as firmicutes, bacteriotides, proteobacteria [28].
On the other hand, to explore finer scales of diversity, OTUs can be defined up to the 99% similarity threshold. For even smaller scales, i.e. strains or exact sequence variant at 100% similarity, reference free method are necessary. By using only ribosomial RNA, the Amplicon Sequence Variant (ASV) is a method to obtain exact sequences using denoising algorithms. In general, this classification is far more reliable than OTUs, and scientific community is shifting to it [30]. Alternatively, shotgun metagenomics procedures that use the whole bacterial genome are the most secure methods of identification, and by using them, also the species metabolic functions can be determined. In Fig. 1.3, we graphically represent bacterial taxonomy and ranks classification in terms of 16s rRNA similarity. As an example, we describe the taxonomy of the marine Prochlorococcus p. marinus, that is the most abundant photosynthetic organism on the planet, and presents also an incredible diversity at the strain level [29, 31, 32]. In this thesis, we will use improperly the word bacterial "species" when referring to OTUs at 97% similarity.

1.3.2 Microbial ecology

Microbes are microscopic organisms, both uni and multi-cellular, like bacteria, archea, protists and some species algae and fungi. Even if microbes have been used by humans in the preparation of food since ancient times, they were discovered for the first time by dutch biologist Leeuwenhoek in 1677 [33].
Bacteria are the most abundant organisms on earth (circa 10^{30} cells [34], total biomass of $70Gt$ of carbon), and they are found all across the biosphere, from human guts to glaciers, from soil to activated sludge. bacteria are a single (prokaryotic) organism, similar to an *open* system able to harvest energy from the environment and capable of converting it into biomass and residuals. Bacteria *grow* by asexual division; the duplication time can vary from 20 minutes for E Coli in laboratory conditions to two years for microbes living in deep sea sediments. Their metabolism is highly diverse, given that they can perform molecular respiration, fermentation and some species are capable even of photosynthesis.
Microbial ecology studies how microbial communities grow, interact, and how species diversity is maintained. This question is particularly relevant for bacteria, because a huge diversity is observed at the thin taxonomic scale of strains [35]. For example, in marine populations of the cyanobacteria prochlorococcus (genus), hundreds of different strains are observed [31]. How can so many strain coexist ?
Many possible ecological processes, or ecological *forces*, can arise in diverse communities, but in general they are a consequence of individuals metabolic properties.
The ecological force that is generally thought as the dominant one is *competition*. It can arise between two individuals when they share a common resource of limited supply [36]. When competition is present, the growth of a species implies a reduction in number of the competing one, possibly leading to extinctions.
In general terms, competition is associated to the "Principle of Competitive Exclusion", stating that "two species or populations cannot inhabit the same niche: one will consistently out-compete the other" [37]. The species niche is an abstract ecological concept consisting of its interactions with its community and the environmental conditions needed for it to stay alive. Hence, if competitive exclusion is in action, there can be just a number of species equal or minor the number of available niches.
Also *mutualistic* interactions between species can arise in bacteria thanks to the by-products of cells metabolism. Indeed, it happens often that two competing species can coexist if

one is able to metabolize the by-product of the other. This phenomenon is called *cross-feding* and is found in many bacterial communities [38, 39, 40]. For example, a couple of strains of the families Pseudomonadaceae and Enterobacteriaceae can survive together because the first is able to growth on acetate, that is a by-product of the second [39].
The metabolic by-products of a species can also *inhibit* the presence of another one, i.e. a behavior called *antagonism*. For example, the bacterium Streptomyces coelicolor is able to produce antibiotics, that are used then to neutralize other species, like E. coli [41].
Predation is also present in prokaryotes, even if it is generally considered rare (just fifteen predatory species have identified) [42, 43]. The highly motile bacterium B. bacteriovorus has been reported to enter in the periplasmic space of other bacteria and to consume on their cytoplasmic contents [44].
Ecological interactions can produce also collective effects such as community functions. Indeed, ecological functions are not just properties of individuals, but more generally they are a collective properties of communities emerging from a complex web of molecular, physiological, and organismal interactions [45], like carbon cycle in closed ecosystems [46], nutrient cycle in soil [47], nitrogen cycle in the ocean [48]. Finally, microbial ecosystems display functional redundancy: even if species presence and abundance varies across communities the number and the type of functions performed is conserved [38, 49]. Even if bacterial communities exhibit an astronomical taxonomic diversity, many bacterial species share the same metabolic functions, suggesting that selection does not act only at the species level.
An additional collective effects arising in bacterial communities are *biofilms*, i.e. communities of bacteria that collectively adhere to surfaces creating a physical medium adapt to interact through signaling and nutrient exchange.
Finally, ecological forces are not just "internal" interactions between species, but also "external" factors like environmental and demographic fluctuations, migrations, spatial structure etc. For example, spatial chaos in meta-populations models is known to enlarge enormously the number of coexisting species with respect to competitive exclusion [50, 51, 52].
Hence, one could ask : Is it possible to disentangle the effects of these different forces? Which one is the main responsible of the existence of biodiversity ? An endless debate over the relative importance of deterministic forces, such as competitive exclusion, or of stochastic ones, such as environmental and demographic fluctuations, in determining the maintenance of biodiversity does exist in the community. In very diverse communities, like the prochlorococcus or the marine plankton [53], the existence of an astronomical number of niches is hipotetic and not completely reasonable . Hence, the relative importance of competitive exclusion can be questioned.
cross-feding seems to have the ability of sustaining diversity, since communities grown on just one carbon source can support up to 40 taxa thanks to metabolites recycle.
Furthermore, bacterial communities have different "diversities" depending on the taxonomic level are which they are considered. How can one choose the correct taxonomic scale ?
Even if the majority of our knowledge of microbial ecology has emerged from simple and controlled laboratory environments, the answer to these fundamental questions can be obtained by interrogating natural communities. In recent years, the human microbiome projects, and similar ambitious programs, have sampled and characterized an astronomical number of communities [54, 55]. Understanding the general patterns emerging in this communities is useful both at fundamental level, but also have important consequences for general health in the case of the human gut microbiome [56]. In the next section we introduce macroecology as tool for tackling this ambitious goal.

Macroecology

Ecological theory has developed a mathematical and universal flavour thanks to the contributions of Robert MacArthur. In the 60s, ecology was dominated by historical explanations, that are surely concrete but also limiting, as the same MacArthur claimed: *"Will the explanation of these facts degenerate into a tedious set of case histories, or is there some common pattern running through them all ?"*[37]. The coronation of this research program is *macroecology*.
Macroecology, is a branch of ecology that studies large scale statistical patterns of species abundance, diversity and interaction across communities [57]. In particular, by comparing statistical patters across very rich communities, macroecology tries to identify universal ecological patterns, also called *macroecological laws*. By connecting universal pattern to mechanistic models, macroecology is also able to disentangle the main ecological forces underpinning the maintenance of diversity. By using just a couple of words, macroecology can be summarized as "studying ecological communities as statistical ensambles".
Practically, macroecology works with the following three methodological steps:

1. Find a universal statistical pattern conserved across communities and eventually across environments.

2. Propose the simplest mechanistic model able to reproduce the pattern

3. Validate the model with other patterns and identify its limits of application.

Necessarily, the macroecological approach takes the hypothesis that in natural communities there exists a hierarchy of patterns that is reflected in statistical significance. A typical macroecological pattern that has been studied in many different communities, from tropical forest to microbes, is the Specees Abundance Distribution (SAD). The SAD measures how common or rare is a species relatively to other species in a community, and has been found to have some universal qualities [58].
On purely theoretical level, it is not possible to measure all the possible interactions present in a community, and hence a simplification is needed. There are two general class of models that study large communities: neutral theory [59], and random interaction models [60, 61, 62, 63].
Neutral theory is an ecological model that neglects differences in trophically similar species and consider just stochastic demographic processes, such as reproduction, death and migration, as the ecological forces present in the community [59]. It was developed by Hubbell ad collaborators by combining ideas from Kimura's neutral theory of evolution [64], MacArthur and Wilson theory of island biogeography [65] and Gould's ideas on null-model [66].
Even if its assumptions are very strong, Neutral theory is able to reproduce some macroecological indicators in tropical forests, such as the Species Abundance Distribution [67] and the community dynamics [68].
On the other hand, random interaction models consider that species interact randomly by treating the ecological network as "quenched disorder". This class of models are was born with May's approach in the resolution of stability-complexity debate [69], and was later generalized thanks to the physics of disordered systems [70, 63, 71] and random matrix theory [72, 61, 62].
Among ecologists, both neutral theory and random models, had polarizing effects: some ecologists were enthusiastic, while a large number firmly opposed it. On the other hand, these theories have attracted the interest of many statistical physicist for their general, macroscopic and stochastic approach [73, 74, 75, 76, 63].
In this thesis we will use principally models that combine the characteristics of the two

classes, i.e. models where disordered interactions amplify or are coupled with stochastic fluctuations [77, 63, 78]. Interestingly, in this class of models different possible ecological "phases" emerge, dominated either by neutrality or by niche competition [78, 79, 80]. In the next Section we introduce the basic mathematical tools to describe macroecology, and complex systems in general, while in Sec.1.6 we finally approach some recent results in microbial macroecology.

1.4 Complex systems

Physics reveals the *hidden* parts of nature. Strange instruments like microscopes, telescopes, spectrometers and thermometers are needed to extend human experience to the inaccessible scales where laws can be revealed.
Nevertheless, beyond all the technological apparatus, the fundamental instrument of physics is mathematics, both at a conceptual ad quantitative level. Without mathematics, physical laws are still hidden. As D.C. Krakauer [81] points out, laws can be hidden in space, time, or in both dimensions. The molecular structure of matter is hidden in space, given that it is detectable at small physical inaccessible scale. On the other hand, electricity and light are hidden in time, given that they travel at a speed not perceptible by our senses. Fundamental particles, and also stars, are hidden both in time and space. The first interacts with each other at a special time span not percetible by humans, while the second are confined enormously far away from us in space-time.
Yet, there are phenomena that are hidden in *plain sight*: complex systems, that are *" the convoluted exhibitions of the adaptive world- from cells to societies"* [81]. The economic growth of a city, the fractal structure of metabolic networks or the human microbiome are something that we perceive, directly or not, every day. Nevertheless, to make these phenomena intelligibile, a technological shift is not sufficient and a conceptual one is needed.
Its objects of study are systems composed by a large number of individual, or elements, from which collective properties can emerge. Philip Anderson, resumed this approach with the iconic sentence *"More is different"*[82], remarking that a complex system is more than the sum of its parts. *Complex systems* physics developed by applying concepts of condensed matter and statistical physics, like emergence, phase transitions and irreversibility, chaos, bifurcations, complex networks and stochastic fluctuations, to "non-physical" systems. In particular, non-equilibrium statistical physics is a field that studies stochastic and irreversible phenomena, being them at the level of molecules, fluids or biological systems. Complex systems, beyond being composed by many components, are generally *open*, i.e. interacting with an environment. The individual interactions, together with the environmental coupling, can drive the system away from equilibrium. In this thesis, we will mainly use the perspective of non-equilibrium statistical physics to describe and study the restless ecology and evolution of biological populations, in particular of bacterial one. Hence, let us gently introduce the main concepts of non-equilibrium statistical physics of complex systems.

1.5 Non-equilibrium dynamics of complex systems

Non-equilibrium statistical physics is a wide and idiosyncratic subject that crosses centuries, different theories and fields of application. This introduction is by no mean complete and general, and we indicate to the reader numerous books and references along the thesis. In general, with non-equilibrium we mean the state of system where currents between different states are present, being them of matter, energy, biomass, or simply probability. A first class of non-equilibria is the one of mascroscopic thermodynamic

systems driven out of equilibrium by boundary conditions, like fluids in contact with two particle reservoirs at different densities of coupled to many thermal baths [83, 84]. An other class of macroscopic non-equilibrium systems are the one characterized by the spreading of some quantity in space, like directed percolation, growing interfaces and the contact process [85] A third class is the one of *mesoscopic* phenomena, i.e. small systems characterized by energy differences of the order of $k_B T$. Examples are colloidal particles, macromolecules, molecular motors and chemical reactions at low density. Finally, complex systems, are out of equilibrium when probability currents between different states emerge. Examples range from ecological populations, to flock of birds and the human brain. For a general introduction see the recent books by Live and Politi [86], Gaspard [87] and the classic by Zwanzig [88]. For non-equilibrium phase transition and percolation see the book by Henkel, Hinrichsen and Lubeck [85] The book by Peliti and Pigolotti is the reference in stochastic thermodynamics [89] and see the book by Gallavotti for mathematical and conceptual foundations [90]. In this thesis we will encounter mainly examples of mesoscopic and complex systems. Luckily enough, the formalism of stochastic processes of finite degrees of freedom describes both groups. In Chapter 5 we will dive deeper into fundamental aspects of non-equilibrium statistical physics of mesoscopic systems, while here we will outline pragmatically the formalism of stochastic processes that will be used during the entire thesis.

1.5.1 Langevin and Fokker-Planck equations

As a general mathematical class of processes, we consider stochastic Markovian ones, i.e. processes without memory [91]. Consider the Markovian stochastic evolution of the N-dimensional vector $\vec{x}$ representing the state of a system, being it the position of a Brownian particle, the abundance of an ecological species or the density of a chemical compound. Its evolution is stochastic due to internal (activity, reproduction) or external fluctuations (interaction with a thermal bath or with an environment). Let's assume that the system evolves in time due to the contribution of a deterministic force $\vec{F}(\vec{c})$ together with a stochastic Gaussian noise ξ_i:

$$\dot{x}_i = F_i(\vec{x}) + G_i(\vec{x})\xi_i(t) \tag{1.1}$$

where

$$\langle \xi_i(t) \rangle = 0 \tag{1.2}$$
$$\langle \xi_i(t)\xi_j(t') \rangle = 2\delta(t-t')C_{ij}, \tag{1.3}$$

being C_{ij} the noise correlation matrix, where the symbols $\langle .. \rangle$ denotes the expectation value over the noise. The coefficient $G_i(\vec{x})$ represents the dependence of the fluctuations on the state of the system; when it is equal to one, the noise is called "additive", while it is "multiplicative" or state dependent in another case. Eq.(1.1) is called "Langevin equation" and was derived in the context of the physics of Brownian motion, and then generalized to general stochastic processes. In that original context, the noise modeled the interaction with a thermal bath, composed by smaller particles like water, while in complex systems it can represent demographic fluctuations in population, environmental noise in ecology or intrinsic fluctuations in stock market. For the moment, let us consider the noise as additive, i.e. $G_i = 1$ for all i in $[1, N]$, and let us rename C as $\hat{D}$, i.e. the diffusion matrix:

$$\dot{x}_i = F_i + \xi_i(t). \tag{1.4}$$

Note that in this simple additive case, by taking the average of Eq.(1.4) the deterministic behavior is obtained directly.
Due to the presence of the stochastic noise, the state of the system is not evolving deterministically, but stochastically, i.e. described by a cloud of probability $P(\vec{x},t)$ [92]. Equivalently to the Langevin equation, the system under consideration can be described by the motion of its probability distribution using the Fokker-Planck equation [93]:

$$\frac{\partial P(\vec{x},t)}{\partial t} = -\vec{\nabla}\cdot\vec{J}(\vec{x},t) \tag{1.5}$$

$$\vec{J}(\vec{x},t) = \vec{F}(\vec{x})P(\vec{x},t) - \vec{\nabla}\hat{D}P(\vec{x},t), \tag{1.6}$$

From Eq.(1.5) it is clear that the probability of a state x_i increases when the current is inflowing ($\vec{\nabla}\cdot\vec{J} < 0$) and decreases when it is out-flowing ($\vec{\nabla}\cdot\vec{J} > 0$).
If the deterministic dynamics presents a fixed point $\vec{x}^*$, i.e. a point where the force is zero $\vec{F}(\vec{x}^*) = 0$, the system will converge to a *stationary state*, i.e. a state where the probability does not depend on time

$$\partial_t P^*(\vec{x},t) = 0. \tag{1.7}$$

From Eq.(1.5) , it is clear that the stationary condition imposes:

$$\vec{\nabla}\cdot\vec{J}^* = 0 \tag{1.8}$$

The condition in Eq.(1.8) leave some freedom to the stationary current but imposes a geometrical constraint. In particular, one distinguish between the case $\vec{J}^* = 0$ and $\vec{J}^* \neq 0$. The zero current condition is called *detailed balance* and it implies that the force derives from a potential V :

$$\vec{J}^* = \vec{F}(\vec{x})P^*(\vec{x}) - \vec{\nabla}\hat{D}P^*(\vec{x}) = 0 \tag{1.9}$$

$$\vec{F} = -\vec{\nabla}\hat{D}\log P^* = -\vec{\nabla}V. \tag{1.10}$$

By consequence, the stationary state is an *equilibrium* one $P^* = P_{eq}$ (ESS). Note that a necessary and sufficient condition for detailed balance can be derived from Eq.(1.9) :

$$\partial_k\partial_i \log P^* = \partial_i\partial_k \log P^* \tag{1.11}$$

$$\partial_k \sum_{j=1}^{N} D_{ij}^{-1}F_j = \partial_i \sum_{j=1}^{N} D_{kj}^{-1}F_j, \tag{1.12}$$

that is a generalized potential condition involving $\hat{D}$ and $\vec{F}$. Finally, note that in the simple case of a diagonal diffusion matrix the equilibrium distribution is easily determined by the the force potential:

$$F_i = -\partial_i V, \qquad P_{eq} \sim \exp\left(-\int d\vec{x}\sum_i^N D_{ii}^{-1}F_i\right) = \exp\left(\sum_i^N D_{ii}^{-1}\partial_i V\right). \tag{1.13}$$

In appendix A, Sec.A.2 we discuss as an example the dynamics of a Browinin particle.
On the other hand, if the detailed balance condition, Eq.(1.9) or Eq. (1.11), is broken, the system will approach a stationary *non-equilibrium* state (NESS). While the equilibrium state is determined just by its stationary distribution P_{eq}, the NESS is described necessary by the couple formed by the stationary distribution and the stationary current (P^*, J^*) [94]. Differently from equilibrium ones, for non-equilibrium stationary state there is not a general derivation method and very few exact solutions exist.

In Chapter 6 we will investigate in deep the general characteristics of NESS and the geometric properties of their current

1.5.2 Multiplicative fluctuations

In complex systems, fluctuations are often multiplicative, i.e. modulated directly by the system state, $G_i = G_i(\vec{x})$. These type of fluctuations are caused both by intrinsic factors, or external ones, such as interactions with an environment, and are known to create very non trivial and purely stochastic effects.
A particular case, is when $G_i = x_i$, i.e. the so called *environmental* noise, or simply multiplicative noise:

$$\dot{x}_i = F_i + x_i\xi_i(t), \tag{1.14}$$

A typical physical example is the motion of a particle in a temperature gradient (see Chapter 5, Sec.5.4)[95, 96]. When multiplicative fluctuations affect a system with a deterministic fixed points, the stationary distribution of $P^*(x)$ presents long tails and exotic scaling laws, like Taylor law (see appendix B, Sec.A.3) [97, 98, 99]. Some examples of this phenomenon are environmental and demographic fluctuations in species abundance in ecosystems [100, 101, 73, 102], demographic fluctuations in social and urban phenomena [103, 104] and protein concentration fluctuations in cells [105].
Furthermore, multiplicative fluctuations can induce the presence of an *absorbing state*. In general, the stochastic evolution of system with multiplicative noise will tend to accumulate probability in regions experiencing low stochastic fluctuations. An examples is thermophoresis, i.e. a particle moving in temperature gradient tends to be localized in region with the lowest temperature [95, 96]. In systems modeling the creation and the elimination of individuals/particles, like chemical reactions and population models, multiplicative fluctuations can be either demographic, i.e. $G \sim \sqrt{x}$ modelling the intrinsic fluctuations of growth and death in a *finite* populations, or environmental, $G \approx x$ modeling, as said above, external fluctuations on growth. In both these situations the dynamics can get trapped in the state $x = 0$. This state is inescapable and fluctuation-free and hence called an *absorbing state* [85, 106, 107].
In the stochastic dynamics of spatial explicit processes, multiplicative fluctuations can cause totally new phenomenon as novel symmetry breaking phase transitions [108, 109, 110, 111], absorbing phase transitions [85, 112], spatio-temporal order [113], synchronization [114, 115] and even self-organization to a critical state [116, 117, 118]. Some examples are auto-catalytic reactions [119], percolation processes [120, 85], epidemic spreading [121], forest fires [122] and spatial ecology [73].
Multiplicative fluctuations are so intimately part of complex systems that they will appear in each Chapter of this thesis.

1.6 Macroecological laws of variation in microbial communities

Why are ecological communities so diverse in both species abundance and composition? Is variation due to the effect of deterministic forces like competition for ecological niches, or is it the results of stochastic environmental fluctuations? This is one of the most controversial issues in microbial ecology.
Recently, a macroecological approach has been put forward to answer to this fundamental question [101, 100, 123, 124, 80, 102]. A common result of these studies is the identifications of universal patterns of species abundance fluctuations across communities,

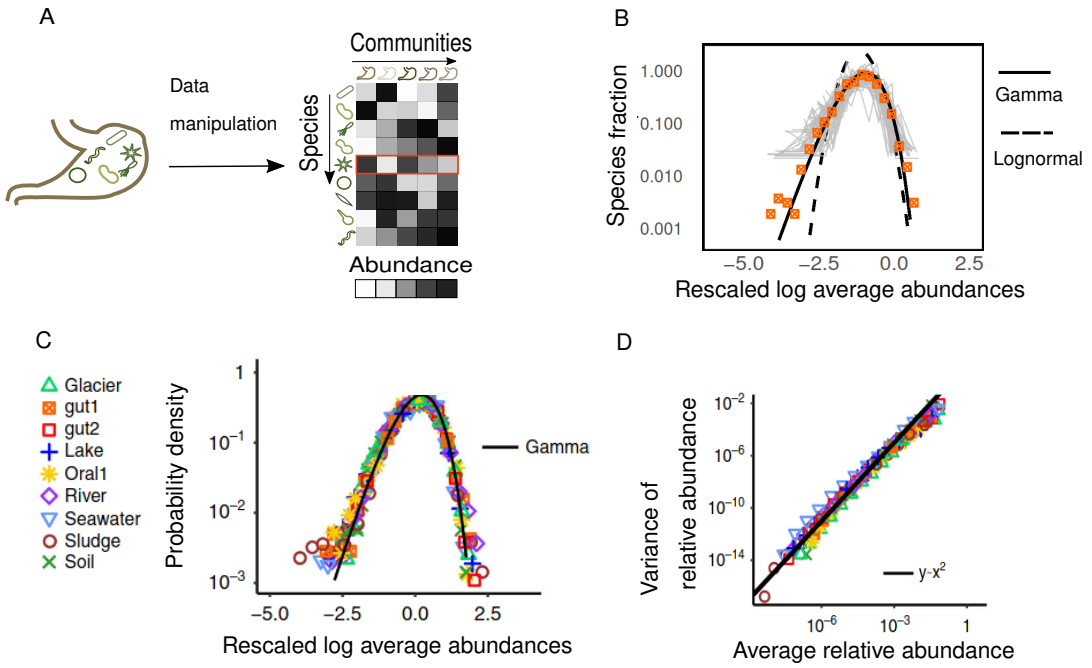


FIGURE 1.4: **Macroecological laws of variation in microbial communities** , figure adapted from [100]. A pictorial illustration of data collection and manipulation in microbial macroecology. Species abundances are sample by different communities (gut in the image) and organized in a table. Data are also manipulated to remove the effect of sampling. B, a row of the table represent the fluctuations of one species abundance across communities. By constructing an histogram of species abundances, rescaled by the mean, and averaging over species, one obtains the Abundance Fluctuation Distribution (gray points single species abundance, orange points average). The distribution is well fitted by a Gamma (solid black line), but not by a Lognormal (dashed line). C If one repeats this procedure across different biomes like, soil, lake, river etc (color legend), the Gamma AFD pattern is conserved. Finally, by comparing the species mean relative abundance and its variance, a robust scaling low emerges, known as Taylor law with exponent 2.

hence suggesting that similar ecological processes are present in different biomes. Furthermore, a common feature of these pattern is their remarkable *multiplicative* nature. In this section, we introduce two of these universal pattern in macroecological laws of variation in microbial communities and show how they can be reproduced by the Stochastic Logistic Model.
Consider the dataset depicted in Fig. 1.4 A, consisting of species (OTU) abundances in different communities organized in a table, where each row represent the abundances of different species, while the column stands for communities in consideration. For example, let us consider that the different communities are various samples of the gut composition of different persons. To understand how much the abundance of each species fluctuates across communities, we can select one row (a species), and plot its abundance histogram across communities (the value of the entire row). This distribution is called the species Abundance Fluctuation Distribution (AFD). In Ref.[100] it is shown that if in this procedures one removes the sampling effects from the datasets, and averages the AFD across species, a clear Gamma distribution emerges. In Fig. 1.4 B we show that the AFD in the gut biome is clearly a Gamma distribution, where the gray lines stand for the single species AFD, the orange points are the species average and the dashed and solid black line a fit with a Lognormal and Gamma distributions respectively [100]. Even more surprisingly, when this operation is repeated across different biomes, the Gamma AFD patterns is conserved (see colored points in Fig. 1.4] C).Finally, when the average and variance of species relative abundances are compared, as second Macroecological law emerges, see Fig. 1.4 D. In particular, the variance of species relative abundances is proportional to the square of the average, following a very omnipresent scaling law in complex system known as Taylor law with exponent two [97].
The fact that the AFD is a Gamma distribution has the non-trivial consequence that absent species have been just not sampled, and have not been eliminated by competitive exclusions. Indeed, Grilli shows that, by assuming the Gamma distribution with finite sample as AFD, the species occupancies are reproduced perfectly. In addition, thanks to the properties of the Gamma distribution and relation between average and variance, to describe each species it is necessary just one parameter, its mean abundance.
Finally, the two macroecological laws, with the addition of third universal regularity on the species mean abundance that is not relevant for this discussion, predicts other macroecological indicators, like the species abundance distribution (SAD). Hence, these two macroecological law suggest that the taxonomic scale of OTUs competitive exclusion is not fundamental and that environmental fluctuations are the dominant ecological force.
To prove this claim, Grilli introduced the Stochastic Logistic Model (SLM) for the dynamics of species abundance x_i:

$$\dot{x}_i = \frac{1}{\tau_i}\left(1 - \frac{x_i}{K_i}\right) + \sqrt{\frac{\sigma_i}{\tau_i}}\xi_i x_i \tag{1.15}$$

where ξ_i is a Gaussian white noise:

$$\langle \xi_i(t) \rangle = 0, \quad \langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t - t'); \tag{1.16}$$

and τ_i is the species growth timescale, K_i its carrying capacity and σ_i the strength of environmental fluctuations (cfr 1.5.2).
The model is composed by a deterministic part that consist of the classic *logistic* growth, and by a *stochastic* part in the form of environmental noise (see Sec.1.5.2). The logistic

equation is a classic equation of population growth in demography, economics and ecology (where it was introduced in 1920 by Pearl and Reed, independently of the work of Verlhust [37]). It is an effective model describing how populations reach a stable equilibrium due to the effect of competition. In ecological terms, τ_i represents the typical growth rate timescale of the species, given that its reciprocal is the actual growth rate, $\lambda_i = \tau_i^i$, and it is the typical time interval in which the population reaches the equilibrium. On the other hand, K_i represents the maximum number of individuals that the environment can support, given a finite number or resources. Therefore, K_i measures also the strength of *intraspecific* competition, and, effectively, encodes all the factors limiting the population, such as space, quantity of resources, effects of abiotic factors etc [3].
The stochastic part has the role of modelling the stochastic fluctuations affecting the population's growth, that can be induced by fluctuations in the concentrations of resources (biotic factors), or rapid change in the value of temperatures, pH etc (abiotic factors).
The deterministic logistic model admits a stable equilibrium population $x_i^* = K_i$, but when stochastic fluctuations are present, the population reaches a stationary state with abundance probability given by a Gamma:

$$P^*(x_i) = \Gamma\left(x|\frac{2}{\sigma_i} - 1, \frac{\sigma_i K_i}{2}\right) = \frac{1}{\Gamma(\frac{2}{\sigma_i} - 1)}\left(\frac{2}{K_i\sigma_i}\right)^{\frac{2}{\sigma_i}-\alpha} x^{\frac{2}{\sigma_i}-2} e^{-\frac{2x}{\sigma_i K_i}}, \tag{1.17}$$

where we have used the Itô discretization (see appendix A Sec.A.3 for the complete derivation). Therefore, the SLM is able to reproduce the first macroecological law. Furthermore, note τ_i has no influence on the stationary distribution and hence it depends just on two parameters (K and σ).
Consider the stationary mean and variance of the Gamma

$$\langle x\rangle = \gamma\theta = \left(\frac{2}{\sigma_i} - 1\right)\frac{K_i\sigma_i}{2}, \quad Var^* = \gamma\theta^2 = \left(\frac{2}{\sigma_i} - 1\right)\frac{K_i^2\sigma_i^2}{4}, \tag{1.18}$$

where we indicate with $\gamma = \left(\frac{2}{\sigma_i} - 1\right)$, $\theta = \frac{K_i\sigma_i}{2}$ distribution shape and scale parameter.
It is trivial to see that for each species the variance is proportional to the square of the mean:

$$\Sigma^* = \left(\frac{2}{\sigma_i} - 1\right)^{-1} \langle x\rangle^{*2} \tag{1.19}$$

and hence, if $\sigma_i = \sigma$ for all species, the second macroecological law, i.e. Taylor law, is reproduced.
Furthermore, in Ref. [100] it is shown that the macroecological laws are also valid in longitudinal time series of biome, as for example, the gut of one single person sampled each day for a year. This "ergodic" equivalence between averaging over different communities and over time is an additional clue on the dominant effect of environmental fluctuations in determining the communities variability.
Finally, let us comment that other studies have confirmed the validity of the SLM in various context such as gut microbiome time series [125, 123], strain dynamics [126] and in taxonomic coarse grained communities[124].
Nevertheless, the SLM, being a model without interactions, fails in reproducing the observed correlation between species abundances. Indeed, the Chapter 2 of this thesis is devoted to study microbial species correlations with a macroecological approach, combining abundances data with phylogenetic analysis (see Sec. 1.3.1), resulting in the discovery of a new macroecological law. Furthermore, we propose a new model able to

reproduce the observed empirical pattern, suggesting that coupled environmental fluctuations are the dominant force at the OTUs scale.

1.7 Bacterial eco-evolutionary dynamics

Darwinian evolution relies on the fundamental principles of reproduction, mutation and selection, and describes how populations change over time and how new forms can branch out from existing ones [13, 127]. The theory of evolutionary dynamics —whose aim is that of formalizing the ideas of Darwinian evolution from a conceptual and quantitative perspective— has become a wide and mature discipline at the crossroad between biology, mathematics and statistical physics (see e.g. [128, 129, 130, 131, 132, 133] and refs. therein). A central problem in evolutionary dynamics is the one of speciation. Historically, the dominant mechanism has been the one of allopatric speciation, suggesting that species are generated when a population is divided geographically in two different groups. Notwithstanding its importance, there are some speciation events, like Darwin's Finches in Galapagos island, where geographic population could not have happened [134].
Another evolutionary mechanism is sympatric speciation, i.e. the production of two species in the same space region caused by *ecological* interactions.
In general, ecological and evolutionary dynamics have been historically regarded as unfolding at broadly separated timescales. However, these two types of processes are nowadays well-documented to intersperse much more tightly than traditionally assumed, especially in communities of microorganisms.
One of the most successful theory in evolutionary dynamics is *Adaptive dyanmics* (AD) and was formulated to unify evolutionary dynamics with realistic ecological scenarios [135]. Adaptive dynamics is based on the following assumptions:

- Individuals reproduce clonally.
- Mutations are infrequent, and natural selection acts quickly. The population can be assumed to be at equilibrium when a new mutant arises.
- The number of individuals with the mutant trait is initially negligible in the large, established resident population.
- Phenotypic mutations occur in small but not infinitesimal steps.

Thanks to these assumptions, Adaptive dynamics consider ecological population at equilibrium, in which a single mutant can arise and eventually take over the entire population. Hence, in this theory, even if both evolution and ecology act on the population, they have very different timescales, and hence happen separately.
Importantly, AD allows for the possibility of "evolutionary branching" –i.e. the split of an initially monomorphic population in two diverse sub-populations— shedding light on how sympatric speciation could happen [136, 137, 135, 138, 139, 140, 141].
Similarly, phenomena such as the evolution of dispersal strategies [142], pathogenicity [143], metabolic preferences [144, 145] and multi-cellularity [146], to name but a few, have been successfully addressed within the context of AD. Moreover, extensions of AD have been developed to include ingredients such as finite-sized populations [147, 148], species interactions [149], sexual populations [139], multi-dimensional phenotypic spaces [150, 151, 152, 153, 154], variable environmental conditions [155], or variability in the evolutionary outcomes [**rcl**], to name but a few.

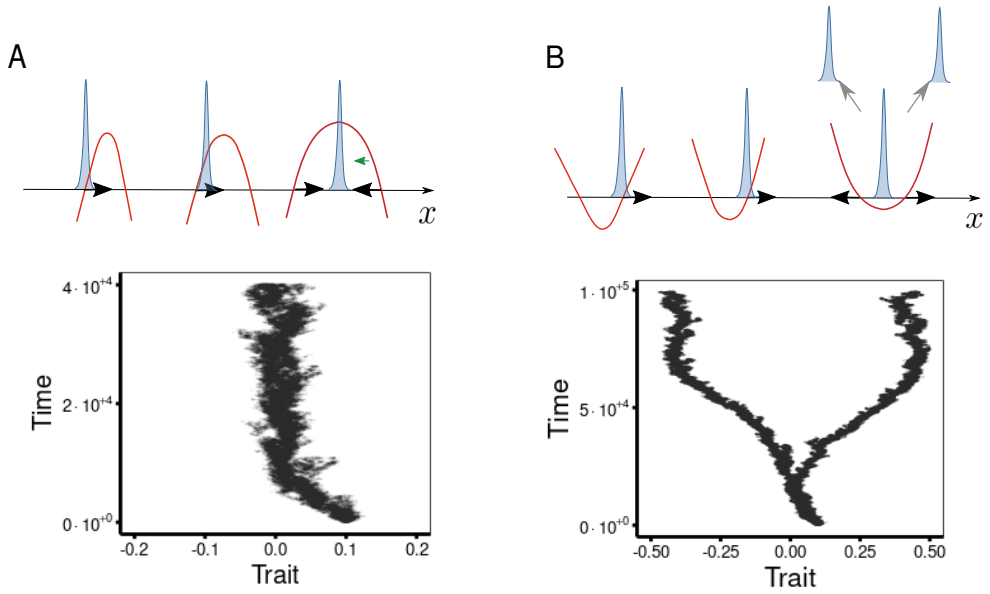


FIGURE 1.5: **Evolutionary stable state and evolutionary branching in adaptive dynamics**. A) *Top* If the invasion fitness has a single maximum, in the model when $\sigma_\alpha > \sigma_K$, the population climbs the fitness to reach it. This case is called "Evolutionary stable state" (ESS).*Bottom*: stochastic simulation of the model in the case of ESS, see B.3.5, Sec.C.5. (B) *Top* On the other hand, when the fitness presents a minimum, the population can get trapped there. To escape from it, evolutionary branching happens and the the population divides into two sub-populations. *Bottom*: stochastic simulation of the model in the case of ESS, see B.3.5, Sec.C.5.

In the next section we will present a paradigmatic model of evolutionary branching and derive the condition for evolutionary branching to happen [156, 157].

1.7.1 Evolutionary branching in adaptive dynamics

One of the most interesting predictions of AD is "Evolutionary Branching": a divergent event that gives rise to two phenotypical different sub populations[156, 158]. Here, we give an introduction to adaptive dynamics by showing how sympatric diversification can happen due to the interplay of growth and competition in a particular mode (see Ref. [156] for the complete derivation).
Let us consider a population characterized by a one dimensional trait x that determines the population capacity of exploiting the environment. This is modeled by the fact that a the population carrying capacity is a Gaussian peaked in a particular trait x_0 with a std. σ_K:

$$K(x) = K_0 \exp\left(-\frac{(x-x_0)^2}{2\sigma_K^2}\right) \tag{1.20}$$

$$\tag{1.21}$$

Furthermore, the population is monomorphic, i.e. all the individual have the same trait, and grows logistically :

$$\frac{dN(x)}{dt} = N(x)\left(1 - \frac{N(x)}{K(x)}\right) \tag{1.22}$$

where x is the phenotypic trait, $N(x)$ the population distribution in phenotypic space and $K(x)$ the carrying capacity distribution. We assume that the population equilibrium $N^*(x) = K(x)$ is reached very rapidly.
Now consider that due to a small mutation rate μ a mutant with trait y will arise and compete with the resident with a kernel $\alpha(x-y)$:

$$\alpha(x-y) = \exp\left(-\frac{(x-y)^2}{2\sigma_\alpha^2}\right). \tag{1.23}$$

Given that the mutants rise in very small numbers and the population is considered at equilibrium, $N(x) = K(x)$, w the dynamics of the mutants is given by:

$$\frac{dN(y)}{dt} = N(y)\left(1 - \frac{\alpha(x-y)K(x)}{K(y)}\right) \tag{1.24}$$

In Eq.(1.24) one can identify the per capita fitness of the mutants, with respect to the resident, as the *invasion* fitness:

$$\frac{dN(y)}{dt} = N(y)f(x,y), \qquad f(x,y) := 1 - \frac{\alpha(x-y)K(x)}{K(y)} \tag{1.25}$$

Now, of we consider the mutations to be small, we can expand the invasion fitness around $x \approx y$:

$$f(x,y) := (1-\alpha(0)) + \frac{\partial f(x,y)}{\partial y}\,|_{x=y}\,(x-y) + \frac{1}{2}\frac{\partial^2 f(x,y)}{\partial^2 y}\,|_{x=y}\,(x-y)^2 + O((x-y)^3), \tag{1.26}$$

and identify the selection *gradient* $D_f(x)$:

$$D(x)_f := \frac{\partial f(x,y)}{\partial y}\,|_{x=y} = \alpha'(0) + \alpha(0)\frac{K'(x)}{K(x)} \tag{1.27}$$

$$\frac{\partial^2 f(x,y)}{\partial^2 y}\,|_{x=y} = -\alpha''(0) + \alpha(0)\frac{K''(x)}{K(x)} - \frac{2(\alpha'(0)K(x) + \alpha(0)K'(x))K'(x)}{K^2(x)} \tag{1.28}$$

It can be shown that the change in population trait is given by the *canonical* equation [147, 159]:

$$\frac{dx}{dt} := \mu D_f(x), \tag{1.29}$$

indicating that the population tends to follow the fitness gradient. In the concrete case of the competition kernel and the carrying capacity given by Eqs.(1.20,1.23, where we set $x_0 = 0$) we have:

$$D_f(x) = \frac{K'(x)}{K(x)}. \tag{1.30}$$

Now, one can search for the fixed point x^* of the canonical Equation (1.29):

$$D_f(x^*) = 0 \to K'(x^*) = 0 \tag{1.31}$$

that hence are *extreme* points of the invasion fitness, in this case the maximum of the carrying capacity $x^* = 0$. Furthermore, one has to check the linear dynamical stability of

the fixed point as:

$$\frac{dD(x)}{dx} \mid_{x=x*} = \frac{K''(x^*)}{K(x^*)} < 0, \tag{1.32}$$

and the so-called evolutionary stability, i.e. that the stationary trait is a fitness maximum such that mutants cannot invade any more:

$$\frac{\partial^2 f(x,y)}{\partial^2 y} \mid_{x=y=x^*} = -\alpha''(0) + \frac{K''(x^*)}{K(x^*)} = \frac{1}{\sigma_\alpha^2} - \frac{1}{\sigma_K^2} < 0. \tag{1.33}$$

These results imply that the population converges always to the carrying capacity maximum, $x^* = 0$, given that it is an extreme point of the fitness. Once there, if the amplitude of competition is larger than the size of the niche, $\sigma_\alpha \geq \sigma_K$, it is an invasion fitness maximum and will be both dynamical and invasion stable ($d_x D_f(x^*) < 0$, $partial_y^2 f(x,y) < 0$), leading to an evolutionary stable state (ESS), see Fig. 1.5 A for illustration.
Nevertheless, if $\sigma_\alpha < \sigma_K$, x_0 it is an invasion fitness minimum, dynamically stable, $d_x D_f(x^*) < 0$, but not evolutionary stable $partial_y^2 f(x,y) > 0$. Hence, once that the population has reached the point $x = 0$, given that it is a fitness minimum, mutants can invade and the population. To escape from the minimum, the population splits into two diverging sub-populations. This the phenomenon called *evolutionary branching*, see Fig. 1.5 B for illustration. In the figure, we also present an individual based simulation of the model, more details will be given in Chapter 3. In resume, adaptive dynamics follows the fitness gradient, in this case reaching the maximum of the carrying capacity. When this attractor is a fitness minimum the population splits in two branches (disruptive selection). This can happen only when the competition kernel is more narrow than the carrying capacity distribution.

1.7.2 Extending adaptive dynamics to bacterial evolution

As discussed in the precedent section, AD has some restrictive assumptions that can be resumed in the timescale separation of evolution and ecology, and hence there is space to generalize AD to full eco-evolutionary scenario, allowing for example for frequent mutations at ecological timescales. Indeed, problems in which ecological and evolutionary changes occur at similar timescales and feedback into each other are ubiquitous in microbiology. Example range from from the emergence of antibiotic resistance and tolerance, to emergence of metabolic strategies.
A possible obstacle relies in fact that microbial evolution has been studied using population genetics, thanks to technological advances enabling the design of high-precision and long-term evolutionary experiments providing access to genetic information and global-fitness measurements of whole populations [11]. In this context, recent efforts have allowed to generalize classic population genetics models to rapid evolution, using e.g. the formalism of fitness travelling waves as well as concepts from statistical physics [160]. On the other hand, microbial phenotypic eco-evolution —that was traditionally left aside owing to the difficulties in measuring single-cell traits [161, 162]— has received reinvigorated attention [163, 145, 164, 165, 166, 167], as a result of technological advances in determining single-cell traits [168, 169, 170] and metabolic functions [171, 172, 45, 173]. A particularly relevant problem is that of the emergence of tolerance to antibiotics by lag, that has been recently shown to emerge very fast in bacterial populations under controlled laboratory conditions [10]. Hence, advancing in the development of mathematical and computational approaches to shed novel light onto eco-evolutionary problems is a challenge of utmost relevance.

The ecological and evolutionary dynamics of large sets of individuals can be naturally addressed from a theoretical perspective using ideas and tools from statistical mechanics. Following this tradition, in Chapter 3 we construct a eco-evolutionary framework for bacterial evolution allowing us to derive "macroscopic" evolutionary equations from a rather general "microscopic" stochastic dynamics representing the processes of reproduction, mutation and selection in a large community of individuals, each one characterized by its phenotypic features.
Furthermore, in Chapter 4 we apply this framework to the relevant problem of the evolution of antibiotic tolerance in bacterial populations.

1.8 Irreversibility in complex systems

The fundamental characteristic of non-equilibrium systems is the breaking of time reversal symmetry, and hence the emergence of irreversibility. Indeed, using statistical physics, irreversibility can be defined as a measure of how much a system breaks the time-reversal symmetry and connected to dissipation in thermodynamic systems. Here we give a very brief introduction of the argument and refer to part III an reference for more details.
Phenomenological non-equilibrium thermodynamics can analyze irreversible phenomena, such as thermo-electric effects, by studying the local change in entropy with the so-called balance entropy equation [174, 175]. In this scenarios, the entropy of a volume element can change due to two contributions. First, there could be a flux of thermodynamic entropy across its boundary, being it positive or negative. On the other hand, the entropy can increase due to the presence of an irreversible phenomena in its interiors that "produce" entropy. This balance equation can be generally represented as:

$$\dot{S} = \dot{S}_{prod} + \dot{S}_{flux}, \quad \dot{S}_{prod} \geq 0, \tag{1.34}$$

where $\dot{S}$ is the change in entropy, while $\dot{S}_{prod}$ and $\dot{S}_{flux}$ are the entropy production and flux. The entropy production is always non-negative due that irreversible phenomena always increase the entropy for the second law of thermodynamics.
Nevertheless, statistical mechanics is needed to related thermodynamic dissipation with irreversibility by constructing a bridge between microscopic mechanical motion and macroscopic thermodynamics.
In classical macroscopic systems, like a molecular gas or a fluid, the microscopic equation of motion are time-reversible, while irreversibility emerge at the macroscopic scale, for example with the appearance of viscosity in the Navier-Stokes equations [90, 176]. Onsager determined that thermodynamic process need to obey some "reciprocal" relations, that emerge at the macroscopic scale as a consequence of the microscopic time-invariance. More generally, Gallavotti and Cohen formulated the celebrated fluctuation theorem, that established a quantitative and parameter free symmetry relation between the stationary state probability of observing a value of the average entropy production rate and its opposite. Hence, the fluctuation theorem provides a direct relation between entropy production an irreversibility in ergodic systems.
In recent years, the theory of stochastic thermodynamics has emerged as non-equilibrium thermodynamic framework for mesoscopic systems that unifies the thermodynamic of irreversible phenomena and non-equilibrium statistical physics [89]. Indeed, stochastic thermodynamics shows that the entropy production appearing in balance Eq. (1.34) derives from the breaking of time symmetry.
This can be easily seen by considering a system described by the Fokker- Planck Eq.(1.5)

in one dimension. One of the main assumptions of Stochastic thermodynamics is the validity of the Shannon information entropy as a thermodynamic quantity that indicate the uncertainty of the system:

$$S = -\int dx P(x) \log P(x). \tag{1.35}$$

Here we will accept pragmatically this assumption, see Ref. [177] for a detailed criticism. By using the Fokker-Planck equation, one can study the entropy dynamics:

$$\dot{S} = -\int dx \partial_t P \log P = \int dx \partial_x J \log P, \tag{1.36}$$

and by integrating by parts and assuming the surface term to vanish, one obtains:

$$\dot{S} = -\int dx J \partial_x \log P. \tag{1.37}$$

Then, by inverting the current definition Eq. (1.5) , one obtains the following relation:

$$\partial_x \log P = \frac{F}{D} - \frac{J}{DP}, \tag{1.38}$$

that inserted in Eq. (1.37) gives a decomposition in two terms:

$$\dot{S} = \int dx \frac{J^2}{DP} - \int dx \frac{JF}{D} = \dot{S}_{prod} - \dot{S}_{flux}. \tag{1.39}$$

The first term can be identified with the entropy production,

$$\dot{S}_{prod} = \int dx \frac{J^2}{DP} \geq 0, \tag{1.40}$$

while the second is the the entropy flux:

$$\dot{S}_{flux} = \int dx \frac{JF}{D}. \tag{1.41}$$

It is clear that $\dot{S}_{prod} \geq 0$, being it is zero just when detailed balance is respected $J = 0$. Furthermore, note that at the stationary state $\dot{S}^* = 0$, and hence the entropy production and flux coincide $\dot{S}^*_{prod} = \dot{S}^*_{flux}$. Similarly to what is done in non-equilibrium statistical mechanics, he entropy production can be estimated directly as measure of irreversibility by using the path-integral method (or in a Master-equation formalism) by uncovering a fluctuation theorem for stochastic dynamics [178, 89]. Even if we do not enter in mathematical details here (see Sec. 5.1 for a brief derivation in stochastic thermodynamics) let us mention the essence of the relation between entropy production and irreversibility. Consider the log of the ration between the probability of a trajectory Γ, and the the probability of the time-reversed one Γ^T, together with the probabilities of the respective initial state. When the average over trajectory and space, one obtains the entropy production:

$$\left\langle \log\left(P(x_0) \frac{P(x_1)P(\Gamma)}{P(\Gamma^T)} \right) \right\rangle_{\Gamma,x,x_0} = \dot{S}_{prod} \tag{1.42}$$

For this reason, the entropy production represents the irreversibility of the system, and hence it has been identified with the second law of thermodynamics.

It is known that thermodynamic systems approach equilibrium by maximizing the Entropy [179]. Is it possible to derive a similar principal for non-equilibrium conditions ? Prigogine's has been one of the most important, and at the same time controversial, researcher in non-equilibrium thermodynamics. In particular, he formulated the disputed Minimum entropy production principle, stating that non-equilibrium systems tend to minimize the entropy production at the NESS. While the assertion is true near equilibrium, its validity far away from equilibrium it is false [180] . In Chapter 5 we will introduce a geometric framework for stochastic thermodynamics enabling us to interpret a particular thermodynamic quantity called the *excess entropy* as a generalization of Prigogine's principle . In particular, we will explore how dissipation induces a chiral symmetry breaking phenomenon in thermodynamics currents during the approach of a NESS.
In stochastic thermodynamics, the entropy production is related to the total dissipation, while $\dot{S}$ to the work performed on the system and $\dot{S}_{flux}$ to the dissipation into the thermal environment, see Sec. 5.1 for more details . In a general complex system, $\dot{S}_{prod}$ represents irreversibility, but it is not directly related with any thermodynamics quantity, and it just quantifies how away from equilibrium is the system. By inverting Eq.(1.39), we interpret also the entropy flux in terms of irreversibility:

$$\dot{S}_{prod} = \dot{S} + \dot{S}_{flux}, \tag{1.43}$$

From these expressions, one can interpret $\dot{S}$ as the part of irreversibility generated by the system dynamics, while $\dot{S}_{flux}$ is a "structural" irreversibility due to the presence of currents and forces in the system. This interpretation can be made more clear when the dynamics of the system has been derived by a microscopic Master equation [89, 181].
In recent years, there is a growing interest in this topic and irreversibility have been studied in many complex systems, like the human brain [182, 183], adaptive evolution [184, 185], active systems[186, 187] and flocks of birds [188, 189].
In Chapter 6 we will follow this ideas and estimate the irreversibility of adaptive evolution in a general way.

Part I

Microbial ecology

Chapter 2

A macroecological law of species interactions in microbial communities

2.1 Introduction

Microbial communities are ubiquitous on earth, from human microbiota, to ocean, soil, and glacial environments [34]. Their widespread presence is paralleled by their complex and highly variable composition, both across time and space [32]. Understanding what are the main drivers, or *"ecological forces"*, shaping the coexistence and stability of microbial communities under changing environmental conditions and perturbations is a fundamental challenge of utmost relevance for, e.g., environmental and health sciences.

Ecological forces can emerge from the interactions between species or between species and the environment, including both biotic and abiotic factors. Experiments in simple and controlled laboratory environments have made it possible to trace the effects of various ecological forces on community composition, often reshaping classical ideas on ecological interactions [190, 191, 192, 193, 194, 195]. For instance, cross-feeding has emerged as a central player in determining community assembly and species coexistence [38, 40]. However, the precise role of different ecological forces in determining composition and variation in more complex natural communities remains mostly unknown. While detailed information about environmental [196, 55, 197] and genetic [198, 199, 200] factors shaping interactions and responses to environmental conditions is sometimes available, we still lack frameworks to infer their quantitative strength and to disentangle the relative relevance of each of the acting ecological forces from available data [201, 202, 203]. Macroecology —i.e. the study of ecological communities through the analysis of global patterns of abundance, diversity, and distribution [57]— stands as a prominent approach to link quantitative ecological models with empirical data of complex and diverse communities [205, 101], see Sec.1.3.2. In particular, in the context of microbial communities, a growing body of evidence reveals that the abundance dynamics observed in microbial communities is characterized by distinctive and reproducible statistical patterns, also known as *macroecological laws* [101, 206, 207, 208, 123]. As we have discussed in Sec.1.6, further evidence shows that, despite the complexity of the underlying "microscopic" dynamics, most of such patterns can be reproduced by relatively simple models —such as e.g. the stochastic logistic model (SLM)— capturing salient features of the underlying ecological forces [206, 207, 208, 123, 209]. However, such simplified models often neglect interactions between species, treating their abundance fluctuations as independent from each other, so that they cannot account for species-correlation patterns. Nevertheless, it

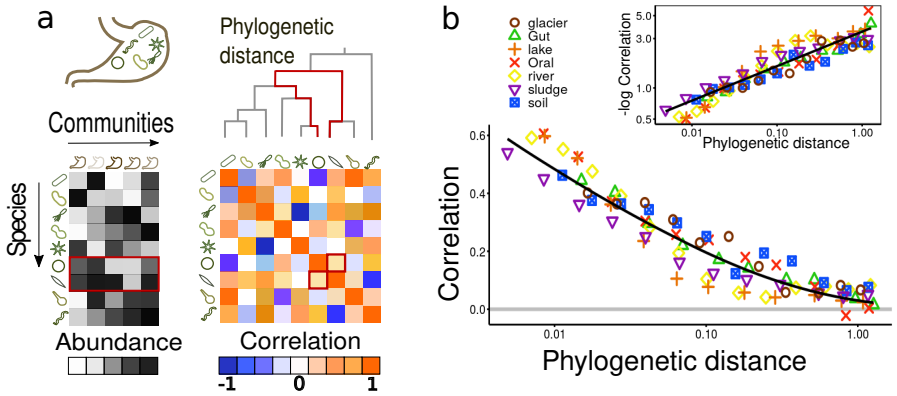


FIGURE 2.1: **(a) Pictorial illustration of the data organization and statistical analyses**. Abundances of different species (i.e. OTU at 97% similarity [204]), for different communities of the same biome (e.g gut of different hosts) are collected, respectively in rows and columns of the left table. The grey scale in the matrix entries stands for the level of abundance with darker shades corresponding to more abundant species. The (symmetric) species-abundance correlation matrix (color coded) is obtained by calculating for each pair of existing species the correlation of abundance fluctuations across communities. Finally, the phylogenetic distance is computed for all possible pairs of species by reconstructing the phylogenetic tree; and than associated with the couple correlation. The abundances, correlation, and phylogenetic distance of two given species are emphasized in red color. **(b) Macroecological law for pairwise correlations as a function of the phylogenetic distance for different biomes.** The correlation of abundance fluctuation averaged over all couples within a given discretized distance bin (colored symbols) decays with the phylogenetic distance (log scale) for all the considered microbiomes (see legend). In particular, each bin in the x-axis includes all couples with a phylogenetic distance within such a discrete bin (each one including at least 10^3 couples for each of the 8 considered biomes; as shown in the appendix B Fig. B.1, the pairs are not uniformly distributed across phylogenetic distances: the vast majority of couples lie in the rightmost bins, with large distances and small pairwise correlation values). The black line represents a stretched-exponential decay, Eq.(B.14) with $\lambda \approx 3.5$. To emphasize the functional dependence, the inset shows the same data but for the negative log of the correlations represented in double logarithmic scale (i.e., a plot in which stretched exponential functions become straight lines; in this case with slope 1/3).

is noteworthy that including species interactions in models such as the SLM does not significantly affect the shape of single-species macroecological patterns. For instance, generalized Lotka-Volterra equations with environmental stochasticity —which reduce to the SLM in the absence of interactions— predict time-series statistics and patterns similar to those of the SLM [208, 207, 123].

On the other hand, it seems clear that the ecological forces shaping community composition and variability can only be unveiled within a macroecological approach by explicitly studying multispecies abundance patterns. For instance, empirically-determined pairwise correlations between species abundances can be partially explained by consumer-resource models with resource fluctuations [209].

One challenge in connecting empirical macroecological patterns with simple yet biologically-grounded models is that not all statistical patterns are equally informative. For instance, it is well known that, in many ecological systems, the empirical shape of the species abundance distribution (SAD) —i.e. one of the most prominent macroecological patterns— can be reproduced by models with very different underlying biological assumptions (such as, e.g. neutral and niche theories [210, 211]). Similarly, multiple mechanisms are expected to determine the observed correlations between species abundance fluctuations.

Pairwise correlations are in fact the result of multiple ecological forces, such as competition, cooperation, cross-feeding, but also of indirect effects through a network of interactions [212]. Analyzing the phylogenetic structure of community composition [213, 214] is a standard approach to disentangling the effects of these alternative assembly mechanisms. This type of approach is generally applied to analyze species (co-)occurrence. For example, shared environmental fluctuations (called "*environmental filtering*" hereon) produce phylogenetic clustering (i.e., similar species share a tendency to be simultaneously present or absent [215]), while exclusion by limiting similarity determines phylogenetic overdispersion (i.e., similar species tend not to be simultaneously present). This type of phylogenetic approaches has been widely applied in plant communities as well as in other systems [216, 217, 218] including, in particular, microbial communities [219]. More generally, phyloecology, which combines phylogenetic relationships with community ecology, has the potential to reveal the processes determining community composition [220, 221]. However, with few notable exceptions focusing on testing neutral models [222, 223], a connection between empirical observations of community ecology based on phylogeny and quantitative predictions of theoretical models is still missing.

Here, our goal is to develop such a connection under the lens of macroecology. In particular, we first elucidate the existence of a new empirical macroecological law that associates pairwise abundance correlations with species phylogenetic similarity. To rationalize such a finding, we formulate three alternative theoretical models —each relying on different ecological forces— all of which reproduce previously studied single-species macroecological patterns [207, 208, 123], but lead to radically different predictions for phylogenetic-dependent correlation patterns. These analyses allow us to conclude that only *environmental filtering* (and not, e.g., species competition) explains the empirically-observed pattern of decaying correlations with phylogenetic distance. Last but not least, we analyze temporal data for a fixed community, showing that the macroecological law also holds quantitatively in this context and that delayed temporal correlations are naturally reproduced by our model with environmental filtering.

2.2 Universal pattern of correlation versus phylogenetic distance

We consider the phylogenetic (or cophenetic) distance, $d_{G,ij}$ (where the subindex G stands for "genetic") for each pair of operational taxonomic units (OTUs) (i, j), by using publicly-available results from 16S ribosomic RNA analyses [224, 204], see Sec.1.3.1 for a brief introduction. This genetic distance exhibits a broad variability across OTU pairs (see appendix B Fig. B.1). For each pair of OTUs, we measure the correlation between the corresponding abundance fluctuations η_{ij} across samples (see Fig. 2.1(a) and Methods). Fig. 2.1(b) illustrates the value of the pairwise correlation η, averaged over the pairs of OTUs at a given phylogenetic distance (where distances are grouped into discrete intervals or bins) for diverse biomes. Remarkably, the resulting averaged correlation is found to decay with the phylogenetic distance, d_G, in a robust way across environments and datasets. In particular, phylogenetically close OTUs (small values of d_G) display a significant positive correlation while, for distant OTUs, the average correlation decreases to zero. We compare this observation with randomized data, obtained by shuffling the position of OTUs on the phylogenetic tree. Such a randomization tree preserves both the statistical properties of the abundances and the property of the tree while removing the relation between the two. The comparison with the randomizations allows us to show that the positive correlations at low phylogenetic distances are significantly higher than what expected by change. We confirmed the robustness of this empirical observation by changing the metric to quantify abundance pairwise correlations, obtaining in all cases

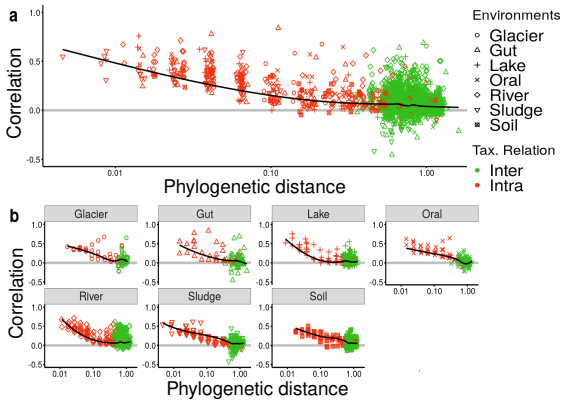


FIGURE 2.2: **Taxonomic Analyses.** a: Correlation between abundance fluctuations versus phylogenetic distance for intra-phylum (red points) and inter (green points) phyla. In total, 29 phyla are considered and each point represents the correlation of one of them, within a certain phylogenetic distance, in a particular biome (shapes). Black lines are averages over both taxas and biomes, weighted by abundances in each considered bin. (b) Same data as in (a) above but plotted separately for each biome.

similar decaying correlation patterns (see Appendix B, Fig. B.2,B.3). At a more quantitative level, the reported decay of the correlation function is well captured on average by a stretched-exponential function [225]:

$$\eta(d_G) = e^{-\lambda d_G^{\chi}}, \tag{2.1}$$

where $\chi \sim 1/3$, as shown in Fig. 2.1(b), so that the decay of the correlation function is slower than exponential. The value of χ and the goodness of fit of the functional form of Eq.(B.14) have both a small degree of variation across biomes. In particular, the best fits of the exponent χ across all the considered biomes —always in the range $0.2 - 0.4$— are reported in sec.B.3.2 of appendix B; the explanation of this variability goes beyond to goals of this manuscript. Let us mention that, we also explored alternative functional forms (e.g. exponential and power-law) for the decay curves (see Tables B3 and B4 in appendix) and observed that, overall, the stretched exponential is the one providing the best description of the patterns. Neverthelesss, note that this is only a phenomenological fit, as we lack a mechanistic understanding of the functional form of the decay. Let us finally remark that the value of $\lambda \approx 3.5$, corresponds to a typical distance for the decorrelation of abundance fluctuations of $d_G \approx 0.3$. To give an approximate taxonomic point of reference, this distance roughly relates the typical distance of species within the same class or order. In order to scrutinize whether the observed pattern is consistent across the phylogenetic tree, we repeated the same type of analyses at the coarser level of taxa, comparing correlations within and between taxonomic orders. Fig. 2.2shows that species from different phyla (i.e. at large phylogenetic distances, see Sec.1.3.1) tend to have, on average, vanishing correlations, while the averaged correlations within the same taxa decay from positive to zero with phylogenetic distance, recovering the pattern in Fig. 2.1 in a consistent way in the vast majority of the observed taxa (see Fig. B.8-B.10). Small deviations to this general pattern appear to be due to specific taxa. In the appendix B we explore the case of soils where a couple of orders are the main drivers of the observed departure from the macroecological law (see Fig. B.9).

These results suggest that the observed correlation pattern showing a stretched-exponential decay with phylogenetic distance is a universal one, not depending on the considered ecological context nor on a particular taxa. Whatever ecological forces are at the origin

of such species-abundance correlations, they manifest themselves regularly and consistently across taxa and environments.

2.3 Models of ecological forces in preference space

Which ecological forces are responsible for the described pattern of abundance correlations across communities? In general, ecological interactions in a network of species could *a priori* create both positive and negative correlations, stemming from both direct pairwise interactions and network effects.

To unravel these conflicting mechanisms, we consider a general population-dynamic model where species grow and may compete for resources in a fluctuating environment. To be more specific, we consider a set of N species whose growth is coupled with a set of R *population-dependent* factors (e.g., resources consumed by the populations) as well as on M other *population-independent* factors (such as resources subject to high influx fluctuations, temperature, pH, salinity, etc.). The key difference between these two sets, is that population-dependent factors are affected by population abundances through consumption and, thus, resource availability explicitly depends on population densities. This dependence (which can be derived from consumer-resource models, as shown in the next section) induces an effective competition. On the other hand, population-independent factors display fluctuations which are mostly insensitive to population densities. Furthermore, we assume that all these $R + M$ factors, both population-dependent and population-independent, are characterized by temporal fluctuations. Population-dependent factors can be interpreted as nutrients, which are depleted by consumption. Population-independent factors could be abiotic factors or nutrients whose dynamics is not governed by population densities. In particular, in Section 2.4 , we explicitly consider a consumer-resource model showing that, with some approximations, resources with high influx rate but low temporal variability behave as population-dependent factors, while the ones with relative low influx but high variability are effectively population-independent.

We consider that the effect of the environment on a given species can be represented as a vector in an abstract preference space, representing its preferences for the R available population-dependent resources and growth responses to the M population-independent factors. Thus, the set of resource preferences characterizing each given species i is represented by a vector $\boldsymbol{b}^i$ in an abstract R-dimensional space of population-dependent factors and by another vector $\boldsymbol{a}^i$ in a M-dimensional space of population-independent factors (see Fig. 2.3(a) for a pictorial illustration). The precise way in which these vectors are generated is described Methods, where diverse algorithms are designed for the cases of high and low-dimensional preference spaces, respectively.

The *per-capita* growth rate of each species is influenced by population-dependent and population-independent factors, both weighted by the corresponding preference vector leading to the following general model

$$\frac{1}{x_i(t)}\frac{dx_i}{dt} = \sum_{\beta=1}^{R} b_\beta^i R_\beta(t) + \sum_{\alpha=1}^{M} a_\alpha^i M_\alpha(t) - \delta \,, \tag{2.2}$$

where $x_i(t)$ is the abundance of species i at time t, $R_\beta(t)$ is the value of population-dependent factor β at time t, $M_\alpha(t)$ is the value of the population-independent factor α, and δ is a constant death rate. This description naturally applies when both population-dependent and population-independent factors can be interpreted as resources, thereby affecting additively the growth rate. As already mentioned, Section 2.4 shows explicitly

how this description emerges from more standard consumer-resource models [226, 227] and how the distinction between population-dependent and population-independent resources stems from their mean values and the amplitude of the fluctuations of their corresponding influxes.

On the other hand, Eq.(2.2) is not adequate to describe the effect of abiotic factors (such as temperature or pH) which are known to affect the growth rate, but not as additive terms; in other words, the growth rate should vanish in the absence of resources, independently of the effect of such abiotic factors. Indeed, if the factors $M_\alpha(t)$ are interpreted as abiotic factors, a more biologically grounded model is

$$\frac{1}{x_i(t)}\frac{dx_i}{dt} = M_i(t)\sum_{\beta=1}^{R} b_\beta^i R_\beta(t) - \delta\,, \tag{2.3}$$

where the effect of abiotic factors influences multiplicatively the growth rate; i.e. they modulate the growth. Clearly, mixed models between Eq.(2.2) and Eq.(2.3), with some population-independent factors affecting additively and other multiplicatively the growth rate are also possible. In the following, without loss of generality, we consider the model of Eq.(2.2) as it leads to more tractable equations. In section 2.4, we explicitly consider Eq.(2.3) and show that all the results obtained in the additive scenario also apply also for the multiplicative one.

We assume the population-independent factors to be subject to stochastic fluctuations

$$M_\alpha(t) = \bar{M}\left(1 + \sqrt{\nu}\zeta_\alpha(t)\right)\,, \tag{2.4}$$

where $\bar{M}$ represents a baseline level, $\zeta_\alpha(t)$ is a (zero-mean unit-variance) Gaussian white noise and the parameter ν quantifies the strength of fluctuations. Similarly, the abundance R_β of each population-dependent resource β fluctuates in time and is reduced by consumption, i.e.,

$$R_\beta(t) = \bar{R}\left(1 + \sqrt{\omega}\varphi_\beta(t) - \gamma\sum_{j=1}^{N} b_\beta^j x_j\right), \tag{2.5}$$

where $\bar{R}$ is a baseline level, γ the consumption timescale, $\varphi_\beta(t)$ a (zero-mean unit-variance) Gaussian white noise, and the parameter ω quantify the amplitude of fluctuations. The choice of Gaussian fluctuations should not be interpreted as an assumption on the shape of empirical resource-fluctuation patterns, which are most likely non-Gaussian (e.g., in the gut microbiome nutrients arrive in batches). It should instead be considered as a coarse-grained description, emerging over longer timescales (e.g., akin to the diffusion limit [92]). In any case, it can be retained as a first step towards the design of more detailed microbiome-specific models.

The model defined by Eq.(2.2) can be effectively written as a generalized Lotka-Volterra model with species competition and fluctuating growth rates:

$$\frac{dx_i}{dt} = x_i\left(\bar{r}_i + \sqrt{\sigma}\xi_i(t) - \sum_{j=1}^{N} C_{ij}x_j\right)\,, \tag{2.6}$$

where $r_i(t) = \bar{r}_i + \sqrt{\sigma}\xi_i(t)$ is a time-dependent and fluctuating growth rate with mean value $\bar{r}_i$; the entries of the competition matrix, C_{ij}, are determined by the overlap in population-dependent factor preferences of species i and j

$$C_{ij} = \bar{R}\gamma\,\boldsymbol{b}^i\cdot\boldsymbol{b}^j, \tag{2.7}$$

and, finally, the noise covariances are $\langle \xi_i(t)\xi_j(t')\rangle = \rho_{ij}\delta(t-t')$ with

$$\rho_{ij} = \frac{\nu \bar{M}^2 \boldsymbol{a}^i \cdot \boldsymbol{a}^j + \omega \bar{R}^2 \boldsymbol{b}^i \cdot \boldsymbol{b}^j}{\nu \bar{M}^2 + \omega \bar{R}^2}. \tag{2.8}$$

The explicit expressions of $\bar{r}_i$,s and ξ as function of the original parameters are:

$$\bar{r}_i = \bar{R} \sum_\beta b^i_\beta + \bar{M} \sum_\alpha a^i_\alpha - \delta \tag{2.9}$$

$$\sqrt{\sigma}\xi_i(t) = \sqrt{\omega}\bar{R} \sum_\beta b^i_\beta \varphi_\beta + \sqrt{\nu}\bar{M} \sum_\alpha a^i_\alpha \zeta_\alpha \tag{2.10}$$

$$\sigma = \nu \bar{M}^2 + \omega \bar{R}^2. \tag{2.11}$$

Observe that growth-rate correlations depend both on population-dependent and population-independent factors ($\boldsymbol{b}$ and $\boldsymbol{a}$) while the effective competition is mediated only by shared population-dependent ones ($\boldsymbol{b}$). In all the variants of the model considered here (A, B and C) only one set of preference vector is needed (either biotic for models A and B, or abiotic for C). Thus, one can quantify the preference similarity or, simply, the "*preference distance*" between species i and j, as the cosine distance between their relevant preference vectors (for simplicity, in the following, we restrict the notation to model C for which abiotic preferences are relevant). The preference distance is defined as:

$$d_{P,ij} \equiv \frac{2}{\pi}\theta = \frac{2}{\pi} \arccos\left(\frac{\boldsymbol{a^i} \cdot \boldsymbol{a^j}}{|\boldsymbol{a^i}||\boldsymbol{a^j}|}\right) = \frac{2}{\pi} \arccos\left(\frac{\boldsymbol{a^i} \cdot \boldsymbol{a^j}}{r_P}\right), \tag{2.12}$$

where the sub-index P stands either for "preference".

Thus, both types of coupling terms in the effective Lotka-Volterra population dynamics —i.e. the elements of the species competition matrix and those of the growth-rate fluctuation covariance matrix— crucially depend on the set of species similarities in preference space. However, depending on the relative strength of both types of couplings, as well as on the distribution of preferences vectors, one can define three different models, depending on which are the dominating ecological forces (see Fig. 2.3 a/b/c):

(A) *Shared fluctuating population-dependent resources.*
If population-independent fluctuations are negligible (i.e. $\nu = 0$), species interactions are determined by a combination of the effect of competition (encoded in the entries C_{ij}) and resource-abundance fluctuations (encoded in the entries ρ_{ij}), which are both proportional to the species resource-preference overlap: $\boldsymbol{b}^i \cdot \boldsymbol{b}^j$.

B) *Shared population-dependent resources and non-overlapping fluctuating population-independent factors.*
If resource fluctuations are negligible (i.e., $\omega = 0$) and population-independent factors preferences are all orthogonal to each other, species experience independent growth rate fluctuations ($\rho_{ij} = \delta_{ij}$) and competition for fixed resources (through the coupling matrix C_{ij}).

(C) *Shared fluctuating population-independent factors with fixed non-overlapping population-dependent resources.*
If population-dependent factor preferences are all orthogonal to each other, there are, essentially, no shared resources. In this case, species experience correlated growth rate fluctuations and no inter-specific competition $C_{ij} = \gamma\ \bar{R}\ \delta_{ij}$. We refer to this force as "*environmental filtering*".

Observe that more general and complex models involving both correlated population-independent factors and resource fluctuations, as well as combinations of the previous limiting cases, could also be constructed, but, for the sake of clarity, here we focus of these three (simplest) ones.

Using extensive numerical simulations and analytic tools like linear noise expansion around the fixed point(see Methods Sec.2.8) we investigate the relationship between pairwise abundance correlations and preference similarities for the three models. In particular, one can define a preference distance, d_P (where the sub-index P stands for "preference") proportional to the angle between preference vectors for each pair of species (with $d_P = 0$ for coinciding vectors and $d_P = 1$ for orthogonal ones). In model (A) and (B) such a distance is calculated over the resource preference $\boldsymbol{b}$, while the vectors of population-independent factors preferences $\boldsymbol{a}$ need to be considered in model (C).

As illustrated in Figure 2.3(a/b/c) the three models give raise to three qualitatively distinct patterns of correlation as a function of preference distance d_P: (A) Shared fluctuating population-dependent resources induce an effective *neutral* behavior, with nearly vanishing correlations across the spectrum of pairwise preference distances. (B) Shared resources and non-overlapping fluctuating population-independent factors produce *negative* correlations at small distances that increase to near-zero values in a monotonic way. (C) Shared fluctuating population-independent factors with fixed non-overlapping resources lead to correlations that decay from *positive* to vanishing values with distance.

These three paradigmatic scenarios apply, in exactly the same way, also for the multiplicative model defined in Eq.(2.3), as shown in sec.2.4. In the next section we delve a bit more on how resources and abiotic factors can behave as population-independent ones.

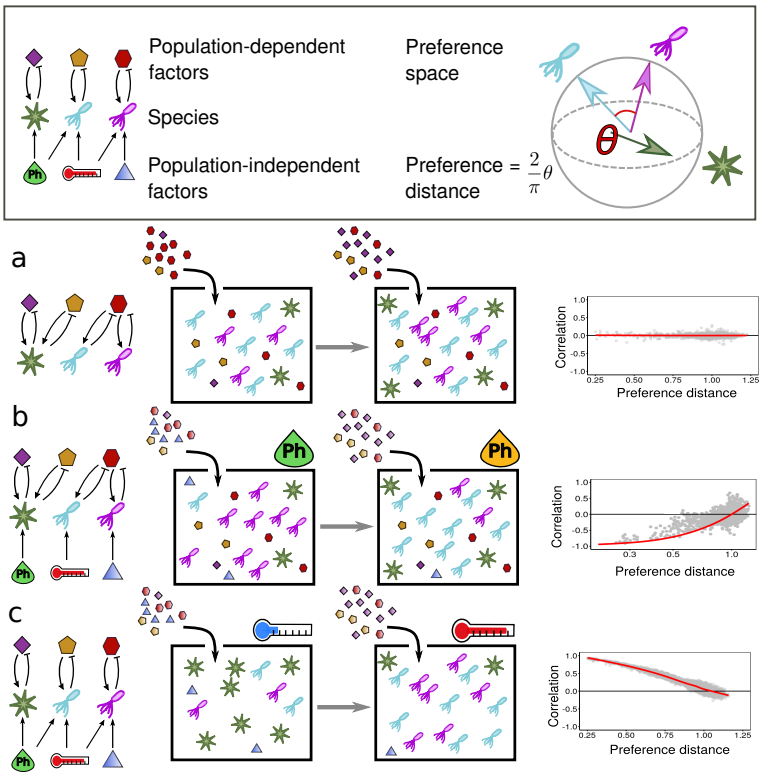


FIGURE 2.3: **(Top) Sketch of the elements of the model.** *Left:* Bacterial species depend upon both *population-dependent factors* such as resource abundances (polygons) and *population-independent factors*,e.g. abiotic variables like temperature, pH, light intensity, etc., and resources with low influx but strong fluctuations (shady polygons). The normal arrows stand for species preferences, while the inhibition arrows symbolize the feedback between population-dependent factors and populations. *Right:* Species preferences are represented as radial vectors in a sphere: positive/negative projections represent positive/negative influence on growth. The (pairwise) preference distance is quantified by the angle between vectors (multiplied by $2/\pi$, see methods). The red and blue species are similar to each other while the green one is more different from them. **(Bottom) Schematic illustration for the three considered scenarios** (models A, B and C) of: (*left*) population-dependent and population-independent factor preferences; (*centre*) model dynamics, and (*right*) stationary correlations as a function of preference distance (with gray dots standing for simulation results and red lines for averages/theory). **(a) Shared fluctuating population-dependent resources.** When species are subjected to the combination of both forces, their effects cancel out leading to an "effective" neutral situation with no correlations. **(b) Shared population-dependent resources and non-overlapping fluctuating population-independent factors.** when two species sharing some resource preference experience an environmental fluctuation, one outcompetes the other, causing negative correlations, increasing monotonically to zero as similarity decreases. **(c) Shared fluctuating population-independent factors with fixed non-overlapping resources.** If two species share some preference for population-independent factors, but not for resources, the follow in a similar way environmental fluctuations, causing positive correlations which decrease with preference distance.

2.4 What are population-independent factors ?

Population-independent resources

In this section we show that by considering a chemostat-like consumer-resource model with fluctuations in the resource-entry rates one can produce radically different dependencies of the species abundances correlation on the preference distance. To sustain this claim, we first sketch how it is possible to derive the general class of models exposed in Sec.3.5 from a general consumer-resource model by using a time-scale separation. In particular , we show that if common resources enter at a slow rate and experience large fluctuations, the species Pearson correlation is positive and decays to zero with the species preference distance. Hence, we show that resources can behave both as population dependent or independent factors, and conclude that the CSLM is a good effective description of the second case.

Model derivation from consumer resource

Consider N species consuming R fluctuating resources :

$$\dot{x}_i = x_i M_i(t) \sum_{\beta=1}^{R} R_j b_\beta^i - \delta_i x_i \tag{2.13}$$

$$\dot{R}_\beta = \lambda_\beta - \mu_\beta R_\beta - \gamma R_\beta \sum_{j=1}^{N} b_\beta^j x_j + \sqrt{\omega}_\beta \varphi_\beta(t), \tag{2.14}$$

with $\langle \varphi_\beta(t) \rangle = 0$ and $\langle \varphi_\beta(t) \varphi_\alpha(t') \rangle = \delta_{\alpha\beta}(t - t')$, where the species i preference for resource β is represented by the component of the vector $\boldsymbol{b}^i$, b_β^i; δ_i is the species death rate, λ_β and μ_β the resource entrance and exit rate, γ the consumption timescale and ω the fluctuation amplitude of the resource noise φ_{beta}. The influence of abiotic factors is modeled by the factor $M_i(t)$ as:

$$M_i(t) = \bar{M} + \sum_{\alpha=1}^{M} \sqrt{\nu} a_{i\alpha} \zeta_\alpha(t), \tag{2.15}$$

where $\bar{M}$ represents some baseline level and $\zeta_\alpha(t)$ is a Gaussian white noise. The parameter ν quantifies the strength of abiotic factors fluctuations.

If the the dynamics of resources entrance and exit is faster than the consumption one, it is possible to use a timescales separation technique to remove the explicit dynamic of resources. These timescale separation is based on the "unified colored noise approximation" in the theory of stochastic processes [228] as well as on recent developments for consumer-resource models [227]. Here, we give a sketch of the result , while we will report the full derivation and more in-deep analyses in a forthcoming publication.

If the species growth time scale is slower than the entrance or exit of resources, one can assume $\dot{R} \approx 0$ and remove the dynamics of the resources focusing on steady-state properties. This leads to the following effective model:

$$\dot{x}_i = x_i \left(M_i(t) \sum_{\beta=1}^{R} b_\beta^i \frac{\lambda_\beta + \sqrt{\omega_\beta} \varphi_\beta}{\mu_\beta + \gamma \sum_{j=1}^{N} b_\beta^j x_j} - \delta_i \right). \tag{2.16}$$

By considering,$\gamma \lll \mu$ and Taylor expanding Eq.(2.16) around $\gamma = 0$ up to first order in γ (neglecting also terms of the order $\gamma\sqrt{\omega}$), one obtains a generalized Lotka-Volterra model

equivalent to the one presented in Sec.3.5:

$$\dot{x}_i = x_i M_i(t) \left(\sum_{\beta=1}^{R} b_\beta^i \left(\frac{\lambda_\beta}{\mu_\beta} + \frac{\sqrt{\omega_\beta}}{\mu_\beta} \varphi_\beta \right) - \gamma \sum_{j}^{N} C_{ij}(t) x_j \right) - \delta_i x_i, \tag{2.17}$$

with the competition matrix specified now by:

$$C_{ij} = \sum_{\beta=1}^{R} b_\beta^i b_\beta^j \frac{\lambda_\beta}{\mu_\beta^2}. \tag{2.18}$$

Resources as population-dependent or independent factors

Here we show how fluctuating shared resource can become population-independent factors and give rise to positive species correlations. Consider that N species consume two different types of resources called R and Q. The first type of resource enters in the system very rapidly at a rate $\lambda_{1,\beta}$, i.e. it is abundant. On the other hand, the resources in the second set enter at a very slow rate $\lambda_{2,\alpha}$, i.e. they are scarce. Furthermore, both sets of resource entry rates fluctuate stochastically in time. Let us consider explicitly the dynamics of the two types of resources and of the species in the consumer-resource framework:

$$\dot{x}_i = x_i \sum_{\beta=1}^{R} R_\beta b_\beta^i + x_i \sum_{\alpha=1}^{Q} Q_\alpha a_\beta^i - \delta_i x_i \tag{2.19}$$

$$\dot{R}_\beta = \lambda_R - \mu_R R_\beta - \gamma R_\beta \sum_{j=1}^{N} b_\beta^j x_j + \sqrt{\omega} \varphi_\beta(t), \tag{2.20}$$

$$\dot{Q}_\alpha = \lambda_Q - \mu_Q Q_\alpha - \gamma Q_\alpha \sum_{j=1}^{N} a_\alpha^j x_j + \sqrt{\nu} \zeta_\alpha(t) \tag{2.21}$$

$$\langle \zeta_\alpha(t) \rangle = 0, \quad \langle \zeta_\alpha(t) \zeta_\beta(t') \rangle = \delta_{\alpha\beta}(t - t'), \tag{2.22}$$

$$\langle \varphi_\beta(t) \rangle = 0, \quad \langle \varphi_\beta(t) \varphi_\alpha(t') \rangle = \delta_{\alpha\beta}(t - t'). \tag{2.23}$$

By using, as above, the timescale separation and expanding in $\gamma \lll \mu$, we obtain the following effective generalized Lotka-Volterra equation:

$$\frac{dx_i}{dt} = x_i \left(r_i(t) - \sum_{j=1}^{N} C_{ij} x_j \right), \tag{2.24}$$

where the entries of the competition matrix are determined by the overlap in resource preferences

$$C_{ij} = \frac{\lambda_R}{\mu_R^2} \gamma \boldsymbol{b}^i \cdot \boldsymbol{b}^j + \frac{\lambda_Q}{\mu_Q^2} \gamma \boldsymbol{a}^i \cdot \boldsymbol{a}^j. \tag{2.25}$$

The time-dependent growth rate can be written as $r_i(t) = \bar{r}_i + \sqrt{\sigma}\xi_i(t)$, where the explicit expressions of $\bar{r}_i$,s and ξ on the original parameters are:

$$\bar{r}_i = \frac{\lambda_R}{\mu_R}\sum_\beta b^i_\beta + \frac{\lambda_Q}{\mu_Q}\sum_\alpha a^i_\alpha - \delta \tag{2.26}$$

$$\sqrt{\sigma}\xi_i(t) = \frac{\sqrt{\omega}}{\mu_R}\sum_\beta b^i_\beta \varphi_\beta + \frac{\sqrt{\nu}}{\mu_Q}\sum_\alpha a^i_\alpha \zeta_\alpha \tag{2.27}$$

$$\sigma = \frac{\nu}{\mu_Q^2} + \frac{\omega}{\mu_R^2} \tag{2.28}$$

and, additionally, the noise correlation $\langle \xi_i(t)\xi_j(t')\rangle = \rho_{ij}\delta(t-t')$ is given by

$$\rho_{ij} = \frac{\nu/\mu_Q^2 \boldsymbol{a}^i\cdot\boldsymbol{a}^j + \omega/\mu_R^2 \boldsymbol{b}^i\cdot\boldsymbol{b}^j}{\nu/\mu_Q^2 + \omega/\mu_R^2}\,, \tag{2.29}$$

Note that at this level, both types of resources can give rise to noise correlations and species competition. Nevertheless, Eq.(2.25) provides us with an intuition on how the Q set of resources can become *population-independent*. While both the noise correlation, Eq.(2.29), and species interaction matrix, Eq.(2.25), depend on μ_Q, only the second one depends upon λ_Q. Hence, if one considers $\lambda_Q \ll \lambda_R$ the Q set of resources should not give rise to competition but only to positive noise correlations. To support this claim, let us consider two simplified scenarios, where for simplicity we take $N = R = Q$:

- Consider that the R resources are not shared between species, i.e. each species eats exclusively one resource of type R, i.e. $\boldsymbol{b}^i\cdot\boldsymbol{b}^j = 0$, while Q resources are shared, as depicted in Fig. 2.5 A. In this case the model reduces to:'

$$\rho_{ij} = \boldsymbol{a}^i\cdot\boldsymbol{a}^j, \quad C_{ij} = \delta_{ij}\gamma\frac{\lambda_R}{\mu_R^2}\gamma|\boldsymbol{b}^i|^2 + \frac{\lambda_Q}{\mu_Q^2}\gamma\boldsymbol{a}^i\cdot\boldsymbol{a}^j. \tag{2.30}$$

  In Fig. 2.5B, we report the stationary Pearson correlation coefficients as a function of the preference distance, as obtained from simulations with $\lambda_Q = 0.1 \ll \lambda_R = 10$. This confirms that the Q type resources are so scarce that they do not induce competition between species (non-negative correlation). Nevertheless, a regime shift happens as a function of ν: while for small fluctuations $\nu = 10^{-2}$ the correlations are null and flat, by increasing ν they assume a pattern of positive correlations decaying to zero. For very strong fluctuations, $\nu = 5$, correlations coincide in magnitude with the analytic value predicted by model C, Eq.(2.150) for $\nu = 5$ (see black line). This transition happens because at small ν positive correlations are balanced by the tiny competition factors, as happens in Model A. However, by increasing ν their strength overcomes competition and dominate the dynamics. Hence, resources with low but highly variable influx rate behave as *population-independent* factors, and one can describe them using the Model C variant of the CSLM.

- Consider that the Q resources are not shared between species, i.e. each species eats exclusively one resource Q, hence $\boldsymbol{a}^i\cdot\boldsymbol{a}^j = 0$, while the R resources are shared, as depicted in Fig. 2.5 C. In this case the model reduces to:

$$\rho_{ij} = \boldsymbol{b}^i\cdot\boldsymbol{b}^j, \quad C_{ij} = \delta_{ij}\gamma\frac{\lambda_Q}{\mu_Q^2}\gamma|\boldsymbol{a}^i|^2 + \frac{\lambda_R}{\mu_R^2}\gamma\boldsymbol{b}^i\cdot\boldsymbol{b}^j. \tag{2.31}$$

In Fig. 2.5 D we show the associated stationary Pearson correlation coefficients as a function of the preference distance obtained from simulations with $\lambda_Q = 10^{-3} \ll \lambda_R = 1$. Observe that the R resources are abundant and induce competition (negative correlations). Nevertheless, a regime shift happens as a function of ω: for small fluctuations, $\omega = 0, 10^{-2}$, the correlations are negative and increase to zero with phylogenetic distance, as predicted by Eq.(2.64) of Model B (black line). By increasing ω,, shared fluctuations balance out competition, leading to null and flat correlations, like Model A. Hence, resources with high influx rate are *population-dependent* factors, and one can describe them using Model A in the case of strong fluctuations or Model B, if fluctuations are absent.

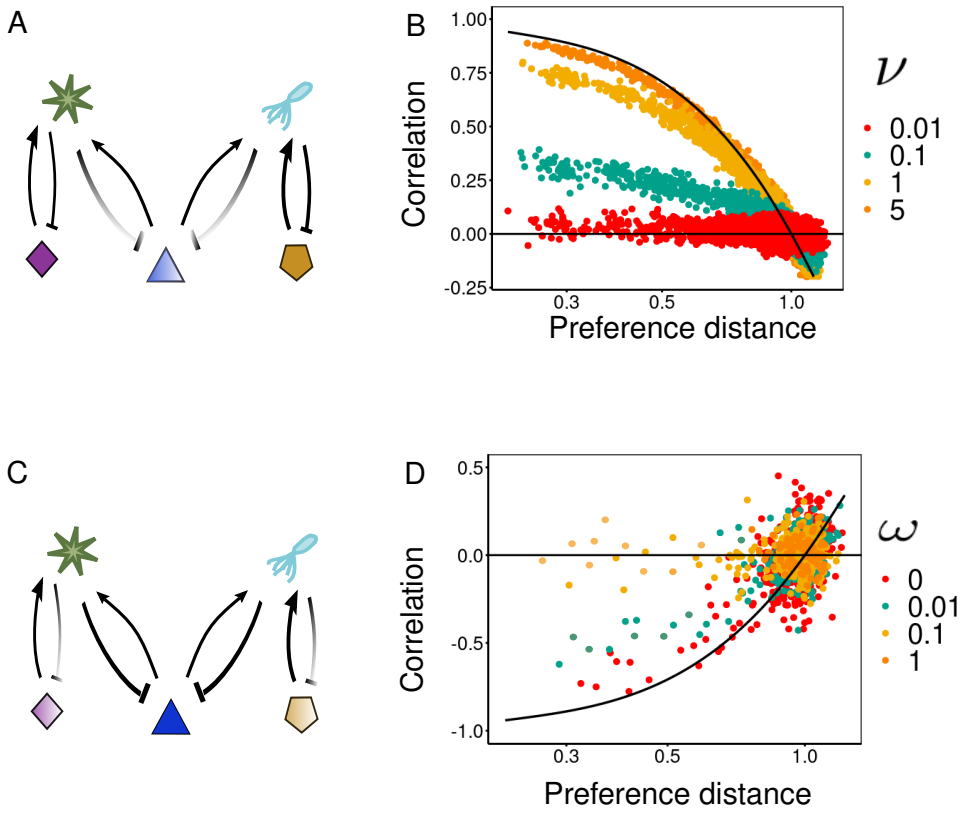


FIGURE 2.4: **Fluctuating resources as either population-dependent or population-independent factors.** Consider that species are limited by two types of resources: one type enters at a very slow rate of influx in the system (light color polygons, Q set) and hence is rapidly diluted; the second set describes resources entering the system very rapidly (dark color polygons, R set), $\lambda_Q \ll \lambda R$. **Top :** Resources as population-independent factors. **A)** We consider a scenario where species share preferences for slow entering resources (light color polygons), but not for very rapidly entering ones (dark color polygons). **B)** These shared resources can produce diverse correlation patterns depending on the interplay of their noise amplitude, ν and species interactions. If ν is low, correlations are flat and almost null. By increasing ν, the correlations reach a positive-to-null decaying pattern that can be approximated by Model C, Eq.(2.150). This behavior is due to the different relative weights of competition and fluctuations: while the first is proportional to λ_Q/μ^2 , the is depends on $\sqrt{\nu}/\mu$ see Eq.(2.30). Hence, by changing ν we are modifying the weight of fluctuations compared with species competition, making the Q resource *population-independent*. **C) Bottom :** Resources as population-dependent factors. Species share preference for rapid entering resources (dark color polygons), but not for very slow entering ones (light color polygons). **D)** These shared resources can produce various correlation pattern depending on the interplay of their noise, ω, with species interactions. If ω is low correlation are negative and increase to zero with preference distance, well described by Model B for zero noise (black line, Eq.(2.64)). By increasing the noise, correlation reach a null and flat pattern that can be approximated by Model A. This happens because the resource entry rate λ_R is proportional to competition but not to stochastic fluctuations, see Eq.(2.31). Hence, By changing ω we are modifying the weight of fluctuations compared with species competition, making the R resource *population-dependent* . *Parameters:* $N = Q = R = 300, \mu_R = \mu_Q = 1, \gamma = 0.1$, *A)* $\lambda_R = 10, \lambda_Q = 10^{-1}, \omega = 0$;*B)*$\lambda_R = 1, \lambda_Q = 10^{-3}, \nu = 0.01$.

Abiotic population-independent factors

To explicit consider population-independent factors as abiotic factors like temperature or Ph, we propose here a modification of the previous models where such factors affect the growth rate in a "multiplicative way". The per-capita growth rate of each species is, as above, influenced both by population-dependent and population-independent factors, weighted by the corresponding preference vector

$$\frac{1}{x_i(t)}\frac{dx_i}{dt} = M_i(t)\sum_{\beta=1}^{R} b_\beta^i R_\beta(t) - \delta\,, \tag{2.32}$$

As above, we assume the population-independent factors to be subject to rapid stochastic fluctuation

$$M_i(t) = \bar{M} + \sum_{\alpha=1}^{M}\sqrt{\nu}a_{i\alpha}\zeta_\alpha(t), \tag{2.33}$$

where $\bar{M}$ represents some baseline level and $\zeta_\alpha(t)$ is a Gaussian white noise and ν quantifies the strength of abiotic factors fluctuations. Similarly, the abundance R_β of each resource β fluctuates over time. In this case, fluctuations are determined by a balance between stochasticity and the consumption by populations:

$$R_\beta(t) = \bar{R}\left(1 + \sqrt{\omega}\varphi_\beta(t) - \gamma\sum_{j=1}^{M} b_\beta^j x_j\right), \tag{2.34}$$

where $\bar{R}$ is a baseline level, γ the consumption timescale, and $\varphi_\beta(t)$ a Gaussian white noise with zero mean and variance 1. Similarly to σ, the parameter ω quantifies the importance of resources fluctuations. Hence, the model defined in Eq.(2.2) reads

$$\begin{aligned}\frac{1}{x_i}\frac{dx_i(t)}{dt} &= \bar{M}\sum_{\beta=1}^{R} b_\beta^i\left(\bar{R} + \sqrt{\omega}\varphi_\beta(t) - \gamma\sum_{j=1}^{N} b_\beta^j x_j\right) + \sqrt{\nu}\bar{R}\sum_\beta b_{i\beta}a_{i\alpha}\zeta_\alpha(t) - \delta \\ &+ \sqrt{\nu\omega}\sum_{\alpha,\beta} a_{i\alpha}b_{i\beta}\zeta_\alpha(t)\varphi_\beta(t) - \sqrt{\nu}\gamma\sum_{\alpha,\beta}\sum_j a_{i\alpha}b_{i\beta}b_{j\beta}x_j\zeta_\alpha(t).\end{aligned} \tag{2.35}$$

First let us note that Model A does not involves population-dependent factors and hence is recovered directly from Eq.(2.32). Second, let us note that if the fluctuation parameters ν, ω and the consumption timescale γ are small, all the terms on the second line are sub-leading and the model is equivalent to the general case defined in the first part of section(3.5). In particular, in this section we will study how considering multiplicative abiotic factors as population-dependent one modifies the results obtained for model B and C. Let us start by considering the modification of model C, obtain from Eq.(2.32) by considering perpendicular population-dependent preferences but shared abiotic ones and $\omega = 0$:

$$\frac{1}{x_i}\frac{dx_i(t)}{dt} = \left(\bar{M} + \sqrt{\nu}\sum_{\alpha=1}^{M} a_{i\alpha}\zeta_\alpha(t)\right)\left(\bar{r}_i - \frac{x_i}{K_i}\right) \tag{2.36}$$

with

$$\bar{r}_i = \sum_{\beta=1}^{R} b_{i\beta}, \quad K_i^{-1} = |\boldsymbol{b}_i|^2\gamma. \tag{2.37}$$

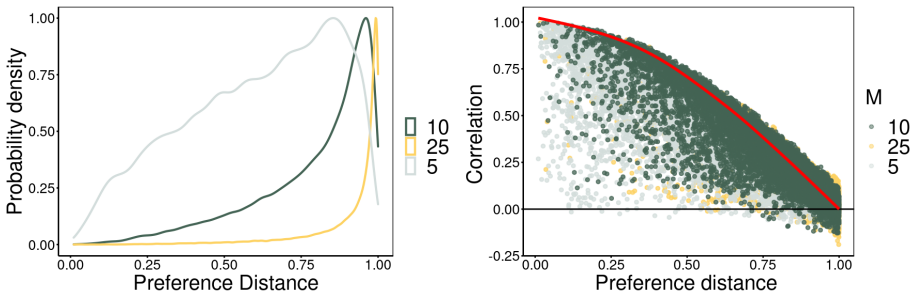


FIGURE 2.5: **Correlation with purely abiotic factors: dependence on the number of abiotic factors.** Study of the model defined by Eq.(2.36) of environmental filtering at the stationary state. Left: preference distance distribution for differente values of M (number of abiotic factors, color coded), $M = 5, 10, 25$ and $N = 300$. Right: Pearson correlation distribution (log-scale) as function preference distance for different values of M (color coded), $M = 5, 10, 25$ and $N = 300$, obtained by the numerical integration of Eq.(2.36). The red line stands for Eq.2.150. As expected, by increasing the number of resources, the correlation distribution variances decreases and the decay is better approximated by the linear noise approximation. Parameter values: $N = 300, q = 0.4, Z = 50N, m = 1, \bar{R} = \bar{M} = 1, \gamma = 0.1, \nu_i = 0.5$, and $t_{fin} = 10^4$

This model and reduces to model C it for $\frac{\sqrt{\nu}}{K} \lll \frac{1}{K}$. To realistically model abiotic factors they need to appear in a number smaller than the number of species, i.e. $N \sim R \ggg M$. In order to obtain a distance distribution with the desired properties in such a scaling regime, we use the evolutionary algorithm exposed in Sec.2.8. Based on a linear-noise perturbation we expect the stationary pattern of correlations to not deviate much from model C. Indeed, Fig. 2.5 shows the numerical solution for the average correlation as function of preference distance still decays in good numerical agreement with the Eq.(2.150).

To resume the validity of our results also in the case of abiotic factors acting multiplicative on the growth rate, in Fig. 2.6 we compare the analytic approximation of the Pearson correlation coefficient in CSLM, Eq.(2.150), with the simulations of model C (A), and its multiplicative modification to include abiotic factors in large (B) or small number (C), Eq.(2.36).

Finally, we also study the modification of Model B to include multiplicative abiotic factors:

$$\frac{1}{x_i}\frac{dx_i(t)}{dt} = \left(\bar{M} + \sqrt{\nu}\sum_{\alpha=1}^{M} a_{i\alpha}\zeta_\alpha(t)\right)\sum_{\beta=1}^{R} b_{i\beta}\left(\bar{R} - \gamma\sum_{j=1}^{N} b^j_\beta x_j\right), \tag{2.38}$$

where the previous equation is obtained from Eq.(2.32) by considering perpendicular abiotic preferences, shared resources preferences and $\nu = 0$. In Fig. 2.7, we compare the analytic prediction of correlations for model B with simulations of different versions of the model. Note that the negative to null correlation pattern is present in all the model versions, and out analytic prediction is a good approximation across a large variation of model scenarios.

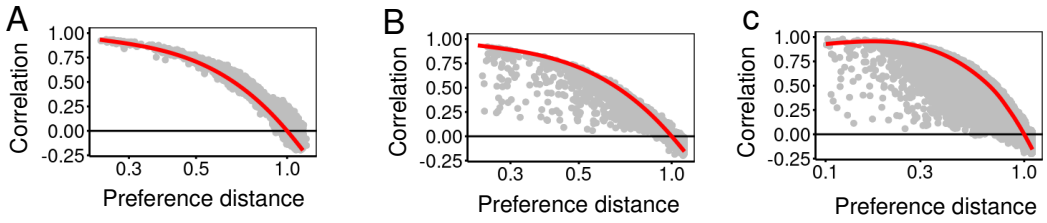


FIGURE 2.6: **Comparison of analytic and simulations for correlations in model C** The red line represents the analytic expectation for the Pearson correlation coefficient, Eq.(2.150), while the gray points the stationary Pearson coefficient for each species couple obtained numerically. In A) we simulated Model C Eq.(2.85), obtaining a very good match between theory and simulations. In B) and C) we simulated the abiotic non-linear version of model C, Eq.(2.36) for a large number of abiotic factors (M=300) or small (M=10). Note that Eq.(2.150) is still a very good approximation in all the cases, but does not capture the fluctuations of the correlation values in the abiotic case B, C. In particular, while in A, for a given distance, the formula capture both the mean and typical correlation, in B) and C) the formula represent the typical correlation but not the average one (that is a bit smaller), as typical of multiplicative processes. Parameter values: $N = 300, M = 300(A, B), 10(C), q = 0.1, Z = 50N, m = 1, \bar{R} = \bar{M} = 1, \gamma = 0.1, \nu_i = 0.5$, and $t_{fin} = 10^5$

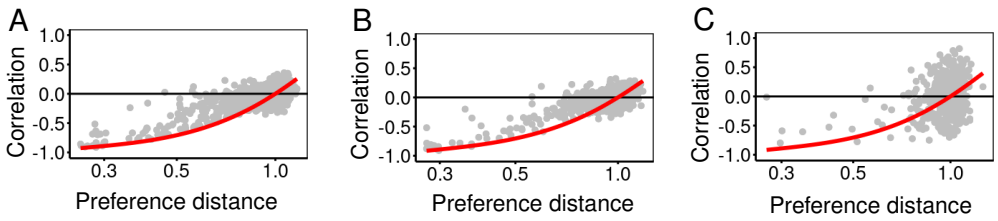


FIGURE 2.7: **Comparison of analytic and simulations for correlations in model B.** The red line represents the analytic expectation for the Pearson correlation coefficient, Eq.(2.64), while the gray points the stationary Pearson coefficient for each species couple obtained numerically. In A) we simulated model B, Eq.(2.60), obtaining a very good match between theory and simulations. In B) and C) we simulated the abiotic multiplicative version of model B, Eq.(2.38) for a large number of abiotic factors ($M = 300$) or small ($M = 10$). Note that Eq.(2.64) is still a very good approximation, but does not capture the fluctuations of correlation values. $N = 300, M = 300(A, B), M = 10(C), q = 0.1, Z = 50, N = 1, m = 1, \bar{R} = \bar{M} = 1, \gamma = 0.1, \omega_i = 0.5$, and $t_{fin} = 10^5$

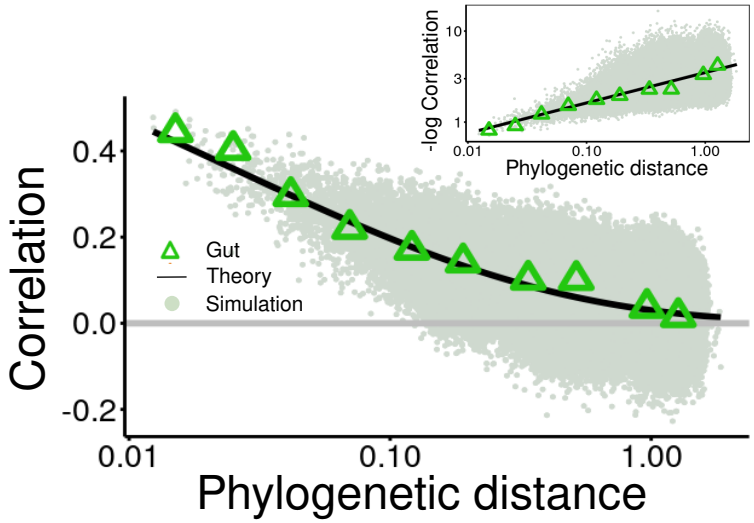


FIGURE 2.8: **The model with environmental filtering reproduces the empirical law.** Correlation values are plotted as a function of the phylogenetic distance both for the gut microbiome data (green triangles for each binarized value) and the simulated computational model (green clouds of points). The analytical expression, Eq.(B.14) with $\lambda = 3.5$ is also plotted (black line). Simulations of the model have been performed, using $N = 300$ species and considering as an input the empirical phylogenetic distance matrix of the gut microbiome, randomly sampling from it the N species. Inset: -log Correlations as a function of phylogenetic distance in double-logarithmic scale, empirically and from the mode, same data as the main figure. For more simulations details see Methods

2.5 Environmental filtering reproduces the correlation decay with distance

In order to make a more quantitative comparison between the previous results and the empirically-determined universal pattern of decaying correlations, it is necessary to specify the relation between the preference distances $d_{P,ij}$ —on which the models rely— and the empirically-determined phylogenetic similarity of actual species, as quantified by their genetic distance $d_{G,ij}$. For this purpose, it seems natural to assume that d_P and d_G are positively correlated, i.e., that phylogenetically close species typically have more similar preferences than distant ones. Under this assumption, the overall trend of the decay in Fig. 2.3 implies that environmental filtering is the process responsible for the empirically-observed decay of correlations (Fig. 2.1). Competition for constant and/or shared fluctuating resources can instead be discarded as the leading mechanism on the basis of the empirically observed pattern. This does not imply that competition is not present, but rather that it does not generate a signal detectable at a phylogenetic level within the present level of resolution.

To make further quantitative progress in the connection between the previous mechanistic modelling approaches, in particular, model C or "environmental filtering", and available phylogenetic data, one needs to define a more precise mapping between preference similarity in the model and empirically-determined phylogenetic distance, i.e. to characterize the functional dependence between d_P on d_G, using information on pairwise correlations.

The previous task is not straightforward: species are coupled to each other within a network of interactions, so that pairs of species cannot be simply analyzed one at the time and, on the other hand, the full set of coupled non-linear equations is intractable. Fortunately, however, as explicitly shown in the Methods sect., one can make further

progress by explicitly mapping model C into a *correlated stochastic logistic model* (CSLM) :

$$\frac{dx_i}{dt} = \frac{x_i}{\tau_i}\left(1 - \frac{x_i}{K_i}\right) + \sqrt{\frac{\sigma_i}{\tau_i}}x_i\xi_i(t), \tag{2.39}$$

where τ_i^{-1} is the growth rate, K_i an effective carrying capacity, σ_i the amplitude of environmental fluctuations, and $\xi_i(t)$ is a Gaussian white noise, with correlations proportional to the preference distance,

$$\langle \xi_i(t)\xi_j(t')\rangle = \delta(t-t')\cos\left(\frac{\pi}{2}d_{P,ij}\right). \tag{2.40}$$

For the sake of simplicity, in the derivation (see Methods), we assumed that the preference space has a large dimensionality, i.e. $M \gg 1$, but this can be shown not to limit the generality of the forthcoming results (see Methods 2.8 for more details).

This mapping is particularly illuminating as the resulting CSLM extends the standard stochastic logistic model (SLM) [207], as it includes correlated growth-rate fluctuations that stem from shared environmental fluctuating resources and that induce non-trivial species correlations. Moreover, it is important to stress that —if species-abundances trajectories are observed individually— there are no statistical differences between the CSLM and the standard SLM. This implies that the CSLM also reproduces (as the SLM does) the three macroecological patterns put forward in [207, 208, 123] (see Methods, 2.8). Thus, the CLSM constitutes an improvement of existing modelling approaches to microbial macroecological laws.

A crucial advantage of Eq.(2.39) (together with and Eq.(2.40)) with respect to the generalized Lotka-Volterra equation is that is can be treated analytically to obtain a mathematical expression linking pairwise species-abundance correlations with their preference distance, $d_{P,ij}$ using a linear approximation around the fixed points:

$$\eta_{ij} \approx \cos\left(\frac{\pi}{2}d_{P,ij}\right) \tag{2.41}$$

. The resulting analytical relationship can be exploited to estimate the preference distance matrix from empirical correlation data, thus allowing us to establish the desired relation between preference distance d_P and phylogenetic distance d_G for every pair of species (see Methods):

$$d_{P,ij} \approx \frac{2}{\pi}\arccos\left(e^{-\lambda d_{G,ij}^{1/3}}\right), \tag{2.42}$$

where λ is a constant. Observe, that Eq.(2.42) is highly non-linear, implying that, as the phylogenetic distance grows, preference distances rapidly saturate to values close to 1; in other words, even phylogenetically similar species tend to have a large preference dissimilarity (i.e. their preference vectors tend to be orthogonal to each other).

By implementing the relation given by Eq (2.42) in the definition of noise correlations Eq.(2.40), we obtain a version of the CSLM, directly relating ecological processes and phylogeny, which allows us to relate the species-abundance pairwise correlations to their empirically measured genetic similarity, $d_{G,ij}$. Actually, given that the macroecological pattern we intend to reproduce is for the averaged correlation at a given (binarized) phylogenetic distance, we dropped the sub-index ij in Eq.(2.42) and use it as a relation between averages (see Methods , Eq (2.146)). In particular, by combining Eq.(2.146) with Eq.(2.41), one obtains exactly Eq.(B.14), i.e. the empirically observed relation between correlation and phylogenetic distance (see Methods).

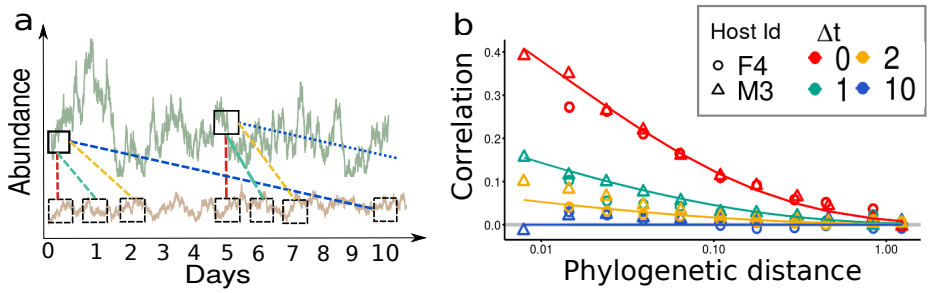


FIGURE 2.9: **(a) Sketch of the time-dependent (longitudinal) correlation data analyses**. Typical time-series for two species (green and brown, respectively) along 10 days. The dashed lines illustrate how equal-time (red) and 1, 2, 10 days delayed correlations (green, yellow and blue respectively) are computed, see Methods for more details. **(b) Macroecological law for temporal data**. Equal time (red), one-day delay (light blue), two-days (yellow) and ten-day delay (blue) symbols represent correlations as a function of the discretized phylogenetic distance (logarithmic scale) for the gut microbiomes of two different hosts (labelled with circles (F4) and triangles (M3), respectively). Solid lines stand for the prediction from the CSLM, Eq.(2.43), averaged over hosts, with timescale parameter $\tau_i = 1$, for $i = 1, .., N$ and $\lambda = 4.5$.

Fig. 2.8 shows that, for the particular case of the human gut microbiome, a computational simulation of the final version of the model captures quite well the averaged decay of pairwise correlations with phylogenetic distance and that the analytical predictions describe accurately such an averaged behavior.

2.6 Macroecological law for temporal data

One important prediction of Eq.(2.39) is that the decay of abundance correlations with phylogenetic distance is caused by shared temporal fluctuations. In order to further test the predictions of Eq.(2.39), we consider longitudinal data from the human microbiome. In particular, we analyzed three human body sites (gut, oral cavity, and hand palms) of two hosts [224]. From these data, we calculate the correlation of species abundance fluctuation η_{ij} as above, but now averaging over time, rather than across individuals (see Fig. 2.9(a)). In particular, Figure 2.9(b) illustrates —for the specific case of the human gut— that the macroecological law of decaying correlation holds also for such temporal data, and that delayed correlations rapidly decay to zero. In particular, the correlations as a function of phylogenetic distance decay on average as a stretched exponential with an exponent close to 1/3, as observed in cross-sectional data.

To further test the CSLM model in its ability to reproduce time-dependent features of species correlations, we also computed delayed pairwise correlations, $\eta_{ij}(\Delta t)$ defined as the correlation between the abundance fluctuations of species i at time t with the abundance fluctuations of species j at time $t + \Delta t$:

$$\eta_{ij}(\Delta t) \approx e^{(1-\frac{\sigma}{2})\frac{\Delta t}{\tau}} \cos\left(\frac{\pi}{2} d_{P,ij}\right); \tag{2.43}$$

(see Methods Sec. 2.8 and Fig. 2.9 (a) for a graphical illustration). Let us remark that, in principle, the value of such a delayed correlation is, in general, not trivially linked to the correlation computed at the same time, as it depends of the specific properties of the dynamics giving rise to species inter-dependencies. Remarkably, as shown in Figure

2.9(b) the CSLM with no additional modification quantitatively reproduces also the temporal delayed correlations for different values of the delay (see appendix B for additional details and analyses) only by setting the growth time scale $\tau_i = 1$ for all species.

2.7 Conclusions and discussion

We have considered both cross-sectional (across communities) and longitudinal (across time) empirical data for the species abundances in microbial communities from many different environments and studied their species-abundance pairwise correlations as a function of pairwise phylogenetic distance, revealing the emergence of an universal macroecological law. This empirical law states in quantitative terms that the average correlation function decays from positive to null values as the phylogenetic distance (or dissimilarity) increases, approximately following a stretched-exponential decay function.

We explored the possible ecological forces shaping species correlations from a theoretical standpoint. In particular, by scrutinizing different ecological models, each one implementing a diverse set of ecological forces between species, we found that the universal correlation pattern cannot possibly be reproduced by competition or exclusion principles. Instead, temporal environmental filtering —i.e. the presence of correlated noise stemming from shared fluctuating factors— as modeled by a correlated stochastic-logistic model (CSLM), explains quantitatively empirical data. Furthermore, time-dependent (delayed) correlations in longitudinal data are also well reproduced by the model.

This novel ecological pattern gives a quantification at the level of phylogenetic signals detectable in taxa-taxa abundance correlation. The pattern, as also shown in Fig. B.5, B.2, and B.3 in appendix B, does not recapitulate the full range of correlations observed in natural communities. In this context, our work complements the research aiming at inferring ecological interactions from correlations, by showing how phylogenetic similarity can be used to disentangle the effects of environmental fluctuations and interactions (such as, e.g., competition).

These results are based on multiple assumptions and their limitations give opportunities for extensions of the current work. First, at a theoretical level, the CSLM reproduces the average correlation at each discrete phylogenetic distance, but not the full distribution around such a mean value (see Fig. 2.18 in Methods). This is because, to be able to connect genetic and preference similarities, we enforced a "mean-field" type of relationship, Eq.(2.42), neglecting variability across pairs of species in the phenotypic-distance-to-preference-distance mapping. On the other hand, in Figure B.5 of appendix B, we show that the variance of the distribution of the empirically-measured pairwise correlations within each distance bin seems to follow a weak decaying power-law pattern with phylogenetic distance, with a diverse decaying exponent characteristic for each analyzed biome. Possibly, these patterns could be used to generate the preference vectors of the model in a more general way, allowing for more variability. Empirical data are not informative enough at the moment to proceed in this direction, and further analyses are required.

It is however important to stress that both the empirical analysis and the model assume a certain degree of niche conservatism . One important assumption of our modeling framework is that ecological similarities are fixed in time and environmentally dependent [229, 230]. In the extreme scenario, in which the ecological strategy is strongly conserved on the phylogenetic tree there would be a $1:1$ mapping between ecological similarity and phylogenetic distance. This strong assumption is however not needed for our analysis, which requires of a much weaker condition: namely, that ecological similarity correlates with phylogenetic similarity. The variability of correlations around the

expected one from phylogenetic distance (shown in Fig. 2.18 in appendix B) should be interpreted in this way. Note that two interpretations of our results are possible. On the most pessimistic side, one could argue that the pattern we discover and the model we propose serve only to describe the phylogenetic signal observed in the correlations, leaving the variation un-explained. Instead, on the most optimistic side, one could argue that the variability observed in the correlations is not a signal of other ecological mechanisms not included in the model, but rather the consequence of the lack of a perfect match between ecological similarity and phylogenetic similarity.

Recent theoretical works (e.g., in the context of consumer-resource models [231]) explored the case of dynamic ecological preferences, where species' preferences are dynamically optimized given an environment. One could envision extensions of our model including dynamical preferences. Nevertheless, these changes in ecological strategies might contribute to the large variation around the mean phylogenetic trend by they should be constrained by the robust pattern of mean correlations reported here.

It is also important to stress that the origin of the stretched exponential behavior and, in particular, its exponent value close to a value 1/3 in the universal pattern of correlations (i.e., Eq.(B.14)) remains unexplained. This type of scaling could be influenced by the scale-invariant, i.e. fractal, structure of phylogenetic trees [232, 233, 234, 235]; further investigations, beyond the scope of the present work, are needed to shed light onto this empirical finding. Furthermore, it is known that a vast class of competitive models can lead to species clustering in trait space [236, 237]. Even if such models produce an "oscillating" pattern of positive and negative correlation, and hence are not sufficient to explain the behavior here reported, their possible extension could be relevant for explaining the phylogenetic distance distribution observed in data (see Fig. B.1in appendix B).

Although environmental filtering has been found to dominate the pattern of species-abundance correlations, the above-mentioned variability could be the result of the complex interplay of other ecological forces. To identify which further forces are relevant and to discriminate their effects it will be important to analyze time-dependent data in a more detailed way as well as to analyze differences in carrying capacities, and correlations between different hosts [123]. Furthermore, an exhaustive analysis of the variations of the correlation pattern across environments and phyla is also needed. Interestingly, Fig. B.12-B.10 in appendix B show that some phyla (e.g. bacteroidetes) follow robustly the pattern, while some others, such as actinobacteria, exhibit wild fluctuations. Indeed, the non-monotonic deviation in the soil biome around distance 0.1 seems to be caused by the actinobacteria phylum and, in particular, by the actinomycetales and gaiellales orders (see Fig. B.9 in appendix B). The fact that the trend of correlation and phylogeny holds across very different environments strongly suggests that the pattern captures an underlying general ecological process, linking phylogeny with ecological similarity and ecological similarity with correlations. Specific environments and specific taxa might have different behaviors, which is reflected in the deviations from the average patterns and in the variability of the fitted parameters of the stretched-exponential. We leave for future work the promising study of deviations across taxa, that could reveal more information on additional interactions responsible for the observed residual correlations.

Another relevant caveat is that our analyses here are limited to the taxonomic resolution of OTUs, clustering together individuals with more than 97% similarity. Recent results suggest that ecological dynamics starts to decouple at much finer phylogenetic resolutions [238]. Moreover, strains seem to still obey the three macroecological laws of variation and diversity valid at species level [126]. These results leave open the question of how ecological forces shape the variation of community composition at finer phylogenetic scales.

On the other hand, from a complementary viewpoint, we analyzed the behavior of correlations at the coarse-grained resolution of phyla. In particular, Fig. 2.2 illustrates that by considering just inter-phyla correlations one cannot observe the stretched exponential decay, that is determined by intra-phyla OTU pairs. Analogously, by extending our analyses to finer phylogenetic resolutions it could be possible to reveal the nature of intra-specific interactions, eventually elucidating the emergence of competition as a key player in determining correlations. Actually, in our view, one should not fix a characteristic taxonomic resolution to have a complete description of complex communities, but, instead, start from individuals (or functional units) and progressively cluster them together at larger and larger coarse-grained scales (i.e. moving across observational scales as customarily done in physics using "renormalization group" tools in statistical physics [239, 240]) as different ecological forces may shape communities at diverse resolution levels [241].

2.8 Methods: detailed analysis of the ecological models

Evolutionary algorithms to generate a wide distribution of preference distances.

If vectors in the preference space –identifying the characteristic of each given species in the different models— are generated in a simple random fashion, they have a large probability to be orthogonal to each other; i.e. vectors with small distances are very unlikely to be randomly generated. In particular, as a consequence of the central limit theorem, for sufficiently large numbers of environmental factors, R, the random vectors $\boldsymbol{b^i}$ tend to be orthogonal to each other, i.e., $D_{P,ij} \approx 0 \quad \forall i,j$, hindering the observation of similar species. To paliate this problem and populate all the space of possible pairwise distances we consider two alternative algorithms.

Algorithm 1: high-dimensional preference space

If the preference space under consideration is high dimensional, i.e. $N \sim R \sim M \ggg 1$, one can employ the following algorithm. We generate the set of M preference vectors $\boldsymbol{b}$ by sampling their component from a Gaussian distribution with mean m/R (m small and positive) and s.t.d. $1/\sqrt{R}$, $\mathcal{N}(m/R, 1/\sqrt{R})$, such that the radius of the overall vector is constant and close to 1 for large M: $r_P^2 = \sum_\beta b_\beta^2 = 1 + \frac{m^2}{R} \approx 1$.

Starting from an initial random distribution of vectors $\boldsymbol{b^i}^0$ —and implementing an evolutionary branching process— generates as an outcome a set of vectors $\boldsymbol{b^i}$ which are distributed across all values of possible cosine distances. The algorithm has the following steps:

1. Sample at random two species i, j, i reproduces and j dies.
2. Replace the pair of vectors by a new pair, specified by: $\boldsymbol{b^i} = q\boldsymbol{b^i} + (1-q)\boldsymbol{\epsilon^i}$, $\boldsymbol{b^j} = q\boldsymbol{b^i} + (1-q)\boldsymbol{\epsilon^j}$, where the $q \in [0,1]$ is the "fidelity" and $p = 1-q$ is the "mutation", and $\epsilon^{i,j}$ vectors sampled from $\mathcal{N}(m/R, 1/\sqrt{R})$.
3. Iterate Z times.

Note that the vectors are automatically kept in the sphere. For large enough values of Z and $q = 0.9$, the population has a small pool of similar individuals, corresponding to a long left tail of the distance distribution (see Fig. B.1).

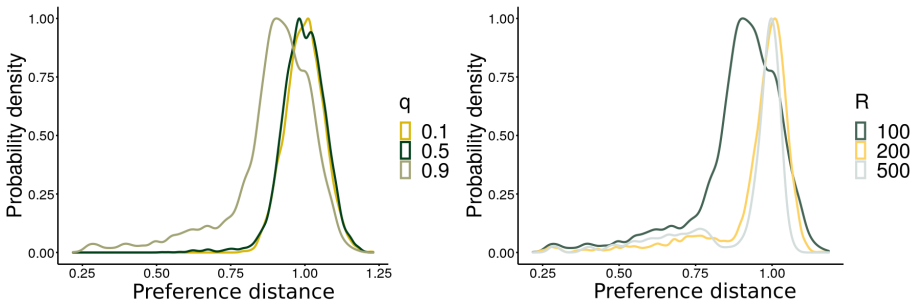


FIGURE 2.10: **Preference distance distribution generated by the algorithm in high-dimensional space.** Left: distribution obtained for different values of q and $R = 100$. Right: distribution with $q = 0.9$ and different values of R. As a consequence of the central limit theorem, the distribution converges to a Gaussian for small values of q and $R \to \infty$. On the contrary, for high values of fidelity, $q = 0.9$, the distribution develops a long left tail, covering all the preference space. As shown in the right figure for any value of q, the variance shrinks by increasing R. In both figures we considered $N = 100$, and iterated the algorithm $Z = 50N$.

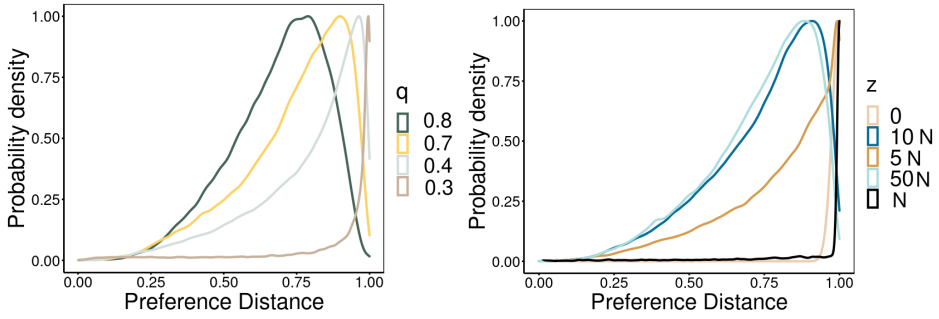


FIGURE 2.11: **Preference distance distribution generated by the second algorithm for low-dimensional preference spaces** Left: distribution for different values of q and $N = 500, M = 10, Z = 10N$. Right: distribution with $q = 0.7$ and different values of Z.

Algorithm 2: low-dimensional preference space

If the preference space is low dimensional in comparison with the species number, i.e. $N \sim R \ggg M \sim 1$, one can employ an alternative algorithm not relying on the central limit theorem. Consider N vectors of dimension M each generated with just one random non-zero entry sampled by taking the absolute value of a normal distribution $N(m/M, 1/\sqrt{(M)})$, with $m > 0$. Then, the algorithm proceeds as follows:

1. Sample at random two species i, j, i reproduces and j dies.
2. Replace both vectors with a mutation of the vector b_i in one random entry, where $q \in [0,1]$ is the fidelity and $p = 1 - q$ is the mutation. $\boldsymbol{b^i} = q\boldsymbol{b^i} + q\boldsymbol{\epsilon}^i$, $\boldsymbol{b^j} = q\boldsymbol{b^i} + (1-q)\boldsymbol{\epsilon}^j$, with $q \in [0,1]$ is the fidelity of reproduction and $\epsilon^{i,j} = (0,..,\epsilon^{i,j}_k,0,..,0)$ with k chosen at random between 1 and M and $\epsilon^{i,j}_k$ sampled by $\mathcal{N}(m/R, 1/\sqrt{R})$.
3. Iterate Z times.

If the fidelity is moderate but and Z is large the distance distribution develops a long tail independently on the dimension of M, see Fig. 2.11

Model 0: Fluctuating and non overlapping factors.

As a first analysis, we use the Stochastic Logistic model as a null expectation for the correlations. Such a model is easily derived if both the population-dependent and population-independent factors preference vectors are perpendicular to each other (for simplicity one component for vector with modulus m)

$$\bar{r}_i = (\bar{R} + \bar{M})m - \delta, \quad C_{ij} = m\bar{R}\gamma\delta_{ij}, \quad \sigma_i = \nu\bar{M}^2 + \omega\bar{R}^2, \quad \rho_{ij} = \delta_{ij}. \tag{2.44}$$

Indeed, such a model describes logistic growth plus environmental fluctuations of N uncoupled species:

$$\begin{aligned} \dot{x}_i &= x_i\left(\bar{r} - \frac{x_i}{K}\right) + x_i\sqrt{\sigma}\xi_i && (2.45)\\ \langle\xi_i(t)\rangle &= 0 && (2.46)\\ \langle\xi_i(t)\xi_j(t')\rangle &= \delta_{ij}\delta(t-t') && (2.47)\end{aligned}$$

where the carrying capacity is $K = (\bar{R}\gamma)^{-1}$. By considering the following ansatz on the parameters:

$$\begin{aligned} \bar{r}_i &= \frac{1}{\tau_i}, && (2.48)\\ K_i &= \tau_i K_i', && (2.49)\\ \sigma_i &= \frac{\sigma_i'}{\tau_i}, && (2.50)\end{aligned}$$

one recovers the standard form of the stochastic logistic model:

$$\begin{aligned} \dot{x}_i &= \frac{x_i}{\tau_i}\left(1 - \frac{x_i}{K_i'}\right) + \sqrt{\frac{\sigma_i'}{\tau_i}}\xi_i(t)x_i && (2.51)\\ \langle\xi_i(t)\rangle &= 0 && (2.52)\\ \langle\xi_i(t)\xi_j(t')\rangle &= \delta_{ij}\delta(t-t'). && (2.53)\end{aligned}$$

Observe that, by definition, the preference distance is always one, and hence this model gives a symmetric distribution of correlations centered in zero, with spurious variance due to finite size effects that converges to zero in the infinite species and resources limit, see Fig. 2.12. See Sec.A.3 in appendix A for the stationary solution of the SLM.

Model A:Shared fluctuating population-dependent factors.

The presence of population-dependent factors (such as, e.g., sugars) in the environment fluctuates due to seasonality, fluxes with external space etc., inducing both competition between species and fluctuating growth. Thus, if population-independent factors are absent, species interactions are determined by a combination of the effect of competition and resources fluctuations:

$$\begin{aligned} \dot{x}_i &= x_i\left(\bar{r}_i - \sum_j C_{ij}x_j\right) + \sqrt{\sigma}\xi_i(t)x_i \\ \langle\xi_i\rangle &= 0, \\ \langle\xi_i(t)\xi_j(t')\rangle &= \delta(t-t')\rho_{ij}; && (2.54)\end{aligned}$$

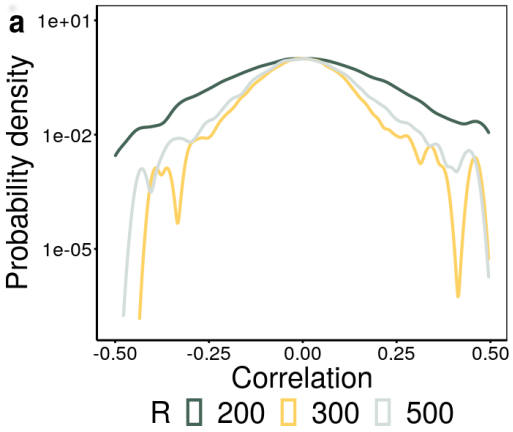


FIGURE 2.12: **Correlations for Model 0.** (a) Stationary probability distribution (log-scale) of correlations for different values of R (color coded) and fixed N. Parameters: $N = 100, K_i = 10.0, \bar{r}_i = 1.0, \sigma_i' = 0.1, q = 0.9, Z = 50N$, and $t_{fin} = 10^4$.

with:

$$\bar{r}_i = \bar{R}\sum_{\beta} b_\beta^i + \bar{M}m - \delta, \quad C_{ij} = \bar{R}\gamma \boldsymbol{b}^i \cdot \boldsymbol{b}^j, \quad \sigma_i = \omega \bar{R}^2, \quad \rho_{ij} = \boldsymbol{b}^i \cdot \boldsymbol{b}^j \tag{2.55}$$

The population-dependent factors preference vector are generated as described in 2.8, Importantly, in the limit $R \gg 1$, the model does not depend on the preference vector but just on the angles between them. Indeed, in such a limit the growth rate depends just on the mean preference

$$\bar{r}_i = \bar{R}R\langle b^i \rangle + \bar{M} - \delta = (\bar{R} + \bar{M})m - \delta. \tag{2.56}$$

Furthermore, the preference vector modulus, i.e. the hyper-sphere radius converges to 1:

$$r_P^2 = \sum_{\beta} b_\beta^2 = 1 + \frac{m^2}{R} \approx 1, \tag{2.57}$$

and hence species interactions depend directly on the preference distance:

$$C_{ij} = \bar{R}\gamma \cos(\theta_{ij}) = \bar{R}\gamma \cos\left(\frac{\pi}{2} d_{P,ij}\right) \tag{2.58}$$

$$\rho_{ij} = \cos(\theta_{ij}) = \bar{R}\gamma \cos\left(\frac{\pi}{2} d_{P,ij}\right). \tag{2.59}$$

Curiously enough, this model leads to a consistent number of extinctions, and the surviving communities are "effectively neutral", i.e. with species pair at almost all distances but with average correlation zero, see Fig. 2.13.

Model B: Shared resources and non-overlapping fluctuating population-independent factors.

We now consider that population-dependent factors do not fluctuate in time, i.e. $\omega = 0$, and that population-independent factors preferences are all perpendicular to each other. Hence species experience independent growth rate fluctuations and competition for resources. Such a model results in competition in preference space for N species and R

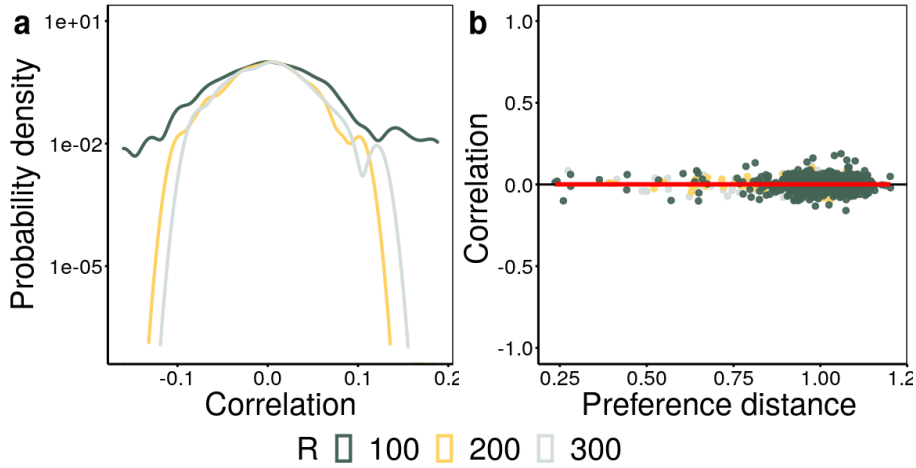


FIGURE 2.13: **Correlations in Model A.** (a) Stationary probability distribution (log-scale) of correlations, and (b) correlations as function of preference distance for different values of R (color coded) and fixed N. The red line in (b) is the correlation averaged over all the presented realizations. As a consequence of limiting similarity various extinction events happen before reaching the stationary state, resulting in the majority of species having distance one, and just a few species in the small distance tail. Note that the correlations do not depend on the species distance, leading to and effective neutral behavior. Parameters: $R = 100, 200, 300$. $q = 0.1$, $Z = 50N$,$N = 100, m = 0.5\bar{R} = \bar{M} = 1, \delta = 0.1, \gamma = 0.3, \omega = 1$.

resources in the form of stochastic Lotka-Volterra equation:

$$\begin{aligned} \dot{x}_i &= x_i\left(r_i(t) - \sum_{j=1}^{N} C_{ij}x_j\right) \\ &= x_i\left(\bar{r}_i - \sum_{j=1}^{N} C_{ij}x_j\right) + \sqrt{\sigma_i}\xi_i(t)x_i, \end{aligned} \tag{2.60}$$

with:

$$\bar{r}_i = \bar{R}\sum_{\beta} b^i_\beta + \bar{M}m - \delta, \quad C_{ij} = \bar{R}\gamma \boldsymbol{b}^i \cdot \boldsymbol{b}^j, \quad \sigma = \nu\bar{M}^2 \quad \rho_{ij} = \delta_{ij}. \tag{2.61}$$

Furthermore, by considering the limit $R \gg 1$ we obtain here a simplified version of the model depending only on the the preference distance matrix $d_{P,ij}$ and not on the vectors $\boldsymbol{b^i}$. Indeed, the deterministic growth rate reads:

$$\bar{r}_i = \bar{R}R\langle b^i\rangle + \bar{M}m - \delta = (\bar{R} + \bar{M})m - \delta, \tag{2.62}$$

$$C_{ij} = \bar{R}\gamma\cos(\theta_{ij}) = \bar{R}\gamma\cos\left(\frac{\pi}{2}d_{P,ij}\right). \tag{2.63}$$

In section2.8 we perform an approximate calculation of the stationary Pearson correlation coefficient, leading to:

$$\eta_{ij}(d_{P,ij}) \approx -\cos\left(\frac{\pi}{2}d_{P,ij}\right). \tag{2.64}$$

predicting a negative to null correlation pattern.‘

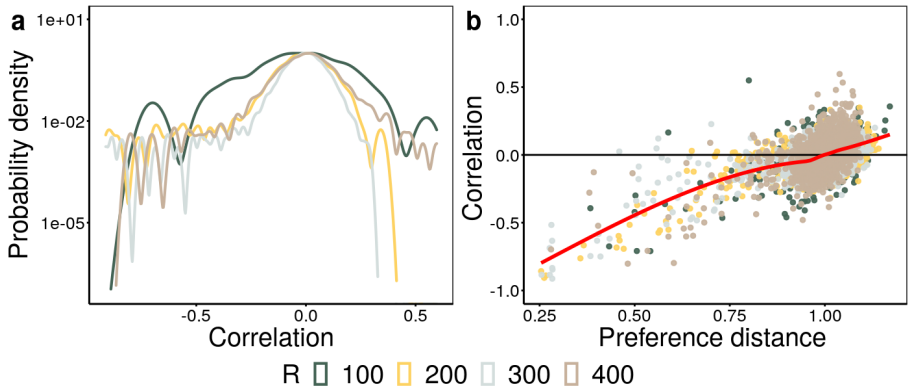


FIGURE 2.14: **Correlations for Model B** (a) Stationary probability distribution (log-scale) of correlations, and (b) correlations as function of preference distance for different values of R (color coded) and fixed N. The red line in (b) is the correlation averaged over all the presented realizations. As a consequence of limiting similarity various extinction events happen before reaching the stationary state, resulting in the majority of species having distance one, and just a few species in the small distance tail. Parameters: $q = 0.1$, $Z = 50N$,$N = 100, m = 0.5\bar{R} = \bar{M} = 1, \delta = 0.1, \gamma = 0.3, \omega = 1$.

Linear approximation around the fixed point

Here we derive the stationary correlation of a couple of species by employing a linear expansion around the fixed point. This approximation assumes that stochastic fluctuations are small and hence the dynamics is localized nearby the deterministic fixed point. We are interested in the relation between the stationary correlation and the interaction coefficients, and hence the preference distance. Hence, for sake of simplicity, but without loss of generality, we consider here the case of gen. LV of two species, interacting through symmetric competition, symmetric noise fluctuations $\sigma_1 = \sigma_2 = \sigma$, carrying capacities $K_1 = K_2 = 1$ and growth rates $r_1 = r_2 = 1$:

$$\begin{aligned} \dot{x}_1 &= x_1 (1. - c x_2 - x_1) + \sqrt{\sigma}\xi_1(t) x_1 \\ \dot{x}_2 &= x_2 (1. - c x_1 - x_2) + \sqrt{\sigma}\xi_2(t)_2, \end{aligned} \tag{2.65}$$

where $c = c_{12} = c_{21}$, cross-diagonal elements of the competition matrix.

By looking at Eq.(2.65), it is clear that there is a deterministic fixed point for each species at:

$$x_1^* = x_2^* = \frac{1}{1+c}. \tag{2.66}$$

Nevertheless, owing to the presence of multiplicative fluctuations it is not possible to perform the analysis around such point in these variables, but it is necessary to go to log abundances, recasting the equations into an additive noise form. In these new variables one can easily perform a linear expansion and, afterward, come back to the original ones. As a first step, one can apply a logarithmic change of variables to make the noise additive:

$$u_i = log(x_i), \tag{2.67}$$

by using the Ito formula Eq.(2.65) becomes:

$$\dot{u_1} = \frac{\dot{x_1}}{x} = (1 - \frac{\sigma}{2} - ce^{u_2} - e^{u_1}) + \sqrt{\sigma}\xi_1(t) \tag{2.68}$$

$$\dot{u_2} = \frac{\dot{x_2}}{x} = (1 - \frac{\sigma}{2} - ce^{u_2} - e^{u_1}) + \sqrt{\sigma}\xi_1(t) \tag{2.69}$$

$$\tag{2.70}$$

The fixed point in this variable is modified by the Ito term:

$$u_1^* = u_2^* = u^* = \frac{1 - \frac{\sigma}{2}}{1 + c}. \tag{2.71}$$

Around it we can perform the linear perturbation analysis:

$$u_i = u^* + \delta u_i \tag{2.72}$$

leading to the following linear equation:

$$\dot{\boldsymbol{\delta u}} = J\boldsymbol{\delta u} + \boldsymbol{\xi} \tag{2.73}$$

where J is the Jacobian of the deterministic part evaluated at the fixed point:

$$J_{u^*} = -\left(\frac{1 - \sigma/2}{1 + c}\right)\begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix} = -a(c)\begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix} \tag{2.74}$$

and diffusion matrix:

$$D = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}. \tag{2.75}$$

Using standard tools in stochastic processes [93], it is easy to see that Eqs.(2.73) at the stationary state converge to bivariate Normal distribution with covariance matrix Σ:

$$P^*(u_1, u_2) = \frac{1}{\sqrt{2\pi det(\Sigma)}} \exp\left(-\frac{1}{2}\boldsymbol{u}^T\Sigma^{-1}\boldsymbol{u}\right) \tag{2.76}$$

$$\Sigma = \frac{1}{Tr(J)det(J)}\begin{pmatrix} D_{22}J_{12}^2 - D_1J_{12}J_{21} + D_{11}J_{22}(J_{11} + J_{22}) & D_{22}J_{11}J_{12} + D_1J_{21}J_2 \\ D_2J_1J_{12} + D_1J_{21}J_2 & D_1J_{21}^2 - D_2J_{12}J_{21} + D_{22}J_{11}(J_{11} + J_{12}) \end{pmatrix} = \tag{2.77}$$

$$= \frac{\sigma}{a(c)(1 - c^2)}\begin{pmatrix} 1 & -c \\ -c & 1 \end{pmatrix},$$

. Now, it is necessary to go back to the original variables by applying the determinant of the transformation:

$$P^*(x_1, x_2) = det\left(\frac{\boldsymbol{du}}{\boldsymbol{dx}}\right)P(u_1, u_2) \tag{2.78}$$

$$= \frac{1}{x_1x_2\sqrt{2\pi det(\Sigma)}} \exp\left(-\frac{1}{2}\log^2 x_1\Sigma_{11}^{-1} - \frac{1}{2}\log^2 x_2\Sigma_{22}^{-1} - \log x_2 \log x_1\Sigma_{12}^{-1}\right).$$

It is immediate to calculate the moments as :

$$\mu_i = u_i^\star \tag{2.79}$$
$$\langle x_i \rangle = e^{\mu_i + \Sigma_{ii}/2} \tag{2.80}$$
$$\langle x_1 x_2 \rangle = e^{\mu_1 + \mu_2 + (\Sigma_{11}+\Sigma_{22})/2 + \Sigma_{12}} \tag{2.81}$$
$$Cov(x_i, x_j) = \left(e^{\mu_i + \mu_j + (\Sigma_{ii}+\Sigma_{jj})/2}\right)\left(e^{\Sigma_{ij}} - 1\right) \tag{2.82}$$
$$Var(x_i) = \left(e^{2\mu_i + \Sigma_{ii}}\right)\left(e^{\Sigma_{ii}} - 1\right). \tag{2.83}$$

In particular, the Person correlation coefficient is:

$$\eta_{12} = \frac{Cov_{12}}{\sqrt{Var_1 Varx_2}} = \frac{e^{\Sigma_{12}} - 1}{\sqrt{e^{\Sigma_{11}} - 1}\sqrt{e^{\Sigma_{22}} - 1}} \approx -c \tag{2.84}$$

Model C: Shared fluctuating population-independent factors with fixed non-overlapping resources.

If resource fluctuations are absent, $\omega = 0$, and population-dependent factors preferences are all perpendicular to each other, i.e. no common resource preferences, species experience dependent growth rate fluctuations and no inter-specific competition:

$$\dot{x}_i(t) = x_i\left(\bar{r}_i - \frac{x_i}{K_i}\right) + \sqrt{\sigma}\xi_i(t)x_i, \tag{2.85}$$

where:

$$\bar{r}_i = \bar{M}\sum_\alpha a^i_\alpha + m\bar{R} - \delta, \quad K_i = (\gamma\bar{R})^(- 1), \quad \rho_{ij} = \boldsymbol{a}^i \cdot \boldsymbol{a}^j \tag{2.86}$$

Hence this model implements the force called "Environmental Filtering". By taking the limit $M \gg 1$, we can recast the model just in terms of preference distance $d_{P,ij}$. The deterministic growth rate reads:

$$\bar{r}_i = \bar{M}\langle a^i\rangle + m\bar{R} - \delta = (\bar{R} + \bar{M})m - \delta; \tag{2.87}$$

while the noise correlation can be rewritten as:

$$\rho_{ij} = \boldsymbol{a}^i \cdot \boldsymbol{a}^j = \cos(\frac{\pi}{2}\theta_{ij}), \tag{2.88}$$

cause

$$|a|^2 = \frac{m^2}{M} + 1 \approx 1. \tag{2.89}$$

Finally, by rescaling the parameter with the growth rate as follow:

$$\bar{r}_i = \frac{1}{\tau_i} \tag{2.90}$$

$$K'_i = K_i\tau_i^{-1} = \frac{(\bar{R} + \bar{M})m - \delta}{\gamma\bar{R}} \tag{2.91}$$

$$\sigma_i = \frac{\sigma'_i}{\tau_i}, \tag{2.92}$$

one obtains the "correlated stochastic logistic" model (CSLM):

$$\frac{dx_i}{dt} = \frac{x_i}{\tau_i}\left(1 - \frac{x_i}{K_i}\right) + \sqrt{\frac{\sigma}{\tau_i}}\xi_i(t)x_i \tag{2.93}$$

$$\langle\xi_i\rangle = 0 \tag{2.94}$$

$$\langle\xi_i(t)\xi_j(t')\rangle = \cos\left(\frac{\pi}{2}d_{P,ij}\right), \tag{2.95}$$

(where the notation has been simplified).

In the following section 2.8, we show that in the linear approximation around the fixed point, the stationary Pearson's correlation reads:

$$\eta_{ij}(d_{P,ij}) = \frac{\exp\left(\cos(\frac{\pi}{2}d_{P,ij})\frac{\sigma}{2-\sigma}\right) - 1}{\exp\left(\frac{\sigma}{2-\sigma}\right) - 1} \approx \cos\left(\frac{\pi}{2}d_{P,ij}\right), \tag{2.96}$$

in the case $\sigma_i = \sigma_j, \tau_i = \tau_j$.

The decay of Pearson's correlation coefficients with preference distance, Eq.(2.96), is

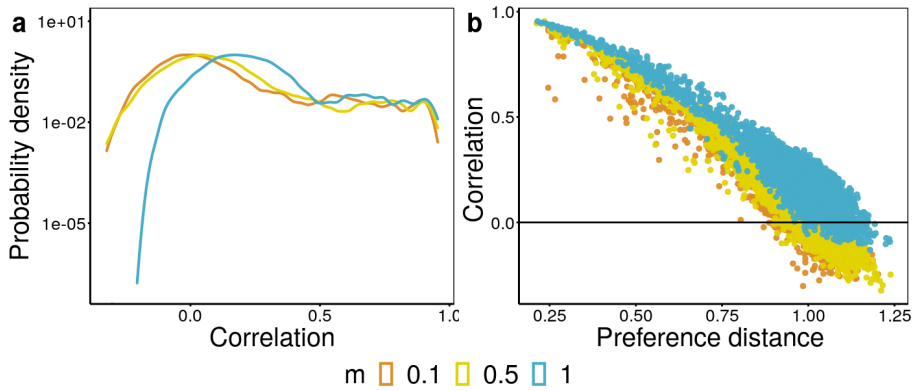


FIGURE 2.15: **Correlation in Model C: dependence on m.** Pearson's correlation distribution (log-scale) (a) and correlations as function preference distance (b) for different values of m (color coded), for $m = 0.1, 0.5, 1$. Parameters: $N = M = 100, q = 0.9, S = 50N, \bar{R} = \bar{M} = 1, \gamma = 0.1, \nu_i = 0.5, t_{fin} = 10^4$.

confirmed by numerical simulations, see Fig.(2.15. Note, in particular, that the deterministic mean preference m divided by the resource number, sets the average correlation, but its decrease implies a large number of extinctions. Furthermore, increasing the number of resources for a constant number of species decreases the mean correlation, its variance (by decreasing the distance distribution one).

In all the models considered here, the effective carrying capacities are constant across species. Here we show how to generalize the calculations to include a carrying capacity distribution. To generalize this it is sufficient to consider the consumption parameter timescale γ_i, to be species dependent. In the present model case each resource is consumed exclusively by one species ($N = R$), hence:

$$R_i(t) = \bar{R}\left(1 - m\gamma_i x_i\right), \tag{2.97}$$

leading to an effective carrying capacity (rescaled by the growth rate):

$$K_i = \frac{\left(\bar{R} + \bar{M}\right) m - \delta}{m^2 \gamma_i \bar{R}} \tag{2.98}$$

To obtain such a pattern, we fixed that the modulus of all preference vectors to one, i.e. $|\boldsymbol{a}_i| = 1$ for $i = 1, 2$ and also fixed preference vector the first species as $\boldsymbol{a}_1 = (1, 0)$. In each simulation (i.e. for each gray point in the figure), we choose the desired preference distance between the two species and derived the component of the second vector as:

$$\begin{aligned} a_2^1 &= \cos\left(\frac{\pi}{2} d_P\right) \\ a_2^2 &= \sqrt{1 - (a_2^1)^2}. \end{aligned} \tag{2.99}$$

In the next section (2.8) we present a more careful explanation on why the pattern is also valid only for two species (at tunable distances).

Linear approximation around the fixed point

Here we derive the stationary correlation of a couple of species by employing a linear expansion around the fixed point. This approximation assumes that stochastic fluctuations are small and hence the dynamics is localized nearby the deterministic fixed point. By

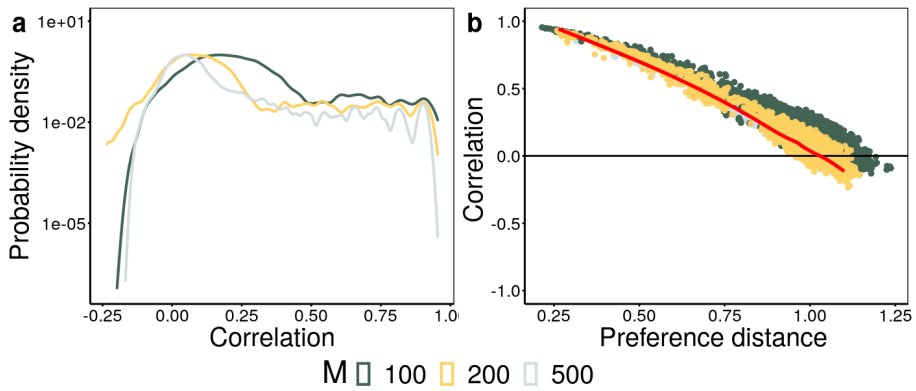


FIGURE 2.16: **Correlation in Model C: dependence on the number of resources.** Pearson correlation distribution (log-scale) (a) and correlations as function preference distance (b) for different values of M (color coded), $M = 100, 200, 500$. The red line in (b) stands for eq.2.96. As expected, by increasing the number of resources, the correlation distribution variances decreases, and the decay is well-approximated by the linear noise approximation. Parameters: $N = 100, q = 0.9, S = 50N, m = 1, \bar{R} = \bar{M} = 1, \gamma = 0.1, \nu_i = 0.5$, and $t_{fin} = 10^4$.

looking at Eq.(2.93), it is clear that the deterministic fixed point for each species is $x_i^* = K_i$. Nevertheless, owing to the presence of multiplicative fluctuations it is not possible to perform the analysis in these variable, but it is necessary to linearize the noise-correlation matrix and recast the equations into an independent additive noise form. In these new variables one can easily perform a linear expansion and, afterward, come back to the original ones. As a first step, one can apply a logarithmic change of variables to make the noise additive:

$$u_i = log(x_i), \tag{2.100}$$

by using the Ito formula Eq.(2.93) becomes:

$$\dot{u}_i = \frac{\dot{x}_i}{x} = \frac{1}{\tau_i}\left(1 - \frac{\sigma}{2}\right) - \frac{e^{u_i}}{\tau_i K_i} + \sqrt{\frac{\sigma_i}{\tau_i}}\zeta_i(t). \tag{2.101}$$

As a second step, one needs to find eigenvalues and eigenvectors of the correlation matrix. Given that we are interested in the two-species correlations, we restrict ourselves here, without loss of generality, to an arbitrary couple of species (i, j). Furthermore, for sake of notation let us rename the parameters as:

$$W_i = \sqrt{\frac{\sigma_i}{\tau_i}}, \tag{2.102}$$

$$\rho_{ij} = \rho = \cos\left(\frac{\pi}{2} d_{P,ij}\right), \tag{2.103}$$

and include these factors in the noise correlation matrix:

$$C_{ij} = \begin{pmatrix} W_i^2 & W_i W_i \rho_{ij} \\ W_i W_j \rho_{ij} & W_j^2 \end{pmatrix}. \tag{2.104}$$

Hence, the eigenvalues of the reduced correlation matrix are:

$$\lambda_{i,j} = \frac{1}{2}(W_i^2 + W_j^2 \pm \sqrt{W_i^4 + 2(2\rho^2 - 1)W_i^2 W_j^2 + W_j^4}), \tag{2.105}$$

and the corresponding eigenvectors

$$v_{i,j} = \left(-\frac{W_j^2 - W_i^2 \pm \sqrt{W_i^4 + 2(2\rho^2 - 1)W_i^2 W_j^2 + W_j^4}}{2W_i W_j \rho}, 1 \right). \tag{2.106}$$

Note that we have to impose $\rho \neq \pm 1$ in order to avoid having a degenerate transformation. For the sake of simplicity let us assume the following ansatz on the parameters:

$$W_i = W_j \rightarrow \tau_i = \tau_j = \tau, \sigma_i = \sigma_j = \sigma. \tag{2.107}$$

In this case, the eigenvalues and eigenvectors reduce to:

$$\lambda_{i,j} = W^2(1 \pm \rho), \tag{2.108}$$

and

$$v_{i,j} = \frac{1}{\sqrt{2}}(1, \pm 1), \tag{2.109}$$

where we have also normalized the vectors. Now, it is possible to decompose the noise as:

$$\zeta_i = \sum_\mu \sqrt{\lambda_\mu} v_{\mu,i} \eta_\mu(t), \tag{2.110}$$

with

$$\begin{aligned} \langle \eta_\mu \rangle &= 0, \\ \langle \eta_\mu(t) \eta_\gamma(t') \rangle &= \delta(t - t') \delta_{\mu,\gamma}; \end{aligned} \tag{2.111}$$

and the dynamical variable as

$$p_\mu = \sum_i u_i v_{\mu,i}. \tag{2.112}$$

In our simple case the variables are:

$$\begin{aligned} p_i &= \frac{u_i + u_j}{\sqrt{2}}, \\ p_j &= \frac{u_i - u_j}{\sqrt{2}}. \end{aligned} \tag{2.113}$$

Hence, the new Langevin equations are

$$\begin{aligned} \dot{p}_i &= \frac{\sqrt{2}}{\tau} - \frac{1}{\sqrt{2}\tau} \left(\frac{e^{\frac{p_i + p_j}{\sqrt{2}}}}{K_i} + \frac{e^{\frac{p_i - p_j}{\sqrt{2}}}}{K_j} + \sigma \right) + \sqrt{\frac{\sigma}{\tau}(1 + \rho)} \eta_i(t), \\ \dot{p}_j &= -\frac{1}{\sqrt{2}\tau} \left(\frac{e^{\frac{p_i + p_j}{\sqrt{2}}}}{K_i} - \frac{e^{\frac{p_i - p_j}{\sqrt{2}}}}{K_j} \right) + \sqrt{\frac{\sigma}{\tau}(1 - \rho)} \eta_j(t), \\ \langle \eta_i(t) \rangle &= 0, \\ \langle \eta_i(t) \eta_j(t') \rangle &= \delta_{ij} \delta(t - t'). \end{aligned} \tag{2.114}$$

Note that the deterministic force admits a scalar potential

$$V(p_i, p_j) = -\frac{\sqrt{2}p_i}{\tau} + \frac{1}{\tau}\left(\frac{e^{\frac{p_i+p_j}{\sqrt{2}}}}{K_i} + \frac{e^{\frac{p_i-p_j}{\sqrt{2}}}}{K_j} + \frac{\sigma}{\sqrt{2}}p_j\right), \tag{2.115}$$

and the minimum of the potential gives the fixed point of the dynamics, namely:

$$\left(p*_i, p_j^*\right) = \frac{1}{\sqrt{2}}\left(ln\left[\left(1-\frac{\sigma}{2}\right)^2 K_i K_j\right], ln\left(\frac{K_i}{K_j}\right)\right). \tag{2.116}$$

Importantly, note that the fixed point, when converted to the original variables, is different from the fixed point of the deterministic dynamics in eq.(2.93):

$$(x_i^*, x_j^*) = (K_i(1-\frac{\sigma}{2}), K_i(1-\frac{\sigma}{2})). \tag{2.117}$$

If one now considers $\sigma^2 \ll 1$ it is interesting to study the linear fluctuations around the fixed point, like in the small-noise approximation:

$$p_i = p_i^* + \delta p_i, \tag{2.118}$$

the Jacobian evaluated at the fixed point reads:

$$J_{p^*} = -\frac{1}{2\tau}(1-\frac{\sigma}{2})\begin{pmatrix} \frac{e^{\frac{p_i+p_j}{\sqrt{2}}}}{K_i} + \frac{e^{\frac{p_i-p_j}{\sqrt{2}}}}{K_j} & \frac{e^{\frac{p_i+p_j}{\sqrt{2}}}}{K_i} - \frac{e^{\frac{p_i-p_j}{\sqrt{2}}}}{K_j} \\ \frac{e^{\frac{p_i+p_j}{\sqrt{2}}}}{K_i} - \frac{e^{\frac{p_i-p_j}{\sqrt{2}}}}{K_j} & \frac{e^{\frac{p_i+p_j}{\sqrt{2}}}}{K_i} + \frac{e^{\frac{p_i-p_j}{\sqrt{2}}}}{K_j} \end{pmatrix}_{p^*} = -\frac{1}{\tau}\left(1-\frac{\sigma}{2}\right)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.119}$$

Therefore, one ends up with the following linearized dynamics

$$\dot{\boldsymbol{\delta p}} = J\boldsymbol{\delta p} + \boldsymbol{\eta}, \tag{2.120}$$

with the Jacobian above and (rescaled) diffusion matrix:

$$D = \begin{pmatrix} D_i & 0 \\ 0 & D_j \end{pmatrix} = \frac{W^2}{2}\begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix}, \tag{2.121}$$

such that

$$\langle \eta_\mu(t)\eta_\gamma(t')\rangle = 2D_{\mu,\gamma}\delta(t-t'). \tag{2.122}$$

We use the following notation for the Jacobian:

$$J = -\begin{pmatrix} J_i & -J_{ij} \\ -J_{ji} & J_j \end{pmatrix}. \tag{2.123}$$

In this scenario, it is know that the stationary distribution is a bivariate Gaussian with covariance matrix[242]

$$\Sigma = \frac{1}{Tr(J)det(J)}\begin{pmatrix} D_j J_{ij}^2 - D_i J_{ij}J_{ji} + D_i J_j(J_i+J_j) & D_j J_i J_{ij} + D_i J_{ji} J_j \\ D_j J_i J_{ij} + D_i J_{ji} J_j & D_i J_{ji}^2 - D_j J_{ij} J_{ji} + D_j J_i(J_i + J_{ij}) \end{pmatrix} \tag{2.124}$$

$$= \frac{\sigma}{2-\sigma}\begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix},$$

$$P^*(\delta p_1, \delta p_2) = \frac{2-\alpha\sigma}{2\pi\sigma\sqrt{1-\rho^2}} e^{-\left(\frac{1}{\sigma}-\frac{\alpha}{2}\right)\left(\frac{\delta p_1^2}{1+\rho}+\frac{\delta p_2^2}{1-\rho}\right)}. \tag{2.125}$$

Now it is necessary to go back to the original variables x_i, taking care of the modulus of the determinant of the Jacobian of the transformation H:

$$\begin{aligned} \boldsymbol{\delta p} &= H(\boldsymbol{x}), \\ \delta p_i = p_i - p_i^* &= \frac{1}{\sqrt{2}} ln\left(\frac{x_i x_j}{\left(1-\frac{\sigma}{2}\right) K_i K_j}\right), \\ \delta p_j = p_j - p_j^* &= \frac{1}{\sqrt{2}} ln\left(\frac{x_1 K_2}{K_1 x_2}\right), \end{aligned} \tag{2.126}$$

$$\begin{aligned} P(\boldsymbol{x}) &= P(\boldsymbol{\delta p}(\boldsymbol{x}))|det(dH)|) = \frac{1}{x_i x_j} P(\boldsymbol{\delta p}(\boldsymbol{x})) \qquad (2.127) \\ &= \frac{2-\sigma}{2\pi x_i x_j \sigma\sqrt{1-\rho^2}} \\ &\quad e^{-\left(\frac{2-\sigma}{2\sigma(1-\rho^2)}\right)\left[ln^2\left(\frac{x_i}{(1-\sigma/2)K_i}\right)+ln^2\left(\frac{x_j}{(1-\sigma/2)K_2}\right)-2\rho ln\left(\frac{x_j}{(1-\sigma/2)K_i}\right)ln\left(\frac{x_j}{(1-\sigma/2)K_j}\right)\right]}. \end{aligned}$$

Note that, for the implicit assumptions of the approximation, this distribution is a log-normal one, while we know that the marginal distribution needs to be a Gamma. This can be appreciated from the moments:

$$\langle x_i \rangle = K_i \left(1-\frac{\sigma}{2}\right) e^{\frac{\sigma}{2(2-\sigma)}} \tag{2.128}$$

$$\langle x_i x_j \rangle = K_i K_j \left(1-\frac{\sigma}{2}\right)^2 e^{\frac{(1+\rho_{ij})\sigma}{2-\sigma}} \tag{2.129}$$

$$\langle x_i^2 \rangle = K_i^2 \left(1-\frac{\sigma}{2}\right)^2 e^{\frac{2\sigma}{2-\sigma}} \tag{2.130}$$

$$Var_i = K_i^2 \left(1-\frac{\sigma}{2}\right)^2 (e^{\frac{2\sigma}{2-\sigma}} - e^{\frac{\sigma}{2-\sigma}}). \tag{2.131}$$

Nevertheless, one can calculate the covariance

$$Cov_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = K_i K_j \left(1-\frac{\sigma}{2}\right)^2 (e^{\frac{(1+\rho_{ij})\sigma}{2-\sigma}} - e^{\frac{\sigma}{2-\sigma}}), \tag{2.132}$$

and finally the Pearson correlation coefficient:

$$\begin{aligned} \eta_{ij} &= \frac{Cov_{ij}}{\sqrt{Var_i Var x_j}} = \frac{e^{(1+\rho_{ij})\frac{\sigma}{2-\sigma}} - e^{\frac{\sigma}{2-\sigma}}}{e^{\frac{2\sigma}{2-\sigma}} - e^{\frac{\sigma}{2-\sigma}}} \qquad (2.133) \\ &= \frac{e^{\rho\frac{\sigma}{2-\sigma}} - 1}{e^{\frac{\sigma}{2-\sigma}} - 1} \simeq \rho_{ij} = cos(\frac{\pi}{2} d_{P,ij}). \qquad (2.134) \end{aligned}$$

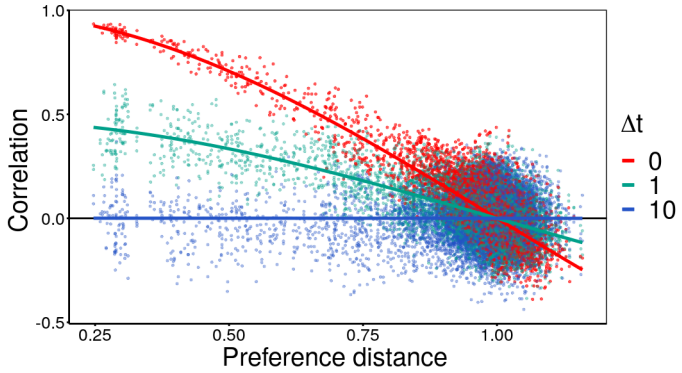


FIGURE 2.17: Delayed correlations as function of preference distance. Points represent species person correlation coefficient at stationary state at equal time (red) and with delay 1 (green), and 10 (blue) for one realization of the model. Solid lines represent the analytically-derived formula Eq.(2.142). Parameters are equal for all species, the variability is given to the variability in preferences distribution, as plotted in Fig. 2.17. Parameters: $N = 100, M = 100, q = 0.9, Z = 50N, m = 0.5, \omega = 0.1, \bar{r} = 1, K = e^2$, and $t_{fin} = 10^4$.

Temporal behavior

Eq.(2.93) can be formally solved exactly, leading to:

$$x_i(t) = \frac{K_i \tau x_i(0) e^{(1-\frac{\sigma}{2})\frac{t}{\tau} + \sqrt{\frac{\sigma}{\tau}} W_i(t)}}{x_i(0) I_i[0,t] + 1}, \tag{2.135}$$

$$W_i(t) = \int_0^t ds \zeta_i(s), \tag{2.136}$$

where $I_i[0,t]$ is the integral of the associated "geometric Brownian motion":

$$I_i[0,t] = \int_0^t ds \exp\left((1-\frac{\sigma}{2})\frac{s}{\tau} + \sqrt{\frac{\sigma}{\tau}} W_i(s)\right). \tag{2.137}$$

The exact integral eq.(2.135) can be used to understand the effect of delay in the dynamics. Namely, assuming that the system is in its stationary state ($t \to \infty$), one can compute heuristically the relation between the abundance at time t and at a later time $t + \Delta t$:

$$\begin{aligned} x_i(t+\Delta t) &\approx K_i \tau e^{(1-\frac{\sigma}{2})\frac{t+\Delta t}{\tau} + \sqrt{\frac{\sigma}{\tau}} W_i(t+\Delta t)} I_i[0,t+\Delta t]^{-1} \\ &= x_i(t) \exp\left((1-\frac{\sigma}{2})\frac{\Delta t}{\tau}\right) \kappa(t,\Delta t), \end{aligned} \tag{2.138}$$

$$\kappa(t,\Delta t) = \exp\left(\sqrt{\frac{\sigma}{\tau}} \int_t^{t+\Delta t} \zeta_i(s)\right) \left(\frac{I_i[0,t]}{I_i[0,t+\Delta t]}\right); \tag{2.139}$$

the function κ in the limit $t \gg \Delta t$ converges to 1, such that

$$x_i(t+\Delta t) \approx e^{(1-\frac{\sigma}{2})\frac{\Delta t}{\tau}} x_i(t). \tag{2.140}$$

By inserting it in the Pearson coefficient formula we obtain:

$$\eta_{ij}(\Delta t) \approx e^{(1-\frac{\sigma}{2})\frac{\Delta t}{\tau}} \eta_{ij}(0), \tag{2.141}$$

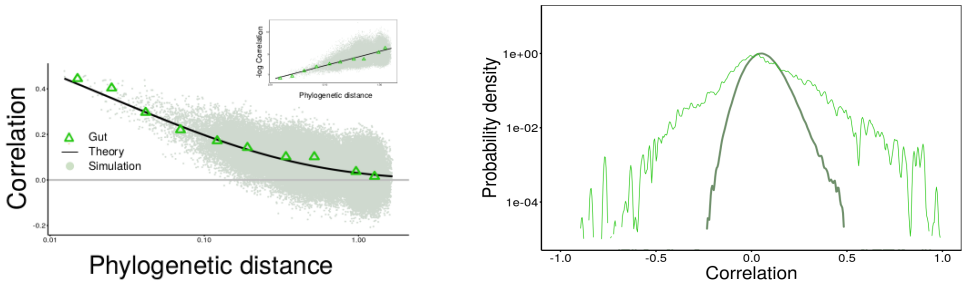


FIGURE 2.18: **CSLM in phylogenetic space for the Gut microbiome** Left: Correlations versus the phylogenetic distance, both for the gut biome data (green triangles), the model simulation (green cloud of points), and analytic formula eq.(2.150)(black line) $\lambda = 3.5$. The model has been simulated 10 times with $N = 300$ species, using as input the empirical phylogenetic distance matrix of the gut, randomly sampling from it N species. Inset: -log Correlations vs phylogenetic distance in log-log scale, empirically and from the model, same data as the main figure. Dark-green points are the Pearson correlation coefficient at the stationary state. Right: Correlation distribution for data and model (log-scale). Light green line is the empirical probability density of correlation for the gut microbiome, while the dark green one corresponds to the correlation distribution of the CSLM (dark green points in left figure.) Carrying capacities are generated log-normally by taking the exponential of random variables sampled by a Gaussian $N(\bar{K}, \sigma_K)$, $\tau_i = \tau$ and $\sigma_i = \sigma$ for $i = 1, .., N$. Parameters: $\tau = 1, \bar{K} = 16.1, \sigma_K = 3.8, \sigma = 1.42, \lambda = 3.5, t_f = 10^4$.

that combined with the linear approximation, Eq.(2.133), finally leads to:

$$\eta_{ij}(\Delta t) \approx e^{(1-\frac{\sigma}{2})\frac{\Delta t}{\tau}} \cos\left(\frac{\pi}{2} d_{P,ij}\right) \tag{2.142}$$

which quantifies analytically the time-delayed correlations.

CSLM in phylogenetic space

To compare the CSLM with empirical data it is necessary to connect the preference distance with the phylogenetic one. Let us $d_{G,ij}$ and $d_{P,ij}$ the phylogenetic and phenotypic distance of the couple of species ij, and $d_G = \langle d_{G,ij} \rangle_{d_{G,ij} \epsilon b}$, $d_P = \langle d_{P,ij} \rangle_{d_{P,ij} \epsilon b}$ the average genetic and preference distance within a given bin. Note that the empirical relationship we aim at reproducing is between the average correlation and the averaged phylogenetic distance in each bin, i.e.:

$$\eta(d_G) = \langle \eta_{ij} \rangle_{d_{G,ij} \epsilon b} = \exp(-\lambda d_G^{1/3}), \tag{2.143}$$

Indeed, we are not interested in the full probability distribution of correlation in one bin, but just in its mean value. Hence, it is sufficient to find a relation between the average distance d_P and d_G, and not between the the full matrices. On the other hand, the CSLM relates the species correlation with their preference distance:

$$\eta_{ij} = \cos(\frac{\pi}{2} d_{P,ij}); \tag{2.144}$$

hence, the average preference distance can be calculated by inverting the formula for the correlation as

$$d_{P,ij} = \frac{2}{\pi} arcos(\eta_{ij}), \tag{2.145}$$

and by taking averages over the couples in the considered bin:

$$d_P(b) = \left\langle d_{P,ij} \right\rangle_b = \left\langle \frac{2}{\pi} \arccos(\eta_{ij}(d_{G,ij})) \right\rangle_b \tag{2.146}$$

To evaluate the last term it would be necessary to know the exact distribution of phylogenetic distance in each bin, but such a distribution is highly non-universal and difficult to quantify. To circumvent the problem, we apply a "mean- field approximation", i.e. we neglect the variance in each bin and consider just its mean:

$$\begin{aligned} d_P(b) &= \left\langle \frac{2}{\pi} \arccos(\eta_{ij}(d_{G,ij})) \right\rangle_b \approx \frac{2}{\pi} \arccos\left(\left\langle \eta_{ij}(d_{G,ij}) \right\rangle_b \right) \\ &= \frac{2}{\pi} \arccos(\eta(d_G)) = \frac{2}{\pi} \arccos\left(e^{-\lambda d_G^{1/3}} \right). \end{aligned} \tag{2.147}$$

To incorporate such a relation in the CSLM let us note once more that we are interested in average pattern of correlation as function phylogeny, and hence it is sufficient to assume Eq. (2.147) true also for the full phylogenetic matrix, i.e.:

$$d_{P,ij} \approx \frac{2}{\pi} \arccos\left(e^{-\lambda d_{G,ij}^{1/3}} \right). \tag{2.148}$$

Putting together all these ingredients, one finally obtains the CSLM in phylogenetic space:

$$\begin{aligned} \frac{dx_i}{dt} &= \frac{x_i}{\tau_i}\left(1 - \frac{x_i}{K_i}\right) + \sqrt{\frac{\sigma_i}{\tau_i}} x_i \xi_i(t) \\ \langle \zeta_i(t) \rangle &= 0 \\ \langle \zeta_i(t) \zeta_j(t') \rangle &= e^{-\lambda d_{G,ij}^{1/3}}. \end{aligned} \tag{2.149}$$

By applying Eq.(2.142) to this specific setting, one finally obtains a theoretical prediction relating correlation and phylogeny:

$$\eta_{ij} = e^{-\lambda d_{G,ij}^{1/3}}, \tag{2.150}$$

$$\eta_{ij}(\Delta t) = e^{-(1-\frac{\sigma}{2})\frac{\Delta t}{\tau}} e^{-\lambda d_{G,ij}^{1/3}}. \tag{2.151}$$

In Fig(2.18) we show that the model correctly predicts the decaying pattern of correlation versus phylogenetic distance. However, it cannot reproduce the full correlation distribution. The correlations are obtained from 10 realizations with $N = 300$ species, the averaged are over 10^3 abundances sampled during the stationary time series every $\delta_t = 10\tau$. To generate the noise correlations, in each realization we sampled N species from the phylogenetic distance matrix of random community of considered biome. Parameters are set to reproduce also the species marginal properties and delayed correlations, see Figs.(B.15-B.18).

Macroecological laws and marginal properties

The CSLM, in the Ito discretization scheme, has a Gamma stationary *marginal* distribution, see Sec.A.3 in appendix A [207, 125]:

$$P^*(x_i) = \frac{1}{\Gamma(\beta_i)} \left(\frac{\beta_i}{\bar{x}_i} \right)^{\beta_i} x_i^{\beta_i - 1} \exp\left(-\beta_i \frac{x}{\bar{x}_i} \right), \tag{2.152}$$

where the average abundance $\bar{x}_i$ and the squared inverse coefficient of variation β_i read:

$$\bar{x}_i = K_i\left(1-\frac{\sigma_i}{2}\right), \tag{2.153}$$

$$\beta_i := \frac{\bar{x}_i^2}{Var_i} = \frac{2}{\sigma_i}\left(1-\frac{\sigma_i}{2}\right)^2, \tag{2.154}$$

respectively, coinciding with the ones obtained for the standard SLM [207]. Hence, the CSLM is able to reproduce the three macroecological laws for diversity and fluctuation, namely:

1. The stationary marginal distribution of species abundances is a Gamma distribution.

2. By fixing $\sigma_i = \sigma$, for all species the Taylor law relating the mean and variances across species is recovered.

3. The mean abundances are distributed as a log-normal just by imposing that the K_i's are log-normally distributed too.

Part II

Eco-evolution of bacterial populations

Chapter 3

Statistical mechanics of phenotypic eco-evolution

3.1 Introduction

Darwinian evolution relies on the fundamental principles of reproduction, mutation and selection, and describes how populations change over time and how new forms can branch out from existing ones [13, 127]. The theory of evolutionary dynamics —whose aim is that of formalizing the ideas of Darwinian evolution from a conceptual and quantitative perspective— has become a wide and mature discipline at the crossroad between biology, mathematics and statistical physics (see e.g. [128, 129, 130, 131, 132, 133] and refs. therein). Diverse theoretical approaches, differing in various aspects, have been proposed to model and rationalize evolutionary phenomena [159] but, rather generically, evolutionary dynamics is formulated building on the theories of stochastic processes and dynamical-systems [243, 244, 64, 245, 246, 247].
Even in such a broad territory common overarching concepts and principles have been found. For instance, G.R. Price in his seminal paper "*Mathematical Theory of Selection*" [248], set the bases for the development of a general and abstract mathematical theory of selection; the so-called "Price equation", which is a keystone in evolutionary dynamics —sometimes called termed the "algebra of evolution" [245, 249, 250]— expresses the rate of abundance change of a given gene or trait as a function of its covariance with the associated fitness function [251, 248, 249, 252, 250, 253, 254, 255]. Later, Page and Nowak showed that the different *"macroscopic"* formulations of evolutionary dynamics are actually equivalent in some way [159] (see also, [256, 257, 149]); however, specific details depend on whether the focus is on genes (*genotypic evolution*) or on phenotypic traits within a population (*phenotypic evolution*).
Population genetics, i.e. the study of genetic variations within a population (or between different populations) together with the evolutionary factors causing such a variation, was developed starting from classical works of Fisher [258, 154], Wright [259], Price [251], Kimura [243, 244] and others, including recent exciting developments (e.g. [169, 260, 261, 262, 160]). On the other hand, quantitative approaches to evolutionary dynamics focusing on phenotypes or traits (rather than on "genotypes") have also extensively developed in the last decades. These include *evolutionary game theory* [263, 264, 265, 266, 267, 268, 269, 246], which aims to study possible equilibria of different populations with fixed traits (or "strategies") as well as *adaptive dynamics* [270, 136, 137, 271, 147, 139, 272, 273], which includes the possibility of mutations (so that phenotypes are not discrete but can change in a continuum) as well as some other relevant ecological ingredients (see below).
A common assumption often considered in these theoretical approaches is that mutations occur at a relatively low pace so that ecological (fast) and evolutionary (slow) timescales are well separated. This assumption which has been historically retained as a very natural one is not, however, universally valid, and it fails, e.g., when dealing with populations

of micro-organisms (see, e.g., [274, 275, 276, 277]).
Microbial populations are the most abundant and diverse ones in the biosphere. Studying their dynamics has become a crucial challenge from many diverse viewpoints including, e.g., environmental and health-related perspectives. As a matter of fact, microbial systems constitute the ultimate frontier and test bench of modern evolutionary theory. Indeed, often, microbial communities evolve very rapidly, with frequent mutations and fast selection. For instance, in viruses and bacteria the mutation rates are astonishing high [278, 279, 280, 281]. As a consequence evolutionary effects cannot be decoupled from ecological ones: they both can occur at similar timescales [282, 283]. As a result, the communities exhibit a very large fine-scale diversity in the form of multiple co-occurring phenotypes [1]. Moreover, such a diversity is nowadays also accessible to experiments owing to technological advances in determining single-cell traits [168, 169, 170] and metabolic functions [171, 172, 45, 285].
All this calls for the development of novel eco-evolutionary frameworks, extending existing ones, to analyze complex microbial communities, distributed in phenotypic space and evolving on ecological time scales.

Adaptive dynamics and its extensions

The standard theory of phenotypic evolution is *adaptive dynamics* (AD) [270, 136, 137, 147, 135, 286, 139, 273], see Sec.1.7 for a brief introduction. In adaptive dynamics a population is assumed to be in a steady state, called "resident type" and small variations of such a type, i.e. "mutants", are assumed to emerge at a very slow rate. In the case that the per-capita growth rate of the mutant within the resident-type population —i.e, its "invasion fitness"— is positive, the mutant is assumed to invade the population and, eventually, become fixated as the new resident type [287].
Within the standard AD framework, mutations are considered to be: (i) rare (which implicitly assumes a large separation of timescales between ecological and evolutionary processes), such that the system has time to re-equilibrate to an equilibrium after each mutation; (ii) small (as a result of which the stationary phenotypic probability distributions are typically Gaussians); (iii) independent of the parental traits; and (iv) not subject themselves to evolution [270, 137, 135, 286, 139].
Since its original formulation, the theory of AD enjoyed a great success. Its main advance was to unify evolutionary dynamics with realistic ecological scenarios [135]. For instance, importantly, AD allows for the possibility of "evolutionary branching" –i.e. the split of an initially monomorphic population in two diverse sub-populations— shedding light on how speciation [136, 137, 135, 138, 139, 140, 141] and diversification in sympatry [286] may come about. Similarly, phenomena such as the evolution of dispersal strategies [142], pathogenicity [143], metabolic preferences [144, 145] and multi-cellularity [146], to name but a few, have been successfully addressed within the context of AD. Moreover, extensions of AD have been developed to include ingredients such as finite-sized populations [147, 148], species interactions [149], sexual populations [139], multi-dimensional phenotypic spaces [150, 151, 152, 153, 154], variable environmental conditions [155], or variability in the evolutionary outcomes, to name but a few. Nevertheless, given the above-mentioned restrictive hypotheses, there is space to generalize AD to describe microbial populations, allowing, e.g., for frequent mutations at ecological timescales.
From the perspective of statistical mechanics, it would be highly desirable to construct a general stochastic individual-based ("many-body" or "many-particle") theory for agents in a community exposed to basic rules of reproduction, mutation, and selection, such

[1] Actually, the concept of "*quasi species*" —rather that that of "species"— might be better suited to describe them [284].

that it could reproduce the above theories in some "macroscopic" limit. The goal is that, starting from a "microscopic" description of stochastic processes acting at the level of interacting individuals (including ecological and evolutionary processes at comparable timescales), one should be able to derive macroscopic probabilistic equations for the evolution of populations and communities by employing the powerful methods of statistical mechanics.
Before proceeding toward this goal, let us remark that diverse approaches have already tackled the previous challenge and that significant advances have already been made in this direction. In particular, we extensively elaborate and rely upon the following seminal works:

- Dieckmann and Law were pioneers in deriving the macroscopic equations of AD using a probabilistic description of the population in phenotypic space. Their approach allows e.g. for the possibility that potentially-successful mutants become accidentally extinct owing to demographic fluctuations before achieving fixation [147].
- Champagnat and colleagues built a rather rigorous mathematical framework, allowing them to derive the "macroscopic" equations of AD starting from "microscopic" underlying birth-and-death stochastic processes [288, 289, 290, 291]. They also generalized ADs conditions in various ways, including mutation properties, population sizes, etc.
- Frey and coauthors [292, 293], developed a formalism to derive a macroscopic equation from the Master equation of an underlying birth-death process in the context of bacterial populations.
- Wakano and Iwasa [148] were pioneers in studying the effect of finite-size fluctuations in the context of adaptive dynamics (see also [291, 141]).

In a similar spirit to these and related approaches, here we develop a general framework, deeply rooted in the views and methods of statistical mechanics, able to generalize adaptive dynamics to eco-evolutionary scenarios. In particular, we present a probabilistic theory of the evolution of *trait distributions* [294] including the effect of selection, arbitrary mutations, and fluctuations stemming from finite population sizes, where ecological and evolutionary processes occur contemporaneously. [2]
The work presented in this chapter contributes to the development of an eco-evolutionary theory for microbial communities, allowing to shed further light on the empirically-observed astonishing diversity in traits and interactions of microbial communities. Our hope is that the present work makes this kind of quantitative approaches to complex eco-evolutionary communities accessible to a broader audience, including physicists, biologists, and ecologists.

Structure of the chapter

In Sec.3.2, we introduce the general eco-evolutionary framework. We start defining a general individual- or agent- based birth-death process involving reproduction, selection, mutation and drift (Sec.3.2). Next, we derive from such a *microscopic* description a *macroscopic* one in the form of a mean-field equation describing the evolution of a general trait distribution in phenotypic space (Sec. C.1.1). Effective equations for the first

[2] Our approach has the same intention and spirit of the more-mathematical framework of Champagnat and collaborators [288, 290], but it is an independent one, based on standard approaches and concepts in physics.

moments of such a distribution are then obtained (Sec. 3.2) and particularized to the case of small mutations (Sec.3.2). This allows to recover the standard theory of AD using a Gaussian approximation for the trait distribution (Sec. 3.3) and, then, to formulate an extended theory "*a la Landau*", including higher moments in the expansion, to go beyond it (Sec. 3.4). For illustration purposes, the general theory is then applied to explain the behaviour of a specific individual-based evolutionary model, including both a fixed ecological niche and competition between individuals. Finally, in Sec. 3.6 we generalize the resulting deterministic theory to include stochastic/demographic effects, stemming from finite-size populations. To close, in Sec.3.7, we discuss our main conclusions as well as possible future developments.

3.2 General framework

For the sake of simplicity and without loss of generality, let us consider —as customarily done in evolutionary dynamics [295]— a population of fixed size, composed of N individuals. As in adaptive dynamics (AD), we choose to focus on a phenotypic description of individuals or "agents". Thus, each of them (say the i-th one, with $i \in [1, N]$) is characterized by a set of phenotypic traits that, in the simplest possible case, can be encoded in a single real-valued variable, x_i. This represents a coordinate in a one-dimensional phenotypic space $\mathcal{P}$. Generalizations to higher-dimensional phenotypic spaces and to populations of variable size are straightforward and will also be addressed in what follows. The population as a whole can be represented by a N-dimensional vector $\boldsymbol{x} = (x_1, x_2, x_3, .., x_N)$, that we call a *phenotypic configuration*. The goal is to describe the dynamics of the probability $P(\boldsymbol{x}, t)$ to observe a population with a given phenotypic configuration, $\boldsymbol{x}$ as a function of time. Note that we changed the vector notation in respect of Sec.1.5 to emphasize that the components are individual traits and to have a more simply notation, see the next section.

Birth-and-death eco-evolutionary process.

The dynamics of $P(\boldsymbol{x}, t)$ can be mathematically described by means of a *Master equation*, which represents the main stochastic processes occurring at an individual or "microscopic" level [296]. In particular, the stochastic dynamics at the individual level relies on the three key ingredients of Darwinian evolution [13, 245] (see the sketch of Fig. 3.1):

- *Reproduction*: Each individual produces (asexually) one offspring at some rate, called "fitness", $f_i(\boldsymbol{x}) \leq 0$, that depends on its phenotype, as well as on the overall system state (i.e. on $\boldsymbol{x}$). In what follows, we restrict ourselves to the case in which the fitness for individuals with trait x_i is composed of (i) an intrinsic growth rate $K(x_i)$ —that is assumed to depend only on the individual trait x_i— and defines an external "ecological niche", specific of the considered environment/conditions, as well as (ii) an additional term that comes from pairwise ecological forces (or "interactions"), $I(x_i, x_j)$ which represent competition, cooperation, and/or more complex ecological forces with all other individuals (e.g. the j-th one with x_j). Thus, finally, the total fitness function $f_i(\boldsymbol{x})$ of individual i can is given by the intrinsic growth rate plus the weighted sum of all the pairwise interactions with other individuals (so that both terms are typically of the same order)

$$f_i(\boldsymbol{x}) = K(x_i) + \sum_{j=1, j\neq i}^{N} \frac{I(x_i, x_j)}{N-1} \tag{3.1}$$

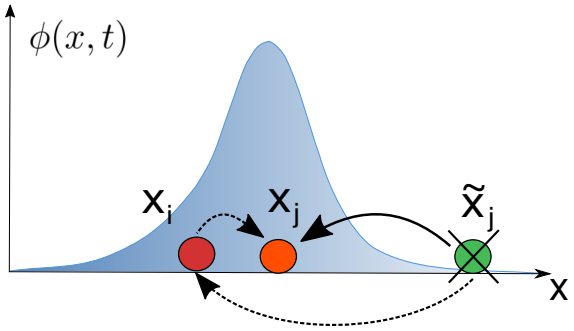


FIGURE 3.1: **Graphical representation of the transition rates characterizing the eco-evolutionary dynamics at a "microscopic" level.** The probability density $\phi(x,t)$ describing the population with phenotypic trait x at a given time t can change as a result of the following stochastic processes: (i) the first one is reproduction with mutation (individual with trait x_i (dark red) generates an offspring with trait x_j relatively close —but not identical— to x_i (light red); (ii) death (an individual with trait $\tilde{x}_j$ (green) is randomly selected to die dies). Mathematically, these processes are fully equivalent to an effective Markov process (as represented by black dashed arrows) in which the individual i performs a *local jump* to the coordinate x_j, while its initial position x_i can be seen as occupied by the removed individual by means of a *non-local* jump. Hence, the composition of the two transitions (birth and death) can be effectively described as a single net jump from $\tilde{x}_j$ to x_j. These combined processes give rise to an evolving distribution that may eventually converge to a stationary one.

or, equivalently

$$f_i(\boldsymbol{x}) = \sum_{j=1, j\neq i}^{N} \frac{f(x_i, x_j)}{N-1} \tag{3.2}$$

where $f(x_i, x_j) \equiv K(x_i) + I(x_i, x_j)$. Higher-order interactions (which can be relevant in this context [297]) could also be straightforwardly implemented by including additional terms in Eq.(C.8)).

- *Selection*. The fact that individuals with larger fitness values are more likely to reproduce allows one to implement natural selection in an indirect way just by imposing a constant population size, N. In particular, as illustrated in Fig.3.1, the model assumes that each time that an individual i produces an offspring, a second individual j (different from i) is randomly chosen and removed from the community [3]. This second individual is selected with certain death probability $d_j(\boldsymbol{x})$ which, in general can depend on the full phenotypic state, $\boldsymbol{x}$, and that, for simplicity, we set to be the same for all individuals, i.e. $d(\tilde{x}_j) = 1/(N-1)$.

- *Heredity and variation*. When an individual replicates, its offspring inherits the phenotypic trait/s of the parent, with some variation or mutation. The probability of a given variation from the mother value x_i to the offspring's one x_j is represented by a generic probability distribution function, $\beta(x_i, x_j)$, called "*mutation kernel*". This distribution can be characterized by its mean $\theta(x_i, \delta)$ and variance $\sigma(x_i, \delta)$, where $\delta = |x_i - x_j|$ and, in some simple cases, β can be assumed to be independent of x_i and depend only on the magnitude of the jump δ.

[3] In other words, this dynamics is a generalization of the Moran process [295].

Thus, the master equation defining the general model (see Fig.3.1) can be written as [296]:

$$\partial_t P(\boldsymbol{x},t) = \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\int_{\mathcal{P}} d\tilde{x}_j\Big[W_i(\boldsymbol{x},\tilde{\boldsymbol{x}}^j)P(\tilde{\boldsymbol{x}}^j,t) - W_i(\tilde{\boldsymbol{x}}^j,\boldsymbol{x})P(\boldsymbol{x},t)\Big] \tag{3.3}$$

where

$$W_i(\boldsymbol{x},\tilde{\boldsymbol{x}}^j) = f_i(\tilde{\boldsymbol{x}})\; d_j(\tilde{\boldsymbol{x}})\; \beta(x_j - x_i) \tag{3.4}$$

is the rate to transition from an initial state $\tilde{\boldsymbol{x}}^j = (x_1,..,\tilde{x}_j,..,x_N)$ to $\boldsymbol{x} = (x_1,..,x_j,..,x_N)$, where $\tilde{\boldsymbol{x}}^j$ differs from $\boldsymbol{x}$ only in the value of the coordinate j, i.e. the individual that is killed and replaced by an offspring of i with mutated trait x_j. Reciprocally, $W_i(\tilde{\boldsymbol{x}}^j,\boldsymbol{x})$ is the transition rate for the reverse process.
Let us remark that the described stochastic process generates a neat flux of probability from the state of the dying individual (i.e. trait $\tilde{x}_j$) to that of the newly generated one, (i.e. x_j), which can be visualized as a jump from the first position to the latter (solid line in Fig.3.1). More precisely, this birth-death process can be further decomposed in two different jumps (dashed lines in Figure 3.1): a non-local one, from $\tilde{x}_j$ to x_i, and a local one from the latter to its offspring x_j. The first (non-local) jump implements the effect of selection while the second (local) one describes mutation. This distinction will be useful for later discussions.

Selection-Mutation mean-field equation

The above Master equation is a very general one but, obviously, difficult to handle analytically. Thus, in order to gain more insight, in what follows we employ some standard approximations to reduce it to a simpler deterministic (mean-field like) equation for the marginal probability to find any individual in a particular state x. The reader who is not interested in the formal aspects of the following mathematical derivations can go directly to Eq.(4.3).
As often done in statistical mechanics, one can assume that individuals with identical traits are, *a priori*, indistinguishable. This equivalence allows one to derive an equation for the individual (or "one-particle") probability density distribution $\phi(x,t)$. More specifically: the *density* of individuals with phenotype x at time t in a given realization of the stochastic process can be simply expressed as

$$\rho(x,t) = \sum_{i=1}^{N}\frac{\delta(x-x_i)}{N}, \tag{3.5}$$

and averaging over possible realizations of the stochastic process one obtains the probability density distribution

$$\begin{aligned}\phi(x,t) &\equiv \Big\langle\sum_{i=1}^{N}\frac{\delta(x-x_i)}{N}\Big\rangle = \frac{1}{N}\sum_{i=1}^{N}\int_{\mathcal{P}^N} d\mathbf{x}\,\delta(x-x_i)P(\mathbf{x},t)\\ &= \int_{\mathcal{P}^{N-1}} dx_2 dx_3..dx_N P(x,x_2,x_3,..,x_N),\end{aligned} \tag{3.6}$$

that is nothing but the marginal probability of the first individual with phenotype $x \equiv x_1$, given that individuals with the same traits are indistinguishable, i.e. that the probability distribution is symmetric with respect to the exchange of individual labels [176, 292].
Taking the time derivative of Eq.(3.6) and plugging into it the master equation, Eq.(3.3),

one readily obtains (see appendix, sec C.1.1):

$$\partial_t \phi(x,t) = \int_{\mathcal{P}^N} dx_2\, dx_3 ... dx_N\, \partial_t P(\boldsymbol{x},t) = \tag{3.7}$$
$$\int_{\mathcal{P}^2} d\tilde{x}\, dy\, [f(\tilde{x},y)\beta(x-\tilde{x})P(\tilde{x},y) - f(\tilde{x},y)P(x,\tilde{x},y)],$$

where the death rate d has been fixed to a constant value. Before proceeding, let us remark that there are two main consequences of individual indistinguishability:
(**i**) The fitness function has been reduced to the pairwise fitness function, $f(x,y)$; i.e. it depends just on the traits of the reproducing individual and another generic "interacting" one.
(**ii**) The resulting simplified master equation, Eq.(3.7), depends only on the two- and three-particle joint probability (rather than in the whole N-particle one) but is not a closed equation for $\phi(x,t)$. In order to obtain a closed equation one needs to do some additional assumption.
In particular, one can consider a mean-field approximation consisting, as usual, in assuming that the joint probability distribution function can be factorized:

$$P(\boldsymbol{x},t) = \prod_{i=1}^{N} \phi(x_i,t), \tag{3.8}$$

which is an exact result in the limit of large-population sizes [176, 292, 293], see appendix sec.C.1.1 for more details. Corrections accounting for finite population size will be explicitly discussed in a forthcoming section. This readily leads to:

$$\partial_t \phi(x,t) = \int_{\mathcal{P}} d\tilde{x}\beta(x-\tilde{x}) f(\tilde{x},t)\phi(\tilde{x},t) - \bar{f}(t)\phi(x,t) \tag{3.9}$$

where we have defined the "*marginal fitness*" associated with state $\tilde{x}$

$$f(\tilde{x},t) \equiv \int_{\mathcal{P}} dy\, \phi(y,t) f(\tilde{x},y) \tag{3.10}$$

and where the population-averaged marginal fitness is

$$\bar{f}(t) \;=\; \int_{\mathcal{P}}\int_{\mathcal{P}} d\tilde{x}dy\, \phi(\tilde{x},t)\phi(y,t) f(\tilde{x},y) = \int_{\mathcal{P}} d\tilde{x}\, \phi(\tilde{x},t) f(\tilde{x},t), \tag{3.11}$$

Importantly, the previous expressions, in particular, Eq.(3.10), encode the idea that the marginal fitness of a given trait, $\tilde{x}$, is frequency-dependent, i.e. the fitness associated with a given trait crucially depends of the distribution of other individuals across the phenotypic space.
Even if the derivation of Eq.(3.9), might look cumbersome, it can be interpreted in a rather straightforward and intuitive way:

- The first term in the right hand side is a positive probability flow into the x state stemming from the probability that an individual with any arbitrary coordinate $\tilde{x}$ is chosen for reproduction and produces a mutated offspring with, precisely, trait x.

- The second term describes the fact that, owing to normalization, reproduction events leading to an increase of probability density for any given trait $x' \neq x$ (which occurs with an average rate $\bar{f}$) reduce the relative probability of finding individuals in state x, i.e. diminish $\phi(x,t)$

General moment equations

The mean-field equation Eq. (3.9) rules the dynamic of the probability density $\phi(x)$. As usual it is convenient to analyze separately the dynamics of the distribution cumulants, μ_i, with $i = 1, 2....$ The mean (population-averaged) value of any possible function $A(x)$ of the trait x is

$$\overline{A}(t) = \int_{\mathcal{P}} dx A(x)\phi(x,t). \tag{3.12}$$

For example, the mean value in trait value is $\overline{x}(t)$ and the mean marginal fitness is $\overline{f}(t)$. Similarly, the (standard) covariance Σ between any two functions $A(x,t)$ and $B(x,t)$ at time t can be expressed as

$$\Sigma[A,B](t) \equiv \overline{A \cdot B} - \overline{A} \cdot \overline{B}, \tag{3.13}$$

which reduces to the variance for $A = B$. From Eq. (3.9) and the previous definitions, the dynamics of the mean value of an arbitrary function $A(x)$ is

$$d_t\overline{A} \quad = \quad \int_{\mathcal{P}^2} dx d\tilde{x}\ A(x)\beta(x-\tilde{x})f(\tilde{x},t)\phi(\tilde{x},t) - \overline{A}(t)\overline{f}(t) \tag{3.14}$$

which can be written in a very compact form (see below) if one defines the covariance between two quantities (here $A(x)$ and $f(x)$) across two consecutive generations:

$$\Sigma_\beta[A,f] \quad \equiv \quad \int_{\mathcal{P}^2} dx d\tilde{x}\ A(x)\beta(x-\tilde{x})f(\tilde{x})\phi(\tilde{x},t) - \overline{A}(t)\overline{f}(t) \tag{3.15}$$

Observe that, while the standard covariance $\Sigma[A,f]$ quantifies the correlation between the quantity A and the fitness f at a given time, $\Sigma_\beta[A,f]$ is the covariance between the fitness of the "mother individual" with some trait $\tilde{x}$ and $A(x)$ evaluated for the offspring with trait x (and the mutation function from $\tilde{x}$ to x is included). Thus, $\Sigma_\beta[x,f]$ stands for the covariance between phenotype and fitness *across a generation*.
Furthermore, note that, in particular, in the absence of variation, i.e. if the mutation amplitude vanishes, $\beta(x-\tilde{x}) = \delta(x-\tilde{x})$, then $\Sigma_\beta[A,f] = \Sigma[A,f]$ and both covariances coincide. Using these definitions, one can finally write down the simple expression

$$d_t\overline{A} = \Sigma_\beta[A,f](t) \tag{3.16}$$

which is nothing but a generalization to any arbitrary quantity $A(x)$ of the *Price equation* [251, 248, 249, 254]:

$$d_t\overline{x}(t) = \Sigma_\beta[x,f] \tag{3.17}$$

describing the dynamics of the mean trait. It states that if a given trait is positively correlated with the fitness function (weighted by the effect of mutation) its mean-value increases (and, as we have shown, the same covariance-dependent type of equation is valid to describe the dynamics of generic quantities).

Diffusive or small-mutation approximation

Further analytical progress can be made by assuming (as done in adaptive dynamics) that the amplitude of mutations, $\delta = |\tilde{x} - x|$ is small, which allows one to perform a (Kramers-Moyal) expansion of Eq.(3.9) in powers of δ (see appendix, sec.C.1.2 and [296] for details):

$$\begin{aligned} \partial_t \phi(x,t) &= \big(f(x,t) - \bar{f}(t)\big)\phi(x,t) + \\ - \quad & \partial_x[\theta(x) f(x,t)\phi(x,t)] + \frac{1}{2}\partial_x^2[\sigma^2(x) f(x,t)\phi(x,t)] \end{aligned} \tag{3.18}$$

where $\theta(x)$ and $\sigma^2(x)$ are the first two moments of the mutation kernel

$$\theta(x) = \int d\delta\, \beta(x,\delta)\, \delta \quad \sigma^2(x) = \int d\delta\, \beta(x,\delta)\, \delta^2, \tag{3.19}$$

where the first one is referred to as "bias", the second in the mutation "amplitude"; higher-order terms can be neglected within this approximation.
Observe that Eq.(4.3) is a linear superposition of the (upper row) replicator equation [298], representing selection and (lower row) a Fokker-Planck type of equation describing mutations as a reaction-diffusion dynamics in phenotypic space. More specifically, the second part is a sort non-linear Fokker-Planck equations or McKean-Vlasov equation [299, 300], as the diffusion function depends itself on the probability-distribution $\phi(x,t)$ through the marginal fitness function, so that it needs to be solved in a self-consistent way.
As a matter of fact, Eq.(4.3) is actually a version of the celebrated *Crow-Kimura (CK) equation* in population genetics [244, 243]. Thus, we call it *generalized Crow Kimura equation* (GCK) as Eq.(4.3) extends the CK equation to phenotypic evolution, clarifies its microscopic foundation, and includes two additional features:

- The fitness function appears in Eq.(4.3) within the partial derivatives, thus coupling reproduction and mutation.
- Eq.(4.3) includes generic mutation functions $\theta(x)$ and $\sigma^2(x)$ that, in general, can be trait-dependent rather than constant coefficients.

Observe that, within this approximation, given that one can expand the function β in power series of δ, it is straightforward to recover after some simple algebra the classical or standard form of the Price equation:

$$d_t \bar{x}(t) = \Sigma[x,f] + \overline{\theta f} \tag{3.20}$$

where the contributions of selection and mutation are decoupled; in particular, the first term encodes *selection* on the mean trait value; i.e. it increases when it correlates positively with fitness increment, while the second term represents the action of biased mutations (and it vanishes if variations are symmetric) [4].
A further simplification is obtained in the case of a constant (trait-independent) mutation rate, for which Eq.(3.20) becomes

$$d_t \bar{x}(t) = \Sigma[x,f] + \theta \bar{f}(t), \tag{3.21}$$

and the equation for the trait variance $\Sigma(t) \equiv \overline{(x-\bar{x})^2}(t)$ reads:

$$d_t \Sigma(t) = \Sigma[x^2, f] - 2(\bar{x} - \theta)\Sigma[x,f] + \sigma^2 \bar{f} \tag{3.22}$$

[4] Let us remark that the zero-covariance condition defines a deterministic evolutionary stable state (ESS) in the framework of evolutionary game theory, where strategies are fixed, i.e. there is no mutation.

where for simplicity in the notation we have omitted time dependencies.
As described in the context of adaptive dynamics (see below) if the variance of the distribution in phenotypic space $\Sigma(t)$ decreases during the dynamics there is a "stabilization" or convergence towards a distribution peaked around the mean-trait value. On the other hand, if the variance increases there might be a "*disruptive-evolution*" mechanism in act, which leads to a broadening in phenotypic space, and possible to non-trivial effects such as an evolutionary branching (see below).
More in general, the equations for higher-order moments form a hierarchy of coupled differential equations. Finding a valid justification to close such a hierarchy is an open general ("moment-closure") problem [301]. In what follows, for the sake of completeness and to set the notation and formalism, we first discuss the Gaussian approximation to the closure problem, i.e. adaptive dynamics, and then —in forthcoming sections— we introduce extensions of it accounting for higher-order moments and a much richer phenomenology.

3.3 Gaussian theory: recovering Adaptive dynamics

In the classical terms of adaptive dynamics individuals around the mean phenotype $\bar{x}$ are called "*residents*" and other possible variations of them are called "*mutants*". As already mentioned, within such a theory, the possible mutations are usually assumed to be small, unbiased, and trait-independent, which greatly simplifies the problem. In particular, under such assumptions, it suffices to study the dynamics of the mean value $\bar{x}$ —which is assumed to coincide with the peak of the underlying distribution— as well as perturbations around it, which is equivalent to considering a Gaussian approximation to describe probabilities in phenotypic space using just the first two cumulants: the mean $\bar{x}(t)$ and the variance $\Sigma(t)$.
More specifically, assuming a fitness function with the form of Eq.(3.10), one can expand $f(x,y)$ in bothn of its arguments around the mean value $\bar{x}$:

$$\begin{aligned} f(x,y) &\approx f(\bar{x},\bar{x}) + f_1^x(\bar{x},\bar{x})(x-\bar{x}) + f_1^y(\bar{x},\bar{x})(y-\bar{x}) \\ &+ \frac{f_2^x(\bar{x},\bar{x})}{2}(x-\bar{x})^2 + \frac{f_2^y(\bar{x},\bar{x})}{2}(y-\bar{x})^2 + f_{11}^{xy}(\bar{x},\bar{x})(x-\bar{x})(y-\bar{x}) + ... \end{aligned}$$

where the subindices of f indicate the number of derivatives, the superindices the variables with respect to which the derivatives are taken:

$$f_{ij}^{xy}(\bar{x},\bar{x}) = \partial_x^i \partial_y^j f(x,y)|_{x=y=\bar{x}}, \tag{3.23}$$

with $i,j = 1,2....$ Higher-order terms in the expansion are neglected within the present approximation.
Averaging over the second variable, one obtains the "marginal fitness" (see Eq.(3.10)):

$$f(x,t) = \int_{\mathcal{P}} dy f(x,y)\phi(y,t) = f(\bar{x}) + f_1^x(\bar{x})(x-\bar{x}) + f_2^x(\bar{x})\frac{(x-\bar{x})^2}{2} + f_2^y(\bar{x})\frac{\Sigma}{2}$$

where $f_2^x(\bar{x}) \equiv f_2^x(\bar{x},\bar{x})$, $f_2^y(\bar{x}) \equiv f_2^y(\bar{x},\bar{x})$.
Defining $f_2(\bar{x}) \equiv f_2^x(\bar{x}) + f_2^y(\bar{x})$, the average fitness (computed as in Eq.(3.11)) can be finally be written as:

$$\bar{f}(t) = f(\bar{x}) + \frac{f_2(\bar{x})}{2}\Sigma \tag{3.24}$$

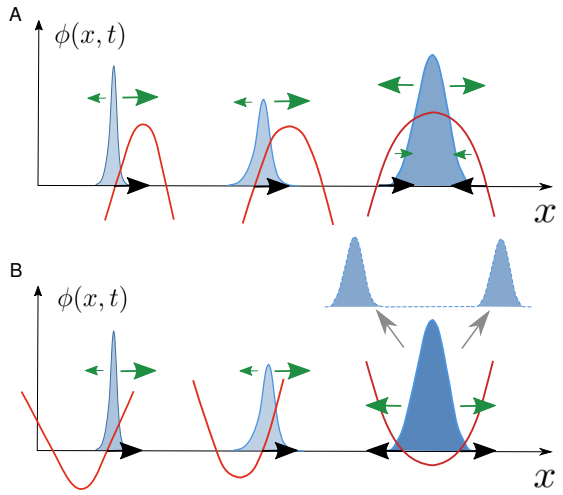


FIGURE 3.2: **Sketch of typical evolutionary trajectories in adaptive dynamics in the case of (top) a stable monomorphic population and (bottom) evolutionary branching.** The distribution of phenotypes (light blue) is represented at various steps/times of the dynamics, together with the fitness function/landscape (red curves) with changes along with the probability distribution. Green arrows mark the intensity of the selection and the black ones stand for the overall selection gradient. **(A)** Evolutionary stable case: the population climbs the fitness landscape and is attracted to a fitness maximum. **(B)** Evolutionary branching: the population is first attracted to a fitness minimum (convergent stable) but then is repelled from it (evolutionarily unstable). Once the branching occur. Let us remark that, the standard theory of adaptive dynamics is equivalent to considering a Gaussian approximation for small, unbiased, and trait-independent mutations in our generalized framework (though it breakdown after, e.g. evolutionary branching, when the population can not be described as a single mono-modal distribution any more).

where both $\tilde{x}$ and Σ can be time dependent functions. Let us caution that $f_2(\bar{x}) \neq \partial_{\bar{x}} f_1(\bar{x})$ but includes also the term $f_2^y(\bar{x})$, as this distinction will be important in what follows.
The previous two expressions for $f(x)$ and $\bar{f}$, respectively, can be plugged into the Price equation, Eq.(3.21), to obtain

$$d_t \bar{x}(t) \quad = \quad f_1^x(\bar{x})\Sigma + \frac{f_2(\bar{x})}{2}\mu_3(t) \tag{3.25}$$

for the case of vanishing bias, $\theta = 0$. Assuming that the phenotypic distribution $\phi(x,t)$ is a Gaussian, the third moment vanishes, i.e. $\mu_3 = 0$, and, hence:

$$d_t \bar{x}(t) = f_1^x(\bar{x})\Sigma(t). \tag{3.26}$$

This last equation is known as the "*canonical equation*" in adaptive dynamics and determines the fate of the mean population trait; $f_1^x(\bar{x})$ is called "*selection gradient*" and determines the direction of the flow, while the variance of the trait distribution controls the "*speed*" of the evolutionary process.
The possible fixed points $\bar{x}^*$ of the previous equation need to be extreme point of the fitness, i.e. $f_1^x(\bar{x}^*) = 0$, and are called "convergent stable" if

$$\partial_{\bar{x}} f_1^x(\bar{x})|_{\bar{x}=\bar{x}^*} < 0, \tag{3.27}$$

so that the mean trait value, $\bar{x}$, is locally attracted to fitness maxima and repelled from fitness minima. In other words, the selection dynamics tends to climb the fitness gradient.

Similarly, for the variance, from Eq.(3.22) one has

$$d_t\Sigma(t) = f_1^x(\bar{x})\mu_3 + \frac{f_2^x(\bar{x})}{2}(\mu_4 - \Sigma^2) + \sigma^2\bar{f}, \tag{3.28}$$

which within the Gaussian ansatz becomes

$$d_t\Sigma(t) = f_2^x(\bar{x})\Sigma^2 + \sigma^2 f(\bar{x}) + \frac{\sigma^2}{2}f_2(\bar{x})\Sigma, \tag{3.29}$$

so that one has a closed set of equations for the evolution of the first two cumulants of the trait distribution.
From these, we can compute the steady-state solution

$$\begin{aligned}\Sigma^* &= \sqrt{-\sigma^2\bar{f}/f_2^x(\bar{x}^*)} \\ &= -\frac{\sigma^2}{2f_2^x(\bar{x}^*)}\Big(\sqrt{\sigma^2 f_2^2(\bar{x}^*) - 4f(x^*)f_2^x(x^*)} + f_2(x^*)\Big),\end{aligned} \tag{3.30}$$

which exists as a real solution only if $f_2^x(\bar{x}^*)$ is negative. Eq. (3.31) also implies that higher fitness peaks (i.e. larger values of $\bar{f}$) lead to larger steady-state variance; this can be understood from the fact that larger fitness means faster reproduction and thus a larger source of mutations and variability. Therefore, the second derivative $f_2^x(\bar{x}^*)$ determines whether a convergent stable point $\bar{x}^*$ is also "*evolutionary stable*" with respect to the introduction of mutants, i.e. if the variance of the distribution also converges to a stable fixed-point value; i.e. an evolutionary stable point needs to be (local) fitness maximum. On the other hand, if $f_2^x(\bar{x}^*) > 0$ there is no stationary stable solution for the variance and, within the present Gaussian approximation, it just grows un-boundedly. More specifically: if $\bar{x}^*$ is an attractor for the mean (i.e. it is convergence stable), but $f_2^x(\bar{x}^*) > 0$ (i.e. it is evolutionary unstable), then x^* is a fitness minimum. In this case —recalling that the fitness function is, in general, a dynamic quantity, that changes together with the distribution– the distribution is repelled from the fitness minimum (as pictorially illustrated in Fig.3.2). In this latter case, the evolutionary process (with fixed mean and ever-growing variance) implies that the distribution splits into a bimodal one, with two diverging peaks, giving rise to the phenomenon of "*evolutionary branching*" [156, 286, 302].
Tus, summing up, under the simplifying assumptions of adaptive dynamics, it is possible to explicitly calculate the conditions for the emergence of either evolutionary stable solutions or evolutionary branching :

- Convergent Stable and evolutionary stable ("non invasible"): x^* is a fitness maximum

$$f_1^x(\bar{x}^*) = 0, \quad \partial_{\bar{x}} f_1^x(\bar{x})|_{\bar{x}=\bar{x}^*} < 0, \quad f_2^x(\bar{x}^*) < 0. \tag{3.31}$$

- Convergent stable but evolutionary unstable (possible branching point): $\bar{x}^*$ is a fitness minimum

$$f_1^x(\bar{x}^*) = 0, \quad \partial_{\bar{x}} f_1^x(\bar{x})|_{\bar{x}=\bar{x}^*} < 0, \quad f_2^x(\bar{x}^*) > 0. \tag{3.32}$$

Nevertheless, it is important to emphasize that in order to unveil how the dynamics proceeds beyond a branching point and, more in general, to explore alternative evolutionary phases and patterns in phenotypic space [**Scheffer**, **Ramos**, 303], it becomes necessary to extend the theory —going beyond the Gaussian approximation— by including, e.g., higher-order terms in the fitness expansion.

3.4 Extended Landau-like theory: beyond adaptive dynamics

In order to extend the previous Gaussian theory (i.e. adaptive dynamics) to allow for a description of the evolutionary dynamics even after e.g. a branching event —when the distribution in phenotypic space can no longer be described as a Gaussian— we now introduce a theory *"a la Landau"* [9, 304] by incorporating higher-order terms in the fitness expansion Eq.(3.23) (see [140]).
Let us recall that the Landau theory of phase transitions uses a parsimony principle combined with symmetry considerations to write down a general functional —a free-energy functional— including only the most important terms in a perturbative expansion of the relevant field needed to describe key aspects of a phase transition.
For instance, in the classical example of an Ising like phase transition –describing the spontaneous breaking of an up-down symmetry— one needs to include only up to quartic terms in the "magnetization field" to derive a theory that quantitatively explains the main features of the transition [9].
Similarly, here we perform an expansion of the "effective fitness" $F(x)$, defined as the marginal fitness $f(x)$ minus its x-independent (constant) part of it:

$$F(x) \equiv f(x) - \tilde{f}, \tag{3.33}$$

which determines the fitness landscape and is the counterpart of the usual free-energy function.
Expanding $F(x)$ in powers of $(x - \bar{x})$ around its mean $\bar{x}$, one obtains (see appendix sec. C.2.1):

$$\begin{aligned} F(x) &= \left(f_1^x(\bar{x}) + \frac{f_{12}^{xy}(\bar{x})}{2}\Sigma(t) + \frac{f_{13}^{xy}(\bar{x})}{3!}\mu_3(t) \right)(x-\bar{x}) + \\ &+ \left(\frac{f_2^x(\bar{x})}{2} + \frac{f_{22}^{xy}(\bar{x})}{4}\Sigma(t) \right)(x-\bar{x})^2 + \frac{f_3^x(\bar{x})}{3!}(x-\bar{x})^3 + \frac{f_4^x(\bar{x})}{4!}(x-\bar{x})^4 + ... \end{aligned} \tag{3.34}$$

where terms above the 4-th order have been neglected for now. Let us remark that this effective fitness determines the fitness landscape but importantly —much as in statistical mechanics— mutations (which play a role analogous to thermal fluctuations, i.e. temperature) are also needed in order to determine the shape of the resulting steady-state probability distribution associated with the GCK equation, Eq.(4.3).
In order to keep the presentation as simple as possible, let us consider for now some simplifying assumptions to $F(x)$, as usually done in the Landau approach. For example, the trait distribution $\phi(x,t)$ can be assumed to be always symmetric around its mean, which implies that all odd central moments, in particular $\mu_3(t)$, vanish. Similarly, one can also assume the interaction term appearing in the fitness function to be symmetric in its two indices, implying that all the cross derivative terms of odd order vanish; e.g. $f_{12}^{xy}(\bar{x}) = 0$. In this way, Eq.(3.35) takes the simpler form:

$$\begin{aligned} F(x) &= f_1^x(\bar{x})(x-\bar{x}) + \left(\frac{f_2^x(\bar{x})}{2} + \frac{f_{22}^{xy}(\bar{x})}{4}\Sigma(t) \right)(x-\bar{x})^2 \\ &+ \frac{f_3^x(\bar{x})}{3!}(x-\bar{x})^3 + \frac{f_4^x(\bar{x})}{4!}(x-\bar{x})^4 + ... \end{aligned} \tag{3.35}$$

Given that the linear term is not changed with respect to its Gaussian counterpart, the stationary condition for the mean value remains unchanged $d_t\bar{x}^* = f_1^x(\bar{x}^*)\Sigma^* = 0$ therefore, its stationary value, $\bar{x}^*$, has to coincide with an extreme point of $F(x)$, $f_1^x(\bar{x}^*) = 0$.

Note, however, that if the fitness function is not symmetric, the term f_3^x may contribute to the dynamics but given that we are assuming a symmetric steady-state distribution it must vanish at stationarity.
For the sake of simplicity and without loss of generality, we set to zero the stationary mean $\bar{x}^* = 0$ and we omit the dependency on it in the fitness derivatives. This leads a simpler form of the effective *stationary* fitness:

$$F(x) = \frac{1}{2}\left(f_2^x + \frac{f_{22}^{xy}}{2}\Sigma\right)x^2 + \frac{f_4^x}{4!}x^4 = \frac{g_2}{2}x^2 + \frac{g_4}{4!}x^2, \tag{3.36}$$

where the functions g_2 and g_4 have been defined in the last line to group the terms proportional to x^2 and x^4, respectively. One can now look for the points of vanishing derivative, corresponding either to fitness maxima or minima, which leads to the trivial solution $x_0^* = 0$ together with a couple of additional extrema,

$$x^{*2}_{1,2} = \frac{-6g_2}{f_4} = -\frac{6}{f_4^x}\left(f_2^x + \frac{f_{22}^{xy}}{2}\Sigma^*\right). \tag{3.37}$$

Let us remark that the location of these two fixed points depends on the steady-state variance Σ^*, something that does not happen in the Gaussian theory. In order to close the moment hierarchy one needs to determine the steady-state variance using some approximation. To do so, one could assume, for instance, a bimodal-Gaussian ansatz for the steady-state distribution. However, in spite of the simplification, this is still quite cumbersome to handle analytically. Thus, we additionally consider the limit of very small mutation amplitudes in which case the bimodal Gaussian can be approximated as the sum of two delta functions,

$$\phi^*(x)_{1,2} = \frac{\delta(x - x^*)}{2} + \frac{\delta(x + x^*)}{2}. \tag{3.38}$$

The associated variance of this distribution is $\Sigma^* = x^{*^2}$ and can be seen as a lower bound for the actual value in the case of larger mutation amplitudes. Plugging this approximation in Eq.(3.37)

$$x_{12}^* = \pm\sqrt{-6\frac{f_2^x}{f_4^x + 3f_{22}^{xy}}}. \tag{3.39}$$

Observe, that the conditions for these two non-trivial solutions to exist are $f_2^x > 0$ —so that the origin is a fitness minimum— and $(f_4^x + 6f_{22}^{xy}) < 0$, to have a positive sign under the square-root. It is easy to see that $x_{1,2}^*$ are maxima if $f_4^x > 3f_{22}^{xy}/2$, a condition that is expected to be always satisfied when the two points exist. Hence, the resulting steady-state distribution is unimodal with a peak at the origin if $f_2^x < 0$ while for $f_2^x > 0$ the origin becomes unstable and a new bimodal stable distribution emerges. As already stressed, in this latter phase, the sign of f_2^x does not fully determine the convexity at the origin of $F(x)$ (i.e. the overall sign of the terms proportional to x^2 in Eq.(3.36), i.e. g_2), due to the presence of the additional term $f_{22}^{xy}\Sigma(t)$, which stems from the interaction kernel as $f_{22}^{xy} = \partial_x^2\partial_y^2 I|_{x=y=\bar{x}}$ and has the same sign as the interaction: positive for cooperation, negative for competition. This dependence may play a non-trivial role during the course of an evolutionary branching as it will be explicitly illustrated in the next Section. In particular, observe that if the population reaches a fitness minimum and starts to branch, the mean-trait value remains fixed in the origin, $\bar{x}^* = 0$ while the variance increases until it possibly reaches a stationary value. This increase in the variance may lead to a progressive change

in the coefficient of the quadratic term, g_2, and hence of the convexity in the origin.
To be more specifically, let us write down the time derivative of the quadratic coefficient in Eq.(3.36) is

$$d_t g_2(t) = f_{22}^{xy}(\bar{x}) d_t \Sigma/2, \tag{3.40}$$

which implies that, e.g., if the variance grows after a branching event, then the fitness barrier separating both of the new attractors can change dynamically, either

- becoming higher in the case in which $f_{22}^{xy}(\bar{x}) > 0$ (e.g., for mutualistic dominant interactions), or
- becoming flatter if $f_{22}^{xy} < 0$ (e.g., for competitive interactions).

In particular, the reduction of the fitness barrier may generate a sort of "*neutral bridge*" in between the two coexisting peaks, leaving the population susceptible to transient invasions in between the two branched sub-populations. Observe that the stationary value of g_2 can be calculated (at least in an approximate way) form its definition and the values of $x_{1,2}^*$ in Eq.(3.39) and $\Sigma^* = x^{*2}$,

$$g_2^* = f_2^x \left(1 - \frac{2 f_{22}^{xy}}{f_4^x + 6 f_{22}^{xy}}\right). \tag{3.41}$$

This shows that $0 < f_2^* < g_2^*$ corresponding to a high barrier for mutualistic interactions and a $0 < g_2^* < f_2^*$, i.e. a flatter intermediate landscape for competitive interactions. Observe also, that, within the present approximation, the sign of g_2 does not change. However, this calculation suggests that if higher non-linearities —such as those appearing in expansions up to 8-th or 10-th order— are included, further corrections to g_2 can appear and they can possibly reverse its overall sign. If this happens and the concavity at the origin changes, a new ecological niche can be generated and it can possibly be repopulated. This last scenario can be realized, for instance, when after a first branching event, each of the two resulting branches converges to a relative fitness minimum, thus leading to a second round of evolutionary branchings and therefore to a total of 4 coexisting sub-populations. If the concavity of the origin has changed, then the two central populations might eventually converge to the origin colliding and repopulating the empty central niche, resulting in "*evolutionary convergence*" and a final stationary distribution with 3 coexisting sub-populations.
A detailed analytical description of the previously described phenomenology is quite intricate as —in order to allow for the possibility of a series of two branching events— it requires to keep terms up to order x^8 or x^{10} in the perturbative expansion of $F(x)$ and the calculation becomes quite cumbersome [305] (see appendix C.2.1for more details).
However, in the next section, we present the numerical solution of an explicit example where the discussed non-trivial effects such as "neutral bridges", "cascades of branching events" and "evolutionary convergence" —all of them beyond the reach of the standard approach of adaptive dynamics— are vividly illustrated.

3.5 Growth-competition model

Model definition

Here we study a simple model that combines characteristics of other models previously studied in the context of adaptive dynamics [156] and in the study of species clustering in phenotypic space [**Scheffer**, **Ramos**, 303], respectively, see also the calculations in Sec.1.7.

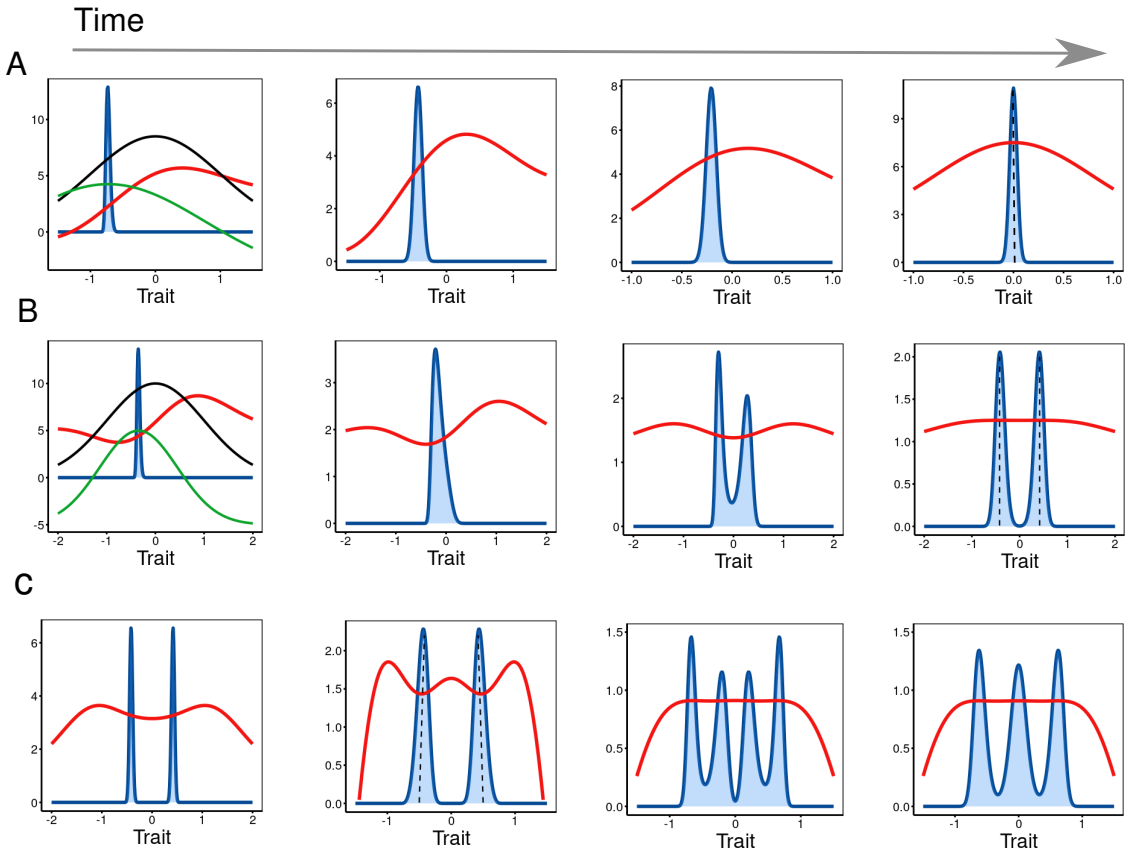


FIGURE 3.3: **Dynamics of the phenotypic distribution (blue) and fitness landscape (red) for different times as obtained from a numerical integration of the full "generalized Crow-Kimura" (GCK) equation, Eq.(4.3) in different cases.** (A) with no branching, (B) one branching event, and a(C) a series of two consecutive branching and a coalescence event. In the first case (A), the competition kernel (green) is wider than the growth one (black), i.e. $\sigma_\alpha > \sigma_K$, producing a fitness landscape with a single maximum (red). The initially peaked distribution moves to the right, in the direction of increasing gradient, thus trying to climb the fitness landscape towards the maximum at $x = 0$, as predicted by theory (dashed vertical line). Observe that the width of the distribution changes across time. In the second case (B), the competition kernel (black) is wider than the growth one, producing a fitness landscape with a two maxima. The population tries to climb the landscape but is trapped in the minimum. To escape and further increase fitness an evolutionary branching happens and the two sub-populations finally reach the the two maxima, as predicted by the theory (dashed lines). Finally, if the competition kernel is way smaller than the growth one (C), after the first evolutionary branching the resulting two sub-populations converge to fitness minima and each one branches again further creating transiently a population with four peaks, that then converges toi a 3-mode ones, once the two central ones coalesce at $x = 0$. Parameter values: $k = 1, \sigma_K = 1$ and $\sigma = 10^{-3}$ in all cases. $\sigma_\alpha = 1.2$ (A), $\sigma_\alpha = 0.8$ (B), $\sigma_\alpha = 0.6$ (C). The initial distribution is localized (delta Dirac function) at x_0, with $x_0 = 0.8$ (A), 0.4 (B), 0.6 (C), respectively.

In particular, we consider a one-dimensional phenotypic space, characterized by a single scalar trait, x. As above, such a trait can also be thought as the individual "niche" coordinate, or as a unidimensional projection of a higher-dimensional phenotypic space.
The trait value x influences the growth rate of individuals, through a growth function $K(x)$ —that in the simplest case can be chosen as a Gaussian centered at $x = 0$— that determines which phenotypes are more likely to grow, in a given environment, thus defining an ecological niche. In addition, individuals with traits x and y compete with each other with a strength that depends on their trait similarity, i.e., on their distance in phenotypic space, $|x-y|$, as specified by certain kernel function $\alpha(x-y)$, that can also be taken to be a Gaussian. Therefore, the total fitness of individual x, conditioned to the existence of individual y is specified by

$$\begin{aligned} f(x,y) &= K(x) - \alpha(x-y) \\ &= k + \exp\left(-\frac{x^2}{2\sigma_K^2}\right) - \exp\left(-\frac{(x-y)^2}{2\sigma_\alpha^2}\right), \end{aligned} \tag{3.42}$$

where $k > 0$ is the basal growth rate (warranting the non-negativity of the overall fitness) and both kernels are Gaussian functions centered at 0 and characterized by standard deviations σ_K and σ_α, respectively.

Adaptive dynamics and beyond

First, we introduce the effective fitness function and identify the possible evolutionary phases. Then, we resort to numerical integration of the generalized Crow-Kimura equation, Eq.(4.3), particularized for the present model. This allows us to gain much insight and visualize the emergent phenomenology. Finally we rationalize the results and quantitatively characterize the trait distribution using the framework introduced in the previous sections. Let us start by calculating the marginal fitness associated width trait x as given by Eq.(3.11), i.e. by integrating the effects of all other possible individuals, y, weighted over the trait distribution:

$$f(x,t) = K(x) - \int_{\mathcal{P}} dy \phi(y,t)\alpha(x-y). \tag{3.43}$$

Observe that while $K(x)$ fosters the concentration of the population around the origin and can be seen as an "attracting force", the competition kernel fosters a kind of "repulsion" between individuals. Moreover, the reach of these two evolutionary "forces" is controlled by their variances: σ_K and σ_α, respectively. The combined effect of these two forces is illustrated in the leftmost panels of Fig.3.3; the black curves represent the external fitness $K(x)$ and have always a maximum at the origin; the green curves stand for the competitive fitness and exhibit a maximum at the peak of the population distribution. Subtracting this last fitness component from the first one, leads to the total fitness function (red curves) which can exhibit a non-trivial shape. To mathematically analyze the dynamics and possible evolutionary phases, we start by considering the first derivative of the marginal fitness to determine the direction of the selection gradient and the location of its maximum, i.e. the attractor for the mean-trait value:

$$f_1^x(x) = -\frac{x}{\sigma_K^2}\exp\left(-\frac{x^2}{2\sigma_K^2}\right) \tag{3.44}$$

so that $x = 0$ is always an extreme point, and the mean value of the distribution converges to 0 along the evolutionary dynamics.

To go beyond standard AD while keeping the calculation as parsimonious as possible, let us expand the relative fitness, $F(x) = f(x) - \bar{f}$, around $x = 0$, including terms up to the 4−th order (as in Eq.(3.36)):

$$F(x) \approx \frac{2f_2^x + f_{2,2}^{xy}\Sigma(t)}{4}x^2 + \frac{f_4^x}{4!}x^4 \tag{3.45}$$

where now

$$f_2^x = \frac{1}{\sigma_\alpha^2} - \frac{1}{\sigma_K^2}, \tag{3.46}$$

$$f_{2,2}^{xy} = -\frac{3}{\sigma_\alpha^4}, \tag{3.47}$$

$$f_4^x = 3\left(\frac{1}{\sigma_K^4} - \frac{1}{\sigma_\alpha^4}\right) \tag{3.48}$$

so that f_2^x and f_4^x have always opposite signs, while the cross-derivative term, $f_{2,2}^{xy}$, is always negative. From Eq.(3.46) one readily sees that the relative fitness has a maximum at the origin $x = 0$ if $\sigma_\alpha \geq \sigma_K$ ($f_2^x < 0$), while the origin is a fitness minimum for $\sigma_\alpha < \sigma_K$ ($f_2^x > 0$). At the transition point $\sigma_\alpha = \sigma_K$ ($f_2^x = 0$), $x = 0$ is still a maximum, given that the only non-vanishing term, f_{22}^{xy}, is negative.
Let us discuss these two cases separately:
(i) No branching: If $\sigma_\alpha \geq \sigma_K$, i.e. the competition kernel has a reach larger than the external fitness, then there is a *stabilizing* evolutionary phase. This is illustrated in the upper row of Fig.3.3 (obtained from numerical integration of the GCK equation): the leftmost panel shows that when the competition kernel (green curve) is wider than the external-fitness (black curve), the resulting total fitness (red curve) is such that the population climbs the (changing) fitness landscape until the moment in which it stabilizes around the fitness maximum at the origin.
(ii) Evolutionary branching: In the case in which $\sigma_\alpha < \sigma_K$ the competition kernel has a reach smaller than the external niche. This scenario is illustrated in the B series of panels of Fig.3.3: observe that after an initial transient the population reaches a fitness minimum, thus leading to a branching event, i.e. to a "disruptive" evolutionary phase as predicted also by adaptive dynamics. After the branching, the two peaks correspond to extreme points of the overall fitness function. Note that the fitness landscape changes its concavity through all the evolution. In particular, after branching its becomes flatter and flatter, as qualitatively indicated by our theory ($f_{22}^{xy} < 0$), leading to what we call a "neutral bridge".
(iii) Multiple branching: Finally, by further decreasing σ_α —as explicitly illustrated in the lower row of Fig.3.3— the two peaks can happen to converge to fitness minima, so that each of them can experience a second-generation of branching events which —at least, transiently— lead to a distribution with four peaks (the position of the secondary branching points as illustrated in the lower row of Fig.3.3C. In particular, the lower panel of Fig.3.3 reveals that the two central peaks (out of the total 4 transient ones) eventually coalesce together at the origin —repopulating the new niche created after the first branching, as discussed above— and generating a 3-peak steady-state distribution. By progressively increasing the value of σ_K or diminishing σ_α one can find a cascade of further branching and coalescence events leading to steady-state distributions with a progressively larger number of peaks.
To summarize the previous results Fig.3.4 shows the phase diagram obtained from as a function of the two parameters (σ_α and σ_K) controlling the fitness function. In particular,

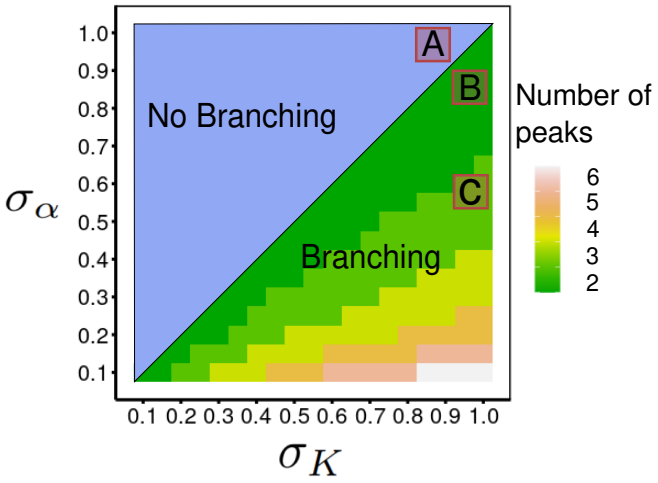


FIGURE 3.4: **Phase diagram for growth-competition model as a function of the two control parameters σ_α and σ_K.** Different colors represent the possible phases: uni-modal one with no branching (blue), while the "diversified phase" below the diagonal line (with colors running from green to reddish and white) corresponds to multi-modal distributions (colors codify the number of peaks). The points marked by letters A, B and C identify typical working points for the analyses shown in Fig.3.3, with none, one or two branching events, respectively. The diagram is obtained by integrating numerically the generalized Crow-Kimura equation, Eq.(4.3). The initial condition, as illustrated in Fig.(3.3), is a delta-Dirac function at $x = 0.1$ in all cases. Other parameter values: $k = 0.1$, $\sigma = 10^{-3}$.

the diversification cascade, consisting in a series of branching events and possible coalescence at the origin may generate steady states with $3, 4, 5, 6, 7$ or 8 peaks —as we have computationally verified— though arbitrarily large number of peaks are expected to be seen for sufficiently small values of σ_α and sufficiently large values of σ_K.
This observation can be easily rationalized: σ_K determines the overall width of the available niche space, so the larger σ_K the larger number of diverse phenotypes that can exist. On the other hand σ_α determines the reach of competition, the smaller it is, the closer two consecutive peaks can be, allowing for a more compact packing of "species".
Finally, it is also noteworthy that by analyzing the time-dependent behavior of the population distribution in situations in which many peaks are expected to occur in the steady state, we observe that in order to reach a final steady state with n peaks —starting from a uni-modal distribution at the origin— the probability distribution goes through all the $m < n$ possible intermediate phases by a series of evolutionary branching events —i.e. a cascade of dynamical phase transitions— possibly with coalescence events in between.

Trait distributions beyond Gaussian theory.

In order to go analytically beyond adaptive dynamics, we use the framework introduced in the previous sections to characterize the properties of the stationary trait distribution in the various phases. When the $\sigma_K \leq \sigma_\alpha$ the population converges to a Gaussian distribution with mean $x = 0$, coinciding with the fitness maximum, and a non-zero variance that can be approximated (see Eq.(3.31)) as:

$$\Sigma^* = \sigma \frac{2\sigma_K^2 - \sigma_\alpha^2}{\sigma_K^2 - \sigma_\alpha^2} \left(\sqrt{\frac{\sigma^2}{4} + \frac{2k}{\sigma_K^2 \sigma_\alpha^2}(\sigma_\alpha^2 - \sigma_K^2)} - \frac{\sigma}{2} \right). \tag{3.49}$$

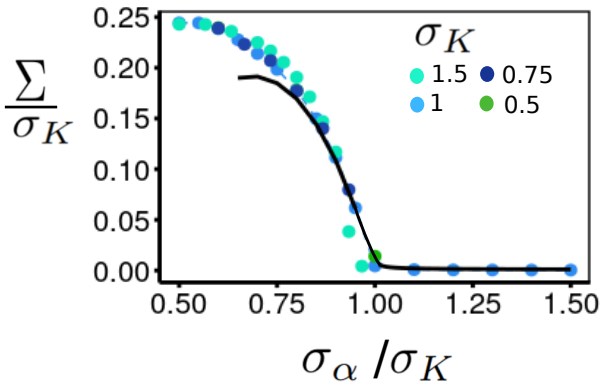


FIGURE 3.5: **Trait variance at stationarity as a function of σ_α (both rescaled with rescaled by σ_K) illustrating the analogy with a phase transition.** The plot shows results from numerical integration of the GCK equation for different values of σ_K (as color-coded in the legend; parameters values as in Fig.3.4). Observe that thanks to the rescaling by σ_K all different cases approximately collapse onto a single curve. The condition $\sigma_K = \sigma_\alpha$ universally determines the transition from monomorphic populations to evolutionary branching. The black line represents the theoretical prediction obtained from Eq.(3.39) ($\sigma_\alpha < \sigma_K$), Eq.(3.49) ($\sigma_\alpha \geq \sigma_K$) or 6-th order theory at $\sigma_K = \sigma_\alpha$. Such a theoretical approximation works relatively well for $\sigma_\alpha > 0.7\sigma_K$ but deviations are evident for smaller values of σ_α, suggesting that higher-order terms are needed in the expansion. Even if the mutation rate is not zero ($\sigma = 10^{-3}$), so that the variance of the monomorphic solution does not vanish, the analogy of this curve with that of the order parameter in a continuous phase transition is evident [9].

Note that at transition point, $\sigma_\alpha = \sigma_K$ this formula ceases to be valid, the variance explodes, and one needs to consider higher-order terms (up to 6-th order at least, see SI) to analytically estimate the steady-state variance.
On the other hand, when $\sigma_K > \sigma_\alpha$, the population —as already discussed— does, at least, an evolutionary branching, and to study it one needs to go beyond the quadratic approximation in the fitness expansion. In particular, the 4-th order expansion can be explicitly worked out in the limit of vanishing mutation amplitudes. In this limit, a double delta-function ansatz, as specified by Eq.(3.39), is justified. Within such an approximation, one can compute the location of the two peaks (as given by Eq.(3.39)):

$$x^* = \pm\sqrt{2\frac{\sigma_K^2\sigma_\alpha^2(\sigma_K^2-\sigma_\alpha^2)}{4\sigma_K^2-\sigma_\alpha^2}}, \tag{3.50}$$

which are real solutions given that, in the present case ($\sigma_K > \sigma_\alpha$), $x = 0$ is a fitness minimum, and are also automatically maxima for the same condition (see appendix B.3.5, Sec.C.2.1).
To increase the precision of the theoretical prediction one can keep terms up to 8-th order in the expansion together with a bimodal Gaussian approximation (i.e. the addition of two symmetric Gaussian distributions) for the stationary distribution. Even if no closed analytical formula is derivable, one can solve numerically equations for the stationary mean $\bar{x}$and variance Σ (dashed vertical lines in Fig. 3.3, see appendix sec.C.2.2for details).
In Fig. 3.5 we summarize the previous analytical derivations by comparing the stationary variance obtained by numerical integration of the GCK equation (color points) and the theoretical approximations (black lines,8-th order for branching, Gaussian theory for

no-branching and 6-th exactly at the transition point) for different parameter values. Observe that —as illustrated in Fig.3.5— by re-scaling both the variance and σ_α by σ_K all the plots collapse to almost the same curve, similarly to what happens for order parameters in second order phase transition [9]. From Fig.3.5, one learns that the theoretical prediction compares well with the numerics for $\sigma_\alpha/\sigma_K > 0.7$ while, below such a limit, further higher-order terms are necessary. In particular, as already discussed, for smaller values of σ_α more peaks appear in the distribution and the bimodal approximation fails.
Notably, the "branching points" for the second generation of evolutionary branching are predicted from the 4th order solution, Eq.(3.39) (dashed vertical lines in Fig.3.3 C). However, beyond such a second generation of branchings the theory cannot be studied in a closed analytical manner in simple terms, as this requires including at least up to the 12-th order terms in the fitness expansion and it becomes rather cumbersome [305]. However, we can heuristically determine the transition line to the multiple branching phase using the numerical evidence. This numerical information can be combined with scaling properties of the variance in Fig. 3.5.
Indeed, by decreasing σ_α the variance reaches a maximum at $\sigma_\alpha = 0.6\sigma_K$, and then decreases. This point coincides with the appearance of the third peak, an the decrease is due to the appearance of the further peaks in the space between the existing ones. Summing up, one can conclude that when the width of the growth kernel, $K(x)$, is broader than the competition one, i.e. when $\sigma_K > \sigma_\alpha$, evolutionary branching emerges in a robust way, without the need of additional requirements, and has intriguing similarities to continuous phase transitions in statistical physics, where the variance plays the role of an order parameter.

3.6 Stochastic finite size fluctuations

Demographic effects, stemming from finite-size populations (also known as "genetic drift" in the context of population genetics) are well-known to have a pivotal role in determining the fate of ecological and evolutionary communities. As a matter of fact, the consequences of demographic fluctuations have been extensively studied in the context of population genetics [64, 280], evolutionary game theory [306, 307, 308, 309] and adaptive dynamics [291, 148, 141]. In particular, it is well established that if the population size is relatively small, there may not be enough the variability for selection to act upon and stochastic effects can deterministic selection mechanisms [310, 148, 141].
In addition to introducing fluctuations around the deterministic behavior, demographic fluctuations have been described to give rise to novel and *a priori* unexpected phenomenology, such as e.g. evolutionary tunneling, population bottlenecks, inversion of the direction of selection, etc. [311, 312, 308].
In order to account for demographic effects in our framework one needs to move away from the infinite-population-size limit and include higher-order terms in an $1/N$ expansion.

Computational evidence of demographic effects

Before doing that, let us start by explicitly illustrating the difference between the previously derived deterministic theory and the actual outcome of evolutionary dynamics for finite populations. For this we show results from computer simulations of the previously defined master equation, Eq.(3.3) —characterizing the actual individual-based dynamics— for the specific case of the growth-competition model introduced in Sec.3.5. For this, we implemented an individual-based Gillespie algorithm as described in detail

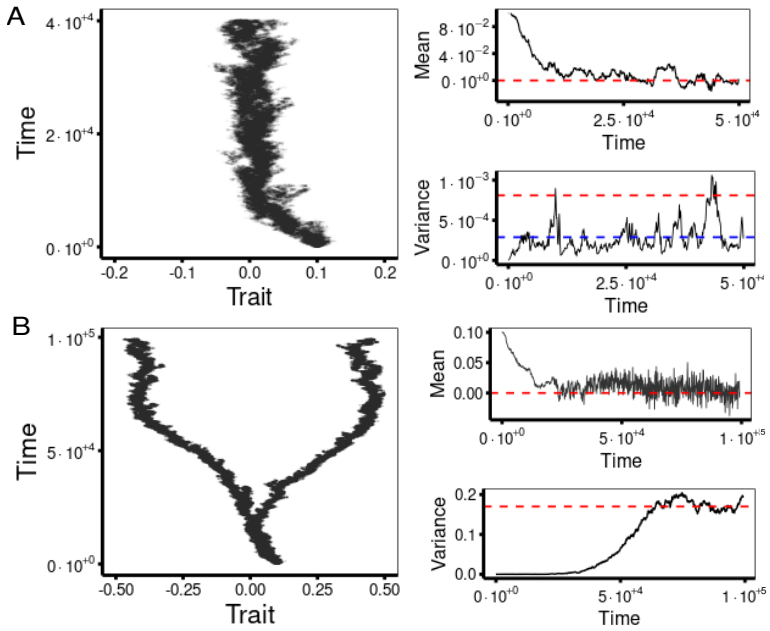


FIGURE 3.6: **Results of simulations of the individual-based growth-competition model.** Typical realization of the model dynamics as a function of time for parameter values in the monomorphic/no-branching phase (Upper panels, A) and in the phase with one branching event (two peaks) (Lower panels, B). Each dot in the left panels corresponds to an individual, thus illustrating the behavior of the population. Instead, the panels to the right show the evolution of the mean population trait (top) and its variance (bottom). The red dashed lines represent the theoretical predictions using Gaussian or Landau theory while the blue one (depicted only for the case in which the Gaussian approximation clearly fails), includes also next-to-leading order stochastic correction, as specified by Eq.(C.116). Note that in case (A) the variance shows the typical large and asymmetric fluctuations characteristic of multiplicative processes [106, 313]. On the other hand, the fluctuations of the mean in (B) increase significantly after the branching event as suggested by Eqs.(3.52). The model has been simulated using the individual-based Gillespie algorithm described in detail the SI. Parameter values: $N = 10^3$, $k = 1.$, $\sigma = 10^{-3}$, $\sigma_K = 1$ and $\sigma_\alpha = 1.2$ (A), 0.8 (B).

in the appendix B.3.5, Sec.C.5.

In particular, Fig.3.6 illustrates the results for the evolution of a finite population of size $N = 10^3$ and a moderate mutation amplitude, $\sigma = 10^{-3}$, both in the case (A) in which a monomorphic population is expected to emerge at the deterministic level (i.e. $\sigma_\alpha > \sigma_K$) and case (B) in which evolutionary branching is deterministically expected to emerge (i.e. $\sigma_\alpha < \sigma_K$).

Observe in the left panels of that the qualitative behavior is in good agreement with the deterministic theory predictions, but there is variability across time. More specifically, note that —as illustrated in the right panels— the mean value and the variance exhibit stochastic fluctuations around their corresponding steady-state averaged values, which roughly coincide with the deterministic expectations (red dashed lines) in three out of four reported plots. In the diverging case, i.e. for the variance in case (A), there are large asymmetric excursions around certain value (blue dashed line) that deviates from the deterministic prediction. Observe, finally that the variance in case (B) exhibits fluctuations whose amplitude grows increase significantly after the population branches out, while the overall variance converges to a value close to the deterministic one Eq.(3.49).

Moreover, in Fig.3.7 we illustrate the actual dynamics of the system —in case (B) in which

evolutionary branching is predicted at a deterministic level ($\sigma_\alpha = 0.8, \sigma_K = 1$)— for relatively small population sizes ($N = 200$ and $N = 1000$, respectively). The figure also illustrates how the results depend on the mutation amplitude σ ($\sigma = 10^{-3}$, $\sigma = 2 \times 10^{-3}$ and $\sigma = 5 \times 10^{-3}$).
The top panels, for $N = 200$, reveal that for small mutations, $\sigma = 10^{-3}$, there is no stable branching (which means that the population remains trapped in a fitness minimum). By increasing σ to 2×10^{-3} a tentative but frustrated branching appears; i.e. one of the branches becomes extinct and the remaining one moves back to the origin. Finally, for 5×10^{-3}, the population is able to generate two branches in a stable way. Similarly, in the bottom panels of Fig.3.7, we consider a large population of $N = 10^3$ in the case in which 3 peaks are deterministically expected to emerge ($\sigma_\alpha = 0.6$ and $\sigma_K = 1$). For $\sigma = 10^{-3}$, much as in the previous case, fluctuations frustrate the emergence of three sub-populations (i.e. the second series of branchings is frustated) and there are just two sub-populations (each of them trapped in a fitness minimum, from where it is not able to escape).
For $\sigma = 2x10^{-3}$ a final state with three sub-populations is reached but —on the contrary of the deterministic predictions (cfr with fig.3.3 C)— branching occurs in an asymmetric way (only in the leftmost part in this specific realization). Finally, for $\sigma = 5 \times 10^{-3}$, a complex multi-branching dynamics appears, where the sub-populations branch asymmetrically and wander in phenotypic space, eventually coalescing or going extinct, but keeping three populations, most of the time.
Summing up, as anticipated, demographic fluctuations have a profound impact on the actual evolutionary/adaptive dynamics that can diverge significantly from the expectations of the deterministic theory.

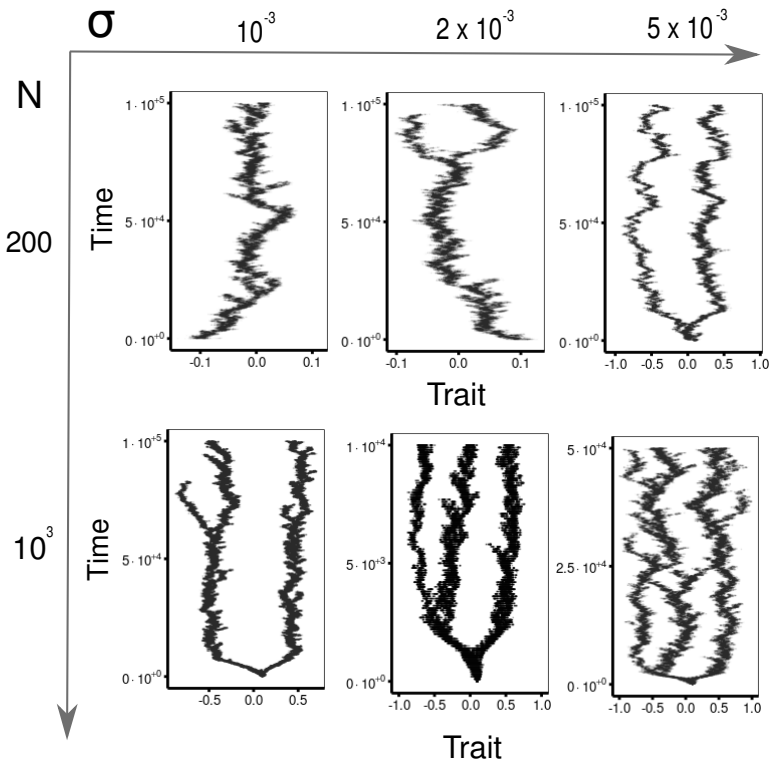


FIGURE 3.7: **Typical stochastic trajectories for the individual-based growth-competition model illustrating the finite-size effects and their dependence of the population size N and the mutation amplitude σ.** The evolutionary trajectories of single individuals are plotted (dots) both for (upper panels) the bi-modal phase ($\sigma_K = 1, \sigma_\alpha = 0.8$) and (lower panels) the 3-modal phase ($\sigma_\alpha = 0.6$). In the first case (upper panels), evolutionary branching is not observed (left) for sufficiently small populations ($N = 200$) and small mutation amplitudes ($\sigma = 10^{-3}$). Branching becomes stochastic, reversible, or "frustrated" (in the sense that one of the branches becomes extinct and the surviving one comes back to the central position) for intermediate values of $\sigma = 2 \times 10^{-3}$ (upper central panel) or, alternatively, by slightly increasing N (not shown). Finally, branching occurs in a stable and reproducible way (upper right panel) for larger values of the mutation amplitude $\sigma = 5 \times 10^{-3}$ (or larger system sizes). A similar phenomenology can be observed (lower panels) in the 3-peak phase (population size $N = 10^3$ in this example): for small mutations (lower left panel) the two branches are trapped in their corresponding fitness minima and when additional branching occur they are reversible or frustrated. Stable branching occurs (lower central panel) for intermediate values of σ (or larger values of N; not shown), while, finally (lower right panel) if σ is very high branches happen but have a finite life time are continuously frustrated and generated.

Stochastic theory for finite populations

To account for the previously reported deviations from the deterministic behavior, in what follows, we derive a generalization of the theoretical approach including finite population sizes.
Finite-size corrections to the deterministic theory are derived performing a size expansion, which is a rather standard procedure in the theory of stochastic processes (details can be found in the appendix sec.C.3).
In what follows, for simplicity in the presentation, we present the resulting equations just for the case of small, unbiased, and trait-independent mutations (for more general cases see the appendix sec.C.3.2).
Under these restrictions one can derive the following stochastic version of the GCK equation for the density in phenotypic space $\rho(x,t)$ (Eq.(3.5):

$$\begin{aligned}\dot{\rho}(x,t) &= \left(f(x)-\bar{f}\right)\rho(x) + \frac{\sigma^2}{2}\partial_x^2 f(x)\rho(x) + \\ &+ \frac{1}{\sqrt{N}}\sqrt{\left(f(x)+\bar{f}\right)\rho(x) + \sigma^2\partial_x^2 f(x)\rho(x)/2}\,\xi(x,t),\end{aligned} \tag{3.51}$$

where ξ is a delta-correlated, zero-mean, unit-variance Gaussian noise.
Note first that in the limit of $N\to\infty$, $\rho\to\phi$ and the original GCK equation is recovered and that the square-root functional form of the correction —stemming from the central limit theorem— is the usual one describing demographic noise in birth-death processes and [296]. Let us also emphasize that the stochastic correction consists of two contributions:

- The first term, $(f(x)+\bar{f})\rho(x)$ quantifies the fluctuations in reproduction and selection (i.e. is associated with the replicator-equation part of the GCK equation).
- The second term, $\partial_x^2\sigma^2 f(x)\rho(x)/2$ describes fluctuations in mutation events (i.e. is associated with the diffusive part of the GCK equation).

From the previous Eq.(3.51) it is straightforward to derive a couple of Langevin equations for the mean and the variance, which are the finite-population counterparts of the deterministic equations Eq.(3.26) and Eq.(3.29), respectively:

$$d_t\bar{x} = f_1^x\Sigma(t) + \frac{1}{\sqrt{N}}\sqrt{f(\bar{x})(2\Sigma+\sigma)}\,\eta_{\bar{x}}(t), \tag{3.52}$$

$$d_t\Sigma = f_2^x(\bar{x})\Sigma^2 + \sigma^2 f(\bar{x}) + \sigma^2\frac{f_2(\bar{x})}{2}\Sigma + \frac{1}{\sqrt{N}}\sqrt{6f(\bar{x})\left(\Sigma^2+\sigma^2\Sigma\right)}\,\eta_\Sigma(t); \tag{3.53}$$

where $\eta_{\bar{x}}$, η_Σ are zero-mean Gaussian white noises (see sec.C.3.2 in the appendix for a detailed derivation).
Observe that the additional noise term in the mean-trait equation, Eq.(3.52), grows with the trait variance Σ, explaining why —as illustrated in Fig.3.6B— fluctuations largely increase after a branching event happens. On the other hand, the noise in the equation for the variance, Eq.(3.53), is proportional to the variance itself, i.e. it is a *multiplicative-noise process* [313, 106]. This gives rise to a number of remarkable features. First of all, it explains the characteristic asymmetric excursions observed in Fig.3.6 (second panel in the right column). Second, remarkably, this multiplicative type of noise implies that there is an "absorbing state", meaning that the variance may become "trapped" in values very close to zero, from which it cannot escape stochastically. Indeed, as we discuss in what follows this latter effect explains the existence of "frustrated branching" as observed in

Fig3.7.
Let us assume for argument's sake that the mean of the trait distribution has already converged to a given stationary value, $\bar{x}^*$, which allows one to have a close Langevin equation for the variance, i.e. Eq.(3.53) with $x = \bar{x}^*$. From such a Langevin equation one can readily write its equivalent Fokker-Planck equation, and from it derive the steady-state probability distribution, which reads

$$P^*(\Sigma) \sim e^{-NV_N(\Sigma)} \tag{3.54}$$

where the effective potential $V_N(\Sigma)$ is

$$V_N = -\frac{2f_2^x}{f(\bar{x}^*)}\Sigma + \left(\frac{f_2^x\sigma^2}{f(\bar{x}^*)} + \frac{1}{N} + 2\right)\log\left(\Sigma + \sigma^2\right) - 2\frac{N-1}{N}\log\left(f(\bar{x})\Sigma\right) \tag{3.55}$$

(see appendix sec.C.3.3 for a full derivation). The crucial point is that, for finite values of N, the potential $V_N(\Sigma)$ exhibits a logarithmic singularity at the origin. In particular, in the limit $\sigma \to 0$ the potential converges to:

$$V_N = -\frac{2f_2^x}{f(\bar{x}^*)}\Sigma + \frac{2}{N}\log\Sigma \tag{3.56}$$

revealing the presence of a negative singularity at small Σ's (while if $\sigma \neq 0$ the singularity is replaced by a well near the origin). This implies that the evolutionary dynamics can become trapped at the potential well around $\Sigma = 0$, inducing possible effects absent in the deterministic limit. The negative singularity at the origin only exists for $\sigma = 0$, while for small values of σ and/or N, there might be a well near the origin and that as σ and N grow the relative weight of the potential well near the origin diminishes and eventually —for sufficiently large sizes and/or mutation amplitudes it disappears.

Is evolutionary branching frustrated?

To be more specific, the potential $V_N(\Sigma)$ is plotted in Fig.3.8 for a relatively small population size $N = 200$ and different values of the mutation amplitude σ (as color coded) for the case in which evolutionary branching is expected deterministically. both in linear and logarithmic scale.
Observe that at large values of Σ, the potential is well approximated by the linear term, that is proportional to $-f_2^x$, leading to a divergence in variance values, as in the case of deterministic branching ($f_2^x > 0$). On the other hand, at small values of Σ we see that the potential can exhibit different behavior depending on the value of σ: a deep well confining the system's state near zero, generating a trapping or "absorbing" state" with very small $\Sigma^* \approx 0$ (orange curves, for $\sigma \sim 0$), or a local shallow minimum with $\Sigma^* > 0$ for intermediate values of σ (dark, light blue and yellow curves, $\sigma = 10^{-2}, 10^{-3}$), and finally, a monotonous curve leading the system to a diverging variance, much as in the deterministic, $N \to \infty$, limit.
To further understand analytically these regimes, we analyzed the number of extreme points of the potential as a function of e.g. the size N and the mutation amplitude σ. (see C.3.3 in the appendix for details). The main conclusion is that, there exist three different regimes depending on the values of $N, \sigma, f(\bar{x}^*)$ and $|f_2^x(\bar{x}^*)|$, that represent the typical scales of drift, mutation, fitness and selection. In particular, to find the extreme points one needs to solve $\partial_\Sigma V_N(\Sigma) = 0$, which leads to :

$$\Sigma^*_{1,2N} = -\frac{\sigma^2}{4}\left(\gamma_N \pm \sqrt{\Delta_N}\right), \tag{3.57}$$

where $\Delta_N = \gamma_N^2 - 8\nu_N$ with

$$\gamma_N = 1 - \frac{f(\bar{x}^*)}{2f_2^x\sigma^2 N}, \quad \nu_N = \sigma^2 f(\bar{x}^*)\left(1 - \frac{2}{N}\right). \tag{3.58}$$

Observe that two new (positive and real-valued) extrema —one minimum and one maximum— absent in the deterministic limit emerge if $\Delta_N > 0$; this condition that can be approximately written as:

$$\Delta_N > 0 \longrightarrow \quad N < \left[N^* \sim \frac{f(\bar{x}^*)}{f_2^x\sigma}\right] \tag{3.59}$$

or, in words, when the N is smaller than a certain threshold value, controlled by the the mutation rate σ, the fitness $f(\bar{x}^*)$ and the selection force f_2^x (see sec.C.3.3 in the appendix for details). Therefore, new "non-deterministic" potential minima may appear for sufficiently small sizes. Equivalently, one can also fix a population size and look for new minima as , e.g. the mutation rate is decreased.
Actually, it is possible to distinguish between three regimes, as graphically recapitulated in Fig.3.9:

- *Deterministic branching.*
  If $N > N^*$, there is no additional extreme point and branching occurs as in deterministic limit (dark blue curve in Fig.3.8A). Thus, this phase is dominated by selection and demographic fluctuations are relatively small. Observe that, exactly at the value $N = N^*$ the potential shows a marginal point (light blue curve in Fig.3.8A and blue region in Fig.3.9), corresponding to the condition $\Delta_N = 0$.

- *Stochastic branching.*
  If $N \lesssim N^*$, the potential develops a relative minimum at Σ_{1N}^* together with a local maximum at Σ_{2N}^* (yellow curve in Fig.3.8A and yellow region in Fig.3.9). Therefore a "*drift barrier*" arises between the low variance minima $\Sigma_{1,N}^*$ and the large variance regime (that asymptotically leads to divergence and evolutionary branching). Therefore, if the variance at the branching point happens to be small, e.g., if the mutation amplitude is small, the population remains trapped for long times in the new minimum. However, possibly fluctuations may drive the population to jump the barrier inducing evolutionary branching stochastically. In this regime finite size fluctuations, selection and mutation are on similar scales.

- *Frustrated branching- trapped state.* If $N \lll N^*$, i.e. $\sigma x \lll N^{-1}$the minimum Σ_{1N}^* converges to zero [5]. In this limit, the potential converges to Eq.(3.56) showing a singularity at the origin and a the infinitely large drift barrier between the minimum and the maximum. The origin has become a trapping point, similar to an absorbing state (see orange curves in Fig.3.8 A and orange region in Fig.3.9). Hence, by decreasing the mutation amplitude σ, the probability of stochastically branch becomes as close to zero as desired and the population does not branch. This regime is thus dominated by drift, and mutations are so small that the population cannot diversify.

To close the loop, observe that these different regimes explain the computational results reported in Fig3.7 for $N = 200$. In particular, in the first plot $N \ll N^*(10^(- 3)) \sim 700)$

[5] This can be seen analytically by Taylor expanding the minimum value around $\sigma = 0$, leading to $\Sigma_{1,N} \approx N\sigma^2 \lll 1/N$, where in the inequality we have used $\sigma(\bar{x}^*)/N$.

and hence the population is trapped in the low-variance state for long times; in the second one $NN^*(2\times10^(-3)) \sim 100$ and the population can branch in a stochastic (reversible) way. Finally, in the rightmost plot $N > N^*(5\times10^(-3)) \sim 100$ and the population branches out in a stable way.
Before concluding, let us emphasize that the fact the genetic drift can hinder population diversification is already known in the context of adaptive dynamics [148, 272, 141]. However, to the best of our knowledge, the phenomenon has not been analyzed in this general mathematical detail nor related to the emergence of absorbing states and new attractors of the evolutionary dynamics [106, 85].
Similarly, if $f_2^x < 0$, i.e. in the case in which no-branching is deterministically expected, one can see Fig.3.8B that there is always a single minimum (corresponding to the only solution $\Sigma^*_{1,N}$ of Eq.(3.57)), but its associated variance changes as a function of N and σ. More specifically, for large values of $N\sigma$ the system behaves almost as in the deterministic case but, as this product decreases, the potential well becomes wider and wider, allowing for larger variability in possible variance values (with smaller mean value, though) i.e. the population develops large intrinsic diversity and temporal variability (see Fig.3.8B). In particular, this justifies the findings reported in Fig.3.7 for the individual-based model, for which the variance was observed to exhibit anomalous large fluctuations, asymmetric around their mean value.

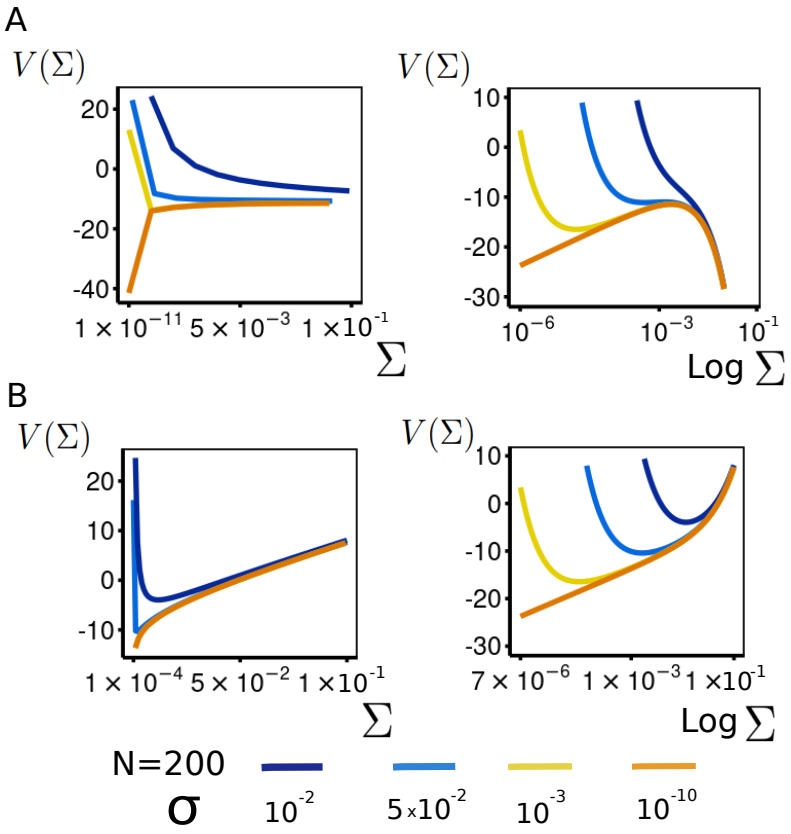


FIGURE 3.8: **Effective potentials, $V_N(\Sigma)$ for finite-size populations and variable mutation amplitudes.** The left panels show potentials in linear scale while the right ones show the same potentials in semi-logarithmic scale; $N = 200$ and different values of σ are employed (as color-coded in the legend). The upper panels (A) stand for a case in which branching is expected at a deterministic level (i.e. for large populations), while the lower ones (B) stand for a case in which a monomorphic population is expected in such a limit. In (A) there are three different regimes depending on the value of σ (and N). In particular, by decreasing the mutation amplitude σ, the population crosses over from a regime where the variance is pushed to infinity (blue curve) —so that branching occurs in an almost deterministic way— to another where a meta-stable minimum arises (yellow curve; as can be better appreciated in the semi-logarithmic plot) and branching occurs after a transient in a stochastic way. Finally, when $\sigma \sim 0$ the potential minimum converges to the origin, generating also an absorbing state (orange curve) in such a limit (and the population remains trapped in a fitness minimum). Similarly, the lower panels reveal a similar transition, but in this case between a situation with a narrow minimum —characterizing a monomorphic population for relatively large values of σ (or N)— and a fluctuation-dominated one, in which the minimum moves progressively closer to zero and the potential becomes flatter, as corresponds to populations with small variance and large fluctuations (see main text). Parameter values: $k = 1.0$, $\sigma_K = 1.0$, $\sigma_\alpha = 0.8$ (A) and $\sigma_\alpha = 1.2$ (B).

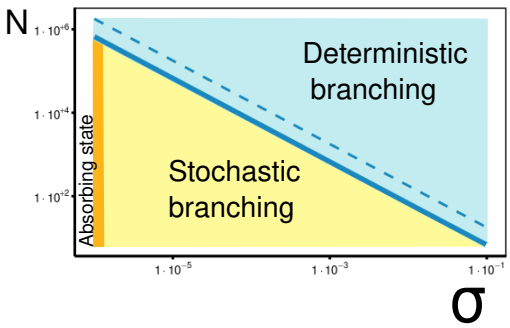


FIGURE 3.9: **Regimes of branching** The figure shows the transition lines between the different regimes of branching: deterministic branching (blue), stochastic branching (yellow) and trapped state (orange). The colors refer to the same regimes plotted in Fig.3.8 The blue region corresponds to deterministic branching, the blue continuous line stands for the critical size N^* that regulates the number of extreme points in the potential. The blue dashed line represents an approximation of the critical size as $N^* \approx f(\bar{x}^*)/(f_2^x \sigma)$. By crossing the blue line, the system enters in the stochastic branching regime, where a metastable low-variance state arises. The probability of reaching the diversified state depend on the size of mutations and the height of the barrier. Finally, when $\sigma \approx 0$, $N^* \to \infty$ the population cannot diversify and is trapped in the low variance state with no escaping probability. Given that the trapping state exist just in the limit $\sigma = 0$ we represent this phase as orange region on the leftmost part of the diagram.

3.7 Conclusions and discussion

Historically, microbial evolution has been studied using population genetics, thanks to technological advances enabling the design of high-precision and long-term evolutionary experiments providing access to genetic information and global-fitness measurements of whole populations [11]. In this context, recent efforts have allowed to generalize classic population genetics models to rapid evolution, using e.g. the formalism of fitness travelling waves as well as concepts from statistical physics [160]. On the other hand, microbial phenotypic eco-evolution —that was traditionally left aside owing to the difficulties in measuring single-cell traits [161, 162]— has received reinvigorated attention [163, 145, 164, 165, 166, 167], as a result of technological advances in determining single-cell traits [168, 169, 170] and metabolic functions [171, 172, 45, 173]. These novel quantitative empirical descriptions of microbial phenotypic diversity —which are crucial for the rapidly-developing field of microbial ecology [297, 314, 80]— call for the design of comprehensive eco-evolutionary theoretical frameworks.

In this chapter we started this ambitious goal with the concrete objective of generalizing adaptive dynamics to eco-evolutionary scenarios. We have performed this generalization using the perspective and the tools of statistical physics, in particular its capacity to connect the miscroscopic scale with the mascroscopic one. As a result, we constructed a general framework for the dynamics of trait distributions, being them Gaussian or not. This frameworks is useful to both clarify many aspects of adaptive dynamics and to go beyond it, for example in studying what happens beyond the branching point. Finally, we have shown how to add stochastic finite-size fluctuations to our framework and unify existing approaches and results in an elegant way. In Fig.3.10, we graphically resume the derivation of the different parts of our evolutionary framework.

Many exciting possibilities open up as follow-ups of the present work. A first purely theoretical extension of our work would be to generalize our derivation to individual based processes with varying population size, like a general birth-death process or a Wright-Fisher model. These generalization would be mathematically not trivial and would increase notably the present comprehension of many-body stochastic processes.

We believe that it is important to use our theory to analyze biologically more structured models, such as consumer-resource ones [163, 144] where different ecological interactions like competition, cross-feeding, environmental fluctuations, etc. maybe simultaneously at work. In particular, environmental fluctuations seem to be the dominant force shaping the statistics of natural microbial communities [297, 80, 124], and very little is known on their influence on evolution [315]. To formulate an eco-evolutionary theory able to predict the evolution of metabolic functions in such complex ecological scenarios is a long-term ambitious goal. To this aim also the influence of spatial effects is emerging as extremely important and will need to be studied [316, 317].

Another direction to further explore the similarities between evolutionary theory and statistical physics is to analyze whether more complex fitness landscapes, such as rugged ones [318, 130, 131], show similarities with other classes of phase transitions, chaos [52, 154], or replica symmetry breaking phenomena [63]. Finally, another promising future direction for this work would be to study the non-equilibrium properties of evolution with our framework and to quantify irreversibility both at a microscopic and macroscopic level, complementing recent studies in the literature [184, 319, 320]. We believe that our work indicate a deep and fruitful relation between evolution and statistical physics that can be further explored by both communities.

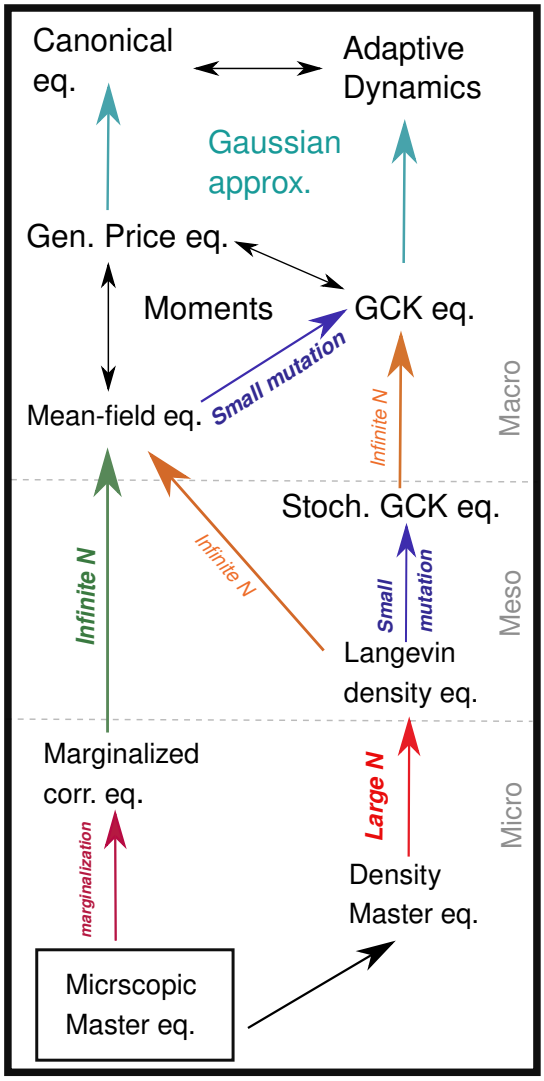


FIGURE 3.10: **Derivation scheme**. The figure represents a scheme that summarized of our framework in its different formulation. Normal arrows represent approximations, while double arrows stay for equivalence. We divide the possible formulations in three category: microscopic (i.e. based on individual rates), mesoscopic (i.e. finite-size stochastic population density) and macroscopic (infinite size population and trait probability density). Starting from the microscopic level (bottom), we formulated an individual based master equation (see Eq.(3.3)). By using marginalization, it is possible to formulate an equation for the 1-particle density $\phi(x)$ that depends on the two and three body probability (purple arrow, see appendix B.3.5, Sec.C.1.1). By using also the mean-field approximation ($N \to \infty$, green arrow), we formulated the mean-field selection mutation equation, that is a macroscpic description of the population, see Eq.(3.9). Finally, by taking the small mutation approximation, we derived the Generalized Crow-Kimura equation (GKC see Eq.(4.3), blue arrow). Alternatively, from the microscopic Master equation, we can derive an alternative Maser equation for the trait density ρ (bottom right, black arrow see appendix B.3.5, Sec.C.3.1). By using a Kramers-Moyal expansion, from the previous description we can derive a Selection-Mutation Langevin equation (red arrow), that if simplified with the small mutation approximate gives a mesoscopic correction to the GCK equation named the Stochastic gen. Crow-Kimura equation (blue arrow, see Eq.(3.51)). These two Langevin equations are *mesoscopic* because they describe a population density in the large but fined-size limit. Finally, by taking the $N \to \infty$ limit in the SGCK equation, one recovers its deterministic version (orange arrow). If one, instead of studying the dynamics of the trait distribution with the GCK equation, considers the dynamic of the population statistical moments, obtains a family of generalized Price equation, that are equivalent to the GCK Eq (double black arrow, see Eq.(3.2)). Finally, if uses a Gaussian approximation for the trait distribution ϕ, recovers the formalism of Adaptive Dynamics (light blue arrows, Sec.3.3).

Chapter 4

Evolution of tolerance in bacterial populations

4.1 Introduction

The extraordinary ability of species to adapt and survive in unpredictably-changing and unfavorable environments is certainly one of the most astonishing features among the many wonders of the phenomenon that we call life. Such adaptations can occur at extremely fast temporal scales thus interspersing ecological and evolutionary processes [321, 322, 323]. A widely spread surviving strategy is *latency* or *dormancy*, i.e., the possibility for organisms to enter a period of reduced metabolic activity and non-replication adopted during adverse environmental conditions [324, 325, 326, 327, 328]. Examples of dormancy can be found across kingdoms, with examples ranging from microorganisms such as viruses, bacteria or fungi [329, 330, 331, 332] to plants [333, 334] and animals [327]. During the latency period the organism is said to be in a *latent or dormant state* and the time it takes to wake up is referred to as "lag time" or simply "lag". Entering and exiting a dormant state are not cost-free processes, since individuals may require of a specific metabolic machinery for performing such transitions and/or the development of specifically-devised "resting structures" [335, 336, 324, 325, 337]. The exit from the dormant state can occur either as a response to environmental signals or cues [324, 325, 338, 330] or, alternatively, in a stochastic way [339, 340, 341, 342]. As a matter of fact, the duration of the lag intervals often varies widely between conspecific individuals and even between genetically identical organisms exposed to the very same environmental conditions [343, 330, 344, 339]. Such a variability is retained as an example of *phenotypic diversification* or *bet-hedging strategy* [345, 346] that confers a crucial competitive advantage in unpredictable and rapidly changing environments, thus compensating the above-mentioned individual costs and providing important benefits to the community as a whole [324, 325, 330, 347, 348, 334].

Although, as already stated, latency is a widespread phenomenon, bacterial communities constitute the most suitable playground for quantitative analysis of latency owing to their diversity, fast life cycle, and the well-controlled conditions in which they can grow and proliferate in the laboratory [349, 11, 350, 351]. Actually, latency was first described by Müller back in 1895 as an explanation for the observed irregularities in the growth rate of bacterial cultures in his laboratory [352]. In recent years it has been realized that bacterial latency is a more complex and rich phenomenon than previously thought. Indeed, paraphrasing a recent review on the subject, the lag phase is "dynamic, organized, adaptive, and evolvable" [330].

Bacterial latency is at the root of *tolerance* to antibiotics as, rather often, bactericidal antibiotics act during the reproduction stage and thus, by entering a dormant state, bacteria become transiently insensitive to antibiotics. Let us recall that bacterial *tolerance* is not to be confused with bacterial *resistance* [353]. While *resistance* refers to the ability of

organisms to grow within a medium with antibiotics, provided these are not in high concentrations, *tolerance* is the ability to transiently overcome antibiotics, even at very high concentrations, provided the exposition time is not too large [354, 343, 353]. The strengths of these two complementary surviving strategies are quantified, respectively, in terms of quantities: **(i)** the *minimum inhibitory concentration* (MIC) of drug that must be supplied to stop the population growth —a quantity that is significantly increased in resistant strains [353, 355, 344, 343]— and **(ii)** the *minimum duration to kill* 99% *of the cells* MDK_{99}, which is increased in tolerant strains [356].

While the importance of bacterial resistance has long been recognized, studies underlining the crucial role played by tolerance are less frequent and more recent [353, 355, 344, 343]. An important caveat is that, while resistance is specific to one or a few antibiotics, tolerance is generically effective for a large diversity of them, leading to survival even under intensive multidrug treatment [353, 344, 343]. Moreover, there exists firm evidence that tolerance is the first response to antibiotic stress [355], facilitating the later appearance of resistance [344]. Therefore, understanding the emergence of tolerance is crucial for the development of more effective therapies aimed at dealing with recalcitrant infections and possibly preventing them. Aimed at shedding light on these issues, here we present an eco-evolutionary approach to analyze the emergence of tolerance by lag in bacterial communities under controlled laboratory experiments. In particular, we scrutinize the conditions under which modified lag-time distributions evolve as a response to stressful environments and investigate the origin of the experimentally-observed broad heavy tails in lag-time distributions (see below).

Beside this specific focus, the present work has a broader breath. The example of rapid evolution of lag-time distributions is used as a test to prove the theoretical framework developed in ch.3. Our framework is similar in spirit to existing approaches such as the theory of *"adaptive dynamics"* and related models in population genetics [357, 358, 359, 360], but aims at reconciling and generalizing them.

As a historical sidenote, let us recall that *adaptive dynamics* (AD) was born as a generalization of evolutionary game theory [361] to allow for a set of strategies that is continuously varying and, upon which selection acts. AD led to the satisfactory explanation of intriguing phenomena such as evolutionary branching [357, 358, 362, 363], speciation [364, 365, 366], diversification [157, 145], the emergence of altruism and cooperation [367, 368], and the evolution of dispersal [369]. Importantly, its foundations are also mathematically well-established [288]. However, in spite of its very successful history, AD in its standard formulation has some limitations that make it not directly applicable to complex situations such as the one we aim at describing here:

(i) First of all, in its standard formulation, populations are considered as monomorphic, i.e. point-like in phenotypic space; thus it does not allow for phenotypically-structured populations (see however [370, 371]).

(ii) The "macroscopic equations" of AD for the populations are not easily connected to microscopic birth-death processes in individual-based models [372].

(iii) Variations are assumed to be small, typically Gaussian-distributed and independent of the parent's phenotypic state.

(iv) Variations are considered to be rare: "after every mutational event, the ecological dynamics has time to equilibrate and reach a new ecological attractor" [373]. In other words, a separation is assumed between ecological and evolutionary timescales, while in microbial communities, such processes may occur in concomitance. Such a convergence of characteristic timescales is the hallmark of eco-evolutionary dynamics [321, 322, 323] and is at the basis of fascinating phenomena such as eco-evolutionary tunneling [373, 374] and other rapid evolutionary phenomena [375, 376, 350, 377, 378] which are difficult to account for in the standard formulation of adaptive dynamics.

A full account of this general theoretical framework is presented the previous chapter3.

The chapter is organized as follows: in the first section we discuss in detail the experimental setup and empirical findings object of our study; then, we introduce a stochastic individual-based model implementing phenotypic variability and inheritability to account for experimental results. We present an extensive set of both computational and analytical results for it, discussing in particular the conditions under which the mathematical results deviate from computational ones. Finally, we discuss the implications of our work both from a biological viewpoint and how it contributes to the understanding of the evolution of heterogeneous phenotypic distributions, as well as from a more general eco-evolutionary perspective.

Empirical observations: rapid evolution of lag-time distributions

For the sake of concreteness, we focus on recent experimental results on the rapid evolution of tolerance in populations of *Escherichia coli* in laboratory batch cultures in Balaban's lab [343]. In particular, a bacterial population is periodically exposed to antibiotics (*amplicillin*) in very high concentrations (much larger than the *MIC*) during a fixed-duration time interval T_a (e.g., $T_a = 3, 5$, or 8 hours). After antibiotic exposure the system is washed and the surviving population is regrown in a fresh medium during a time interval T (with $T = 23h - T_a$). The antibiotics/fresh-medium cycle is iterated at least 8 or 10 times. Results are averaged over 2 experimental realizations for each T_a and the resulting maximal carrying capacity is about 10^9 individuals (we refer to [343] for further biological and experimental details).

Once the cycles are completed, Fridman et al. [343] isolated some individuals from the surviving community and by regrowing them in a fresh medium they found that the distribution $P(\tau)$ of lag times τ —i.e. the time individual dormant cells take to start generating a new colony after innoculation into a fresh medium— changes from its ancestral shape to a modified one, shifted towards larger τ values. More specifically, the mean value grew to a value that approximately matches the duration of the antibiotic-exposure time interval, T_a (see [343]). This modified lag-time distribution entails an-Moreover, the distribution develops right-skewed heavytails, revealing the presence of individuals with anomalouslylarge lag times. increase in the survival probability under exposure to ampicillin but, also, to antibiotics of a different bactericidal class such as *norfloxacin*, for the same time period. Furthermore, mutations were identified in diverse genes, some of them known to be related with regulatory circuits controlling the lag-time distribution, such as the toxin–antitoxin one [379]. Subsequently, after many cycles, the population was also observed to develop resistance to ampicillin [343]. Thus, the conclusion is that non-specific tolerance —stemming from lag— emerges in a very rapid way as a first adaptive change/response to antibiotic stress. More in general, these results reveal that the adaptive process is so fast that ecological and evolutionary processes occur at comparable timescales [378, 380, 381, 382, 383].

The experimentally-determined lag-time distributions reveal another intriguing aspect that —to the best of our knowledge— has not been extensively analyzed so far: their variance is also significantly increased as T_a grows and, related to this, the resulting mean value of the distribution is always larger than its median [384]. This is an indication that, as a matter of fact, the empirically-obtained lag-time distributions are skewed and exhibit heavy tails, including phenotypes with anomalously-large lag times —much larger than T_a—, especially for large T_a's. This observation is surprising as, under such controlled lab conditions, one could naively expect to find lag-time distributions sharply peaked around the optimal time value, T_a, since, ideally, the best possible strategy would

be to wake-up right after antibiotics are removed and any further delay comes at the price of a reduction of the overall growth rate or fitness. Fridman *et al.* proposed that the increase in the variance might suggest a past selection for a bet-hedging strategy in natural unpredictable environments; however, anomalously-large lag-time values were not present in the original wild-type population. The authors also suggested that there could be constraints at the molecular level imposing the mean and the variance of the lag-time distribution to increase concomitantly [343, 379], a possibility that inspired us and that we will carefully scrutinize from a theoretical and computational perspective in what follows.

4.2 Computational Model

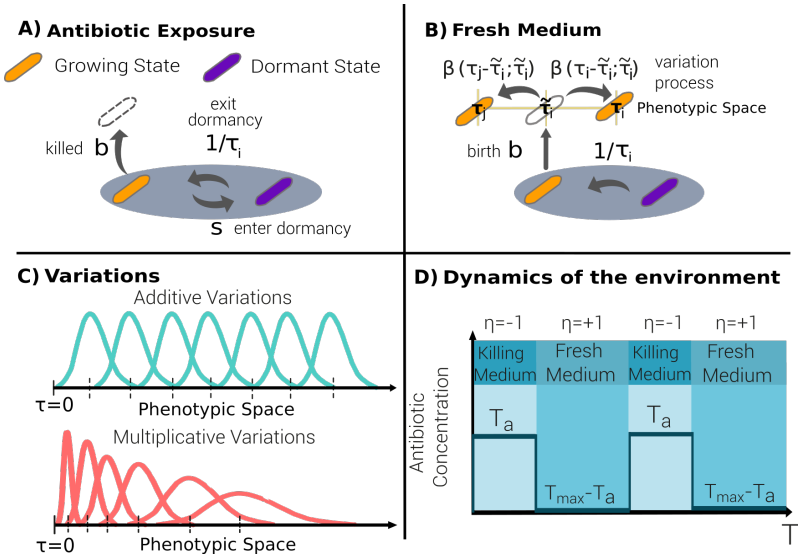


FIGURE 4.1: **Sketch of the main ingredients of the individual-based stochastic model**. Each individual bacterium (i) is characterized by its phenotypic state, lag time τ_i and experiences demographic processes. **(A)** In the presence of antibiotics, bacteria can stochastically switch between the dormant and the growing state (at transition rates s and $1/\tau_i$, respectively); growing individuals can also attempt reproduction (at a "birth" rate b) and be immediately killed by the action of antibiotics (as bactericidal antibiotics usually act during duplication attempts). **(B)** In the fresh medium, dormant bacteria can wake up at a rate $1/\tau_i$, that depends on their intrinsic (phenotypic) lag time; on the other hand, growing cells can reproduce asexually by duplication; the resulting offspring inherit the characteristic time scale with some variation, as specified by a function β. **(C)** Two possible types of variation functions β: in the additive case (top), the standard deviation is constant, i.e. independent of the initial state τ_i, while in the multiplicative case (bottom) the standard deviation is assumed to grow linearly with the parent's lag time τ_i. **(D)** Sketch of the environmental variation, alternating periodically between antibiotic exposure (time T_a) and a fresh medium ($T_{max} - T_a$).

Aimed at shedding light onto these empirical findings, here we propose an individual-based stochastic model for phenotypic adaptation in which each single individual cell can be either in an "awake" or in a "dormant" state [10, 342, 385, 386] (see Fig 4.1 for a sketch of the model). Mimicking the experimental protocol of Fridman et al.— a population of such individuals is exposed to alternating adverse and favorable conditions with durations T_a and $23h - T_a$, respectively (aMoreover, the distribution develops right-skewed heavytails, revealing the presence of individuals with anomalouslylarge lag times. function $\eta(t)$ labels the environmental state at any given time t: $\eta(t) = -1$ in the presence of antibiotics and $\eta(t) = +1$ in the fresh medium).

The model assumes that each awake cell is able to sense the environment and respond to it by regulating its state: they can sense the presence of antibiotics and enter the dormant state at rate s, while such a machinery is assumed to be turned off during dormancy. In appendix D, Sec. D.5, we also consider a generalization of the model in which awake individuals can also enter the dormant state as a response to other sources of stress such as starvation [387]. Indeed, the wake-up is assumed to occur as a result of a Markovian stochastic process; each individual bacteria i is *phenotypically* characterized by its intrinsic typical mean lag time τ_i meaning that, it wakes up stochastically at a constant transition rate $1/\tau_i$. Therefore, the time t at which a dormant cell actually resumes growth is a random variable distributed as $P(t|\tau_i) = e^{-t/\tau_i}/\tau_i$, with mean value τ_i [388, 91]. In the last section we discuss recent alternatives to Markovian processes, i.e. including some form of "memory", which can give rise to non-exponential residence times, to describe this type of waking-up phenomena [389, 390].

Awake individuals are exposed to stochastic demographic processes: they attempt asexual reproduction (i.e., duplication) at a constant birth rate b and die spontaneously at rate d (that we fix to 0 without loss of generality). Reproduction attempts are successful in the fresh medium while, in the presence of antibiotics, they just lead to the parent's death and its removal from the community. Following this dynamics, the population can freely grow, until its size reaches a maximal carrying capacity K. Once this limit has been reached, the population enters a saturated regime, within which each new birth is immediately compensated by a random killing (much as in the Moran process [295]).

Importantly, in parallel with the above demographic processes, the model implements an evolutionary/adaptive dynamics. The phenotypic state τ_i of each successfully dividing individual is transmitted, with possible variation, to its progeny. In particular, the two offspring resulting from duplication have phenotypic states $\tau_i + \xi_1$ and $\tau_i + \xi_2$, respectively, where ξ_1 and ξ_2 are the phenotypic stochastic variations, sampled from some probability distribution, that we generically call $\beta(\xi; \tau_i)$ and that, in the more general case, can be state-dependent, i.e. depend on τ_i. More specifically, we implemented two different variants of the model, depending of the standard deviation of the probability distribution $\beta(\xi; \tau_i)$:

- The *additive* model, with a standard deviation, α_A, common to all phenotypes.
- The *multiplicative* model, with a state-dependent standard deviation, $\alpha_M \tau_i$, for individuals with intrinsic lag time τ_i, where α_M is a constant (see Methods).

Observe that in the multiplicative case, the larger the parent's lag time the larger the possible amplitude of variations, in a sort of rich-get-richer or Matthew-effect mechanism, well-known in the theory of stochastic processes to generate heavy tails [**Sornette**, 391, 392, 393, 394, 112, 395, 110, 396, 397]. As a motivation for this choice, let us mention that there is solid evidence that the genetic circuits involved in the regulation of the lag-time distribution (such as the toxin-antitoxin one), can indirectly produce this type of fluctuations at the phenotypic level [379]. Furthermore, similar phenotypic-variation kernels have been argued to arise from non-linear effects in the way genotypic changes (mutations) are manifested into phenotypicMoreover, the distribution develops right-skewed heavytails, revealing the presence of individuals with anomalouslylarge lag times. variability (see e.g. [398, 399]).

4.3 Analytical (mean-field) theory

In this section we apply the theoretical framework of Chapter 3 to our problem. Before delving into computational analyses of the model, let us present a mathematical framework allowing us to obtain theoretical insight. Readers not particularly interested in analytical approaches can safely skip this section, and just be aware that it is possible to mathematically understand all the forthcoming computational results.

The previous Markovian stochastic individual-based model is mathematically defined as a "many-particle" Master equation ruling the time evolution of the joint probability-distribution functions for the whole set of all "particles" (i.e., cells). The resulting master equation can be simulated computationally by employing the Gillespie Algorithm (see below and appendix D, Sec.D.1.2, for details) [388, 91, 400]. However, as it is often the case for such many-particle Master equations, it is hard to handle analytically in an exact way. Thus, in order to gain quantitative understanding beyond purely computational analyses, here we develop an approximation —which becomes exact in the limit of infinitely large population sizes [401, 402]— that allows us to derive aMoreover, the distribution develops right-skewed heavytails, revealing the presence of individuals with anomalouslylarge lag times. macroscopic (or "mean-field") description of the stochastic model in terms of the probability density of finding an individual at any given phenotypic state, τ (i.e. the "one-particle" probability density), see chapter3 for details.

A first step toward the derivation of a *macroscopic equation* relies on a marginalization of the many-particle probability-distribution function to obtain a one-particle probability density (see appendix D Sec.D.2). The resulting marginalized distribution function encapsulates the probability density $\phi(\tau,t)$ that a randomly sampled individual at time t has lag time τ. This probability —that needs to be normalized, so that $\int_0^\infty \phi(\tau,t)d\tau = 1$— can be decomposed in two contributions $\phi(\tau,t) = \phi_G(\tau,t) + \phi_D(\tau,t)$ representing, respectively, the relative fraction of individuals in growing (G) and dormant (D) states. Observe that these two densities are not probability distributions and thus they are not normalized to unity separately. In the limit of infinitely-large population sizes, the evolution of the probability density for individuals in the growing state, $\phi_G(\tau,t)$ is ruled by the following equation (details of the derivation can be found in appendix D, Sec.D.2):

$$\begin{aligned}\partial_t\phi_G(\tau,t) &= \frac{1+\eta(t)}{2}\left[-b\phi_G(\tau,t) + 2b\int_0^\infty d\tilde{\tau}\beta(\tau-\tilde{\tau};\tilde{\tau})\phi_G(\tilde{\tau},t)\right.\\ &- \left. b\phi_G(\tau,t)\int_0^\infty d\tilde{\tau}\phi_G(\tilde{\tau},t)\right]\\ &- \frac{1-\eta(t)}{2}\left[b\Big(1-\int_0^\infty d\tilde{\tau}\phi_G(\tilde{\tau},t)\Big)+s\right]\phi_G(\tau,t) + \frac{1}{\tau}\phi_D(\tau,t). \qquad (4.1)\end{aligned}$$

Even if this equation might look cumbersome, its different terms have a rather intuitive interpretation:

- *In the fresh medium* (terms proportional to $1+\eta(t)$): (i) the first term represents the negative probability flow stemmingMoreover, the distribution develops right-skewed heavytails, revealing the presence of individuals with anomalouslylarge lag times. from growing individuals with generic phenotypic trait τ that reproduce (at rate b) and change to any other arbitrary phenotypic state; (ii) the second represents the positive contribution of reproducing individuals (at rate b) with any arbitrary trait $\tilde{\tau}$, for which one of the two resulting offspring jumps to τ (controlled by the function $\beta(\tau-\tilde{\tau};\tilde{\tau})$); (iii) the third *selection term* stems from the normalization of the overall probability density: if the population size grows because any individual

with arbitrary trait $\tilde{\tau}$ reproduces successfully (at rate b), then the relative probability to observe phenotype τ decreases to keep the overall probability-density conserved.

- *In the presence of antibiotics* (terms proportional to $1-\eta(t)$): (i) the first term represents the rate at which growing individuals that attempt reproduction (at rate b) are killed by antibiotics; (ii) the second term is a selection term, fully analogous to the above-discussed one: when any arbitrary individual dies the overall probability density increases. (iii) the third term represents the outflow of individuals entering the dormant state at rate s.
- *In both environments* (no dependence on $\eta(t)$): the only term, proportional to the rate $1/\tau$, describes the probability inflow stemming from dormant individuals that become awake.

Similarly, the equation for the density of individuals in the dormant state is

$$\partial_t \phi_D(\tau,t) = -\eta(t) b \phi_D(\tau,t) \int_0^\infty d\tilde{\tau} \phi_G(\tilde{\tau},t) - \frac{1}{\tau}\phi_D(\tau,t) + \frac{1-\eta(t)}{2} s \phi_G(\tau,t) \quad (4.2)$$

where the first (selection) term stems from the overall probability conservation when the population either grows or shrinks (negative or positive signs, respectively), and the remaining two terms have the opposite meaning (and signs) of their respective counterparts in Eq.(4.1).

In order to make further analytical progress, in the case in which variations are assumed to be small, it is possible to introduce a further ("diffusive" or "Kimura") approximation as often done in population genetics as well as in adaptive or evolutionary mathematical approaches [403]. More specifically, one can perform a standard (Kramers-Moyal) expansion of the master equation by assuming that jumps in the phenotypic space are relatively small [388, 91], i.e. expanding the function *beta* in Taylor series around 0. After some simple algebra (see appendix D, Sec.3.2) one obtains a particularly simple expression for the overall probability distribution:

$$\begin{aligned} \partial_t \phi(\tau,t) &= \eta(t)\left[f(\tau,t) - \bar{f}(t)\right]\phi(\tau,t) \\ &- (\eta(t)+1)\left[\partial_\tau \theta(\tau) f(\tau,t)\phi(\tau,t) - \frac{1}{2}\partial_\tau^2 \sigma^2(\tau) f(\tau,t)\phi(\tau,t)\right] \end{aligned} \quad (4.3)$$

where the " *effective fitness function*" $f(\tau,t) \equiv b\phi_G(\tau,t)/\phi(\tau,t)$ and its population average $\bar{f}(t) = \int_0^\infty d\tau f(\tau,t)\phi(\tau,t)$ have been introduced, and where $\theta(\tau)$ and $\sigma^2(\tau)$ are the first and second cumulants of the variation function β (in first approximation we can assume $\theta(\tau) = 0$, while $\sigma^2(\tau) = \alpha_A^2$ for the additive case and $\sigma^2(\tau) = \alpha_M^2\tau^2$ for the multiplicative case). Observe that, remarkably, this last equation is a generalization of the celebrated continuous-time *Crow-Kimura equation* of population genetics [404], also called *selection-mutation equation* [405, 406]. In particular, notice that the dynamics of the probability density is exposed to the combined action of the process of selection (first term in Eq.(4.3), which is nothing but the *replicator equation* [361, 246]) and mutation, as specified by the drifts in the second line. This type of equations, combining replicator dynamics with Fokker-Planck type of terms —even if with a slightly different interpretation— have been also studied by Sato & Kaneko and Mora & Walzak [407, 408, 409]. The main —and crucial— differences between Eq.(4.3) and the standard Crow-Kimura equation are:

- The fitness function appears in the mutation terms —whereas in the standard Crow-Kimura equation the diffusion term would read $\partial_\tau^2\phi(\tau)$— thus correlating reproduction rates and mutation amplitudes. Observe that here variations are always

associated with reproduction events, as typically in bacteria and viruses, in such a way that a higher fitness rate implies a higher mutation rate.

- There is a general dependence on the cumulants of the variation kernel that, in general, can be trait-dependent and asymmetric.

These generalizations are essential ingredients to capture the essence of our Markovian model as we will see and, to the best of our knowledge, have not been carefully analyzed in the past. From here on, we refer to Eq.(4.3) as the *generalized Crow-Kimura (GCK) equation*.

4.4 Model results

In order to scrutinize whether the proposed adaptive stochastic model can account for the key empirical findings of Fridman *et al.* [343], we perform both (i) extensive computational simulations and (ii) numerical studies of the mean-field macroscopic equation, Eq.(4.3).

- Computational simulations rely on the Gillespie algorithm [400], which allows us to simulate exactly the master equation defining the stochastic model. In all cases, we consider at least 10^3 independent realizations to derive statistically-robust results. Without loss of generality and owing to computational costs, the maximal population size or carrying capacity is fixed to $K = 10^5$.
- On the other hand, for analytical approaches, in spite of the relatively simple form of Eq.(4.3) owing to its non-linear nature and to the time-variability of environmental conditions $\eta(t)$, it is not possible to solve it analytically in a closed way and, thus, it becomes mandatory to resort to numerical-integration schemes. In particular, from this equation —or, more precisely, from integration of its two additive components: Eq.(1) and Eq.(2)— one can derive the time-dependent as well as the asymptotic lag-time distributions and, from them, monitor the leading moments or cumulants as a function of time.

Further details of both computational simulations and numerical integration of the macroscopic equation can be found in the Methods section as well as in the appendix D. In what follows we present together both types of analyses, underlining where the mean-field approach works well and where its predictions deviate from direct simulations.

Transient dynamics: determining variational amplitudes

Parameter values in the model are fixed to agree as much as possible with the empirical ones measured by Fridman et al. [343] (see Methods). In particular, we used (i) the same set of environmental-period durations T_a and $T_{max} - T_a$ as in the antibiotics/fresh-medium cycle, (ii) the experimentally measured reproduction rate in the fresh medium, (iii) the empirical "falling-asleep" rate s, as well as (iv) the same number of antibiotic cycles (ten) as in the experimental setup. Initially all individuals are assumed to have small intrinsic lag-time values of τ. In particular, we consider a truncated normal distribution with mean value and variance as in the actual ancestral population in the experiments ($\langle\tau\rangle^{exp.} = 1.0 \pm 0.2\ h$).

Employing this set of experimentally-constrained parameter values and initial conditions, we ran stochastic simulations in which the whole population expanded and then shrank following the periodically alternating environments. Along this dynamical cyclic

process the distribution of τ values across the population varies in time; in particular, we monitored the histogram of τ values and obtained the corresponding probability distributions right at the end of each antibiotic cycle, just before regrowth, as in the experiments.

Fig 4.2A and 4.2B show the evolution of the mean (i.e. the first cumulant, K_1) across cycles, while Fig 4.3A and 4.3B illustrate the full distribution and higher-order cumulants after 10 cycles. Observe that the value of K_1 after 10 cycles depends on the choice made for the only remaining free parameter, i.e. the variation-amplitude parameter α_A or α_M, for the additive or multiplicative versions of the model, respectively. In order to tune either of these parameters, we imposed that $K_1(10)$ reproduces in the closest way the experimentally determined values, as measured right before the $10th$ regrowth cycle. This tuning procedure leads to $\alpha_A = 0.16(1)h$ and $\alpha_M = 0.048(1)$ for the additive and multiplicative cases, respectively (parentheses indicate uncertainty in the last digit); these are the values that best reproduce the empirical findings in the sense of least-square deviation from the available empirical data for different T_a's (see Fig 4.2C).

Let us remark that both variants of the model are able to reproduce the key experimental feature of generating mean lag times close to T_a (observe, however, that there is always a small deviation in the case $T_a = 8\ h$, for which even experimentally, $K_1 \approx 10\ h > T_a$). Nevertheless, as illustrated in Fig 4.3A and 4.3B there are significant differences between the two variants. In particular, the additive model fails to reproduce the following empirical observations:

- T_a-dependent variances,
- large differences between median and mean values, and
- strongly skewed distributions with large tails.

For instance, in the experiments, for $T_a = 8\ h$, the difference between the mean and the median is $1.1(1)\ h$ while in the additive model is $0.15\ h$, i.e. about one order of magnitude smaller. Furthermore, in the experiments, lag times of up to $30\ h$ are observed, while in the additive model values above ≈ 15 are exponentially cancelled; i.e. they have an extremely low (negligible) probability to be observed. This is also illustrated in Fig 4.3D where the second and third cumulants (variance and skewness) of the distribution after 10 cycles are plotted as a function of T_a. Observe that both cumulants remain almost constant, revealing the absence of heavy tails for large values of T_a.

On the other hand, the multiplicative model is able to reproduce not only the experimental values of the mean but also —with no additional parameter nor fine tuning— (i) the existence of large lag-time variances that increase with T_a, (ii) the above-mentioned large differences between the mean and the median (1.3(1) in this case), as well as (iii) heavily skewed lag-time distributions that strongly resemble the empirically measured ones (see Fig. 2 in Fridman *et al.* [343]). In particular, lag times of the order of $30\ h$ have a non-negligible probability to be observed for $T_a = 8\ h$ within the multiplicative version of the model, after 10 cycles. The resulting probability and the corresponding cumulants (see Fig 4.3D depend strongly on T_a.

Importantly, the previous results are quite robust against changes in the model. In particular, if growing cells are allowed to switch to dormancy in response of starvation, the mean lag time increases, as expected, but the qualitative shape of the lag-time distribution remains unchanged (see sec.D.5). Hence, just by modifying accordingly the parameter α allows one to recover the same conclusions.

As a word of caution let us emphasize that the distributions in Fig 4.3 are not obtained exactly in the same way as the experimental ones. The first are distributions of

characteristic times τ (inverse of intrinsic transition rates) while the second are the actual lag times t measured after regrown in a fresh medium. Actually, the characteristic time τ, in our model, is just a proxy for the actual time that it takes for the colony formed by such an individual to be observable or detectable in actual experimental setups. Below we discuss this issue more extensively as well as the possible limitations it implies and extensions of the modelling approach to circumvent them.

Let us also underline that Fig 4.3 reports not only the results of direct simulations but also the theoretical predictions (dashed lines) derived from numerical integration of the macroscopic equations for the two different cases. The agreement with simulation results is remarkably accurate; the origin of the existing small discrepancies will be analyzed in detail in a forthcoming section.

Thus, the main conclusion of these computational and theoretical analyses is that *state-dependent (multiplicative) variability is needed in order to account for the empirically observed key features of the lag-time distributions emerging after a few antibiotic/fresh-medium cycles.* Once this variant of the model is chosen, a good agreement with experimental findings if obtained by fitting the only free parameter: the amplitude of variations.

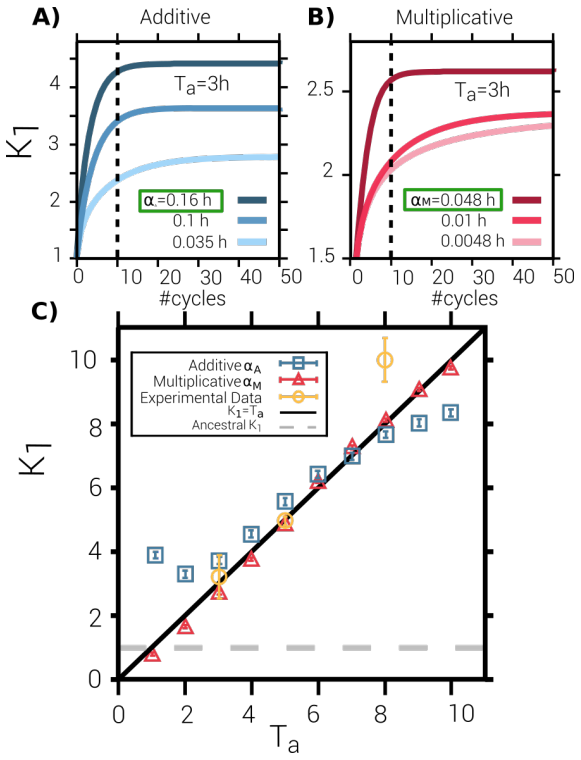


FIGURE 4.2: **Tuning the only free parameter to match empirical results.(A-B)** Mean of the lag-time distribution K_1, measured right at the end of each antibiotic cycle, as a function of the number of cycles. Computational results are shown for antibiotic duration $T_a = 3h$ for both the additive (A) and the multiplicative (B) versions of the model. The different curves (color coded) correspond to different mutation amplitudes α_A (in A) and α_M (in B), respectively. The dashed vertical lines indicate the $10th$ cycle, when experiments stop. Remarkably, the mean lag time strongly depends on the mutational amplitude, both in the transient regime and in the asymptotic state. We implement an algorithmic search to tune the only free parameter (either α_A or α_M) to best fit the experimental mean lag times for all values of T_a together and, in particular, their experimentally-reported linear dependence on the antibiotic exposure time T_a (see Methods). **(C)** Mean of the lag-time distribution as a function of T_a for the model (squares for additive and triangles for multiplicative versions of the model) tuned to reproduce experimental values (yellow symbols). While empirical data are available for $T_a = 3, 5$ and $8h$, the model can be analyzed for generic values of T_a. The solid line indicates the linear dependence between the mean and lag-time distribution, $K_1 = T_a$, while the horizontal dashed line represents the mean lag time of the ancestral population. Parameter values: $K = 10^5$, $\alpha_A = 0.16h$, $\alpha_M = 0.048$, $b = 2.4h^{-1}$, $d = 3.6 \cdot 10^{-5}h^{-1}$, $s = 0.12\ h^{-1}$, $T_{fresh} = 23h - T_a$, 10 cycles (see Methods).

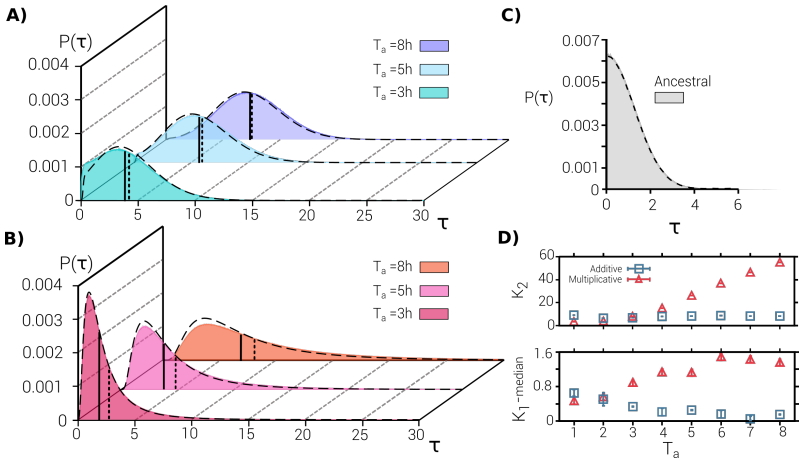


FIGURE 4.3: **Lag-time probability distributions: theory and simulations. (A-B)** Lag-time distribution after 10 cycles as obtained in the simulation of the individual-based model in both the additive (A) and the multiplicative (B) case, for different antibiotic-exposure periods, $T_a = 3, 5$ and $8h$ (marked with different colours). Solid and dotted vertical lines indicate, respectively, the median and mean of the corresponding distributions (a large separation between these two indicators reflects asymmetries in the distribution such as the emergence of a heavy tail to the right). Dashed lines represent results from the numerical integration of the GCK equation, Eq.4.3, using the same parameters and external conditions. Observe that the multiplicative model generates much larger tails, reproducing the experimental phenomenology better than the additive one. **(C)** Initial lag-time distribution mimicking the experimentally observed one for the ancestral population. **(D)** Variance, K_2, and difference between mean and median, $K_1 - median$, of the lag-time distribution as a function of T_a in the additive (blue squared symbols) and multiplicative (red triangular symbols) versions of the model. K_2 grows with the antibiotic exposure time in the multiplicative case, while in the additive case it remains nearly constant. The difference between the mean and the median is very small in the additive case, while it increases with T_a almost monotonically in the multiplicative one. In summary, the multiplicative model generates a distribution with a variance that grows with the mean, as well as heavy tails, reproducing well the key experimental findings. Parameter values are as in Fig 4.2.

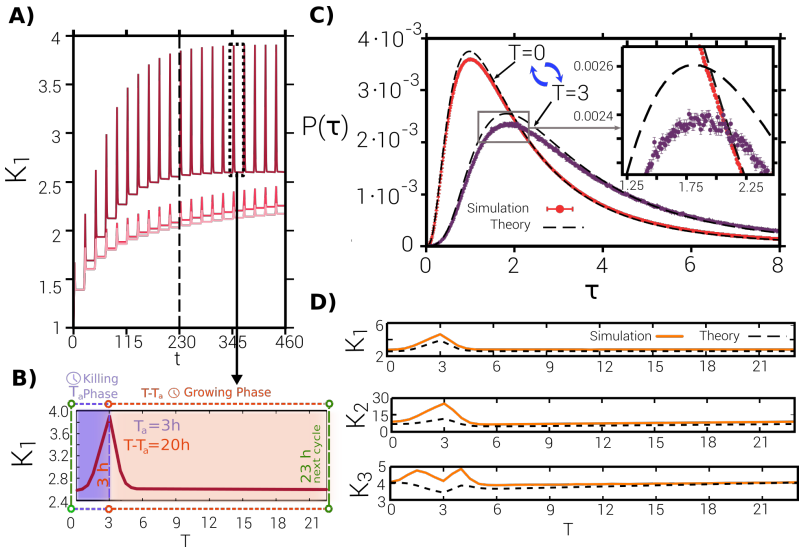


FIGURE 4.4: **Characterization of the asymptotic state in the multiplicative version of the model** . **(A)** Approach to the dynamic asymptotic state for the multiplicative case, as resulting from the integration of Eq.(1) and Eq.(2) for $T_a = 3h$ (t indicates overall time as measured in hours). The different curves correspond to three different values of the variation amplitude (from bottom to top: $\alpha_M = 0.0048, 0.01, 0.048$). The difference between this plot and Fig 2 is that K_1 is measured at different times within the cycle and not just right at the end of antibiotic exposure (see (B)). The vertical dashed line marks the $10th$ cycle at which the experiment stopped. Observe that the steady-state mean value, the oscillations amplitude, and the relaxation time depend on the variation amplitude α_M. **(B)** Mean lag time within a single cycle ($T \in [0, 23]h$.) in a asymptotic state. During the killing phase (antibiotic exposure), i.e. for $t < T_a$, the mean lag time increases to maximize the number of dormant individuals; then, in the fresh medium the mean relaxes back to the initial value. **(C)** Lag-time probability distribution —as derived from theory (dashed lines) and computationally (solid lines)— at the start of the cycle (leftmost curve) and when antibiotics are removed (rightmost curve); in the asymptotic state the system oscillates between these two limiting probability distributions, both of them exhibiting heavy tails. **(D)** Evolution of the three first cumulants K_1, K_2, K_3 (mean, variance, and skewness, respectively) within a asymptotic cycle (both theoretical and computational results are shown). Observe in (C) and (D) that the theory correctly predicts the properties of the distribution but there are some small errors due to finite size effects.

Asymptotic state

Even if experimental results are available for a fixed and limited number (10) of antibiotic-exposition cycles, the already-calibrated model allows us to scrutinize the possible emergence of asymptotic states after a much-larger number of cycles. In other words, it is possible to go beyond the experimental limits and analyze the fate of the population. In this sense, the experimental results can be seen as a "transient adaptation" to the environment, while the evolutionary cycle would be completed only when an asymptotic (evolutionary stable) state is reached. Let us remark that the asymptotic state is necessarily a *periodic* one, as the phenotypic distributions vary at different instants of the cycle, i.e. the asymptotic distributions —measured at arbitrary times within the cycle— exhibit periodic oscillations in its shape, tracking the perpetual environmental cyclic changes. This is illustrated in Fig 4.4, showing results obtained by numerically integrating the macroscopic equations, Eq.(1) and Eq.(2). First of all, it shows periodic oscillations of the mean lag time K_1; as shown in panel (A) it first increases from its initial value $K_1 = 1$ and then, eventually, reaches an oscillatory steady state. More specifically, as clearly seen in the zoomed plot of panel (B), within the steady state, the maximum mean value within each cycle is reached right before antibiotics removal. This is an expected result as in the first part of the cycle, i.e. during the "killing phase", the presence of antibiotics induces a selective pressure towards increasing the mean lag-time value because delaying the exit from the lag phase provides protection from the antibiotics. On the other hand, in the fresh medium (growing phase) the selective pressure quickly reduces the mean lag time to foster fast growth and increased fitness. Thus, summing up, the periodic alternation of environmental conditions induces a stable periodic change in the mean lag-time value.

Actually, it is not only the mean that changes periodically, but the whole probability distribution that varies cyclically. This is illustrated in Fig 4.4C and 4.4D which shows computational and theoretical results for the lag-time probability distribution and its first cumulants, K_1, K_2 and K_3, for the multiplicative case (similar plots for the additive case are shown in appendix D). Observe, in particular, in panel C, that the distribution oscillates between two extreme or limiting cases corresponding to the times of antibiotics inoculation and antibiotics removal, respectively. This effect can be more vividly seen in the videos that we have produced, see sec.D.7

Let us also highlight that the probability distributions exhibit non-Gaussian tails and are right-skewed. In particular, to make these observations more quantitative, Fig 4.4D shows the variance, K_2, and the skewness, K_3, along the cycle in the steady state. Notice also the very-good —though not perfect— agreement between computational results and theoretical estimates (dashed lines in Fig 4.4C and 4.4D).

Furthermore, let us emphasize that, importantly, the amplitude of the variations —as controlled by the parameter α_M (or, similarly, α_A for the additive case)— has a non-trivial effect on both the transient and the asymptotic behavior. In particular, the value of such amplitude not only affects the mean value of lag times after 10 cycles —as illustrated by the plateau of the oscillations in Fig 4.4A and 4.4B— but also (i) its asymptotic value, i.e. the mean lag time, (ii) the amplitude of the oscillations across a cycle in the steady state, and (iii) the relaxation time to the asymptotic state (i.e., the speed of evolution). This is due to the heavy tails of the distribution: increasing the amplitude of variations directly increments the variance of lag times, but this also enlarges the left-skewness of the distribution, feeding-back to the mean value. Therefore, the eco-evolutionary attractor is shaped both by selection and mutation, departing from the classical evolutionary scenario, as e.g. in adaptive-dynamics, in which the amplitude of the variations just affects the variance of the resulting distribution but not the overall attractor.

Finally, we complement our observations with the population structure dynamics,

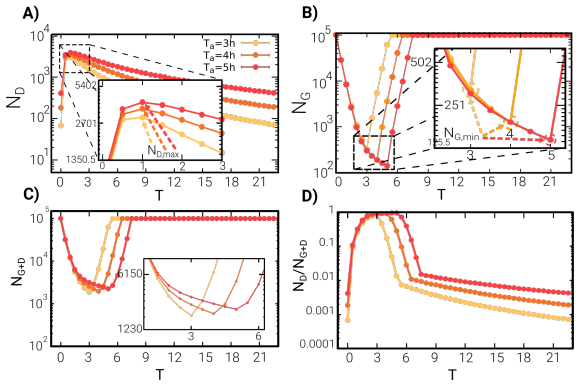


FIGURE 4.5: **Population dynamics** . Abundances N_D and N_G of dormant **(A)** and growing **(B)** cells, respectively, along a full cycle in the asymptotic regime (reached after only three cycles) for the multiplicative model (the curves are the result of averaging over many independent realizations for different T_a as color coded; observe the semi-logarithmic scale). Dormant cells abundances reach a maximum value after approximately one hour of exposition to antibiotics, almost independently of T_a, and then start a slow decrease, while N_G exhibits an opposite trend: it rapidly decreases and reaches a minimum at $T = T_a$ (see inset), after which it grows exponentially fast until the carrying capacity is reached. **(C)** The total number of cells $N = N_G + N_D$ is plotted along the cycle: for all values of T_a the absolute minimum is reached near T_a (as clearly seen in the inset). **(D)** The fraction of dormant cells relative to the total number is maximal nearby T_a and decreases when antibiotics are removed.

i.e. proportion, minimum and maximum of dormant and awake cell numbers, in Fig 4.5. In particular, panels (A) and (B) show the abundances of dormant and awake cells as function of time along an asymptotic cycle. Observe that the number of dormant cells reaches a maximum, $N_{D,max}$, after one hour, independently of the antibiotic duration time T_a, while its height is proportional to this parameter. On the other hand, the position of minimum of growing cells number, $N_{G,min}$, scales with T_a and its magnitude decreases correspondently. In Fig 4.5 we also show and discuss the dependence the total number of cells $N = N_G + N_D$ (panel C) as well as the relative fraction of dormant individuals along a full cycle in the stationary state (which is reached after onley a few (three) antibiottic cycles).

For the sake of completeness, let us also emphasize that both versions of the model are able to generate MDK_{99} values that grow as a function of the number of antibotic cycles, converging to an asymptotic-state value; at the end of the tenth cycle simulations compare well with empirical observations for different values of T_a (see FigsD.10,D.11 ; observe that the largest difference appears for $T_a = 8$, a case for which also K_1 deviates slightly from T_a in the experiments).

Deviations between theory and simulations: finite-size effects

Thus far, we have reported results stemming from computational analyses of the individual based model as well as from numerical integration of the associated macroscopic theory, i.e. the GCK equation. Small but systematic discrepancies between theory and simulations are evident, see for example Fig 4.4C and 4.4D. Let us here discuss the origin of such differences.

The theoretical approach relies on two different approximations: (i) on the one hand it considers the small-variation approximation to include just the first two moments of the variation function (i.e. a diffusion approximation); (ii) on the other hand, in order to derive the macroscopic GCK equation, one needs to neglect correlations between individuals, a type of mean-field approximation that, as usual, is expected to be exact only in the

infinite population-size limit [388, 91]. In appendix D, Sec.D.4, we show computational evidence that the small mutation approximation is not a significant source of errors; hence, the discrepancies necessarily stem from finite-size effects. Indeed, in the present experimental set up, there is a bottleneck at the end of each antibiotics cycle, when there is a small number of surviving individuals, thus limiting the validity of the mean-field approximation in such a regime. As a matter of fact, one can clearly see from Fig 4.4D that the largest discrepancies appear around the end of the killing phase, when the population is the smallest. Note also that the main features of the dynamics in phenotypic space are reproduction and variation: i.e., offspring are similar to their progeny. But reproduction events occur only within the awake (growing) sub-population; the full population-size, involving also dormant ones, is not the most relevant quantity to gauge finite-size effects. Therefore, in order to minimize the discrepancies between theory and simulations it does not suffice to consider larger population sizes: even for huge values of the carrying capacity K, we find that the population at the end of the killing phase is always rather small and, hence, exposed to large demographic fluctuations, i.e. to finite-size effects.

To put these observations on more quantitative bases, we define a parameter δ as the deviation between theoretical and computational results for the mean lag-time value after antibiotic exposure and monitor its dependence on the minimal size of the awake population (i.e. right at the end of the antibiotic phase). Fig 4.6 illustrates that: (A) the deviation grows with the antibiotic-exposure time T_a, whereas (B) the minimum awake subpopulation size decreases with T_a. Combining these two pieces of information one can see (C) that the deviation parameter δ decreases as the minimum subpopulation of awake individuals increases. Unfortunately, the convergence to zero of this last curve is very slow, and thus, it is computationally very expensive to remove finite-size effects.

Finally, let us remark that we leave for future work the formulation of an extension of the mathematical theory accounting for finite-size effects [410, 411, 148], including corrections to the GCK equation.

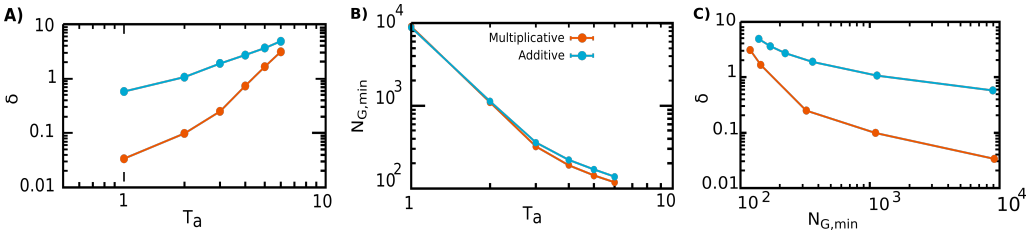


FIGURE 4.6: **Analysis of the deviations between simulations and theory.** The parameter δ is defined as the difference between the mean lag times (right at the end of the antibiotic cycle) in the theoretical approach and in computer simulations. **(A)** Double-logarithmic plot reporting the dependence of the error parameter δ on the antibiotic time exposure T_a for both the additive (blue dots) and the multiplicative (red dots) versions of the model; in either case, the larger the exposure time the larger the error. **(B)** Minimum number of awake individual during the cycle in the asymptotic regime, $N_{G,min}$, as function of T_a in double-logarithmic scale. As expected, the larger the exposure time the smaller the number of surviving individuals. **(C)** Combining the data from (A) and (B) it follows that δ decreases with increasing $N_{G,min}$, meaning that deviations between theory (expected to be exact for infinitely large population sizes) and computational results stem from finite-population-size effects. Notice that errors are smaller in the multiplicative version of the model.

4.5 Conclusions and discussion

Summary of results and conclusions. We have presented a mathematical and computational model to quantitatively analyze the emergence and evolution of tolerance by lag in bacteria. Our first goal was to reproduce the main results reported in the laboratory experiments of Fridman *et al.* [343] in which the authors found a very fast evolution of tolerance by lag in a community of *Escherichia coli* bacteria periodically exposed to an antibiotics/fresh-medium cycle. In particular, after a relatively small number of such cycles, there is a clear change in the individual-cell lag-time distribution with its mean value evolving to match the duration of antibiotic exposure. This is remarkable, and demonstrates that tolerance by lag is the first and generic strategy adopted by bacteria to survive under harsh environmental conditions such as the presence of antibiotics. A second key empirical finding is that concomitantly with the evolution of the mean lag time, also the variance of lag times is significantly increased for longer antibiotic-exposure periods: i.e. the harsher the conditions the more diversified the lag times within the population. More generally, the full lag-time distribution becomes wider and develops a heavy tail for sufficiently large times. This means that there exist individual phenotypes that are clearly sub-optimal under the strictly controlled laboratory conditions and most-likely reflects a bet-hedging strategy, preparing the community to survive under even harsher conditions (i.e. longer stressful periods).

To shed light onto these observations we developed a stochastic individual-based model assuming that individuals are characterized by an intrinsic lag time, setting the "typical" time at which such individual stochastically wakes up after dormancy. This phenotypic trait is transmitted to the progeny with possible variation. By considering a protocol analogous to the experimental one (i.e. alternating antibiotic exposure and fresh medium growth) the model is able to produce a distribution of characteristic lag times across the population that reproduces quite well the empirical results in all cases by tuning a single parameter value. In particular, the emerging lag-time distributions have a mean that matches the period of antibiotic exposure T_a, an increase of the mean and variance with T_a, as well as a large difference between the mean and the median, which result from the appearance of heavy tails in the lag-time distributions. Nevertheless, it is important to underline that the distributions that the model generates are just a proxy for the empirically-determined ones, where the actual times in which individual bacteria give rise to new and detectable (i.e. visible with the available technology) colonies are measured.

Importantly, in order to account for all the above empirical phenomenology, the model needs to assume *multiplicative variations*, i.e. that the variability between the parent's trait and those of its offspring increases (linearly) with the parent's lag time: the larger the parent's lag time the larger the possible variation. This multiplicative process — at the roots of the emerging heavy tails in the lag-time distribution— resembles the so-called *rich-get-richer* mechanism of the *Matthew effect* [98, 392, 393, 110, 391]. This type of variations implements an effective dependence between the parent's trait value and the variation amplitude, that was hypothesized as a possible mechanism behind the experimental results and that could stem from a highly non-linear map between genotypic changes and their phenotypic manifestations [343, 379].

Notably, our analyses reveal that the amplitude of variations affects not just the variance (K_2) of the resulting lag-time distribution, but also its mean (K_1) as well as other

higher-order cumulants such as the skewness (K_3). This is in blatant contrast with standard approaches to evolutionary or adaptive dynamics, in which the "mutational amplitude" only influences the "broadness" (K_2) of the distribution of traits in phenotypic space, but does not alter the attractor of the dynamics (e.g. K_1). Thus, the introduction of state-dependent (multiplicative) variability constitutes a step forward into our understanding of how simple adaptive/evolutionary processes can generate complex outcomes.

Let us finally mention that our model describes rapid evolution, where ecological and mutational time scales are comparable. This interplay between ecological and evolutionary processes is explicit in the asymptotic state: it is not an "evolutionary stable state" but a *"non-equilibrium evolutionary stable state."* By non-equilibrium we mean that the detailed-balance condition —a requirement of equilibrium states [412]— is violated and thus, there are net probability fluxes in phenotypic space. These correspond to adaptive oscillations in phenotypic space. Key properties of such a state (oscillation plateaus, amplitudes, etc.) depend on the mutational amplitude, i.e., the amplitude of variations determine the eco-evolutionary attractor. In chapter6 we will scrutinize much in depth non-equilibrium characteristic properties, such as non-vanishing entropy-production of these type of complex eco-evolutionary processes [184].

Advantages and limitations of the phenotypic-modeling approach. As already underlined, the present model assumes adaptation at a phenotypic level. Is this a biologically realistic assumption? The answer to this question, in principle, is affirmative but some caveats are in order.

First of all, let us recall that a large part of the theoretical work on evolutionary dynamics and adaptation developed during the last decades focuses on phenotypic adaptation. For instance, in the theory of adaptive dynamics, individuals are always characterized by some phenotypic trait or set of traits which is subject to selection and transmission to the progeny with variation [357, 358, 359, 364, 365] (see also e.g. [381, 382]). In general, this is the most parsimonious way of modeling adaptation as the details of the genotypic-phenotypic mapping are usually highly non-linear or simply unknown (see e.g. [399, 413, 414, 415, 416]).

On the one hand, adaptation beyond genetic changes, for example epigenetic adaptation, is a well-documented phenomenon in the bacterial world [417] and is the focus of intense research activity [418, 419, 346, 420, 421]. For instance, recent work explores "the evolutionary advantage of heritable phenotypic heterogeneity", which suggests that evolutionary mechanisms at a phenotypic level, such as the ones employed in our approach, might be biologically favored with respect to more-standard genetic mechanisms, under certain circumstances [422]. In particular, such phenotypic variability can provide a faster and more flexible type of response than the one associated with traditional genetic mutations.

Nevertheless, it is important to underline that Fridman *et al.* found empirical evidence that —in their specific setup— genetic mutations were always present in the evolved strains. In particular, they found mutations in genes controlling the so-called toxin-antitoxin circuit, mediating the response to antibiotic stress [343]. This regulatory circuit is known to lead to "multiplicative fluctuations" in the lag-time distribution at the phenotypic level [379]. Thus, strictly speaking, our modeling approach constitutes an effective or phenomenological approximation to the more complex biology of this problem.

This observation opens promising and exciting avenues for future research to shed light on how broad probability distributions of lag times —possibly with heavy tails—

can be actually encoded in phenotypic or genetic models. Actually, scale-free (power-law) distributions of bacterial lag times have been recently reported in a specifically-devised experimental setup [423]. Similarly to our conclusions, this work also emphasizes that a broad distribution of individual-cell waking-up rates is needed to generate non-exponential decays of the overall lag-time distribution.

Similarly, another exciting possibility would be to develop computational models akin to the phenotypic one proposed here but implementing genetic circuitry; i.e. models where the phenotype is the (possibly stochastic) outcome of an underlying regulated genetic process and where the object of selection are not specific lag times but their whole distributions as genetically encoded.

Future developments and perspectives. In future research, we would like to further delve onto several aspects, both biological and theoretical, of the present work. As a first step, we leave for forthcoming work the analysis of the pertinent question of how similar systems respond to randomly fluctuating environments as opposed to periodically changing ones; do they develop heavier tails to cope with such unpredictable conditions in a sort of bet-hedging strategy? How do the statistical features of the environmental variability translate into the emerging lag-time distributions? [349, 390, 424, 347, 425].

From a more theoretical perspective, we leave for an impending work the application of the generalized-Crow-Kimura macroscopic equation accounting for demographic fluctuations, see Sec.3.6. Within this context, treating the variation-amplitude itself as an evolving trait is also a potentially fruitful route for further studies.

Finally, as a long-term project we plan to develop models and analytical approaches, similar to the ones explored here, but focusing on genetic evolution, employing explicit genotypic-phenotypic mappings, rather than just on phenotypic changes. In particular, by introducing this further layer of complexity it would be possible to generate more general types of single-cell lag-time distributions, not limited to exponential ones as the purely Markovian approach considered here. Let us recall that a more general stochastic non-Markovian framework —i.e., including memory effects (see e.g. [389, 390, 426])— is a challenging goal that promises to be very pertinent and relevant for many diverse problems in which the control of time is important.

4.6 Computational Methods

Numerical values of the parameters

In order to fix parameter values we employed the experimental values and measurements in [343] as closely as possible. The number of bacteria involved in the experiment reaches values of the order of $\sim 10^9$; however this number is prohibitively large for computer simulations and we fixed a maximum carrying capacity of $K = 10^5$, verifying that results do not depend strongly on such a choice (see finite-size effects section). Initially the number of cells in the growing state is fixed to be equal to the carrying capacity K; thus no cell is initially in the dormant state). The doubling time of both the ancestral and the evolved populations is $25 \pm 0.3\,\mathrm{min}$; thus on average every single bacteria attempts reproduction at a rate $b = 1/25\,\mathrm{min} = 2.4\ h^{-1}$. The death rate for (natural) causes (i.e. other than antibiotics) is $d = 3.6 \cdot 10^{-5}\ h^{-1}$. The awakening rate is given by the inverse of the characteristic time $a = 1/\tau$ [10]. The initial condition (ancestral or wild population) was randomly sampled from a truncated Gaussian peaked at $\tau = 0$. Since the empirical ancestral distribution is narrow and close to the origin [343] (mean lag time $\langle\tau\rangle_0^{exp.} = 1 \pm 0.2\ h$) we fix the standard deviation of the truncated Gaussian distribution

to $\sigma = 1\ h\ 16$ min in such a way that $\langle\tau\rangle_0^{sim.} \sim 1\ h$. Neither the exit rate from dormancy s nor the amplitude of the mutations, α_A and α_M, can be experimentally measured, but we can fix them indirectly (the rest of the parameters are kept fixed with the values specified above). First, s can be chosen using the experimental information that for the ancestral population $MDK_{99} \sim 2.55h$. Hence, we leave the (simulated) ancestral population in the antibiotic phase until the 99% becomes extinct; averaging over different initial conditions, we found that $s = 0.12\ h^{-1}$ is a good approximation. To fix the constants α_A and α_M we performed simulations for diverse values of such parameters and looked for those that best reproduce the experimental tendency after 10 exposure cycles for the different exposure times under consideration (in particular, we performed a least-square deviation analysis to match the straight-line $\langle\tau\rangle = T_a$ when performing a linear interpolation for all T_a's). A systematic sweep of the values of the first two significant digits led us to $\alpha_A = 0.16h$ and $\alpha_M = 0.048$.

Variation Functions

We consider two different variation kernels for lag-time variations $\delta = \tau - \tilde{\tau}$: the additive one, $\beta_A(\delta;\tilde{\tau})$ and multiplicative one, $\beta_M(\delta;\tilde{\tau})$. Both of them are probability density functions of δ, normalized in $[-\tilde{\tau},\infty]$ and may depend on the initial phenotype $\tilde{\tau}$. In particular, the additive case reads: $\beta_A(\delta;\tilde{\tau}) = e^{-\frac{\delta^2}{2\alpha_A^2}}/Z_A(\tilde{\tau})$ with $Z_A(\tau) = \alpha_A\frac{\sqrt{\pi}}{\sqrt{2}}Erfc(-\frac{\tau}{\sqrt{2}\alpha_A})$ (where $Erfc$ stands for the complementary error function), while in the multiplicative case, we consider $\beta_M(\delta;\tilde{\tau}) = e^{-\frac{\delta^2}{2\alpha_M\tilde{\tau}}}/Z_M(\tilde{\tau})$ with $Z_M(\tau) = \alpha_M\tau\sqrt{\frac{\pi}{2}}Erfc(-\frac{1}{\sqrt{2}\alpha_M})$.

Measuring lag-time distributions

In order to determine lag-time distributions, we computed histograms in phenotypic space, as discretized in bins of size $\Delta\tau = 10^{-2}$ and averaged over many realizations of the process. In the asymptotic steady state, similar histograms were computed at different times along the antibiotic/fresh-medium cycle (e.g. right after antibiotic inoculation or after antibiotic removal). To obtain results for the transient state we determined the histogram after running for 10 cycles. On the other hand, to determine the steady state, we started measuring after 300 cycles (to make sure that a steady state has been reached) and then collect statistics up to cycle 1500, at intervals of 10 cycles to avoid correlations. We repeated the process for 30 realizations and calculated the histogram as well as the associated cumulants.

Numerical Integration of the macroscopic equation

The parameter set and initial condition for numerical integration of the mean-field equations are the same as specified above. Numerical integration was carried out using the finite differences method. In particular ∂_t was approximated using first order forward differences, ∂_τ^2 using second-order centered differences, and integrals were approximated as Riemann sums [427]. The numerical integration steps used in the figures are: $h_t = 10^{-6}$ and $h_\tau = 10^{-2}$. Note that, when we calculate the probability distributions during the simulation, we must use the same bin size to be able to correctly compare with the theoretical distributions later. We used absorbing boundary conditions $\phi(\tau = 0, t) = \phi(\tau = \tau_{max}, t) = 0$, where τ_{max} is the limit of the phenotype space considered for the numerical integration, in particular: $\tau_{max} = 60h$.

Part III

Irreversibility in complex systems

Chapter 5

Dissipative symmetry breaking

5.1 Stochastic thermodynamics: a geometrical perspective

Stochastic thermodynamics is branch of non-equilibrium statistical physics that studies the thermodynamcis properties of *mesoscopic* systems, like colloidal particles, electronic circuits, macromoleculas and chemical reactions at low densities [89, 428, 178]. In these systems are characterized by an energy scale comparable with the thermal energy $k_B T$, where k_B is the Boltzmann constant and T the temperature. The system is in contact with an *equilibrium* thermal bath, and their interaction must be weak and negligible. The interaction between the bath and the system generates stochasticity in the time evolution, as in Brownian motion. The non-equilibrium conditions are given by the external force driving the system, that provides it energy in the form of stochastic work. Also, the system dissipate energy into the bath in the form of stochastic heat. Given the stochastic nature of the system, stochastic thermodynamics is able to define such thermodynamic observables as fluctuating quantities and along a single trajectory. Notably, the stochastic heat and work are related to the system irreversibility and entropy production.
Stochastic thermodynamics is the results of the hybridization between different research line in statistical physics like:the study of Brownian motion founded by Einstein [429], the discovery of Onsager-reciprocal relations[179, 430] and Fluctuations Theorems by Gallavotti, Cohen and other authors[431, 432, 433], the Brussels school of thermodynamics and its founder Prigogine[175, 434], early development in NESS by Lebowitz and Bergman[435], the thermodynamics of Langevin eqs. by Sekimoto [428] and the network thermodynamics of Schnakenberg[436].
A general aspect of non-equilibirum statistical physics is that a non-equilibrium stationary state (NESS) needs to be characterized by the stationary currents flowing in the system, in addition to the states probability [94]. It is known that these currents generally produce a rotational thermodynamics force in the system, in addition to a potential force typical of equilibrium. Even if many analogies with electromagnetism and gauge theories have been proposed [437, 438], a general understanding of the geometrical and topological order produced by currents is lacking (see this reference for a parallel tentative [130]).
In the next sections we will present an original perspective on stochastic thermodynamics that shed light onto the geometrical and topological properties of stationary currents, making connection between thermodynamics and symmetry breaking.

Constant diffusion

Geometrical and topological properties of current velocities

Consider an overdamped driven-diffusive system whose evolution is described by the following equation (in the covariant formulation):

$$\dot{x}^\mu = F^\mu + \xi^\mu(t) \,, \tag{5.1}$$

where $\vec{x} = (x^\mu)$, $\mu = 1, \ldots, d$, is the state of the system, e.g., position in real space, $\vec{F}$ a general force and ξ^μ a Gaussian white noise with variance $\langle \xi^\mu(t)\xi^\nu(t')\rangle = 2\sigma D^{\mu\nu}\delta(t-t')$, modelling its interaction with a thermal bath. The dissipation-fluctuation relation is in general violated, hence the system will eventually reach a non-equilibrium stationary state (NESS) [439]. Here, σ controls the amplitude of the noise and, in a thermodynamic context, it can be interpreted as the temperature. For simplicity, we consider the diffusion matrix to be state-independent. Later on, we will generalize our framework to multiplicative noise. Let us remind to the reader that in the covariant formalism [440] any vector, or matrix, element is represented with a up greek index, i.e x^μ, $D^{\mu\nu}$ or $\partial^\mu = \frac{\partial}{\partial x_\mu}$. Analogously, the element of the inverse of a matrix is represented with down indices: $(\hat{D}^{-1})^{\mu\nu} = D_{\mu\nu}$. Any index repetition , necessarily one up and one down, means a sum over such index; for example the scalar product between two vectors $\vec{\alpha}$ and $\vec{\beta}$ reads:

$$\vec{\alpha}\cdot\vec{\beta} = \alpha^\mu\beta_\mu = \sum_{\mu=1}^{d}\alpha^\mu\beta^\mu. \tag{5.2}$$

Note that in this case the matrix for the scalar product is the trivial flat one. The Fokker-Planck equation (FPE) associated with Eq. (5.94) is:

$$\partial_t P = -\partial_\mu\left(F^\mu P - \sigma D^{\mu\nu}\partial_\nu P\right) = -\partial_\mu J^\mu \,. \tag{5.3}$$

If respected, the steady state is an equilibrium one; alternatively the system is away from equilibrium. The stationary solution of the system admits a solenoidal stationary current, i.e.,

$$\partial_\mu J^{*\mu} = 0 \tag{5.4}$$

Following the pioneering works of Graham and Tél [441, 442, 443] and recent developments in Macroscopic Fluctuation Theory [84, 444], the steady-state solution in the weak noise limit can be written as:

$$P^*(\vec{x}) = \frac{e^{-V(\vec{x})/\sigma}}{Z} \,, \tag{5.5}$$

where $V(\vec{x})$ is a non-equilibrium potential, or quasi-potential. Despite its analytical intractability, there are numerous numerical and empirical methods to estimate $V(\vec{x})$, where the potential is exactly defined as the minimum of the action along a trajectory to create a fluctuation $\vec{x}$ [84]:

$$V(\vec{x}) \quad = \quad V(\vec{x}^*) + \min\left(\frac{1}{2}\int_{-\infty}^{0} dt\left[\frac{1}{2}(\dot{\vec{x}}-\vec{F})\hat{D}^{-1}(\dot{\vec{x}}-\vec{F}) - \vec{\nabla}\cdot\vec{F}\right]\right)$$

where $\vec{x}^*$ is the minimum of the potential, corresponding to the deterministic fixed point. Using the introduced relations, it is possible to decompose the thermodynamic force $\vec{F}$

into a conservative contribution given by the gradient of the quasi-potential and a dissipative one proportional to the stationary current:

$$J^{\mu *} = F^{\mu} P^{*} - \sigma D^{\mu\nu} \partial_{\nu} P^{*} \tag{5.6}$$

$$F^{\mu} = J^{\mu *} P^{-1} - D^{\mu\nu} \partial_{\nu} V = v^{\mu *} - D^{\mu\nu} \partial_{\nu} V, \tag{5.7}$$

where we have defined the current velocity as:

$$v^{\mu} = \frac{J^{\mu}}{P}. \tag{5.8}$$

Given that the quasi-potential contributes with a gradient, and the stationary current is a solenoidal field [445], one would be tempted to assume that also the velocity is solenoidal, obtaining in this way a complete geometrical decomposition of the force. Such an assumption is not always true, as it can be seen by recasting the stationary condition, Eq. (5.4) in terms of the velocity:

$$\nabla \cdot v^{*} = \frac{1}{\sigma} v^{*} \cdot \nabla V \tag{5.9}$$

Hence, the divergence of $\vec{v}^{*}$, i.e. its propensity to be a sink, is just the scalar product between the velocity and the gradient of the potential. To uncover the geometric properties of the system, we employ the quasi-potential ansatz, Eq. (5.5), to solve the Fokker-Planck Equation in perturbative orders of σ [442]:

$$\partial_t P^{*} = \sigma^{-1} P^{*} \left(F^{\mu} + D^{\mu\nu} \partial_{\nu} V \right) \partial_{\mu} V - \left(\partial_{\mu} F^{\mu} + D^{\mu\nu} \partial_{\mu} \partial_{\nu} V \right) P^{*} = 0. \tag{5.10}$$

The first contribution is of order $1/\sigma$:

$$(F^{\mu} + D^{\mu\nu} \partial_{\nu} V) \partial_{\mu} V = 0 \tag{5.11}$$

Graham interprets this equation as a Hamilton-Jacobi equation with position $q^{\mu} = x^{\mu}$ and moment $p^{\mu} = \partial^{\mu} V$, and can be ideally solved with the method of characteristics. Furthermore, inspired by the work of Jona-Lasinio and collaborators [446], we interpret it as a perpendicularity condition for the stationary current velocity:

$$\vec{v}^{*} \cdot \vec{\nabla} V = 0. \tag{5.12}$$

This relation tells us that the stationary velocity is always perpendicular to the gradient of the potential, i.e. it is tangent to the potential height lines. As a consequence, as already observed in [445], the fixed point of the velocity must coincide with the potential minimum.

Then, the second (and last) contribution is of order σ^{0}:

$$\partial_{\mu} F^{\mu} + D^{\mu\nu} \partial_{\mu} \partial_{\nu} V = 0, \tag{5.13}$$

and naturally imposes that the stationary velocity is also a divergence free field, i.e. a solenoidal field:

$$\vec{\nabla} \cdot \vec{v}^{*} = 0, \tag{5.14}$$

such that in three dimension it can be written, like the current, as the the curl of a vector field $\vec{B}$:

$$\vec{v}^* = \vec{\nabla} \times \vec{B}. \tag{5.15}$$

In particular, note that the vorticity of the stationary velocity can be easily calculated:

$$\vec{w}^* = \vec{\nabla} \times \vec{v}^* = \vec{\nabla} \times \vec{F} + \vec{\nabla} \times (D\vec{\nabla}V), \tag{5.16}$$

where the second contribution is non-zero for anisotropic diffusion. In the case of a diagonal diffusion matrix in two dimension, one obtains:

$$\vec{w}^* \quad = \vec{\nabla} \times \vec{F} + (D_{xx} - D_{yy})\partial_x \partial_y V, \tag{5.17}$$

indicating that the vorticity depends on the curl of the force and on the asymmetry of the diffusion coefficients in the two directions. Finally, decomposing the velocity at any time in its stationary and relaxation part, we obtain:

$$v^\mu(t) = v^{\mu *} + \tilde{v}^\mu \tag{5.18}$$

$$\tilde{v}^\mu = -\sigma D^{\mu\nu}\partial_\nu \log \frac{P(x,t)}{P^*} = -\sigma D^{\mu\nu}\partial_\nu \phi(x,t) \tag{5.19}$$

where we defined:

$$\phi(x,t) = \log \frac{P}{P^*} = \log P(x,t) + \frac{V(x)}{\sigma}. \tag{5.20}$$

Hence, the relaxation part $\tilde{v}$ is geometrically a gradient and goes to zero at stationarity. Finally, let us note that the general thermodynamics force can be decomposed in a scalar part generated by the quasi-potential, and a curl one, emerging from the stationary non-equilibrium current:

$$\vec{F} = \vec{v}^* + \sigma \hat{D}\vec{\nabla} \log P^* = \vec{\nabla} \times \vec{B} - \hat{D}\vec{\nabla}V \ . \tag{5.21}$$

The result presented here clarifies a long-lasting comparison between the geometrical properties of non-equilibrium thermodynamics and electromagnetism. Indeed, equilibrium thermodynamics present only a gradient field, like the electric one, while non-equilibrium conditions add a solenoidal field, analogous to the the magnetic one. References to this analogy are scattered across the literature: for example see the introduction to non-equilibrium statistical physics by T.Chou, K.Mallick and R. Zia [94], the works on non-equilibirum landscape by J. Wang and collaborators [437] and various works in MFT [447].

Thermodynamics

In this section, we show that various thermodynamic quantities are naturally expressed in terms of the velocities and have an interesting geometrical meaning. First of all, consider the system entropy, or Shannon entropy:

$$S(t) = -\int dx P(x,t) \log P(x,t). \tag{5.22}$$

In stochastic thermodynamics the Shannon entropy has physical meaning, beyond the probabilistic/informational one. Its thermodynamics value comes from general "ergodic principle" that quietly lies at the base of stochastic thermodynamics. Formally, the probability P represents at the same time the probability of the system state, say the position

of a Brownian particle, and the density of a gas of Brownian non-interacting particles in the thermodynamic limit. This come from the fact that an average over time of the position of a single brownian particle is equivalent to the mean position of the particles in the imaginary brownian gas. Hence, the current in the FP equation, can be interpreted as a current obtained by the average velocity of the single particle in a particular point of space. This point is quite controversial and not sufficiently discussed in the literature, but see [177] and chapter 9 in[89] for a critic of the Shannon entropy and [448] for a possible solution in the "informational" interpretation of thermodynamics.
The change in the system entropy can be always decomposed into a total non-negative part and the environmental dissipation:

$$\begin{aligned}\dot{S}(t) &= -\int dx \partial_t P(t) \log P(t) = \int dx \vec{\nabla} \cdot \vec{J} \log P = -\int dx \vec{J} \cdot \vec{\nabla} \log P \\ &= \int dx \frac{J^\mu D_{\mu\nu} J^\nu}{P} - \int dx J^\mu D_{\mu\nu} F^\nu = \dot{S}_{prod}(t) - \dot{S}_{flux}(t),\end{aligned} \tag{5.23}$$

where the total entropy production is

$$\dot{S}_{prod}(t) = \int dx \frac{J^\mu D_{\mu\nu} J^\nu}{\sigma P} = \langle \vec{v} \circ \vec{v} \rangle \geq 0, \tag{5.24}$$

and the entropy flux to the environment is

$$\dot{S}_{flux}(t) = \langle \vec{v} \circ \vec{F} \rangle, \tag{5.25}$$

with the diffusive scalar product indicated by the asterisk in which the diffusion matrix acts as a metric [440], i.e.:

$$\vec{\alpha} \circ \vec{\beta} = \alpha^\mu D_{\mu\nu} \beta^\nu; \tag{5.26}$$

and

$$\vec{\alpha} \circ \vec{\alpha} = |\vec{\alpha}|_D^2. \tag{5.27}$$

The average is performed over the probability:

$$\langle O \rangle = \int O(x) dx P(x,t). \tag{5.28}$$

While the total entropy production represent the dissipation of the system and the environment together, the environmental one quantifies the dissipation into the bath. At the steady state the change in system entropy is zero and the only dissipation is the environmental one. Alternatively, the total entropy production can can be decomposed in two non-negative parts (the so-called adiabatic decomposition [434, 449, 450]):

$$\dot{S}_{prod} = \langle \vec{v} \circ \vec{v} \rangle = \langle (\vec{\tilde{v}} + \vec{v}_s t) \circ (\vec{\tilde{v}} + \vec{v}_s t) \rangle = \langle \vec{v}^* \circ \vec{v}^* \rangle + \langle \vec{\tilde{v}} \circ \vec{\tilde{v}} \rangle + \langle \vec{v}^* \circ \vec{\tilde{v}} \rangle + \langle \vec{\tilde{v}} \circ \vec{v}^* \rangle, \tag{5.29}$$

where the last line is zero because

$$\begin{aligned}\langle \vec{v}^* \circ \vec{\tilde{v}} \rangle &= -\frac{\sigma}{2}\int dx P(x,t) v^{\mu *} D_{\mu\nu} D^{\nu\lambda} \partial_\lambda \phi(x,t) = -\frac{\sigma}{2}\int dx P(x,t) v^{\mu *} \partial_\mu \phi(x,t) \\ &= -\frac{\sigma}{2}\int dx \frac{P(t) J^{\mu *}}{P^*}\left(\frac{\partial_\mu P}{P} + \frac{\partial_\mu V}{\sigma}\right) \\ &= \frac{\sigma}{2}\langle \vec{\nabla}\cdot\vec{v}^* \rangle - \frac{1}{2}\langle \vec{v}_{set}\cdot\vec{\nabla} V\rangle = 0, \end{aligned} \tag{5.30}$$

for the perpedicularity conditions (Eq.s5.12 and 5.14). Hence, we obtain the following expression:

$$\begin{aligned}\dot{S}_{prod} &= \langle \vec{v}\circ\vec{v}\rangle = \langle |\vec{v}|^2_D\rangle = \langle \vec{v}^*\circ\vec{v}^*\rangle + \langle \vec{\tilde{v}}\circ\vec{\tilde{v}}\rangle \\ &= \langle |\vec{v}^*|^2_D\rangle + \langle |\vec{\tilde{v}}|^2_D\rangle = \langle |\vec{\nabla}\times\vec{B}|^2_D\rangle + \sigma^2\langle |\vec{\nabla}\phi|^2_D\rangle. \end{aligned} \tag{5.31}$$

The two contributions corresponds to housekeeping (or adiabatic) and excess (or non-adiabatic) components [450]

$$\dot{S}_{prod}(t) = \dot{S}_{hk}(t) + \dot{S}_{ex}(t) \tag{5.32}$$

$$\dot{S}_{hk}(t) = \langle \vec{v}\circ\vec{v}^*\rangle = \langle \vec{v}^*\circ\vec{v}^*\rangle = \dot{S}_a(t) = \langle |\vec{\nabla}\times\vec{B}|^2_D\rangle \geq 0 \tag{5.33}$$

$$\dot{S}_{na}(t) = \langle \vec{\tilde{v}}\circ\vec{\tilde{v}}\rangle = \sigma^2\langle |\vec{\nabla}\phi(x,t)|^2_D\rangle = \dot{S}_{ex}(t) \geq 0 \tag{5.34}$$

where in the first line we have used the perpendicularity condition, Eq.5.30. Notice that the housekeeping part is zero only at equilibrium and corresponds to the stationary entropy production. Conversely, the excess part is zero at any NESS and corresponds to the definition given by Prigogine (see [434, 449]). Hence, we have derived that the housekeeping entropy is just given by the non-conservative or curl part of the force, while the excess one by the time-dependent, scalar contribution.

Finally, let us note that the time derivative of the average potential gives the balance between the house-keeping entropy production and the entropy flux to the environment, i.e.,

$$\frac{d_t\langle V\rangle}{\sigma} = \langle \vec{v}(t)\cdot\vec{\nabla} V\rangle = \langle \vec{v}\circ\vec{v}^*\rangle - \langle \vec{v}\circ\vec{F}\rangle = \dot{S}_{hk} - \dot{S}_{flux}, \tag{5.35}$$

and hence must be zero at stationarity.

Trajectory thermodynamics

In this section we show that in stochastic thermodynamics is possible to use thermodynamic quantities along single trajectories. Consider again the Langevin description of the system:

$$\dot{x}^\mu = F^\mu + \xi^\mu(t), \tag{5.36}$$

$$\langle \xi^\mu(t)\rangle = 0, \tag{5.37}$$

$$\langle \xi^\mu(t)\xi^\nu(t')\rangle = 2\sigma D^{\mu\nu}\delta(t-t'). \tag{5.38}$$

where ξ^μ is a Gaussian white noise, F^μ the deterministic force and $D^{\mu\nu}$ is the diffusion matrix (the average is done over the realizations). We define the following generic current [451, 452]:

$$\dot{R} = \vec{c}_R(\vec{x})\circ\dot{\vec{x}}, \quad c^\mu_R(x) = \frac{\partial R}{\partial x^\nu} D^{\nu\mu}. \tag{5.39}$$

where $\vec{c} = \vec{c}_R(\vec{x})$ is the vector field that determines the current. By substituting the Langevin dynamics in $\dot{\vec{x}}$using the Ito formula, we obtain:

$$\dot{R} = \vec{c} \circ \dot{\vec{x}} = \frac{\partial R}{\partial x^\nu}\dot{x^\nu} + \frac{1}{2}\frac{\partial^2 R}{\partial x^\mu \partial x^\nu} D^{\mu\nu} = \vec{c} \circ \vec{F} + \sqrt{2}\vec{c} \circ \vec{\xi} + \vec{\nabla} \circ (\hat{D}\vec{c}) \tag{5.40}$$

$$= \vec{c} \circ \vec{v} + \frac{\vec{\nabla} \circ (\hat{D}\vec{c}P)}{P} + \sqrt{2}\vec{c} \circ \vec{\xi}. \tag{5.41}$$

Thanks to the explicit equation for the dynamics of the current we can take the average and obtain:

$$\langle \dot{R} \rangle = \int d\vec{x} P(\vec{x}, t) \int d\vec{\xi} P(\vec{\xi}, t) \dot{R} = \langle \vec{c} \circ \vec{v} \rangle. \tag{5.42}$$

We can now consider some concrete thermodynamics observables along a trajectory, like the total entropy production and the environmental dissipation:

$$\dot{s}_{prod} = \dot{s} + \dot{s}_{flux} = -\partial_t \log P|_\Gamma + \dot{\vec{x}} \circ \vec{v}|_\Gamma \tag{5.43}$$

$$\dot{s}_{flux} = \dot{\vec{x}} \circ \vec{F} = \dot{\vec{x}} \circ \vec{v}^* - \dot{\vec{x}} \circ \hat{D}\vec{\nabla} V = \dot{s}_{hk} + \dot{s}_{ex}, \tag{5.44}$$

$$\dot{s}_{\mathrm{hk}} = \dot{\vec{x}} \circ \vec{v}^* \tag{5.45}$$

$$\dot{s}_{\mathrm{ex}} = -\dot{\vec{x}} \circ D\vec{\nabla} V \tag{5.46}$$

where $\dot{s}_{hk}$ the environmental entropy related to housekeeping and $\dot{s}_{ex}$ the excess one related with the potential part. By taking the averages we obtain that:

$$\dot{S}_{\mathrm{m}} = \langle \dot{s}_{\mathrm{m}} \rangle = \langle \dot{\vec{x}} \circ \vec{F} \rangle = \langle v \circ \vec{F} \rangle \tag{5.47}$$

$$\dot{S}_{\mathrm{hk}} = \langle \dot{s}_{\mathrm{hk}} \rangle = \langle \dot{\vec{x}} \circ \vec{v}^* \rangle = \langle v \circ \vec{v}^* \rangle \tag{5.48}$$

$$\dot{S}_{\mathrm{ex}} = \langle \dot{s}_{\mathrm{ex}} \rangle = -\langle \dot{\vec{x}} \circ D\vec{\nabla} V \rangle = -\langle v \cdot \vec{\nabla} V \rangle \tag{5.49}$$

If the Einstein relation $\sigma = k_B T$ is valid, the entropy production in the environment can be related to the heat flow in the thermal bath:

$$\dot{s}_{flux} = \frac{1}{k_B T}\dot{\vec{x}} \circ \vec{F} = \frac{\dot{q}}{k_B T}, \tag{5.50}$$

where we have rescaled the diffusive scalar product by the temperature. Following Hatano and Sasa [453] and the seminal work of Sekimoto [428], one can decompose the heat flow into the excess and housekeeping part:

$$\dot{q} = \dot{q}_{ex} + \dot{q}_{hk} = \dot{q}^\Delta + \dot{q}^* \tag{5.51}$$

$$\dot{q}_{ex} = \dot{q}^\Delta = -\dot{\vec{x}} \circ \hat{D}\vec{\nabla} V \tag{5.52}$$

$$\dot{q}_{hk} = \dot{q}^* = \dot{\vec{x}} \circ \vec{v}^* = \dot{\vec{x}} \circ \vec{\nabla} \times \vec{B}. \tag{5.53}$$

By taking the averages, we obtain:

$$\dot{Q}_{hk} = \dot{Q}^* = \langle \dot{q}_{hk} \rangle = \langle \vec{v} \circ \vec{\nabla} \times \vec{B} \rangle \tag{5.54}$$

$$\dot{S}_{hk} = \frac{\dot{Q}_{hk}}{k_B T} \tag{5.55}$$

$$\dot{Q}_{ex} = \dot{Q}^\Delta = -\langle v \cdot \nabla V \rangle \tag{5.56}$$

$$\dot{S}_{ex} = \frac{\dot{Q}_{ex}}{k_B T}. \tag{5.57}$$

In agreement to what we have found in the precedent section.

Fluctuation relations

Equivalently to the Langevin equation, the dynamics can be represent with the path-integral representation and the Onsager-Machlup action [179, 430, 432]. Consider a trajectory Γ in a time interval $[0, t]$, its probability in the Stratonovich discretization is [454]:

$$P(\Gamma) \sim e^{-\mathcal{S}[\Gamma]}, \qquad \mathcal{S}[\Gamma] = \frac{1}{2}\int_0^t ds\left[\frac{1}{2}(\dot{\vec{x}} - \vec{F})\hat{D}^{-1}(\dot{\vec{x}} - \vec{F}) - \vec{\nabla}\cdot\vec{F}\right], \tag{5.58}$$

where we have not reported the Jacobian of $P(\Gamma)$ and $S[\Gamma]$ is Onsager-Machlup action. Now, we consider the time-reversed trajectory $\tilde{\Gamma}$, obtain by inverting the time sign $s \to t - s$. Given that this transformation reverse the sign of the velocity, we can divide the Onsager action in a time-symmetric and time-asymetric part:

$$\mathcal{S}[\Gamma] = \mathcal{S}_s[\Gamma] + \mathcal{S}_a[\Gamma], \tag{5.59}$$

$$\mathcal{S}_s[\Gamma] = \frac{1}{2}\int_0^t ds\left[\frac{1}{2}\dot{\vec{x}}\hat{D}^{-1}\dot{\vec{x}} + \frac{1}{2}\vec{F})\hat{D}^{-1}\vec{F} - \vec{\nabla}\cdot\vec{F}\right], \tag{5.60}$$

$$\mathcal{S}_a[\Gamma] = -\frac{1}{2}\int_0^t ds\dot{\vec{x}}\hat{D}^{-1}\vec{F}. \tag{5.61}$$

Hence, when one takes the ration of probability of the time directed and of reversal trajectory just the time-asymmetric part survives, and it corresponds to the change int entropy flux along the directed trajectory:

$$\log\frac{P(\Gamma)}{P(\tilde{\Gamma})} = -2\mathcal{S}_a[\Gamma] = \int_0^t ds\dot{\vec{x}}\hat{D}^{-1}\vec{F} = \Delta s_{flux} = \int_0^t ds\dot{\vec{x}}\circ\vec{\nabla}\times\vec{B} - \Delta V, \tag{5.62}$$

where in the last equality we have used the geometric decomposition of the force. On the other hand, if one takes into consideration also the probability of the initial states of the trajectory, x_0 and x_t, one obtains the entropy production along the trajectory:

$$\log\left(\frac{P(x_0)P(\Gamma)}{P(x_t)P(\tilde{\Gamma})}\right) = \Delta s_m + \Delta s = \Delta s_{prod} \tag{5.63}$$

Hence, the entropy production is a measure of the trajectory irreversibility. From Eq.(5.63) it is easy to verify that

$$\langle e^{-\Delta s_{prod}}\rangle = 1, \tag{5.64}$$

also known as integral fluctuation relation [178], from which by using the Jensen inequality, it derives the second law of thermodynamics:

$$\langle \Delta s_{prod}\rangle \geq 0 \tag{5.65}$$

Note that, for Eq.(5.63), the entropy production as the following symmetry, called an involution:

$$\Delta s_{prod}[\Gamma] = -\Delta s_{prod}[\tilde{\Gamma}]. \tag{5.66}$$

At the stationary state the entropy production coincides with the entropy flux, and the following steady-state fluctuation relation emerges from Eq. (5.63):

$$\frac{P(-\Delta s_{prod})}{P(\Delta s_{prod})} = e^{-\Delta s_{prod}}, \tag{5.67}$$

where we have used that, for a general probability distribution of the entropy production it holds that

$$\langle \delta(\Delta s_{prod} - s) e^{-\Delta s_{prod}} \rangle_\Gamma = \langle \delta(\Delta s_{prod} - s) \rangle_{\tilde{\Gamma}} = \langle \delta(\Delta s_{prod} + s) \rangle_{\tilde{\Gamma}}, \tag{5.68}$$

together with Eq.(5.66), see Ref.[89] for more details. Eq. (5.67) is a relation between the stationary state probability of observing a value of the average entropy production rate and its opposite, and hence express the same statement of the original fluctuation theorem by Gallavotti and Cohen.

Multiplicative noise: Geometrical properties of velocities

Consider a general overdamped system where the diffusion coefficient depend on the variables, i.e. with multiplicative noise:

$$\begin{aligned} \dot{x}^\mu &= F^\mu(x) + G^{\mu\nu}(x)\xi_\nu(t) && (5.69)\\ \langle \xi^\mu(t) \rangle &= 0, && (5.70)\\ \langle \xi^\mu(t)\xi^\nu(t') \rangle &= 2C^{\mu\nu}\delta(t-t'). && (5.71) \end{aligned}$$

with $\hat{G}$ the matrix ruling the multiplicative fluctuations, that for simplicity we consider diagonal $\hat{G}(\vec{x}) = \mathrm{diag}(g_1(\vec{x}), g_2(\vec{x}), \ldots, g_N(\vec{x})$, and $\hat{C}$ the correlation matrix of the Gaussian noise ($\vec{x}$-independent). In the following, we employ an α-dependent discretization of the noise in order to encode all possible prescription. Indeed, $\alpha = 0$ corresponds to the Stratonovich case, while $\alpha = 1$ to the Ito prescription. The corresponding Fokker-Planck equation reads:

$$\begin{aligned} \partial_t P &= -\partial_\mu J^\mu && (5.72)\\ J^\mu &= F^\mu P - \alpha \partial_\nu D^{\mu\nu} - D^{\mu\nu}\partial_\nu P, && (5.73) \end{aligned}$$

where the diffusion matrix is constructed as follows:

$$\hat{D} = \hat{G}^T \hat{C} \hat{G}. \tag{5.74}$$

In this general case, Eq. (5.5) is not valid anymore and the stationary solution in the small noise limit is:

$$P^*(\vec{x}) = \eta(\sigma) z(\vec{x}) e^{-\frac{V}{\sigma}}. \tag{5.75}$$

The stationary velocity is defined now as:

$$\begin{aligned} v^{\mu *} &= F^\mu + D^{\mu\nu}\partial_\nu V - \alpha\sigma\partial_\nu D^{\mu\nu} - \sigma D^{\mu\nu}\partial_\nu \log z \\ &= (v^{(0)})^{\mu *} + \sigma (v^{(1)})^{\mu *}. && (5.76)\\ (v^{(0)})^{\mu *} &= F^\mu + D^{\mu\nu}\partial_\nu V && (5.77)\\ (v^{(1)})^{\mu *} &= -\alpha\partial_\nu D^{\mu\nu} - D^{\mu\nu}\partial_\nu \log z. && (5.78) \end{aligned}$$

As a first comment, note that the non-equilibrium potential V should have its minimum in the fixed point of the force (because in the deterministic limit the velocity must be zero), so that $\vec{v}^{(0)*} \approx 0$ in the fixed point, and the v_1 would be dominant. On the other hand, if the Ito prescription is used, the maximum of the full probability distribution will not coincide with the deterministic one, and will be shifted for the presence of the Ito term. Furthermore, note that one can identify the terms that generate the vorticity in the current:

$$\begin{aligned} w &= \vec{\nabla} \times \vec{v}^* = w_0 + w_1 = \vec{\nabla} \times \vec{v}_0 + \sigma \vec{\nabla} \times \vec{v}_1 && (5.79)\\ w_0 &= \vec{\nabla} \times \vec{F} + (\vec{\nabla} \times D)\vec{\nabla} V && (5.80)\\ w_1 &= -\vec{\nabla} \times \vec{\nabla} D - \vec{\nabla} \times D \vec{\nabla} \log z && (5.81) \end{aligned}$$

It means that in the special case where the diffusion matrix is diagonal (but x dependent) the first order velocity has no vorticity:

$$\begin{aligned} w_0 &= \vec{\nabla} \times \vec{F} && (5.82)\\ w_1 &= 0. && (5.83) \end{aligned}$$

Now, We solve order by order in σ the stationary Fokker-Planck equation. The dominant σ^{-1} term does not change with respect to the additive case:

$$(F^\mu + D^{\mu\nu}\partial_\nu V)\,\partial_\mu V = 0, \quad \vec{v}^{(0)*} \cdot \vec{\nabla} V = 0. \tag{5.84}$$

This condition, as in the constant case, tells us that the first order velocity is tangent to the height lines of the non-equilibrium potential. Assuming that the potential has just one minimum, by using the Gauss' law, we can show this geometrical property of $\vec{v}^{(0)}$ coincides the fact that it also has zero divergence. Consider the volume Ω enclosed by a single level surface of the potential V. The vector pointing outward in each point of the boundary is the gradient of the potential. From the Gauss' law:

$$\int_\Omega \vec{\nabla} \cdot \vec{v}^{(0)*} d\Omega = -\int_{\Sigma=\partial\Omega} d\Sigma\; \vec{v}^{(0)*} \cdot \frac{\vec{\nabla} V}{|\vec{\nabla} V|} = 0. \tag{5.85}$$

Given that for any point in the phase space there exists a contour line passing through it, the divergence is always zero, i.e. $\vec{\nabla} \cdot \vec{v}^{(0)*} = 0$.

The σ^0 order has two new terms with respect to the additive scenario:

$$\begin{aligned} & (F^\mu + D^{\mu\nu}\partial_\nu V)\partial_\mu \log z \\ + \;& \partial_\mu \left(F^\mu + \frac{D^{\mu\nu}}{2}\partial_\nu V \right) + \left(D_{\mu\nu}\partial^\nu \log z + \alpha \partial^\nu D_{\mu\nu} \right) \partial^\mu V = 0 && (5.86)\\ = \;& \vec{\nabla} \cdot \vec{v}^{(0)*} + \vec{v}^{(0)*} \cdot \vec{\nabla} \log z - \vec{v}^{(1)*} \cdot \vec{\nabla} V = 0 && (5.87) \end{aligned}$$

Finally, the order σ reads:

$$\begin{aligned} & \partial_\mu \left(\alpha \partial_\nu D^{\mu\nu} + D^{\mu\nu}\partial_\nu \log z \right) + \left(D^{\mu\nu}\partial_\nu \log z + \alpha \partial_\nu D^{\mu\nu} \right) \partial_\nu \log z \\ = \;& \vec{\nabla} \cdot \vec{v}^{(1)*} + \vec{v}^{(1)*} \cdot \vec{\nabla} \log z = 0. && (5.88) \end{aligned}$$

To go further and investigate the geometric properties of the first order velocity, let's recall that by a simple change of variable it is possible to transform the multiplicative noise of the Langevin equation into an additive one. If $\hat{\Lambda}$ is the Jacobian of the transformation, i.e., $\vec{x}' = \hat{\Lambda}\vec{x}$, this mapping is done by requiring that $\hat{\Lambda} = \hat{G}^{-1}$. In these new coordinates,

all the results obtained before straightforwardly hold:

$$P' = \frac{e^{-V(x')/\sigma}}{Z}. \tag{5.89}$$

Hence, by writing down the transformation of P, we have:

$$\begin{aligned} P &= P'\left|\hat{\Lambda}\right| \\ P &= \frac{1}{Z}|\hat{\Lambda}|e^{-V(x'(x))/\sigma} \end{aligned} \tag{5.90}$$

where $|\cdot|$ is the determinant. Hence, we identify:

$$z = |\hat{\Lambda}| = |\hat{G}|^{-1}. \tag{5.91}$$

As a consequence, the first order velocity is proportional to the gradient of the transformation plus a α-dependent term coming from the change of coordinates:

$$\begin{aligned} (v^{(1)*})^{\mu} &= -\alpha\partial_\nu D^{\mu\nu} - D^{\mu\nu}\partial_\nu \log|\Lambda| && (5.92)\\ &= -\alpha\partial_\nu D^{\mu\nu} + D^{\mu\nu}\partial_\nu \log|G| \\ &= \alpha\partial_\nu D^{\mu\nu} + D^{\mu\nu}\partial_\nu \psi && (5.93) \end{aligned}$$

5.2 Symmetry and irrversibility: an intrduction

How out-of-equilibrium systems self-organize to reach a stable non-equilibrium stationary state (NESS) is an unsolved enigma. Classical general results dictating how NESS should be approached, such as Onsager's minimum dissipation principle [179, 430] and the Glansdorff-Prigogine criterium [455], have attracted recent attention in various contexts [447, 449, 456]. However, an agreement on a unified and solid physical picture is still lacking, leaving open the problem of finding first principles valid away from equilibrium. Nevertheless, the last decades witnessed the discovery of fluctuation theorems [431, 457, 432, 458, 433], universal relations holding arbitrarily far from equilibrium. The core message of these results is that non-equilibrium systems break time-reversal symmetry, preserving a weaker version of it that quantifies dissipation.

Symmetry breaking mechanisms are also considered at the heart of the emergence of self-organization away from equilibrium, following the inspiration of Prigogine's idea of "dissipative structures" [459, 460, 461]. Although a strict connection was, and is, still debated, recent developments showed that macroscopic systems can spontaneously break some symmetries due to collective effects [462, 463] or rare large fluctuations [464, 465]. From a broader and more fundamental perspective, how dissipation induces a symmetry breaking in the *trajectory* space is a fascinating topic still largely unexplored.

Here, we build a bridge between symmetry breaking and dissipation in mesoscopic non-equilibrium systems, i.e., ruled by Stochastic Thermodynamics. This framework encompasses several well-known experimentally realizable examples, from molecular machines [178] to chemical reaction networks [466]. In recent years, an increasing wealth of studies is investigating the non-equilibrium features of these systems, such as entropy production [450, 451, 181, 467], current fluctuations [468, 469], and dissipation-driven asymmetries [470, 471, 472]. Starting from a Langevin dynamics with constant diffusion, we write down a general non-equilibrium functional whose minimization gives the correct stationary state. We show that this is composed of two terms: the first one is the total entropy production and quantifies dissipation, while the second one is minus the

housekeeping dissipation. Most importantly, this second contribution is related to broken chiral symmetries in the trajectory space and coincides with the dissipation along cycles (in the simplest case). As a consequence, any stationary state can be understood as the minimum dissipative state compatible with the existence of fluxes along preferential directions, which are absent at equilibrium. Remarkably, this statement corresponds to the minimization of the excess entropy production, which appears to be a more fundamental principle than the standard second law for non-equilibrium systems [450, 453, 428, 456, 473, 474, 475]. Finally, we unravel the connection between dissipation and chiral symmetry breaking in systems with multiplicative noise. This allows us to derive the most general form of the non-equilibrium functional and generalize the principle of excess entropy production minimization.

5.3 Lyapunov functional

Consider an overdamped driven-diffusive system whose evolution is described by the following equation:

$$\dot{\vec{x}} = \vec{F} + \vec{\xi}(t) \,, \tag{5.94}$$

where $\vec{x}$ is the state of the system in dimension d, e.g., position in real space, $\vec{F}$ a general force and $\vec{\xi}$ a Gaussian white noise with variance $\langle \xi_i(t)\xi_j(t')\rangle = 2\sigma D_{ij}\delta(t-t')$, with $i,j = 1,..,d$. The dissipation-fluctuation relation is in general violated, hence the system will eventually reach a non-equilibrium stationary state (NESS) [439]. Here, σ controls the amplitude of the noise and, in a thermodynamic context, it can be interpreted as the temperature. First, we consider the diffusion matrix $\hat{D}$ (with elements D_{ij}) to be state-independent. The Fokker-Planck equation (FPE) associated with Eq. (5.94) is:

$$\partial_t P = -\vec{\nabla}\cdot\left(\vec{F}P - \sigma\hat{D}\vec{\nabla}P\right) = -\vec{\nabla}\cdot\vec{J} \,. \tag{5.95}$$

The stationary solution of the system admits a solenoidal stationary current, i.e., $\vec{\nabla}\cdot\vec{J}^* = 0$ (see [445] for a detailed discussion). Following the pioneering works of Graham and Tél [441, 442, 443] and recent developments in Macroscopic Fluctuation Theory [84, 444], the steady-state solution in the weak noise limit reads:

$$P^*(\vec{x}) = \frac{e^{-V(\vec{x})/\sigma}}{Z} \,, \tag{5.96}$$

where $V(\vec{x})$ is a quasi-potential that can be rigorously defined and estimated via path-integral methods defined as the minimum of the action along a trajectory that generates a fluctuation $\vec{x}$[84]:

$$V(\vec{x}) \quad = \quad V(\vec{x}^*) + \min\left(\frac{1}{2}\int_{-\infty}^{0} dt\left[\frac{1}{2}(\dot{\vec{x}}-\vec{F})\hat{D}^{-1}(\dot{\vec{x}}-\vec{F}) - \vec{\nabla}\cdot\vec{F}\right]\right)$$

Despite its analytical intractability, there are numerous numerical and empirical methods to estimate $V(\vec{x})$. It is worth noting that, in the presence of additive noise, the deterministic fixed point of Eq. (5.94) coincides with the minimum of this potential. Noticeably, Eq. (5.96) is exact for any linearized stochastic system. In the following, we will consider the current velocity as main thermodynamic object:

$$v^\mu = F^\mu - \sigma D^{\mu\nu}\partial_\nu \log P \tag{5.97}$$

and use their geometrical properties derived in sec.5.1 to understand the properties of the NESS. In particular, note that the thermodynamic force admits a natural decomposition

in terms of stationary current and quasi-potential:

$$F^{\mu} = v^{\mu *} + D^{\mu\nu}\partial_{\nu}V \tag{5.98}$$

Hence, by understanding the geometrical properties of $\vec{v}^*$ it will be possible to decompose the force into a dissipative and conservative part. As shown in sec.5.1 by plugging Eq. 5.96 into the FPE, expanding up to the zeroth order in σ, and equating terms of the same order, we get two consistency equations:

$$\vec{v}^* \cdot \vec{\nabla} V = 0 \qquad \vec{\nabla} \cdot \vec{v}^* = 0 \,, \tag{5.99}$$

The first relation states that the $\vec{v}^*$ must be tangent to the potential contour lines, while the second one dictates that it also has to be a solenoidal field. An informative way to express these geometrical properties of $\vec{v}^*$ in the pedagogical case of a $3D$ system is $\vec{v}^* = \vec{\nabla} \times \vec{B}$, i.e., the curl of a field. Most importantly, these features lead to the following decomposition of the force in a solenoidal (*dissipative*) and gradient (*conservative*) part [476]:

$$\vec{F} = \vec{\nabla} \times \vec{B} - \hat{D}\vec{\nabla} V \,. \tag{5.100}$$

We will focus on $3D$ systems in this manuscript only for clarity of notation, but the findings of this Letter hold true in any dimension. In fact, later on we will also present a $2D$ Brownian gyrator as an example. For stochastic systems, the Kullback-Leibler divergence between $P(\vec{x}, t)$ and the stationary distribution $P^*(\vec{x})$ is known to be a Lyapunov function of the dynamics [477], and is defined as:

$$D_{KL}(t) = \int d\vec{x}\; P(\vec{x}, t) \log\left(\frac{P(\vec{x}, t)}{P^*(\vec{x})}\right) . \tag{5.101}$$

Note that, in thermodynamic context, $\sigma = k_B T$, so that such a functional is akin to a dynamic free energy:

$$\mathcal{F} = k_B T D_{KL}(t) = \langle V \rangle - k_B T S(t). \tag{5.102}$$

As a consequence, its time-derivative has to be non-positive and vanishes at the steady-state. Thus, the correct stationary solution of Eq. (5.95) can be found by minimizing the following functional with respect to P (including also the normalization constraint):

$$\mathcal{G} = -\frac{dD_{KL}}{dt} = \langle \vec{v} \circ \vec{v} \rangle - \langle \vec{v} \circ \vec{v}^* \rangle = \langle \vec{v} \circ \vec{v} \rangle - \langle \vec{v} \circ \vec{\nabla} \times \vec{B} \rangle \,, \tag{5.103}$$

that highlights how the geometrical properties of $\vec{v}^*$ enters the game. Here, $\vec{\alpha} \circ \vec{\gamma} = \sigma^{-1} \vec{\alpha}^T \hat{D} \vec{\gamma}$ and $\langle \cdot \rangle = \int \cdot P d\vec{x}$. The first interesting observation is that $\mathcal{G}$ can be written solely in terms of the velocity of probability currents, which appears to be the most natural quantity to characterize out-of-equilibrium dynamics and NESS.

$$D_{KL}(t) = \int d\vec{x}\; P(\vec{x}, t) \log\left(\frac{P(\vec{x}, t)}{P^*(\vec{x})}\right) , \tag{5.104}$$

We can go further with the computation, noticing that the first contribution on the r.h.s. of Eq. (5.103) coincides with the total entropy production, while the second one quantifies the housekeeping heat dissipation [453]:

$$\dot{S}_{\text{tot}} = \langle \vec{v} \circ \vec{v} \rangle \qquad \dot{S}_{\text{hk}} = \langle \vec{v} \circ \vec{\nabla} \times \vec{B} \rangle \tag{5.105}$$

This equivalence can be obtained starting from the trajectory-dependent formulation. In sec.5.1 we have shown that the entropy production along a trajectory in the form of heat is

$$\dot{s}_{\rm m} = \dot{\vec{x}} \circ \vec{F} \tag{5.106}$$

using the Stratonovich prescription [428, 178, 476]. A part of $\dot{s}_{\rm m}$ is due to the housekeeping heat, i.e., the one necessary to maintain the steady-state distribution,

$$\begin{aligned} \dot{s}_{\rm m} &= \dot{\vec{x}} \circ \vec{F} = \dot{s}_{\rm hk} + \dot{s}_{\rm ex} && (5.107)\\ \dot{s}_{\rm hk} &= \dot{\vec{x}} \circ \vec{v}^{\rm st} = \dot{\vec{x}} \circ \vec{\nabla} \times \vec{B} && (5.108)\\ \dot{s}_{\rm ex} &= -\dot{\vec{x}} \circ D\vec{\nabla} V && (5.109) \end{aligned}$$

Consider now the functional along a single stochastic trajectory Γ:

$$f[\Gamma] = -\log\left(\frac{P(\vec{x},t)}{P^*(\vec{x})}\right)\Big|_{\Gamma}. \tag{5.110}$$

Its average coincides with the Kullback-Leibler divergence and the time derivative reads:

$$\begin{aligned} g[\Gamma] &= \frac{df[\Gamma]}{dt} = -\frac{d\log P}{dt}\Big|_{\Gamma} - \dot{x}^{\mu}\partial_{\mu}\log P|_{\Gamma} + \dot{x}^{\mu}\partial_{\mu}\log P^*|_{\Gamma} \\ &. \end{aligned} \tag{5.111}$$

By using the Fokker-Planck equation, we have:

$$\begin{aligned} g[\Gamma] &= -\frac{d\log P}{dt}|_{\Gamma} - \dot{x}^{\mu}D_{\mu\nu}v^{\nu}|_{\Gamma} + \dot{x}^{\mu}D_{\mu\nu}v^{*}\nu|_{\Gamma} \\ &= -(\vec{\nabla}\cdot\vec{v} - \sigma\vec{v}\cdot\vec{\nabla}\log P)|_{\Gamma} + \dot{\vec{x}}\circ\vec{v}|_{\Gamma} - \dot{\vec{x}}\circ\vec{\nabla}\times\vec{B}|_{\Gamma}. \end{aligned} \tag{5.112}$$

By using Eqs.(5.3) it is immidiate to see that:

$$g[\Gamma] = \dot{s}_{prod}(t) - \dot{s}_{\rm hk}(t), \tag{5.113}$$

where the first term is the entropy production and the second one is the current surviving at the steady state that depends on the solenoidal part of the force. By using the average formula for currents Eq.The average of $\dot{s}_{\rm hk}$ over trajectories readily gives the second term in Eq. 5.103. As a result, the housekeeping dissipation emerges from the non-conservative part of the force and, as we will show later, is linked to the tendency of performing cycles in the trajectory space. In summary, the total entropy production is related to the "modulus" of the velocity, while the housekeeping dissipation emerges from the non-conservative tendency of performing cycles in the trajectory space. Hence, in analogy to [450], we have:

$$\mathcal{G} = \dot{S}_{\rm tot} - \dot{S}_{\rm hk} = \dot{S}_{\rm ex} \geq 0\,. \tag{5.114}$$

The correct steady state of a general non-equilibrium dynamics is given by the minimum excess entropy production. Eqs.5.103 and 5.105 give a geometrical meaning to this principle and constitute the first result of this Letter. The first principle providing the correct stationary state of a general non-equilibrium dynamics is the minimization of the excess entropy production. This minimum $\dot{S}_{\rm ex}$ criterion resembles the Glandsdorff-Prigogine principle and trivially corresponds to the minimum entropy production in the equilibrium case [175, 475, 449]:

$$D_{KL}(t) = \int d\vec{x}\, P(\vec{x},t) \log\left(\frac{P(\vec{x},t)}{P_{\rm eq}(\vec{x})}\right), \tag{5.115}$$

$$.\mathcal{G}_{eq} = -\frac{dD_{KL}}{dt} = \langle \vec{v} \circ \vec{v} \rangle = \dot{S}_{\rm tot} \tag{5.116}$$

In our framework, we can improve the physical interpretation of this result. The first term of $\mathcal{G}$ in Eq. (5.103) only dictates that the system tends to minimize its total dissipation, as for equilibrium relaxation phenomena. The second contribution emerges only out of equilibrium, hence encoding an additional dissipation stemming from the properties of the steady-state velocity. Indeed, Eq. (5.103) explicitly shows that this extra dissipative term manifests into a symmetry breaking in the trajectory space, since the velocity tends to maximize the dissipation along preferential directions. In other words, the NESS is the least dissipative state compatible with a velocity that is maximally aligned with the closed force lines of B, thus accounting for a tendency to circulate and a consequent stationary dissipation into the environment.

To better characterize the connection between dissipation and emergent symmetry breaking in NESS, we study the probability to observe a closed trajectory. The Onsager-Machlup action [454] using the Stratonovich prescription for the system in Eq. (5.94) is:

$$\mathcal{S}(\Gamma) = \frac{1}{2}\int_0^t d\tau \left[\frac{1}{2}\left(\dot{\vec{x}} - \vec{F}\right)^T \hat{D}^{-1}\left(\dot{\vec{x}} - \vec{F}\right) - \vec{\nabla}\cdot\vec{F}\right], \tag{5.117}$$

where Γ is a trajectory of duration t, from $\vec{x}_0$ to $\vec{x}_f$ along which all quantities have to be evaluated. The asymmetric part of the action, $\mathcal{S}^{\rm a}$, is related to the dissipation along a trajectory, which is also equal to the ratio between the probability of Γ and its time-reversal, $\tilde{\Gamma}$ [178]. By considering any closed trajectory $\Gamma_\circ$, i.e., where $\vec{x}_0 = \vec{x}_f$, the terms accounting for the initial and final states of the trajectory vanish, thus we have:

$$\mathcal{S}^{\rm a} = \log\frac{P(\Gamma_\circ)}{\tilde{P}(\tilde{\Gamma}_\circ)} = \int_0^t d\tau\, \dot{\vec{x}} \circ \vec{F} = \int_{\Gamma_\circ} d\vec{x} \circ \vec{\nabla} \times \vec{B}\,. \tag{5.118}$$

This establishes a clear connection between the propensity of performing closed trajectories in a preferential direction and the housekeeping dissipation. Indeed, $\dot{S}_{\rm hk}$ enters into the non-equilibrium functional counterbalancing the entropy production minimization, Eq. (5.114), hence breaking the chiral symmetry (i.e., clockwise or counter-clockwise) that is present at equilibrium. This finding clarifies the main result of this Letter, providing also a broader context for recent results about a topological fluctuation theorem [478] and gauge symmetries in thermodynamics [438, 448, 479].

The proposed framework extends beyond the case of additive noise to multiplicative noise scenarios, where the Glansdorff-Prigogine principle has not been formulated. Notably, a state-dependent diffusion coefficient emerges in systems affected by thermal gradients [471, 472] or fluctuating environments [480, 481]. Moreover, finite-size fluctuations are important in fluctuating hydrodynamics [482] and field theories [483, 484].

Consider the following generic Langevin equation :

$$\dot{\vec{x}} = \vec{F} + \hat{G}(\vec{x})\vec{\xi}(t)\,, \tag{5.119}$$

with $\vec{\xi}$ Gaussian white noises with correlation matrix $\hat{C}$, $\hat{G}$ the state-dependent part of the diffusion matrix that we consider diagonal for simplicity, and hence a total diffusion matrix $\hat{D} = \hat{G}^T\hat{C}\hat{G}$. For $\hat{G} = \hat{1}$ (the identity matrix), we go back to Eq. (5.94). This choice

allows us to write the following stationary solution of Eq. (5.119) in the weak-noise limit:

$$P^*(\vec{x}) = z(\vec{x})\eta(\sigma)\frac{e^{-V(\vec{x})/\sigma}}{Z} \,. \tag{5.120}$$

Here z and η are functions of space and noise amplitude, respectively, whose form depends on the model. The first important observation is that the maximum of the probability distribution does not coincide with the deterministic fixed point, because of the presence of a space-dependent noise. Moreover, note that the stationary velocity now contains the sum of a zeroth and a first order term in the noise amplitude (see sec.5.1):

$$\vec{v}^* = \vec{v}^{*(0)} + \sigma\vec{v}^{*(1)} \tag{5.121}$$

By plugging Eq. (5.120) into the corresponding FPE and expanding in orders of the noise amplitude, we can investigate the topological properties of $\vec{v}^*$. In sec.5.1 we have shown that:

$$\begin{aligned} \vec{v}^{*(0)} \cdot \vec{\nabla} V &= 0 \\ \vec{\nabla} \cdot \vec{v}^{*(0)} &= \vec{v}^{*(0)} \cdot \vec{\nabla} \log z + \vec{v}^{*(1)} \cdot \vec{\nabla} V \\ \vec{\nabla} \cdot \vec{v}^{*(1)} &= \vec{v}^{*(1)} \cdot \vec{\nabla} \log z \,. \end{aligned} \tag{5.122}$$

In particular, the zeroth order velocity is tangent to the height lines of V, as in the additive case. In precedent section (Eq. 5.85) we have shown that this condition actually implies that the $\vec{v}^{*(0)}$ is, again, a solenoidal field, i.e.

$$\vec{v}^{*(0)} = \vec{\nabla} \times \vec{B}. \tag{5.123}$$

Furthermore, the geometrical properties of the first order velocity can be unveiled by using a specific change of variables. Here, we consider Stratonovich integration, even if our result can be derived for any prescription, as we show in sec.5.1. We remind that a change of variables would allow to map a multiplicative noise into an additive one, restoring the results we obtained before, but in the transformed space. In particular, choosing the Jacobian of the transformation $\vec{x} \to \vec{x}'$ as $\hat{\Lambda} = \hat{G}^{-1}$, the diffusion matrix of the transformed dynamics is equal to $\hat{C}$. By deriving how the probability distribution and the velocity transform under this change of variables, we determine that

$$z(\vec{x}) = |\hat{\Lambda}| = |\hat{G}|^{-1} \tag{5.124}$$

, where $|\cdot|$ indicates the determinant. As a consequence,

$$\vec{v}^{*(1)} = \hat{D}\vec{\nabla}\psi \tag{5.125}$$

, where $\psi = \log|\hat{G}|$, i.e., a gradient field that contributes with a new term in the functional (see sec.5.1):

$$\mathcal{G} \;=\; \langle \vec{v} \circ \vec{v} \rangle - \langle \vec{v} \circ \vec{v}^* \rangle = \langle \vec{v} \circ \vec{v} \rangle - \langle \vec{v} \circ \vec{\nabla} \times \vec{B} \rangle - \sigma \langle v \vec{\nabla} \psi \rangle - \alpha\sigma \langle \vec{v} \circ \vec{\nabla} D \rangle. \tag{5.126}$$

In analogy to Eq. (5.103), the first contribution quantifies the total entropy production, which tends to be minimized as the system goes toward stationarity. The second and third terms amount to the dissipated heat to maintain the steady-state and depends on the symmetries that are broken in the trajectory space. These terms counterbalance the entropy production minimization and play a role analogous to the housekeeping heat in

the additive noise case. In particular, the second term accounts for the heat dissipated along the solenoidal part of the force, while the third contribution is proportional to the derivative of $|G|$ and vanishes in the limit of additive noise. An intuitive understanding of this last term might come considering a system subjected to a thermal gradient, $T(\vec{x})$. In this case, it can be readily shown that $\vec{\nabla}\psi \propto T^{-1}\vec{\nabla}T$, resembling an additional dissipation arising from thermophoretic effects and due to the necessity of transporting heat [95]. A general physical interpretation of ψ might be particularly challenging to find, since the multiplicative noise arises in a wide variety of systems, and might be the topic of future investigations. Eq.5.126 constitutes the second main result of this section. In this more complex scenario, the non-equilibrium functional $\mathcal{G}$ does not coincide with the excess entropy production, and its thermodynamic properties are not known a priori. We find that it can always be interpreted as the sum between the excess entropy production and an additional *thermophoretic* term, named after its meaning in thermally-driven systems and indicating the accumulation of probability in regions of low noise. This result generalizes the Glandsdorff-Prigogine principle to generic (linear and non-linear) stochastic systems away from equilibrium. Let us comment that the thermodynamics of system with temperature gradients in general setting defies a description in terms of overdamped dynamics, as Eq.(5.69), and hence the principle, Eq.(5.126), in this scenario is not fully correct. [96, 485, 95]. Nevertheless, in sec.5.4 we provide an example of a system with a temperature gradient and study it in the *underdamped* limit finding an interesting functional with analogies to Eq.(5.126).

5.4 Examples

Here, we present two pedagogical examples where the chiral symmetry breaking and the emergence of preferential cyclic trajectories accountable for the housekeeping entropy production are easy to visualize.

Driven Brownian particle on a torus

Consider a Brownian particle confined in a torus by a quadratic potential, $U(x,y)$, and driven along the torus itself by a constant driving force, f. Here, f breaks the detailed balance, and the stationary polar flux will induce a topological symmetry breaking in the system. Let us start with the description of the Brownian motion:

$$\begin{aligned} \dot{x} &= f_x(x,y) - \partial_x U(x,y,z) + \sqrt{2D}\,\xi_x(t) \\ \dot{y} &= f_y(x,y) - \partial_y U(x,y,z) + \sqrt{2D}\,\xi_y(t) \\ \dot{z} &= f_z(x,y) - \partial_z U(x,y,z) + \sqrt{2D}\,\xi_z(t) \end{aligned} \tag{5.127}$$

where f_i is the component of the driving force along i and $\xi_i(t)$ is a Gaussian white noise with unit variance. These equations can be rewritten in polar coordinates (ρ,ϕ,z), with $\rho = \sqrt{x^2+y^2}$ and $\phi = \arctan(y/x)$. Employing the Ito's formula for the change of variable, we obtain:

$$\begin{aligned} \dot{\rho} &= -\gamma(\rho-\rho^*) + \frac{D}{\rho} + \sqrt{2D}\,\xi_\rho(t) \\ \dot{\phi} &= \frac{f}{\rho} + \frac{\sqrt{2D}}{\rho}\xi_\phi(t) \\ \dot{z} &= -\gamma z + \sqrt{2D}\,\xi_z(t) \end{aligned} \tag{5.128}$$

with $U = (\gamma/2)(\rho - \rho^*)^2 + (\gamma/2)z^2$, f only acts along ϕ, and D/ρ is the usual Ito term. The resulting Fokker-Planck equation is:

$$\partial_t P = (\mathcal{L}_\rho + \mathcal{L}_\phi + \mathcal{L}_z)P \tag{5.129}$$

where we defined:

$$\begin{aligned} \mathcal{L}_\rho &= \partial_\rho \left(\gamma(\rho - \rho^*) - \frac{D}{\rho} \right) + D\, \partial_{\rho\rho} \\ \mathcal{L}_\phi &= \partial_\phi \left(-\frac{f}{\rho} \right) + \frac{D}{\rho^2}\, \partial_{\phi\phi} \\ \mathcal{L}_z &= \partial_z \left(\gamma z \right) + D\, \partial_{zz} \end{aligned} \tag{5.130}$$

The motion along z is decoupled, while the other two equations can be solved imposing that the flux only flows along ϕ. The steady-state is:

$$P^* = Z^{-1} \exp\left(-\frac{\rho(\rho - 2\rho^*)\gamma + \gamma z^2 - 2D \log \rho}{2D} \right) \tag{5.131}$$

Hence, in the small noise limit, we have:

$$P^* \simeq Z^{-1} \exp\left(-\frac{\rho(\rho - 2\rho^*)\gamma + \gamma z^2}{2D} \right) \tag{5.132}$$

which is of the form outlined in the main text. The flux only acts along ϕ, and it is equal to:

$$\vec{J}^* = Z^{-1} \exp\left(-\frac{\rho(\rho - 2\rho^*)\gamma + \gamma z^2}{2D} \right) f\, \mathbf{i}_\phi \tag{5.133}$$

where $\mathbf{i}_\phi$ is the versor indicating the coordinate ϕ. Hence, the stationary velocity reads:

$$\vec{v}^* = -\frac{f}{\rho}\, \mathbf{i}_\phi \tag{5.134}$$

As expected, $\vec{v}^*$ is a solenoidal field and indicates that the housekeeping heat stems from cyclic trajectories running across the entire torus in a preferential direction. In this example, the chiral symmetry breaking is explicitly linked to the symmetry of the non-conservative driving, making immediate to identify the origin of the term in $\mathcal{G}$ that tends to be maximized in the NESS.

Brownian Gyrator

The second example is a $2D$ Brownian gyrator, i.e., a diffusive particle in a confining potential subjected to two reservoirs at different temperatures, T_x and T_y, each one acting along one direction.
Consider the two dimensional motion in the $x - y$ plane of a particle under the effect of a parabolic potential U and two baths at temperatures $T_x = T$ and $T_y = T(1 + \delta)$, each one acting along a different direction:

$$\begin{aligned} \dot{x} &= -\partial_x U(x, y) + \xi_x(t) \\ \dot{y} &= -\partial_y U(x, y) + \xi_y(t) \\ \langle \xi_i(t) \rangle &= 0 \\ \langle \xi_i(t) \xi_j(t') \rangle &= 2T_i \delta_{ij} \delta(t - t'), \end{aligned} \tag{5.135}$$

where the potential is $U(x,y) = \frac{x^2}{2} + \frac{y^2}{2} + uxy$, with $|u| < 1$ to confine the particle near the origin. The system approaches the following stationary solution [486]:

$$P^*(x,y) = Z^{-1} e^{-\frac{V}{T\eta}} \tag{5.136}$$

$$V(x,y) = \frac{\gamma_1}{2}x^2 + \frac{\gamma_2}{2}y^2 + u\gamma_3 xy, \tag{5.137}$$

where we introduced $\circ to$ $\circ \eta = 1 + \delta + \frac{u^2}{4}\delta^2$

$\gamma_1 = 1 + \delta - \frac{u^2}{2}\delta$

$\gamma_2 = 1 + \frac{u^2}{2}\delta$

$\gamma_3 = (2+\delta)$(5.138)$\circ$ Note that in recover the typical form of the potential Eq.(5.5) it is sufficient to rename $\sigma = T\eta$. Note that the stationary solution respects the non-equilibrium potential ansatz, where T is the parameter regulating the noise, while δ and u affect the detailed balance condition that reads as follows:

$$T_x \partial_x F_y = T_y \partial_y F_x \to u\delta = 0. \tag{5.139}$$

Due to the anisotropy of the temperatures, detailed balance is broken and a current emerges with a velocity:

$$v_x = \partial_x U - \eta^{-1}\partial_x V \tag{5.140}$$

$$v_y = \partial_y U - \eta^{-1}(1+\delta)\partial_y V. \tag{5.141}$$

Its divergence reads:

$$\vec{\nabla}\cdot\vec{v}^* = \frac{1}{\eta}\left(2\eta - \gamma_1 - (1+\delta)\gamma_2\right) = 0, \tag{5.142}$$

thus the vorticity is:

$$w^* = \partial_x v_{yst} - \partial_y v_{xst} = \frac{u\gamma_3(T_y - T_x)}{\eta} = -\frac{u\delta\gamma_3}{\eta}$$

$$= \frac{4T_x u}{4T_x T_y + (T_x - T_y)^2 u^2}\left(T_x^2 - T_y^2\right). \tag{5.143}$$

Interestingly, the chiral symmetry breaking comes from the simultaneous presence of a temperature difference and an interaction coefficient, u. Since this model can be solved analytically for all times, in Fig. 5.1a we show the temporal evolution of $\dot{S}_{\rm tot}$ and $\dot{S}_{\rm hk}$, highlighting that the steady state coincides with a minimization of the total entropy production compatible with the maximum dissipation due to chiral symmetry breaking. While converging to stationarity, the system selects a region of the space where to concentrate probability fluxes, resulting in a stationary vorticity (see Fig. 5.1b).

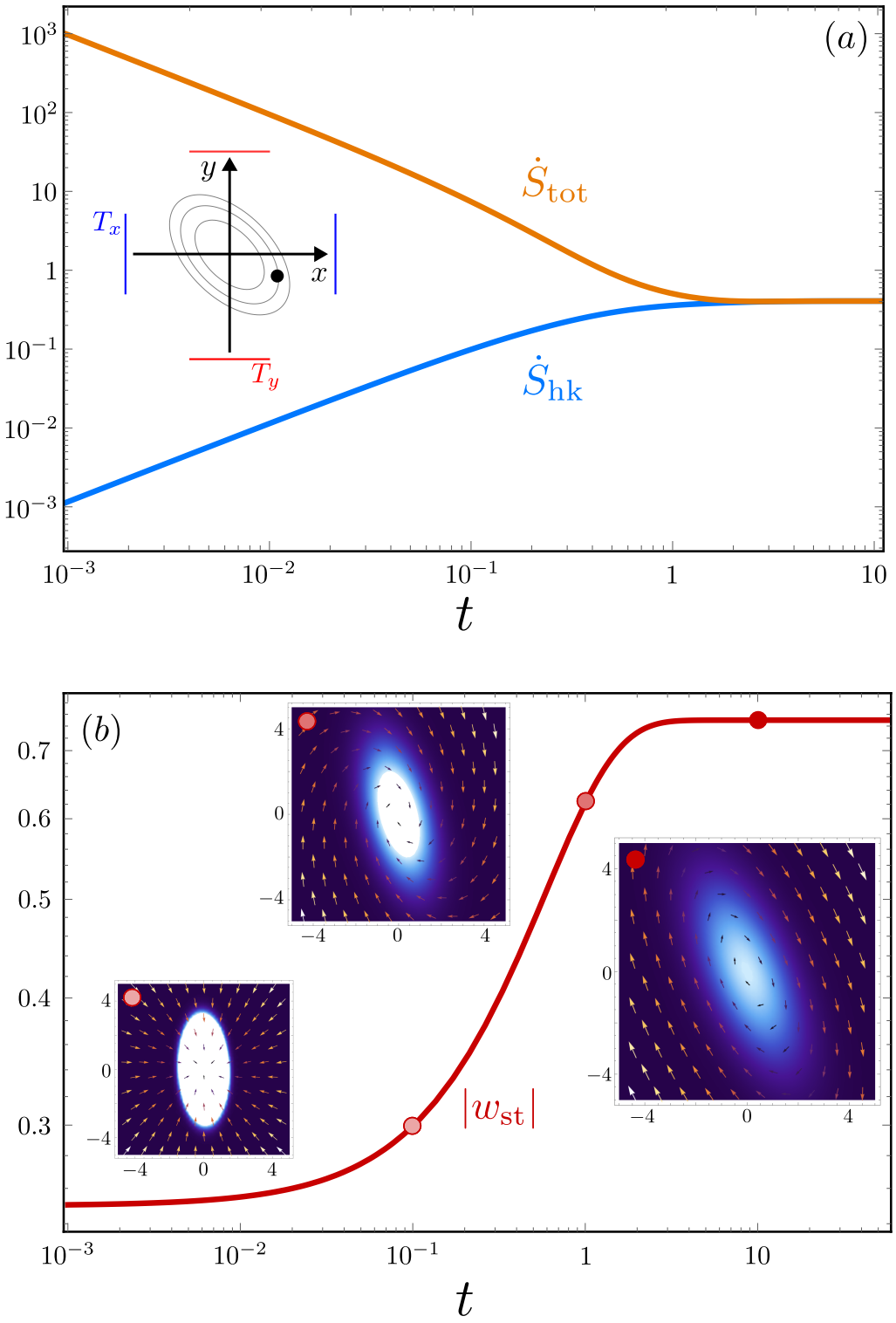


FIGURE 5.1: Dissipative symmetry-breaking in a Brownian gyrator. (a) A sketch of the model is presented. As a function of time, we show that the system tends to minimize $\dot{S}_{\rm tot}$, while maximizing $\dot{S}_{\rm hk}$, which is the dissipation associated with the symmetry breaking in the trajectory space. At the NESS, $\dot{S}_{\rm tot} = \dot{S}_{\rm hk}$. (b) The modulus of the vorticity is plotted. The inset show the pdf (in color-scale) and the velocity vector field (bigger arrows corresponds to a stronger field) as time increases. Insets clearly indicate that the tendency to rotate in a preferential direction increases towards the NESS.

Motion in temperature gradient

Finally, we consider as a third example the motion of a particle in a temperature gradient. The motion of a particle, here considered free for simplicity, has to be described in the entire phase-space (x, w). Hence, a $1D$ dynamics turns into a $2D$ dynamics, when velocities and positions are considered. In this case, including also the presence of a temperature gradient, the Langevin equations describing the system are:

$$\dot{x} = w \tag{5.144}$$

$$\dot{w} = -\frac{\gamma}{m}v - \frac{1}{m}\partial_x U + \frac{1}{m}\sqrt{2\gamma T(x)}\xi(t) \tag{5.145}$$

where γ represents the friction, m the mass, $\partial_x U$ a (confining) potential, and the diffusion coefficient satisfies the Einstein's relation $D(x) = \gamma T(x)$ at each point in space. Notice that the temperature gradient translates into a multiplicative noise. The associated Fokker-Planck equation is then:

$$\partial_t P + w\partial_x P = \frac{1}{\tau}\partial_w \left(wP + \frac{\partial_x U}{\gamma}P + \frac{T(x)}{m}\partial_w P \right) \tag{5.146}$$

where $\tau = m/\gamma$ is the characteristic friction time. When $\tau \to 0$, velocities equilibrate infinitely faster, the system reaches the overdamped limit and can be described only through its dynamics in the position space. Performing a time-scale separation up to the first order, as outlined in [95], we obtain the following solution:

$$P_{\tau st} = e^{-\frac{mw^2}{2T}} \left[\Phi_0 - \tau \left(w\partial_x \Phi_0 + \left(\frac{mw^3}{6T^2} + \frac{w}{T} \right) \Phi_0 \partial_x T + \frac{w}{T}\Phi_0 \partial_x U \right) \right]$$

where $\Phi_0(x,t) = \phi_0(x,t)\sqrt{2\pi T(x)/m}$, with ϕ_0 satisfying the following $1D$ Fokker-Planck:

$$\partial_t \phi_0(x,t) = -\partial_x \left(\frac{\partial_x U}{\gamma}\phi_0(x,t) \right) + \partial_{xx} \left(\frac{T(x)}{\gamma}\phi_0(x,t) \right) \tag{5.147}$$

that is usually solved imposing a zero-flux conditions in the x-space. Notice that this solution does not depend on the employed prescription, since the x-dependent multiplicative noise acts only on w. It turns out that the overdamped Langevin equation should be treated according to the Ito prescription. There are other limits in which a different prescription has to be taken for consistency, see [95, 485]. As an additional remark, $\Phi_0(x,t)$ provides the correct solution only in the bulk since boundary terms cannot be consistently computed using a time-scale separation analysis, as we did.

The form of the steady-state of $P_{\tau st}$ is similar to the one considered in Eq.(5.5), with some differences. First of all, we are not in the weak noise limit. Second, to match all the steps outlined above, we need to perform the computations from scratch since we have to consider different (physically relevant) small parameters, such as τ and the magnitude of the thermal gradient.

Close-to-equilibrium regime

To investigate how the multiplicative noise affects the geometric properties of the system, let us consider the simple setting of a linear gradient, $T(x) = T_0(1 + \alpha x)$, with $\alpha \ll 1$. To avoid unnecessary lengthy calculations, we also consider the case $U = 0$. We already know all the symmetries holding when $\alpha = 0$, as the system has a constant diffusion coefficient.

Writing the stationary solution as follows:

$$P_{\tau st} = e^{-V(x,w)}\left(f_0 + \tau f_1\right), \tag{5.148}$$

the implicit equation resulting from the stationarity condition is:

$$\vec{\nabla}\cdot\vec{v}^* + \vec{v}^*\cdot\vec{\nabla}\log\left(f_0+\tau f_1\right) = \vec{v}^*\cdot\vec{\nabla}V \tag{5.149}$$

where $\vec{v}^* = \vec{J}^*/P^*$ at stationarity as before.

Analyzing each contribution separately, we find that all of them are $\mathcal{O}(\alpha)$, hence the multiplicative noise affects to the same degree all contributions, thus no symmetry is perturbatively preserved - for small enough α - in the presence of thermal gradient. Moreover, even for $\tau \to 0$, i.e. in the overdamped limit, the symmetries are broken.

Nonequilibrium functional and entropy production Let us now find the functional in this particular case. The expression of the dissipated heat for an underdamped system in a temperature gradient only depends on the irreversible flux [95, 96]. However, computing the Lyapunov function of the dynamics, both reversible and irreversible contributions appear in the functional. In full generality,

$$\begin{aligned} \vec{v}_{\text{rev}} &= \left\{w, -\frac{1}{m}\partial_x U(x,t)\right\}, \\ \vec{v}_{\text{irr}} &= \left\{0, -\frac{1}{\tau}w - \frac{1}{m\tau}T(x)\partial_w \log(P(x,w))\right\} \end{aligned} \tag{5.150}$$

We notice is that $v_{\text{revst}} = v_{\text{rev}}$, thus we have:

$$\mathcal{G} = \langle v_{\text{irr}}\circ v_{\text{irr}}\rangle - \langle v_{\text{irr}}\circ v_{\text{irrst}}\rangle - \langle v_{\text{rev}}\circ(v^*_{\text{irrst}} - v_{\text{irr}})\rangle \tag{5.151}$$

The entropy production for underdamped systems with temperature gradients has been derived in various works, and depends just on the irreversible velocity [96, 487]:

$$\dot{S}_{prod} = \langle v_{\text{irr}}\circ v_{\text{irr}}\rangle, \quad \dot{S}_{hk} = \langle v_{\text{irr}}\circ v_{\text{irrst}}\rangle \tag{5.152}$$

and correspond to the second and third terms. The new extra extra term is:

$$-\langle v_{\text{rev}}\circ(v^*_{\text{irrst}} - v_{\text{irr}})\rangle = \tau\int dxdv\frac{\partial_x U(x,t)}{T(x)}\left(J_{\text{irr}} - J_{\text{irrst}}\right) \equiv \frac{\dot{Q}(t)}{T(x)} - \frac{\dot{Q}^*}{T(x)} \tag{5.153}$$

which relates to the *excess* of power converted into a change of internal (kinetic) energy without being dissipated into the environment, i.e., $\int dxdvFJ$. In the case of driven systems, this quantity is also identified as the instantaneous energy input. Hence, by summing up the functional is:

$$\begin{aligned} \mathcal{G} &= \langle v_{\text{irr}}\circ v_{\text{irr}}\rangle - \langle v_{\text{irr}}\circ v_{\text{irrst}}\rangle - \langle v_{\text{rev}}\circ(v^*_{\text{irrst}} - v_{\text{irr}})\rangle \\ &= \dot{S}_{prod} + \dot{S}_{hk} + \frac{\dot{Q}(t)}{T(x)} - \frac{\dot{Q}^*}{T(x)} = \dot{S}_{ex} + \frac{\dot{Q}_{ex}}{T(x)} \end{aligned} \tag{5.154}$$

If we split v in its reversible and irreversible part, it is easy to pinpoint the geometrical properties of both components in the case of a free particle, for the sake of simplicity, as

before. Indeed, we have that all terms scales linearly with α, except for:

$$\begin{aligned} \vec{\nabla} \cdot \vec{v}_{\text{rev}} &= 0 \\ \vec{v}_{\text{irrst}} \cdot \vec{\nabla} \left(f_0 + \tau f_1 \right) &\simeq \mathcal{O}(\alpha^2 \tau) \end{aligned} \tag{5.155}$$

5.5 Conclusions and discussion

Our approach sheds light on the role of topological aspects in NESS. Indeed, the housekeeping dissipation is associated with the onset of vortex structures around the zero-current point that acts as an *emerging* topological defect (see Fig. 5.1b). Vortices arise around the deterministic fixed point only in the presence of additive noise, otherwise complex structures (e.g., dipole currents) might emerge [445]. Moreover, beyond the simple $2D$ scenario, zero-current manifolds might appear and act as defects, hence generating non-trivial vortex structures at NESS and enriching even more the presented picture [478].

The idea that some kind of selection can naturally take place out-of-equilibrium due to dissipation-driven processes is gaining momentum in the field of physical chemistry [471, 472] and evolutionary dynamics [488]. Our framework might provide a useful tool to tackle this problem. At stationarity, probability fluxes will be focused in determined regions of the variable space, $\vec{x}$. If each x_i represents a species in a fictitious space, the emergent chiral symmetry breaking can reflect the onset of preferential cycles involving some of them. Additionally, these cycles are intimately connected with the housekeeping dissipation and could be explored using large-deviation theory [489]. This perspective intriguingly resembles the idea of *hypercycles* in the context of the origin of life [490]. In order to push forward this analogy, we leave for future investigation the extension of the framework discrete-state dynamics (e.g., master equations), in line with preliminary works in this direction [456, 474].

Active matter represents another possible field of application of the presented framework. In the simplest case of an active Brownian particle [491], for example, a solution can be readily found at all times and it closely resembles the one of the $2D$ Brownian gyrator. The conceptual leap, in this context, is that emergent dissipative cycles arise in the space describing the particle and the bath, making the physical interpretation harder [492]. Future works might also explore this direction of research.

To summarize, we unraveled the connection between dissipation and chiral symmetry breaking in NESS of stochastic mesoscopic systems. We showed that the housekeeping dissipation is intimately connected to the tendency of performing cycles in the trajectory space. This quantity tends to be maximized in the NESS and counterbalances the entropy production minimization. In the most general scenario of multiplicative noise, the non-equilibrium functional shows two dissipative symmetry-breaking contributions. The first accounts for cycles in a transformed space, while the second can be seen as a thermophoretic dissipation. This results generalizes the principle of excess entropy production minimization, extends and clarifies the physical meaning of the Glansdorff-Prigogine principle, and paves the way to understand selection phenomena in different contexts as a result of a symmetry breaking process driven by non-equilibrium dissipation.

Chapter 6

Irreversibility in adaptive evolution

6.1 Introduction

Evolution is an intrinsically non-equilibrium process: selection, mutations and genetic drift shape constantly the population composition, causing diversification, extinctions and, possibly, an increase in complexity [319, 275, 493]. Most importantly, evolution is characterized by a strong irreversibility and a "weak" direction of time. Evolution of complex traits is in general considered irreversible for the accumulation of neutral and beneficial mutations, making the ancestral traits less fit in respect to new ones. Nevertheless, trait reversibility has been observed for complex ones, such as multicellularity, causing controversy in the field [494]. Historically, the possibility of a direction of time in evolution has been studied in comparison with the second law of thermodynamics. In particular, Fisher formulated his celebrated "Fundamental theorem of natural selection", stating that an evolving population will maximize its fitness [495, 496]. The "theorem" derives easily from the Price equation, raising some criticisms, it has influenced deeply the scientific community and has been generalized in a number of ways [497, 245, 498]. One of the main criticisms is the definition of fitness: instead of being considered as an emerging ecological quantity, in population genetic fitness is generally considered as an advantage parameter [499]. In this way, relative fitness is a dimensionless quantity, which is calculated in evolution experiments as the ratio of the growth rate of the derived type to the one of its ancestral competitor during direct competition. Notably, in many experiments, such as the Lenski long term one, it has been found to increase with time [500].
John Maynard Smith, another father of modern evolutionary theory, had a different perspective on the topic. In particular, in his essay " Time in the Evolutionary Process " [263] he dismissed the fundamental role of Fisher theorem due to the difficulties in defying and measuring fitness, and concluded:
"Fisher's theorem cannot help us to put an arrow on evolutionary time. Yet it is in some sense true that evolution has led from the simple to the complex: prokaryotes precede eukaryotes, single-celled precede many-celled organisms, taxes and kineses precede complex instinctive or learnt acts. I do not think that biology has at present anything very profound to say about this. If there is a "law of increasing complexity", it refers not to single species, as does Fisher's theorem, but to the ecosystem as a whole. The complexity of the most complex species may increase, but not all species become more complex.". This perspective led him to formulate with E. Szathmary the celebrated theory of the major transitions in evolution [493]. They pointed out that during the history of life many irreversible transitions happened from a group of independent individuals to a collective entity, such as the origin of chromosomes and eukaryotes and the transition from unicellular to multicellular organisms. As he says in the citation, he did not believed in a rigid law of increasing complexity, but more on a common efficient evolutionary way of transmitting information and divide labor. At the moment no simple quantitative theory of major transitions exists.
Going back to the parallelism between between thermodynamics and evolution, in recent

years many researchers have studied the non-equilibrium and irreversible properties of adaptation [501, 184, 319, 502]. In particular, first Mustonen and Laessig [184], and recently Rao and Leibler [319] applied key-results of stochastic thermodynamics to adaptation. In the first pioneering work, a non-negative quantity called "Fitness flux" consisting of the sum of adaptation along the trajectory, is identified as driving irreversibility in the context of stochastic molecular evolution . In the deterministic limit of an infinite population size, $N \to \infty$, the fitness flux converges to Fisher-theorem. The limitation of this work lies in the fact that mutations are not considered as a source of stochasticity, and everything relies on the role of genetic drift. Furthermore, the derivation is intrinsically macroscopic, i.e. consisting of the non-equilibrium properties of a Langevin/Fokker-Planck such as system, while in non-equilibrium statistical physics it is fundamental to study irreversibility at the microscopic individual based level and then apply a "coarse-graining" procedure [181, 177]. In the second work, Rao and Leibler studied the non-equilibrium properties of evolution at a microscopic level, and identified the so called non-conservative evolutionary forces, but do not study the consequences on irreversibility, and furthermore, their example is not a concrete nor standard scenario.
In this chapter our aim is to study the non-equilibrium and irreversible properties of evolution using the framework introduced in Sec. 3.2, deriving rigorously the macroscopic irreversibility from the microscopic one. Furthermore, we understand the relation between irreversibility and information [502, 498, 503], and study what happens to irreversibility during diversification phenomena, or "evolutionary transitions", such as evolutionary branching. In Sec.6.2 we recall the main characteristic of our evolutionary framework and use it to derive general evolution equations for quantities of interest. Sec. 6.3 is dedicated to the study of adaptation with information theory tools, that will be connected with microscopic irreversibility in the next section 6.4. To have a concrete understanding of irreversibility in evolution in Sec. 6.5 we study various typical examples such as stabilizing evolution and evolutionary branching. In Sec. C.4 we report all the technical details on the mean-field derivation of the entropy production. Finally, in Sec.6.6 we comment on possible stochastic generalizations and compare our approach to the work by Mustonen and Laessig. Furthermore, we propose an interpretation of the stationary entropy production in terms of "population efficiency" and comment on open questions and possible generalizations.

6.2 Introduction to the framework

In this section we study the irreversibility of adaptive evolution using the framework introduced sin Sec. 3.2. In particular, we consider the evolution of a population of N individuals, each represented by a one dimensional trait continuous trait, x. The dynamics involves reproduction, represented by the fitness function $f(x_i, \widetilde{\mathbf{x}}^j)$, random death with rate $d = 1/(N-1)$ and mutation given by the kernel $\beta(x - \tilde{x})$. Hence, it is represented by the following N-body Master Eq.:

$$\partial_t P(\mathbf{x}, t) = \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \int_{\mathcal{P}} d\widetilde{x}_j \Big[W_i(\mathbf{x}, \widetilde{\mathbf{x}}^j) P(\widetilde{\mathbf{x}}^j, t) - W_i(\widetilde{\mathbf{x}}^j, \mathbf{x}) P(\mathbf{x}, t) \Big] \tag{6.1}$$

where

$$W_i(\mathbf{x}, \widetilde{\mathbf{x}}^j) = f(x_i, \widetilde{\mathbf{x}}^j) \;\; \beta(x_j - x_i) \; d(\widetilde{x}_j) f(x_i, \widetilde{\mathbf{x}}^j) \; \beta(x_j - x_i)/(N-1) \tag{6.2}$$

is the transition rate to go from a state $\widetilde{\mathbf{x}}^j = (x_1, .., x_i, .., \widetilde{x}_j, .., x_N)$ (with just one coordinate differing from those of $\mathbf{x}$) to $\mathbf{x} = (x_1, .., x_i, .., x_j, ..x_n)$, and $W_i(\widetilde{\mathbf{x}}^j, \mathbf{x})$ is the rate for the reverse process.

As explained in Sec. 3.2, one can derive the evolution probability density of finding *any* individual with phenotype x at time t, $\phi(x,t)$ using a coarse-graining procedures composed by mean-field and small mutation approximation (see Eq.(4.3)):

$$\partial_t \phi(x,t) = \left(f(x,t) - \bar{f}(t)\right)\phi(x,t) - \partial_x \theta(x) f(x,t)\phi(x,t) + \frac{1}{2}\partial_x^2 \sigma^2(x) f(x,t)\phi(x,t), \tag{6.3}$$

where the first term of the Eq. models selection, while the second and third mutation. For its similarity with a celebrated equation in population genetics, we named Eq.(6.3) the Generalized Crow-Kimura equation (GCK).

The second term can be written as the divergence of the "mutational" current:

$$\partial_t \phi(x) = (f(x) - \bar{f})\phi(x) - \partial_x j(x), \tag{6.4}$$

with $j(x,t) = \theta(x) f(x,t)\phi(x,t) - \frac{1}{2}\partial_x \sigma^2(x) f(x,t)\phi(x,t)$, suggesting a similarity with the Fokker-Planck equation. For simplicity, in the following sections we consider $\theta = 0$, $\sigma^2(x) = \sigma^2$. We are interested in the dynamics of any average general quantity $A(x,t)$,:

$$d_t \overline{A} = \partial_t \overline{A} + \overline{\dot{A}} = \int dx A\dot{\phi} + \int dx \dot{A}\phi \tag{6.5}$$

where the first term is the change in the average value of the observable, and the second one is it's the average of dynamic change. Using Eq. (6.3) the time evolution of the expectation value of any observable, i.e. the first term, is a Price equation [251] with the following form:

$$\partial_t \overline{A} = \int dx A\dot{\phi} = \Sigma[f, A] + \overline{v\partial_x A}, \tag{6.6}$$

where the first term is the covariance between the quantity A and the fitness:

$$\Sigma[f, A] = \int dx \phi(x) f(x) A(x) - \bar{f} \int dx A(x)\phi(x), \tag{6.7}$$

while the second one is obtained by integrating by parts (assuming the surface term to be zero) and by defining the mutation velocity

$$v = \frac{j}{\phi} = -\frac{\sigma^2}{2}\partial_x f - \frac{\sigma^2}{2} f \partial_x \log \phi \tag{6.8}$$

$$-\int dx \partial_x j(x) A(x) = \int dx j(x) \partial_x A = \int dx \phi(x) \frac{j(x)}{\phi(x)} \partial_x A \overline{v\partial_x A}. \tag{6.9}$$

The formulation of the GCK Eq. in terms of a currents clarifies that, at a stationary state, selection and mutation balance each other, defining a "mutation-selection balance":

$$(f^* - \bar{f}^*)\phi^* = \partial_x j^*, \quad \Sigma^*[f, A] = \overline{v\partial_x A}. \tag{6.10}$$

Using our physical intuition one could claim the mutation-selection balance to be out of equilibrium, given that there is a non-zero mutational current j^*. Indeed, in the next

sections we rigorously verify this claim.

6.3 Information entropy dynamics

To study evolution irreversibility we start by considering the dynamics of the Shannon entropy, as usually done in Stochastic Thermodynamics [89]. The informational entropy has been proposed by Shannon in the context of signal communication [504]. Its function is to estimate the information value of a message in a source-channel-receiver system [505]. Indeed, one can define the surprise or information rate of a certain event x with probability $P(x)$ as:

$$s(x,t) = -\log P(x), \quad S = -\int dx P(x)\log P(x) = \bar{s}, \tag{6.11}$$

obtaining the Shannon entropy is just the mean surprise rate. With this definition, $s \approx 0$ and very little information is obtained when the event is very likely. On the other hand, if the event is very rare, $s > 0$, information is gained when observed. Apart from communication theory, Shannon's entropy has been widely used also in the biological sciences. In the context of ecology, the Shannon entropy has been used to quantify the diversity or the complexity of the system [506], while in the context of bet-hedging its play a critical role in correctly choosing a good evolutionary strategy [247]. Furthermore, it has been also used in the context of mathematical evolution, where many researchers claim that information is the natural language of evolution [498, 503, 250, 502].
For its similarity with Gibbs entropy, the informational entropy has been used as a generalization of the thermodynamic one in stochastic thermodynamics [89], used in thermodynamic systems that store information [507]. In addition, the Shannon entropy can be used as an ingredient to study irreversibility in non-equilibrium statistical physics [89, 433]. Indeed, in non-equilibrium statistical physics it is known that for a general Markov process, described either by a master or a Fokker-Planck equation, the change in Shannon entropy can be always decomposed in a non-negative contribution $\dot{S}_{prod}$ and a second component $\dot{S}_{flux}$:

$$\dot{S} = \dot{S}_{prod} - \dot{S}_{flux}, \quad \dot{S}_{prod} \geq 0, \tag{6.12}$$

where $\dot{S}_{prod}$ quantifies how much detailed balance is broken, and hence how irreversible is the system [433, 89]. As a consequence, the first step to understand irreversibility in evolution is to study the dynamics of its Shannon entropy.
In the following sections, we will use the Shannon entropy to estimate the effect of selection, mutation and neutrality in evolution. Selection is the force that tends to shape the population trait distribution or in a particular way that would increase its fitness. On the other hand, mutations and drift tend to disorder the population traits, creating diversity but also lowering the fitness by inserting randomness in the trait distribution. Hence, we define the population information rate $s(x,t)$ and its average as:

$$s(x,t) = -\log\phi(x,t) = s(x,t) \tag{6.13}$$

$$\bar{s} = -\int dx\phi(x,t)\log\phi(x,t) = S(t). \tag{6.14}$$

The change in information rate in time, $\dot{s}$, has zero average owing to the conservation of probability:

$$\dot{s} = -d_t \log \phi = -\frac{\dot{\phi}}{\phi}, \quad \bar{\dot{s}} = -\int dx \phi(x) \frac{\dot{\phi}(x)}{\phi(x)} = 0. \tag{6.15}$$

The time derivative of the Shannon entropy is:

$$\partial_t \bar{s} = \dot{S} = \Sigma(f, s) + \overline{v \partial_x s}, \tag{6.16}$$

where we have used Eq.(6.6). The first term is the effect of selection on the entropy

$$\dot{S} = \Sigma(f, s), \tag{6.17}$$

and tends to concentrate the distribution around the fittest individuals.
On the other hand, the second term can be rewritten as:

$$\overline{v \partial_x s} = -\int dx \phi(x) v(x) \partial_x \log \phi(x) = \frac{\sigma^2}{2} \int dx \partial_x f(x) \phi(x) \partial_x \log \phi(x), \tag{6.18}$$

where we have renamed $\sigma^2/2 \to \sigma^2$ for notation's sake and without loss of generality. By inverting Eq.(6.8) we obtain the following relation:

$$-\partial_x \log \phi(x) = \frac{v}{\sigma^2 f(x) \phi(x)} + \partial_x \log f(x); \tag{6.19}$$

that inserted in Eq.(6.18) gives:

$$\overline{v \partial_x s} = \dot{S}_{mut} + \dot{S}_{type}, \tag{6.20}$$

$$\dot{S}_{mut} = \overline{\frac{v^2}{\sigma^2 f}}, \quad \dot{S}_{type} = \overline{v \partial_x \log f}. \tag{6.21}$$

The first term $\dot{S}_{mut}$ is non-negative and interestingly corresponds to the typical entropy production of driven-diffusive systems [89]. It can be interpreted as the production of variability by mutations, i.e. an "entropic effect" that decreases the strength of selection but also makes possible the accumulation of information. To interpret the second one, $\dot{S}_{type}$, consider the time evolution of the average logarithmic fitness by using Eq.(6.6):

$$\partial_t \overline{\log f} = \Sigma(f, \log f) + \overline{v \partial_x \log f}, \tag{6.22}$$

hence, the second term in Eq.(6.20) $\dot{S}_{type}$ represents the mutation effect on logarithmic fitness. In particular, its sign indicates the "type" of mutation, i.e. deleterious ($\dot{S}_{type} < 0$) or beneficial ($\dot{S}_{type} > 0$).
Rewriting everything together one obtains:

$$\dot{S} = \Sigma(f, s) + \overline{\frac{v^2}{\sigma^2 f}} + \overline{v \partial_x \log f}. \tag{6.23}$$

At the stationary state, $\dot{S} = 0$ hence there is no net change in entropy. The $\dot{S}_{type}$ term is negative, because from Eq.(6.22) it is clear that:

$$-\Sigma(f, \log f) = \overline{v \partial_x \log f} \leq 0, \tag{6.24}$$

meaning that on average mutations reduce the fitness, i.e. they are deleterious. Summing up, at stationary state Shannon-entropy balance is:

$$\Sigma(f,s) + \overline{v\partial_x \log f} = -\overline{v\partial_x s} = -\frac{\overline{v^2}}{\sigma^2 f} \leq 0 \tag{6.25}$$

$$\overline{v\partial_x \log f} \leq -\Sigma(f,s). \tag{6.26}$$

Therefore, mutations increase the total entropy and lowers the fitness by introducing new traits, while selection balances to zero the change in entropy by eliminating some traits (the less fit ones) and amplifying the presence of others (the fittest ones).
Finally, it is quite interesting to study a last information-theoretic quantity, the *Jeffrey's* divergence, that is an estimation of the distance between two probability distributions infinitesimally close in time:

$$\mathcal{J} = \int dx \frac{\dot{\phi}^2}{\phi} \geq 0. \tag{6.27}$$

In evolutionary terms it measures how much will the next generation be different from the present one. It can be written as minus the change of the average time derivative of the information rate $\dot{I}$ using Eq.(6.6)(see Eq.(6.15).):

$$\mathcal{J} = \int dx (d_t \log \phi)\dot{\phi} = -\partial_t \bar{\dot{s}} = -\Sigma[f,\dot{s}] - \overline{v \cdot \partial_x \dot{s}} \tag{6.28}$$

By using the definition of $\dot{s}$ Eq(6.15) in $-\Sigma[f,\dot{s}]$ one obtains that it equivalent to the change in mean fitness:

$$-\Sigma[f,\dot{s}] = -\int dx f\dot{s}\phi + \bar{f}\bar{\dot{I}} = \int dx f\dot{\phi} = \partial_t \bar{f} \tag{6.29}$$

By using Eq.(6.6)the change in mean fitness, also know as *Fisher* theorem of natural selection, reads:

$$\partial_t \bar{f} = \Sigma[f] + \overline{v\partial_x f}. \tag{6.30}$$

By joining Eq.(6.29) and Eq(6.28) one obtains that

$$\mathcal{J} = -\partial_t \bar{\dot{s}} = \partial_t \bar{f} - \overline{v \cdot \partial_x \dot{s}} \geq 0 \tag{6.31}$$

$$\tag{6.32}$$

Hence, the Jeffrey's divergence, Eq.(6.31), measures how fast is the changes in information/surprise, showing that it is proportional to the change in mean fitness. In this section, we have shown that change in the population Shannon entropy is able to distinguish between the effects of selection and mutations, and most importantly it has some intriguing connections with irrversibility. To understand better these connections, in the next section we calculate the system irreversibility starting from the underlying miscroscopic process.

6.4 Irreversibility: from micro to macro

By looking directly at the change in the population entropy derived in the precedent section, Eq.(6.23), one could be tempted to identify the system irreversibility, or entropy production, as the mutational non-negative contribution $\dot{S}_{mut}$. Instead of performing the

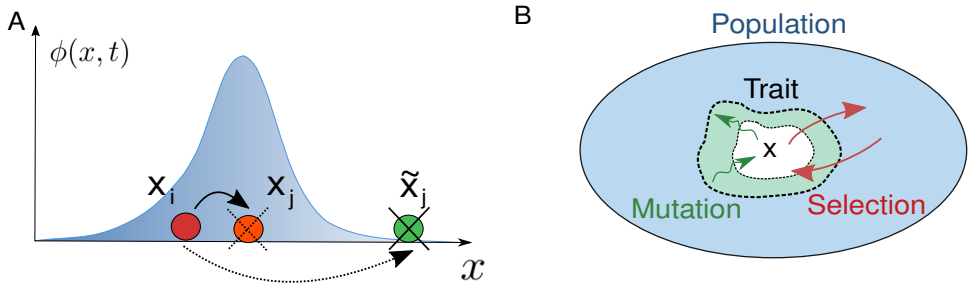


FIGURE 6.1: **A)Illustration of the detailed balance** Consider a typical process, bold arrows, where the individual $\tilde{x}_j$ is sampled to die, while individual x_i produces an offspring with trait x_j. The reversed process is depicted in dashed arrows and consist of the death of individual x_j, and the reproduction of x_i with an offspring $\tilde{x}_j$. Given that $\tilde{x}_j$ can be any individual in the population, its trait can be very different from x_i. In this scenario the reversed transition is very unlikely, unless the mutations are very big. **B) Adaptive population as an open system** In this figure we illustrate conceptually that each trait can be considered an "open" system interacting with the whole population. Interactions can be local as mutations, i.e. each trait produces a mutational current in neighbor traits, or non-local such as selection, given that the relative fitness involves the mean population fitness $\bar{f}$. The balance of these local and non-local current determines the population stationary state, that will a non-equilibrium one unless both are zero.

aforementioned arbitrary identification, in this section we rigorously derive the entropy production from the underlying microscopic process. Let us call N dimensional vector $\boldsymbol{x}$ representing the state of the population the *microstate*. Let us start from the miscroscpic master Eq.(D.2), and rewrite it in terms of probability currents:

$$\partial_t P(\mathbf{x},t) \;=\; \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\int_{\mathcal{P}} d\tilde{x}_j\Big[J_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) - J_i(\mathbf{x},\tilde{\mathbf{x}}^{\mathbf{j}})\Big] \tag{6.33}$$

with $J_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x})$ is the probability current from the the microstate $\tilde{\mathbf{x}}^{\mathbf{j}}$ and the microstate $\mathbf{x}$ obtained by the reproduction of the $i-$*individual*

$$J_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) = W_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x})P(\tilde{\mathbf{x}}^{\mathbf{j}},t) - W_i(\mathbf{x},\tilde{\mathbf{x}}^{\mathbf{j}},)P(\mathbf{x},t). \tag{6.34}$$

The dynamics converges to an equilibrium stationary state P^*_{eq}, i.e. a stationary state without probability currents, if and only if the detailed balance condition is verified:

$$J_i^*(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) = J_i^*(\mathbf{x},\tilde{\mathbf{x}}^{\mathbf{j}}) \tag{6.35}$$

for all $i,j \in [1,N]$. In detailed balance is not fulfilled, the population converges to a non-equilibrium stationary state (NESS). If one considers the microscopic Shannon entropy density

$$S^{micro} = -\frac{1}{N}\int d\mathbf{x} P(\mathbf{x},t)\log P(\mathbf{x},t), \tag{6.36}$$

and computes its change in time using Eq.(D.2) and simple manipulations, it is possible to identify two contributions (see the appendix Sec. C.4.2 and Ref. [433] for a detailed derivation):

$$\dot{S}^{micro} = -\frac{1}{N}\int d\mathbf{x}\partial_t P(\mathbf{x},t)\log P(\mathbf{x},t) = \dot{S}^{micro}_{prod} - \dot{S}^{micro}_{flux} \tag{6.37}$$

with

$$\dot{S}^{micro}_{prod} = \frac{1}{2N}\sum_{i,j=1,j\neq i}^{N}\int d\mathbf{x}d\tilde{x}_j J_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x})\log\frac{J_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x})}{J_i(\mathbf{x},\tilde{\mathbf{x}}^{\mathbf{j}})} \tag{6.38}$$

$$\tag{6.39}$$

$$\dot{S}^{micro}_{flux} = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\int d\mathbf{x}d\tilde{x}_j J_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x})R(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}), \tag{6.40}$$

with $R_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x})$ is the so called, *evolutionary force* [319]:

$$R_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) = \log\frac{W_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x})}{W_i(\mathbf{x},\tilde{\mathbf{x}}^{\mathbf{j}})}. \tag{6.41}$$

$\dot{S}_{prod}$ is the entropy production while $\dot{S}_{flux}$ is the entropy flux. It is easy to note that $\dot{S}_{prod}$ is the Kullback-Leibler divergence of the current between two miscrostates and the reversed one, and hence it is strictly related to the detailed balance condition. As a consequence it is always non-negative: if $\dot{S}_{prod} = 0$ detailed balance is respected, while for $\dot{S}_{prod} > 0$ it is broken. Hence, it represents the measure of the systems irreversibility [433, 89].
On the other hand, the entropy flux, $\dot{S}_{flux}$, measures the strength of evolutionary force, i.e. log ratio of forward and backward transition probability between couples microstates. If $R_i \neq 0$, one of the two transitions is probabilistically favored, and hence a current is generated. In particular, note that the evolutionary force can be decomposed into two terms :

$$R_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) = R_{i,f}(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) + R_{i,\beta}(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}), \tag{6.42}$$

$$R_{i,f}(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) = \log\frac{f_i(\widetilde{\boldsymbol{x}}^{j})}{f_i(\boldsymbol{x})}, \qquad R_{i,\beta}(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) = \log\frac{\beta(\tilde{x}_j - x_i)}{\beta(\tilde{x}_j - x_i)}, \tag{6.43}$$

where $R_{i,f}$ is the microstates difference in log fitness, while $R_{i,\beta}$ is the log ration of the probability of mutating from one state to the other.
$\dot{S}_{prod}$ and $\dot{S}_{flux}$ have been extensively studied in the context of non-equilibrium statistical physics and stochastic thermodynamics . In such a context, the Master equation models the dynamics of a "thermodynamic system" (i.e. a Brownian particle, a molecular motor or a chemical reaction network) in contact with an environment, typically a thermal bath ore many of them [89]. In non-equilibrium regimes, these mathematical quantities have been interpreted as thermodynamic dissipation and work. In a different context, these quantities cannot, in principle, be directly and generally linked to thermodynamics, and need to be interpreted within the context of information theory. This should not be understood as limitation, but, on the contrary, as a more general structure under which both thermodynamics and evolution are funded.
To shed light on the evolutionary meaning of irreversibility, in Sec. C.4 we perform a mean-field/coarse graining calculation of the entropy production in the limit of $N \to \infty$. Note that the mean-field limit of the entropy production is always smaller than the original one [181].
As a result, by applying the mean-field approximation in the $N \to \infty$ limit to the microscopic entropy production we find a macroscopic decomposition in terms of the time

derivative of the Shannon entropy and the entropy flux :

$$\dot{S}^{micro}_{prod} \xrightarrow[N\to\infty]{} \dot{S}_{prod} = \dot{S}_{flux} + \dot{S} \geq 0. \tag{6.44}$$

The microscopic change in Shannon entropy converges to the macroscopic one obtained by using GCK Eq, see Eq.(6.20):

$$\dot{S}^{micro} \xrightarrow[N\to\infty]{} \dot{S} = \dot{S}_{sel} + \dot{S}_{mut} + \dot{S}_{type} \tag{6.45}$$

where

$$\dot{S}_{sel} = \Sigma(f,s), \qquad \dot{S}_{mut} = \overline{\frac{v^2}{\sigma^2 f}} \geq 0, \qquad \dot{S}_{type} = \overline{v\partial_x \log f} \tag{6.46}$$

$\dot{S}_{sel}$ and $\dot{S}_{mut}$ represent the effect of selection and mutation on entropy, respectively . On the other hand, $\dot{S}_{type}$ represents the effect of mutation on the log of the fitness function, and its sign corresponds the *average* "type" of mutation ($\dot{S}_{type} > 0$ beneficial, $\dot{S}_{type} < 0$ deleterious, see Sec.6.3 for more details).
In the $N \to \infty$ limit the micrscopic entropy flux, $\dot{S}^{micro}_{flux}$, is composed just by the mutation contribution of the evolutionary force Eq.(6.41) (see appendix B.3.5 for more details):

$$\dot{S}^{micro}_{flux} \approx \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\int d\mathbf{x}d\tilde{x}_j J_i(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x})R_{i,\beta}(\tilde{\mathbf{x}}^{\mathbf{j}},\mathbf{x}) \xrightarrow[N\to\infty]{} \dot{S}_{flux}, \tag{6.47}$$

where

$$\dot{S}_{flux} = \dot{S}_{div} = \frac{1}{2\sigma^2}\left(d_t\Sigma + (\Sigma-\sigma^2)\bar{f}\right). \tag{6.48}$$

$\dot{S}_{div}$ represents the contribution of diversity to irreversibility: the first term in Eq.(6.48) is proportional to the change in time of trait variance $d_t\Sigma$, while he second term is proportional to $(\Sigma - \sigma^2)$, the non-mutational part of the variance. Furthermore, both terms are directly proportional to the mean fitness $\bar{f}$ and inversely proportional to the mutation amplitude σ^2. Hence, on the contrary of the components of $\dot{S}$, $\dot{S}_{div}$ it is the sum of the influence of different forces, making its interpretation more involved. To clarify the nature of $\dot{S}_{div}$ term, let us remark that it derives from microscopic contribution of the evolutionary force accounting for asymmetry in the mutation probabilities. As an example, consider the typical process depicted in Fig.6.1 with bold arrows: an individual with trait $\tilde{x}_j$ dies and is removed from the population, while another individual x_i produces an offspring with trait x_j, obtained from the father trait plus a small variation. To observe the reverse process represented with dashed arrows, x_j has to die, while x_i has to produce an offspring with the trait that was removed by selection, $\tilde{x}_j$. Given that $\tilde{x}_j$ could have been *any* individual in the population, if the variance of population distribution is big or/and the mutation are small, observing such a process is *extremely* unlikely. Hence, $\dot{div}$ quantifies the fact that individuals events are irreversible because mutation cannot always revert the act of selection.
More generally, while $\dot{S}$ takes into account the irreversibility generate by the dynamics, $\dot{S}_{div}$ is not zero even when evolution "stops", and therefore has to do with a kind of "structural" irreversibility. Indeed, the stationary entropy production

$$\dot{S}^{*}_{prod} = \dot{S}^{*}_{div} = \frac{(\Sigma-\sigma^2)\bar{f}}{2\sigma^2} \geq 0, \tag{6.49}$$

is directly proportional to the variance, i.e. the population diversity, and to the mean fitness, i.e. the typical reproduction rate, and inversely proportional to the mutation amplitude. Therefore, even if the population reaches a stationary distribution, the individual reproduction events are irreversible: the stationary state is maintained by a global balance between selection and mutation (see Eq.(6.63)). Selection acts as non-local force favoring the fittest individuals, while the mutation current j^* produces a kind of diffusion in trait space. Summing up, we can interpret the entropy production as stemming by two different terms: $\dot{S}_{prod} = \dot{S} + \dot{S}_{flux} \geq 0$,

- $\dot{S}$ quantifies the irreversibility due to net changes in the population. It is composed by three terms, accounting for the change in entropy given by selection, mutation and the interplay of the two:

$$\dot{S} = \dot{S}_{sel} + \dot{S}_{mut} + \dot{S}_{type} \tag{6.50}$$

  At a stationary state $\dot{S} = 0$, the population converges to a given distribution of fix traits with no net change in entropy.

- On the other hand, $\dot{S}_{flux}$ quantifies the currents running *through* the population. Indeed, at a stationary state, it is the only contribution to the entropy production. It quantifies that for the presence and balance of selection and mutation, traits are constantly removed (selection) and introduced(mutation) in the population, similarly to what happens in a resetting process [508]). In Fig.6.1 B we conceptually illustrate the view of population as an "open system": each trait is interacting constantly with the whole population locally with mutation and non-locally through selection. The balance of these two currents determine the population stationary state.

6.5 Examples

Neutral evolution

The simplest example, and yet surprising one, that we consider is neutral evolution. In this scenario, the fitness function is independent of the trait, $f(x) = f = const$, and hence there is no selection , and hence it is useful to disentangle which non-equilibrium effects are due to mutations and which to selection. The GCK Eq in this case is simply:

$$\partial_t \phi(x,t) = f\sigma^2 \partial_x^2 \phi(x,t); \tag{6.51}$$

assuming an initial condition $\phi(x,0) = \delta(x)$, the exact solution is equivalent to a Wiener process [91, 92]:

$$\phi(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 ft}} e^{-\frac{x^2}{2\sigma^2 ft}}. \tag{6.52}$$

The moments of the distribution are trivially:

$$\bar{x} = 0 \tag{6.53}$$

$$\Sigma_t = \sigma^2 ft; \tag{6.54}$$

and the mutational velocity reads:

$$v(x,t) = -\sigma^2 f \partial_x \log \phi(x,t) = \frac{x}{t}. \tag{6.55}$$

The population entropy at each time is:

$$S(t) = \frac{1}{2}\log(2\pi\Sigma_t) + \frac{1}{2}, \tag{6.56}$$

and hence grows *logarithmically* in time. During neutral evolution the population entropy grows making the population explore and occupy all the accesible the trait space.z Given that the fitness does not depend on the trait, the only non-zero term in the change in Shannon entropy. Eq.(6.20) is the mutational one:

$$\dot{S} = \dot{S}_{mut} = \frac{\overline{v^2}}{f\sigma^2} \geq 0, \tag{6.57}$$

by using the explicit form of the velocity, Eq.(6.55), and of the variance, Eq.(6.53) it reads:

$$\dot{S}_{mut} = \frac{\Sigma(t)}{\sigma^2 f t^2} = \frac{1}{t}. \tag{6.58}$$

Hence, the change in Shannon entropy decreases with time as $1/t$. The the entropy flux, Eq.(6.48), can also be calculated using the exact solution of the variance, Eq.(6.53) :

$$\dot{S}_{flux} = \dot{S}_{div} = \frac{d_t \Sigma}{2\sigma^2}\bar{f} + \frac{\Sigma - \sigma^2}{2\sigma^2}\bar{f} = \frac{f}{2} + \frac{f(ft-1)}{2} = \frac{f^2 t}{2}, \tag{6.59}$$

and it grows linearly with time. Hence, the entropy production can calculated by joying Eq.(6.58) and Eq,(6.59)

$$\dot{S}_{prod} = \dot{S} + \dot{S}_{div} = \frac{1}{t} + \frac{f^2 t}{2} \tag{6.60}$$

Given that $\dot{S}$ decaying hyperbolically in time, while the entropy flux, grows linearly with time, here exists a critical time $t^* = \frac{\sqrt{t}}{f}$ where for $t < t^*$ the first contribution dominates, while later the second one is leading. Even more interestingly, there is no dependence on the mutation strength but only on the fitness rate, probably a consequence of the fact that the dimension of mutation is trait and not time i.e. $[\sigma] = [x]$. In Fig. 6.2, we plot the total entropy production Eq.(6.36) (red line) and its mutational component (orange line), confirming the analytical results just discussed.

Stabilizing selection

In the following example, we consider a general fitness emerging from pairwise interactions, and expand around its maximum with a Gaussian approximation of the trait distribution, as done in Sec. 3.3:

$$\begin{aligned} f(x) &= \int_{\mathcal{P}} dy f(x,y)\phi(y,t) = f(\bar{x}) + f_1^x(\bar{x})(x-\bar{x}) \\ &+ f_2^x(\bar{x})\frac{(x-\bar{x})^2}{2}, \end{aligned} \tag{6.61}$$

$$\bar{f} = f(\bar{x}) + \frac{f_2(\bar{x})}{2}\Sigma. \tag{6.62}$$

As shown in Sec. 3.3, if $f_2^x < 0$ the distribution converges to a Gaussian with mean trait the fitness maximum x^* and variance given by Eq. 3.31. At the stationary state, a mutation-selection balance is established, and hence there is a local mutational current in the system, balanced by the *non-local* selection term:

$$(f - \bar{f}^*)\phi^*(x) = \partial_x j^*(x) \tag{6.63}$$

$$j^*(x) = -\frac{\sigma^2}{\partial_x} f(x)\phi^*(x) \tag{6.64}$$

$$= -\frac{\sigma^2}{2}\Big[f(x^*) + f_2(x^*)(x-x^*)\Big(1 - \frac{(x-x^*)^2}{2\Sigma^*(x)}\Big)\Big]\phi^*(x). \tag{6.65}$$

In Fig. 6.3 A we plot the stationary mutational current, Eq.(6.64) (orange curve), the stationary relative fitness $(f(x) - \bar{f})$ (dark blue curve) together with the stationary distribution (black curve) in the concrete scenario of the model introduced in Sec. 3.5. From Eq.(6.63) it is clear that in the region with positive relative fitness there is an exit of probability for the mutational current ($\partial_x j^* > 0$, central region of the distribution), while in the trait with $f < \bar{f}$ there is an entrance of probability ($\partial_x j^* < 0$, tails of the distribution). These two regions are separated by traits with the typical population fitness, $f = \bar{f}$, that have zero current (i.e. the derivative of the current is zero and $f = \bar{f}$).
By using the stationary solution for the trait variance, and assuming for simplicity $f_2 \approx f_2^x$, we can calculate all the components of the stationary entropy production:

$$\dot{S}^*_{prod} = \dot{S}^*_{sel} + \dot{S}^*_{mut} + \dot{S}^*_{type} + \dot{S}^*_{div} = \dot{S}^*_{div} \tag{6.66}$$

$$\tag{6.67}$$

Starting from the selection contribution, it easy to see that it is negative

$$\dot{S}^*_{sel} = \Sigma(f,s) = \frac{f_2^*(x^*)}{2}\Sigma^* \leq 0, \tag{6.68}$$

implying that stabilizing selection tends to reduce the population entropy. The mutation contribution is positive, balancing the selection contribution:

$$\dot{S}^*_{mut} = \frac{\overline{v^2}}{f\sigma^2} \geq 0. \tag{6.69}$$

On the other hand, the contribution of interplay between selection and mutation is negative, as always

$$\dot{S}^*_{type} = -\frac{\sigma^2}{2} f_2^x(x^*)^2 \int dx \frac{(x-x^*)^2}{f(x)}\phi(x) + \frac{\sigma^2}{2} f_2(x^*) \tag{6.70}$$

Finally, the entropy flux reads:

$$\dot{S}_{div} = \bar{f}\frac{\Sigma^* - \sigma^2}{2\sigma^2} \approx \frac{f(x^*)}{\sigma} \geq 0, \tag{6.71}$$

In Fig. 6.2 B, we report the dynamics of the total entropy production (red curve), and of the components of the Shannon entropy change (color lines in the inset plot) in the specific scenario of the model. Note that the entropy production grows till saturating to a stationary value corresponding to $\dot{S}^*_{div}$. Similarly to the neutral case, $\dot{S}^*_{mut}$ decreases with time as $\approx 1/t$ until reaching a tiny positive value. The selection contribution instead decreases with time to a stationary *negative* value, similarly to the $\dot{S}^*_{type}$ contribution. Note that $\dot{S}^*_{type}$ is negligible with respect to the other contributions and is reported alone in the inset plot (green line). Its negative value indicates that mutations are in average deleterious. This can be seen again in Fig. 6.3 A: the dominant central part of the trait distribution has $(f - \bar{f}) > 0$, indicating that, to balance this tendency, the mutation current tends to exit this region, lowering the fitness it average fitness.

Hence, in the case of stabilizing selection, mutation increases the entropy, while selection tends to decrease it.

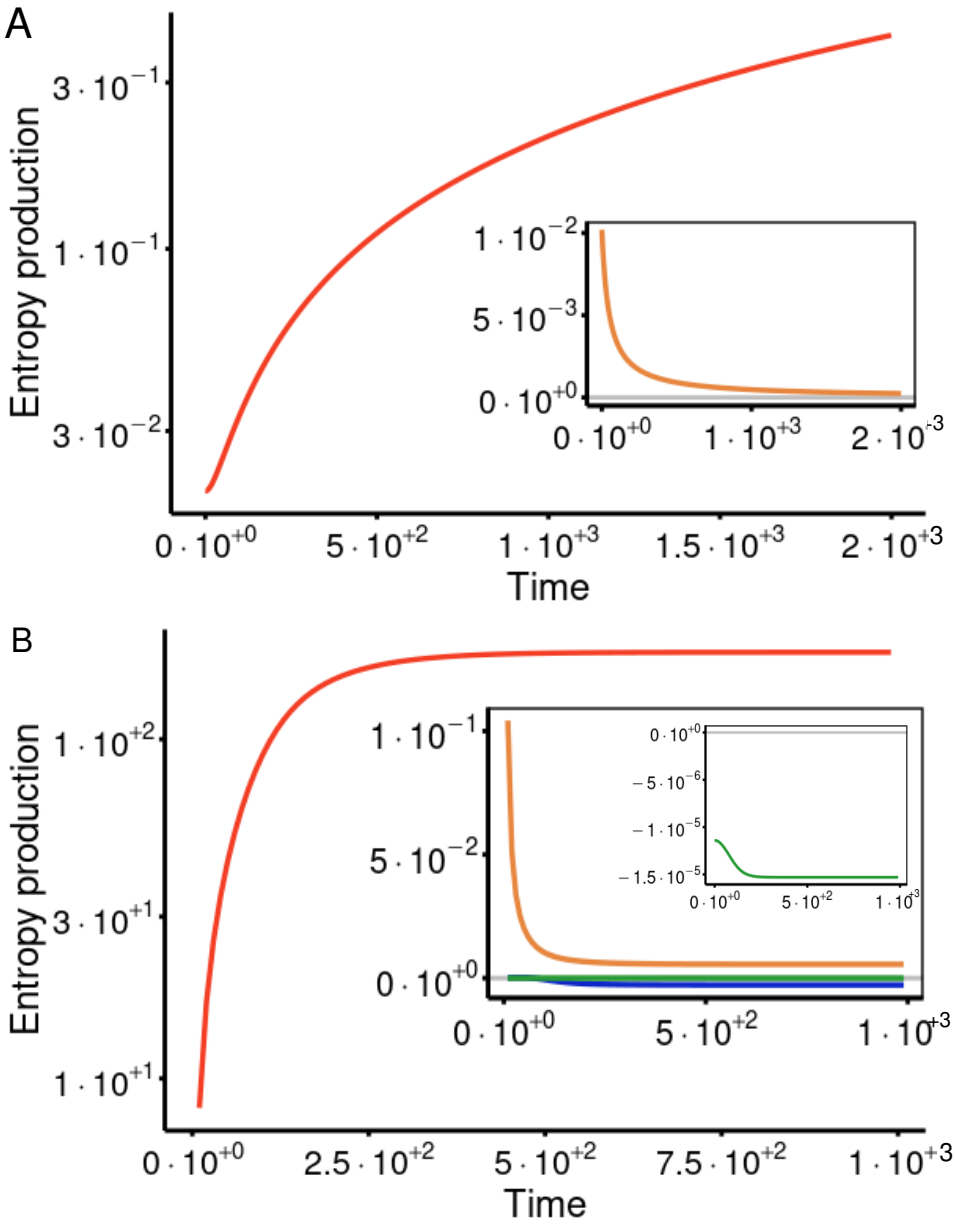


FIGURE 6.2: **Entropy production for the neutral (A) and stabilizing selection (B) cases.** The red line represent the total entropy production as a function of time, that is dominated by $\dot{S}_{div}$. In each panel, in the inset we represent the components of the Shannon entropy change: $\dot{S}_{mut}$ (orange),$\dot{S}_{sel}$ (blue) and $\dot{S}_{type}$ (green). A) Note that the total entropy production increases indefinitely with time in the neutral case, while the mutational and type contributions decay with time to zero. B) In the stabilizing selection case we considered as a specific example the model described in Sec.3.5. In this case, the total entropy production saturates to the stationary positive value given by $\dot{S}_{div}$. The mutational contribution behaves similarly to the neutral case, while the type and selection contributions decay to a negative stationary value. The $\dot{S}_{type}$ contribution is shown also in inset, and is extremely small but negative. *Parmeters:* $A) f = 10^{-2}, \sigma = 5 \cdot 10^{-3}; B) \sigma_K = 1, \sigma_K = 1, \sigma_\alpha = 1.2, k = 0.1, \sigma = 10^{-2}$.

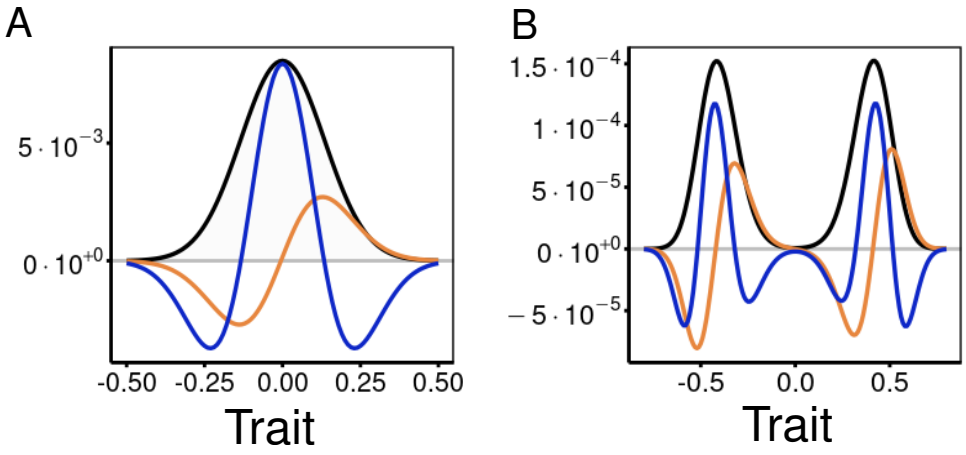


FIGURE 6.3: **Mutational current and relative fitness for stabilizing selection and evolutionary branching** In both A) and B) we plot the stationary mutational current Eq.6.64 (orange curve), the relative fitness $(f(x) - \bar{f})$ (dark blue curve) together with the stationary trait distribution (light blue), in the case of the model exposed in Sec.3.5. A) In the stabilizing selection scenario the population converges to a normal distribution with peak at zero. The mutational current (orange line) is negative for negative traits and positive for positive ones, with an extreme point in each region. Its derivative determines the mutational flux in the system and is equal to the opposite of the weighted relative fitness (dark blue lines). Hence, in the region between the minimum and the maximum of the current the relative fitness is positive and the mutational current is exiting from it ($\partial_x j > 0$). On the other hand, in the tail of the distribution the relative fitness is negative and balanced by an influx of probability due to mutations. B) In the case of evolutionary branching a similar structure of probability fluxes emerges around each of the two peaks. Around the peaks, relative fitness is positive and mutations out-flowing, while in the tails and in the separation region the relative fitness is negative and mutations in-flowing. *Parameters*: A) $\sigma_K = 1, \sigma_K = 1, \sigma_\alpha = 1.2, k = 0.1, \sigma = 10^{-2}$;B)$\sigma_K = 1., \sigma_\alpha = 0.8, k = 0.1, \sigma = 10^{-3}$.

Evolutionary branching

In this section, we analyze the entropy production in the scenarios of one or more evolutionary branching. As disxcussed in Sec.3.3 evolutionary branching happens in general when $f_2^x < 0$, i.e. the population is attracted by a fitness minimum, and once there it separates in two sub-populations. Sadly, no general analytic expression of the stationary distribution is available (except the vanishing mutation limit). Hence, to study the irreversibility of this phase we need to rely only on numerical integration of the growth-competition model (see Sec. 3.5). As a start, in Fig. 6.3 B) we have plotted the mutational current (orange curve), the relative fitness (blue curve) together with trait distribution (light blue curve) in this particular case. Clearly, the system is out of equilibrium because there is a current flowing through it, even if balanced by the non-local action of selection. Similarly to what happens in the stabilizing selection case (the A panel), around each peak there is non-vanishing mutation current, that decreases to zero in the separation space between them.
In Fig. 6.4 we present the numerical results for the total entropy production (A), and the contributions to the entropy change in two different temporal moments(B,C). Note that the entropy production saturates very rapidly when the population branches, at $t \sim 750$, to the very high value given by $\dot{S}_{div}$. On the other hand, the components of the entropy change present a very interesting non monotonic behavior: at the branching point the selection entropy production increases till reaching a very sharp positive maximum, while the mutational one decrease to a (positive) minimum (blue and orange curves respectively in Fig.6.4 B). After the branching, the selection contribution decreases to a negative minimum, to finally increase again to a very small stationary negative value (blue curve in Fig. 6.4B,C). Interestingly, we deduce that when $\dot{S}_{sel} > 0$ selection is disruptive, while when $\dot{S}_{sel} < 0$ it is stabilizing. On the other hand, the mutation entropy presents a maximum in correspondence to the selection minimum, decreasing then to a tiny positive value (orange curve). Finally, in the inset of Fig. 6.4 A, we separately plot the $\dot{S}_{type}$ contribution, given that it extremely smaller than the other ones (green line). Before the branching point $\dot{S}_{type}$ is positive and increasing, indicating that the mutations in average tend to increase the fitness. Nevertheless, after the branching point, it tuns slightly negative, similarly to the stabilizing selection case. These results reveal that irreversibility increases after branching, and that the various contributions of the entropy production clearly correlate with the different evolutionary regimes and transitions.
Thanks to the notable richness of evolutionary phases of the considered model, we can also study what happens to the entropy production when further branching event occur. In this scenario a first branching happens, leading to two sub-populations, and then each of them branches again producing four sub-populations. Later, the two central subpopulations coalesce producing a stationary trait distribution with three peaks (see Sec. 3.5 for more details). In Fig.6.5 we plot the entropy production (A) and the entropy flux component (B,C) for this double series of branching case. Note that the entropy production grows rapidly after the first branching reaching a plateau, and then increases even more at the second branching. In B and C we plot the component of the entropy change before and after the second branching point. After the first branching the selection contribution has, as explained before, a deep negative minimum (B). Then, it increases rapidly reaching a second positive maximum corresponding to the second branching point and finally relaxes to a negative value (C). The mutational contribution, after experiencing a positive maximum (B), relaxes to stationary positive value (C). In the inset of Fig.6.5 A, we also plot the type contribution $\dot{S}_{type}$, showing that similarly to precedent case, at each branching point it decays very rapidly, first to a positive value and finally to a negative one. Hence, the entropy production increases in during a branching event, and

the entropy change contributions present a highly non-trivial behavior showing maxima or minima in correspondence to these events, similarly to what has been observed near phase transitions [509, 510, 188] and bifurcations [185].

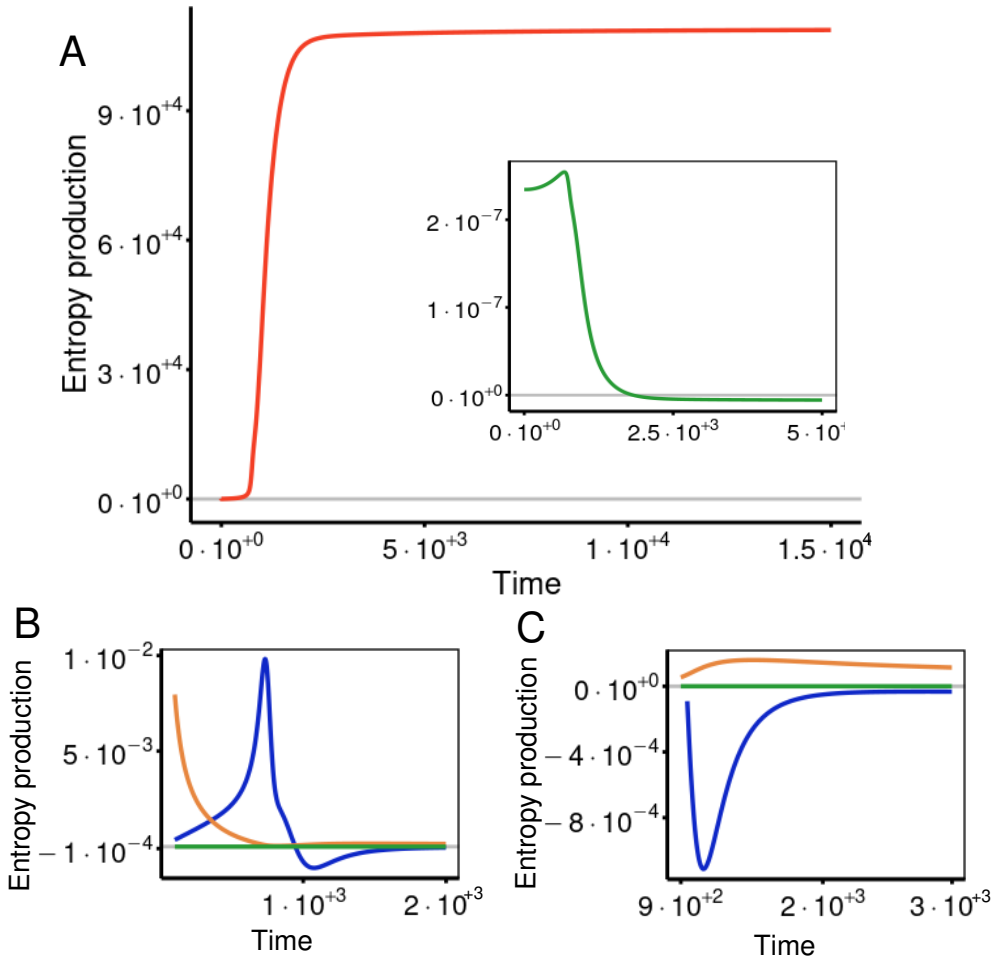
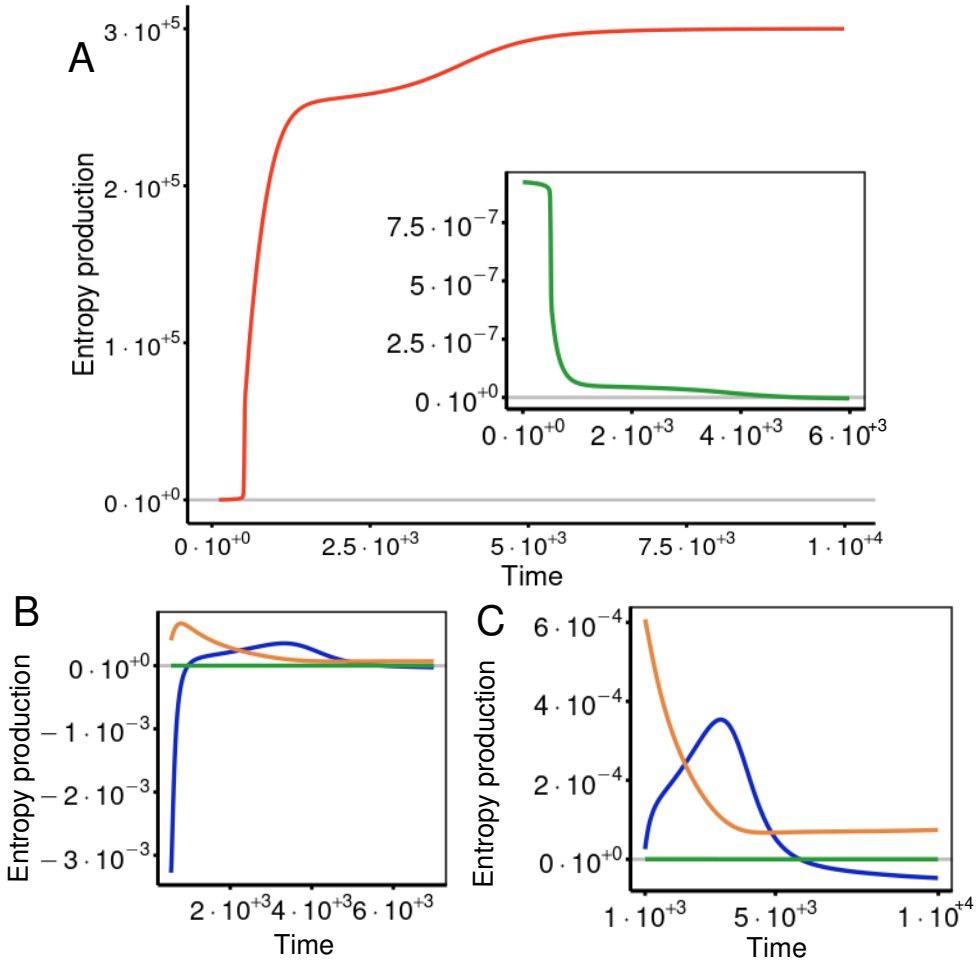


FIGURE 6.4: **Entropy production in the evolutionary branching phase** In A) we plot the entropy production as a function of time (red line), while in B and C we report the component of the entropy change in two different time shots nearby the branching point. The colors represent the different terms in the entropy production: $\dot{S}_{mut}$ (orange),$\dot{S}_{sel}$ (blue) and $\dot{S}_{type}$ (green). Note that in A the entropy production increases rapidly at the branching point,$t \sim 700$, and then reaches a stationary value correspondent to $\dot{S}_{div}$. On the other hand, in B we can appreciate the non-monotonic behavior of $\dot{S}_{mut}$ and $\dot{S}_{sel}$: while the first has positive minimum at the branching point, the second shows there a positive maximum with a cuspid form. Notably, the C panels show that after the branching the mutational contribution increases till a maximum and than relaxes to stationary value, while the selection one reaches a deep negative minimum and than increases to a stationary negative value. Finally, in the inset of the A panel we plot the $\dot{S}_{type}$ contribution, showing that at the branching point it decays from positive not negative values. *Parameters*:$\sigma_K = 1., \sigma_\alpha = 0.8, k = 0.1, \sigma = 10^{-3}$.



$Parameters: \sigma_K = 1., \sigma_\alpha = 0.6, k = 0.1, \sigma = 10^{-3}$.

$Parameters: \sigma_K = 1., \sigma_\alpha = 0.6, k = 0.1, \sigma = 10^{-3}$.

FIGURE 6.5: **Entropy production in the multiple (two) evolutionary branching phase** In A) we plot the entropy production as a function of time, while in B and C we report the component of the entropy flux in two different time shots nearby the second branching point. The colors represent the different terms in the entropy production: $\dot{S}_{mut}$ (orange),$\dot{S}_{sel}$ (blue) and $\dot{S}_{type}$ (green). Note that in A the entropy production increases rapidly at the branching point,$t \sim 600$, reaching a first plateau. At the second branching point it increases again saturating finally to $\dot{S}_{div}$. On the other hand in B and C we can appreciate the non-monotonic behavior of $\dot{S}_{sel}$ (blue) and $\dot{S}_{mut}$ (orange). After the first branching, the selection term increases from a deep negative minimum to positive maximum corresponding to the second branching event (B and C). At later times, it relaxes to a negative stationary value (C). On the other hand, the mutational contribution after a positive maximum (B) decreases to a positive minimum and then converges to a stationary value. Finally, in the inset of the A panel we plot the $\dot{S}_{type}$ contribution, showing that at the first branching point it decays to a positive plateau, and later during second branching to tiny negative values.
$Parameters: \sigma_K = 1., \sigma_\alpha = 0.6, k = 0.1, \sigma = 10^{-3}$.

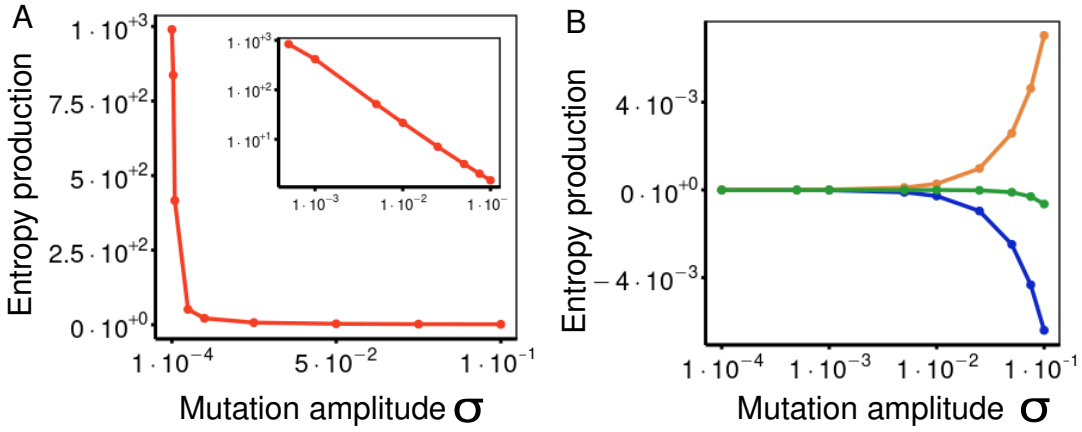


FIGURE 6.6: **Dependence of entropy production on the mutation amplitude σ in the case of stabilizing selection**. A the stationary entropy production $\dot{S}_{div}$ as a function of σ in linear scale in the main plot and log-log one in the inset. The stationary entropy decays as σ^{-1}, producing a kind of trade-off dependency: for high mutations the population is weakly irreversible, but by decreasing σ the entropy production increases very rapidly. B) Change in entropy components as a function of σ : $\dot{S}_{sel}$ (blue curve),$\dot{S}_{type}$ (orange curve) and $\dot{S}_{type}$ (green curve). Note that the mutation component increases with σ while the type and selection one decrease with it. *Parameters*:$\sigma_K = 1, \sigma_K = 1, \sigma_\alpha = 1.2, k = 0.1$

6.6 Conclusions and discussion

In this chapter we have studied the irreversible properties of evolution, both in a general fashion and in various scenarios such as neutral evolution, stabilizing selection and evolutionary branching. We have found that evolution is in general out of equilibrium due to the contemporary presence of a mutation current and selection differences. Even if these forces balance at a stationary state, their different effects on the population drive the system away from equilibrium. Furthermore, we have found the different contributions to the entropy production correlate with the evolutionary regimes and evolutionary transitions that the population is experiencing. Even if our results are promising, we feel to have just scratched the surface of this interdisciplinary subject. Many possible avenues open in front of use. Let us introduce some of them starting from the open question emerging from our results. First, it would be interesting to prove in general that the stationary covariance between the surprise rate s and the fitness is negative or zero $\dot{S}^*_{sel} = \Sigma(f, s) \leq 0$.

Stochastic generalizations A possible future direction would be to study the effects of finite size drift on evolutionary irrversibiliy. Stochastic effects are fundamental and very well studied in evolution. In Sec. 3.6 we have generalized the framework to accommodate for stochastic fluctuations. Yet, this possibility make us face a dilemma in trying to calculate their effects on irreversibility. Following our derivation we could use the SGCK Eq. to derive the fluctuations of Shannon entropy of the trait density:

$$S = -\int dx \rho(x) \log \rho(x) \tag{6.72}$$

$$\dot{S} = \partial_t \langle s \rangle = \Sigma(f, s) + \overline{v \partial_x s} + \int dx \sqrt{\frac{(f(x) + \bar{f})\rho(x) + \sigma^2/2 \partial_x f(x) \rho(x)}{N}} s(x) \xi(x, t) \tag{6.73}$$

If we approximate the fitness function to first and neglect the mutational contribution in the stochastic coefficient (as done in appendix Sec.3.6), the Langevin Eq. (6.73)can be simplified as:

$$\dot{S} = \partial_t \langle s \rangle = \Sigma(f, s) + \overline{v \partial_x s} + \sqrt{\frac{2 f(\bar{x}) \overline{s^2}}{N}} \eta(t), \tag{6.74}$$

revealing the presence of multiplicative fluctuations in the entropy. This fluctuating evolution of the entropy could be used to search for universal properties of the entropy production, as done in stochastic thermodynamics [451], and to derive uncertainty relations and bound for the entropy production such as the "Time-information uncertainty relation" [511, 512, 513].
another possibility would to consider the Onsager-Machlup action corresponding to the SGCK [187]:

$$\mathcal{S} = -\frac{N}{2} \int_0^T dt \int dx \frac{(\dot{\rho} - (f - \bar{f})\rho + \partial_x j)^2}{(f + \bar{f})\rho - \partial_x j} \tag{6.75}$$

and to study its time asymmetric part as a quantification of irreversibility:

$$\mathcal{S}^- = N \int_0^T dt \int dx \frac{\dot{\rho}((f - \bar{f})\rho - \partial_x j)}{(f + \bar{f})\rho - \partial_x j}. \tag{6.76}$$

This approach is the one taken by Mustonen and Laessig [184]. It is evident the quantities described in Eq.(6.74) and Eq.(6.76) are radically different. While the first in the deterministic limit converges to the mean-field limit of the microscopic change in entropy, the second, as shown in [184], recovers Fisher-Theorem, or more in general is related with the Jeffreys divergence. A possible direction of research would be to try to relate Eq.(6.74) with the change in the population microstates, hence clarifying its physical meaning. At the same time, Eq.(6.76) can be seen as the macroscopic approximation of the irreversibility of a Master equation in the space of trait frequencies, as explained in appendix B Sec.3.6. In future works we will explore in detail this dilemma and understand the differences in these two measures of irreversibility.

Irreversibility vs mutation trade-off and the error catastrophe While the components of the change in Shannon entropy are clear expressions of the different forces acting during evolution, the dominant stationary entropy production is the combinations of different quantities:

$$\dot{S}^*_{div} = (\Sigma - \sigma^2)\frac{\bar{f}}{2\sigma^2} = [diversity] \times \frac{[mean\ fitness]}{2[mutation\ strength]}, \tag{6.77}$$

making its interpretation not straightforward. In Sec. 6.4 we have shown that it emerges from the continuous balance between the local mutational current and the non-local effect of selection, or, at a microscopic level it is the consequence of the incessant generation and removal of individuals. In Fig. 6.6 A we plot $\dot{S}^*_{div}$ as a function of σ in the concrete example of stabilizing selection. Figure 6.6 A shows that its dependence on σ creates a kind of trade-off: for large mutations the system is weakly irreversible, while for small mutations irreversibility is extremely large. This behavior give as a clue on the interpretation of $\dot{S}^*_{div}$ as a "population efficiency". For high mutations the trait distribution is very wide, many sub-fit traits are maintained by mutations. Even if the average fitness is not high, the

population can explore easily vast regions of trait space. On the contrary, for small mutations the distribution is very peaked around the fittest individual, and hence efficiency is very high, but diversity is low. Selection is very strong, causing high irreversibility. This trade-off is analogous to phenomenon known as the Eigen error catastrophe [514, 493], where mutation rate cannot overcome a critical vale to keep the population "efficient", with more a majority of fittest individuals. In future works we would like to explore more in detail this interpretation and understand how different organisms, such as bacteria, viruses or eukaryotes, react to this trade-off. It would be very intriguing to study the irreversbility in models based in population genetics, such as fitness waves [169, 160, 275] or neutral networks [515]. In genetic system it would be possible to relate evolutionary information, i.e. the genetic patterns maintained by selection, with irreversibility and complexity. Finally, it would be interesting to consider systems of molecular replicators where fitness could be related to thermodynamic or to energetic properties [516, 517, 518, 519]. Is evolutionary irreversibility related to thermodynamic dissipation ?

Chapter 7

General conclusions an perspectives

In this thesis we have studied collective phenomena in microbial ecology and evolution, using the theory of (non-equilibrium) statistical physics. In this section, we derive some general conclusions and comment on possible upcoming developments. Before flying to future perspectives let us briefly summarize the results reported in this thesis:

- In Chapter 2, we have revealed the emergence of a new macroecological law in microbial communities. This empirical law states in quantitative terms that the average pairwise correlation function decays from positive to null values as the phylogenetic distance increases, approximately following a stretched-exponential decay function.
  By defining a stochastic ecological model (the CSLM), we have shown that coupled environmental (*multiplicative*) fluctuations, also known as environmental filtering, are responsible for such universal pattern. By connecting ecological preferences with species phylogentic distance, we can formulate a model where phylogenetic trees can be directly used. Last but not least, we analyze temporal data for a fixed community, showing that the macroecological law also holds quantitatively in this context and that delayed temporal correlations are naturally reproduced by our model with environmental filtering. Therefore, the results reported in this Chapter allow us to conclude that only *environmental filtering* (and not, e.g., species competition) explains the empirically-observed pattern of decaying correlations with phylogenetic distance, and hence is probably the dominant ecological force in natural microbial communities.

- In Chapter 3, we introduced an eco-evolutionary framework for bacterial traits. In particular, we descrobed adaptive evolution as the dynamics of the population trait distribution in phenotypic spaces, synthetized by the socalled Generalized Crow-Kimura equation (GCK). By modelling explicitly selection, mutations and finite-size drift we generalize the celebrated theory of adaptive dynamics to eco-evolutionary scenarios in. With our formalism, we can exactly describe the population and trait distribution also after an evolutionary branching, and in particular, predict the peak of the two subp-populations. Finally, we generalized our description to include stochastic finite-size effects, revealing the possible frustration of evolutionary branching. The frustration is due to the peculiar multiplicative fluctuations of the trait variance, that can create a metastable or an absorbing state in the dynamics. Hence, the work presented in this Chapter contributes to the development of an eco-evolutionary theory for microbial communities, allowing to shed further light on the empirically-observed astonishing diversity in traits and interactions of microbial communities. Our hope is that the present work makes this kind of quantitative approaches to complex eco-evolutionary communities accessible to a broader audience, including physicists, biologists, and ecologists.

- In Chapter 4 we applied the aforementioned eco-evolutionary framework to the emergence of antibiotic tolerance by lag- time adaptation in bacteria. Recent laboratory experiments on E. Coli have shown that bacteria adapt their lag time to survive to antibiotics in a dormant state [10]. The resulting lag time population distributions show a direct proportionality between the mean lag time and its variance. We identified this characteristic as typical of multiplicative processes, and designed an evolutionary model where the offspring lag variation is proportional to the mother cell value. The model can effectively be described by the GCK equation with a new "*multiplicative* variations term", and it is able to reproduce the empirical lag distributions. Hence, the work presented in this Chapter shed light on the evolution of antibiotic tolerance, and furthermore it represent an example of a general class of evolutionary problems where our framework can be applied.
- The last part of the thesis was devoted to non-equilibrium statistical physics, and, in particular, in Chapter 5 we illustrated the general geometrical and topological properties of velocity currents in NESS. These geometrical properties can be explained as symmetry breaking induced by irreversibility and dissipation. While in dimension equal or large than two, stochastic system with additive noise generate divergence free current velocities, the addition of *multiplicative* noise induces more complicated geometrical situations. In particular, in this case gradient-like velocities push the system to regions of the phase space with low fluctuation, such as absorbing states. Remarkably, thanks to this geometric framework, we can give a clear thermodynamic interpretation of the excess entropy and relate it with Prigogine principle, solving a debated issue with a long-history.
- To conclude, in Chapter 6 we studied the irreversible nature of evolution, using the framework introduced in 3 both in a general fashion and in various scenarios concrete such as neutral evolution, stabilizing selection and evolutionary branching. We found that evolution is in general out of equilibrium due to the contemporary presence of a mutation current and selection differences. Even if these forces balance at a stationary state, their different effects on the population drive the system away from equilibrium. Furthermore, we found the various contributions to the entropy production correlate with the evolutionary regimes and evolutionary transitions that the population is experiencing. In the discussion, Sec.3.7, we commented on a possible stochastic generalization of the Shannon entropy change, and in particular we remarked its *multiplicative* nature. Hence, the work presented in this Chapter shed light onto the fundamental aspect of time in adaptation, and reveal the role of mutations and selection in creating irreversibility during Darwinian evolution.

We have emphasized the important role of multiplicative fluctuations that indeed are present in *all* the Chapters of this thesis. Now, let us conclude with some general perspectives and possible future directions of research.

Microbial ecology: a problem with many scales of diversity

In the article "Problems in Physics with many Scales of Length", Kenneth G. Wilson introduced the Renormalization group as tool for studying phenomena generate by the co-presence of many scales [239]. The paradigmatic example is a critical point in a continuous phase transition, where scale invariant fluctuations are generated [9].
By applying a coarse-graining procedure, the renormalization group gives the possibility of exploring physical phenomena at different scales and individuating scale-free phenomena.

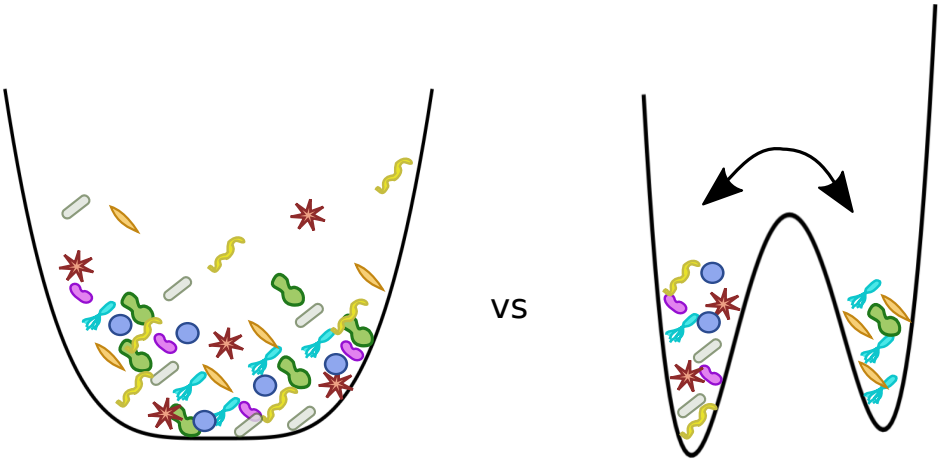


FIGURE 7.1: **Ecological attractor**. Two possible ecological attractors are depicted in a pictorial way. On the left, a single noisy attractor while on the right a multiple attractors scenario.

As we have discussed in the introduction and in Chapter 2, microbial ecology is a problem where various diversity scales exist. Different ecological forces manifest themselves in each scale, and depending on the particular question, an operational scale must be chosen. On the other hand, some other patterns seem to be conserved across scales. To clarify this issue, phylogenetic coarse-graining methods have been developed both in theoretical models [241, 520], and in empirical datsets [124]. In particular, the Gamma AFD is conserved both at the finer scale of strains [126], and in a continuous coarse-graining to higher taxa[124].
On the other hand, in Sec.2.2 we have shown that the phylogenetic signal of correlations is lost at the coarse-grained scale of phyla. In particular, Fig. 2.2 illustrates that by considering just inter-phyla correlations one cannot observe the stretched exponential decay, that is determined by intra-phyla OTU pairs. Analogously, by extending our analyses to finer phylogenetic resolutions it could be possible to reveal the nature of intra-specific interactions, eventually elucidating the emergence of competition as a key player in determining correlations [238, 126].
Furthermore, the oscillating decay of the correlation variance, see Fig.B.5, suggests that also the residual correlations, i.e. the ones not captured by the stretched exponential decay, show a kind of "heterogenous" scale invariance, typical of multifractals [521].
Hence, we propose to analyze the species correlation pattern across diversity scales in future works. As a first step, we will repeat the correlation vs phylogeny analysis at the finer level of ASV, and then perform a coarse-graining in sequence space, similarly to the Kadanoff renormalization group [240, 124]
Alternatively to the taxonomic coarse-graining, we could use a correlation coarse-graining method, recently developed in the field of neuroscience, inspired by the combination of the renonrmalization group with principal component analysis [522, 523, 524, 525].
By studying the correlation pattern at different diversity scales, we could understand which are the dominant ecological forces across taxonomic resolutions

The nature of the ecological attractor

Which are the sources of variation in microbial communities? A general, and probably oversimplified, answer distinguishes between stochastic and deterministic effects. In the language of non-equilibrium physics, this distinction can be translated in two different

characteristic of the dynamics attractor, i.e. of the stationary probability distribution.
As a first expectation, the ecological dynamics could converge to a unique globally stable attractor, but continuously perturbed by large stochastic fluctuations (see Fig. 7.1 left). In this case, the variability is a pure consequence of fluctuations, similarly to what happens in models such the (C)SLM.
On the other hand, it could be that, at the stationary state, multiple deterministic community attractors are present, with fluctuations make possible the transitions between them (see Fig. 7.1 right). In this case, variability is given by the transition between these multiple states, similarly to what happens to some class of random interaction models, like the GLVM with demographic fluctuations in the low noise regime [63, 71].
Both the phylogenetic pattern of correlations analyzed in Chapter 2, and the macroeclogical laws discussed in Sec.1.6 emerge equivalently by averaging over different communities or in a single one by averaging over time. This ergodic property, i.e. the equivalence of time and ensamble average, suggests that the correct scenario is the single attractor one, as depicted in Fig.7.1 left, given that ergodicity is generally broken in the presence of many possible attractors, such as in disordered systems [526, 527, 528]. Nonetheless, in laboratory experiments of microbial community assembly, alternative stable states are found [39, 529], suggesting also the validity of the second picture.
Furthermore, an analysis of species abundance dissimilarity in time series of the human gut have revealed a mixed scenario: while some species are stable around a stationary value, some others experience transitions to alternatively states [123].
In synthesis, we can hypothesize that the strong fluctuations present in natural communities tend to destroy alternative states, but, as often i in real phenomena, the truth lies somewhere in between.

The evolution of ecological interactions

In Chapter 3 we introduced an eco-evolutionary framework for bacterial traits. By reconsidering a classic model from adaptive dynamics and speciation [157, 303], we showed how one or multiple evolutionary branchings can happen by the interplay of growth and competition. In particular, the number of niches, or of "species", produced by evolution increases with the strength of competition.
Nevertheless, bacterial populations are subject to many differences ecological forces emerging from resource consumption, like cross-feeding and environmental fluctuations. Hence, in future works we will combine the framework derived in Chapter 3 with the ecological phenomena studied in Chapter 2. In particular, we will study the evolution of ecological interactions in a fluctuating consumer-resource model. As discussed briefly in Sec. 2.4, we have devised a method to derive an effective GLVM from the consumer-resource one. This method will let us perform some analytical predictions on the evolution of metabolic preferences, and hence of ecological interactions , going beyond recent works in the field[144]. In generally used adaptive models such as the one we have analyzed in Sec. 3.5, one can derive a criterion for evolutionary branching in terms of the strength of competition and the size of the available niche. On the other hand, with our method we are able derive these two parameters from the resources entry and exit rate and species metabolic properties. Hence, under which circumstances does branching happen in this scenario?
Furthermore, we hypothesize that after a first evolutionary diversification induced by competition, as the one studied in Sec. 3.5, cross-feeding could evolve to reduce competition even more.
Finally, evolutionary diversification in the presence of environmental fluctuations is a new research direction promising recent results on the emergence of cooperation [315].

The universal strategy of dormancy: from bacteria to plants

In Chapter 4, we studied the evolution of bacterial lag as a strategy to survive in presence of cycles of antibiotics. This strategy is called dormancy, and is found across biological levels in different forms, from bacteria and viruses to plants and animals [530]. In general terms, it consists of entering a "dormant" low- metabolic but highly resistant state to cope with sub-optimal environmental conditions. In large populations, often exists a small subset of dormant individual, or "seed bank", that, even if inactive, store information and impart memory giving rise to complex and multiscale dynamics and interactions involving the population or even full ecosystems. Examples range from extended lag phase in bacteria to tolerate the presence of antibiotics during many hours, dormancy in seeds to prevent germination in dry period of days or months and even decade long metastatic dormancy in tumor cells. While the biological and physiological causes of dormancy vary across systems, its adaptive value could be universal.
The adaptive model introduced in Chapter 4 was defined for bacterial dormancy, but it can easily be extended to more general situations. In particular, it would be relevant to add memory effects to the model, enabling the implementation of non-exponential waiting times distributions. In the original context of evolution of tolerance, it would make possible to model the noisy effects of the toxin-antitoxin circuit, hence generating an individual lag variability. Apart from this concrete case, this modification would be useful for studying the evolution of dormancy under general circumstances, and in particular for analyzing the emergence of bet-hedging strategies [342].

Selection and irreversibility away from equilibrium : a hint for the origin of life ?

In Chapter 5, we discussed how irreversibility is associated with a chiral symmetry breaking in the stationary current velocity of general mesoscopic systems in a NESS. In the simplest case, this symmetry breaking consists in the appearance of vortex-like configurations in the velocity field, rotating clock or anticlockwise. Following a recent results in the thermodynamics of chemical systems [471], one could interpret this symmetry breaking as a "selection" of a particular configuration with respect to the others. In general, one could frame non-equilibrium statistical physics as theory of selection in the trajectory space: if one considers two possible transition between two states, while at equilibrium all directed and inverse transitions are equally probable, in non-equilibrium conditions one (the time directed one) is preferred to the reversed one.
This perspective could be applied to some physical-chemical scenarios of the origin of life, bridging thermodynamics with evolution. In particular, several types of molecules happen to be selected away from equilibrium, like handedness or homochiraly [531], or the preference for furanose against pyranose in nucleid acids chemical components [472]. More generally, the appearance of cycling currents in non-equilibrium systems echoes the chemical hypercycles theorized by Eigen as one of the basic element in proto-cellular organization [490].
To discern real connections between the chemistry of the origin of life and dissipative symmetry breaking, in the future we will extend our analysis to discrete stochastic systems, that comprehend chemical reaction networks.

Irreversibility in evolution: can it be quantified in data ?

Finally, in Chapter 6, we have derived a general formula for estimating the irreversibiliy in adaptive evolution using our eco-evolutionary framework. Interestingly, if one considers an experimental evolving population, and is able the estimate the population trait

diversity, the average fitness and mutation rate, the entropy production rate can be estimated by data. Nevertheless, the majority of high quality evolutionary data are in terms of genetic sequences and not in traits [11], and hence in the future we will derive with the entropy production in population genetics models, even if with the same mean-field procedure. Furthermore, in recent studies the full genetic networks of viral populations are becoming available, together with some genotypic-phenotypic maps [515, 416, 532]. Viral populations have very large size, but also big mutation rates, making them an ideal candidate to study irreversibility.

Finally, in the last section, we draw some personal and philosophical conclusion from this travel across the boundaries of physics and biology.

Epilogue: from individual to collective phenomena

In the book "What is Life ?", Erwin Schrödinger laid the foundations of contemporary biophysics by analyzing biological phenomena with the lenses of physics [533]. His central question is stated in the introduction : *" How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry ?"*.
In the book, Schrödinger goes through different concepts and phenomena such as molecular chaos, Darwinian evolution, the physical origin of mutations, and, finally, the nature of biological inheritance. In discussing this last argument, Schrödinger theorized that the genetic material must be molecule, but a particular one that he define as "aperiodic crystal", able to encode a large quantity of information. Even if some theories of heredity were yet in circulation, Schrödinger's intuition can be considered a theoretical prediction about the properties of DNA, and it stimulated a vigorous interest in researching the molecular basis of genetic. Indeed, Watson and Crick acknowledged Schrödinger as source of inspiration.
In the last, less known, part of the book titled "On determinism and free will", Schrödinger flies to more metaphysical reflection on the nature of consciousness, influenced principally by Schopenhauer and the Indian Upanishad. The incipit of the Chapter is: *"As a reward for the serious trouble I have taken to expound the purely scientific aspects of our problem sine ira et studio, I beg leave to add my own, necessarily subjective, view of the philosophical implications.*
We think the usage of a scientist exposing is philosophical ideas to be a very healthy and necessary exercise. Indeed, the majority of great physicist of the past century, such as Einstein, Bohr, Pauli, Heisenberg, and, of course Schrödinger, wrote extensively about their philosophical. Hence, with humility, we propose here some general reflections.
We are not interested in the nature of consciousnesses, but we share with Schrödinger the necessity of not interpreting the individual as the basis of nature. Indeed, collective ecological relations, lateral genetic transfer and symbiosis in evolution invite us to rethink the concept of individual. Let us guide the reader through some interesting examples illustrate this necessity.
In Sec.1.3.1 we reported briefly that during biological evolution, many endosymbiotic events happened. For example, the eukaryotic mitochiondria are the remnants of proteobacterium captured by, probably, an archea. Viruses had also an important role in shaping life as we see it. In 2001, the International Human Genome Sequencing Consortium established that the 8% of the genome of mammals, and hence of humans, consists of remnants of ancient viruses, while up to the 40% is composed by repetitive strings of genetic letters that is also thought to have a viral origin [534, 17]. The ancient viruses

were endogenous, as the modern HIV, and, to reproduce, had to insert their genetic material into their host's genome. Some of them were able to infect germ cells, like eggs or sperm, and transmitted their genome to future generations [535]. These hybrid part of the genome is often beneficial, and in mice has been found to give protection against similar viruses. Nevertheless, the most surprising fact is that this viral genome is responsible for the for a key cellular process involved in development of the mammalian placenta [536]. T. Heidmann, the group leader of aforementioned research on the placenta, resumed these facts in a provocative way :"Our genes or not only *our* genes. They are also retroviral genes" [17].

Ecological interactions between species of different domains can be so tight to put in crisis our concept of individual. Lichens are composite organism that arise from algae or cyanobacteria living among filaments of multiple fungi species in a symbiotic relationship. The fungi benefit from the nutrients produced by the algae or cyanobacteria via photosynthesis. On the other hand, the algae or cyanobacteria benefit by being protected from the environment by the filaments of the fungi[537]. Symbiosis is also extremely frequent in marine environments, especially as chemosynthetic symbioses between bacteria and invertebrates, where the latter are the primary producers, providing most of the organic carbon needed for the animal host's nutrition [538]. In a recent study [539], it has been reported a three-way symbiosis in a tropical grass plant. This particular plant grows in high soil temperatures thanks, it was thought, to a symbiotic fungus living in its leaves. Instead, the responsibility for heat tolerance is of a virus that has infected the fungus.

Humans have also evolved intimate symbiotic relationships with gut microbes (microbiome), that influence the host's health [540, 541, 56]. Furthermore, recent evidence suggests that the present species are an inheritance of our ancestors, concluding the the gut microbiome is heritable [542].

On a more general level, all plants and animals live in symbiosis with its microbiome, composed by bacteria, archea, fungi and protists. Hence, we are not an individual, but an *ecosystem*, or in a more specific term an *holobiont*. As M. Sheldrake states in his book *Entangled life* [537], microbial communities make us understand that "individual" is just a scientific category with some limits of applicability: *"Biology, the study of living organisms, has transformed into ecology, the study of interactions between living organisms."*

In this perspective, statistical physics and complex systems are useful scientific disciplines to comprehend the new collective world emerging from the relations of an astronomical number of living beings. If we accept that an individual is not an island, a new tangled world appears in front of our eyes.

Appendix A

Stochastic processes

A.1 Itô-Stratonovich dilemma

Stochastic differential equations (SDE)like the Langevin one are mathematical defined only in the discrete time limit, from which the continuous formulation can be derived. This definition is problematic in the moment of treating integrals involving the noise, i.e. the so called *stochastic integrals*.
Let us integrate Eq(1.1), for simplicity in one dimension, over a time interval $[t_0, t]$, obtaining:

$$x(t) = x(t_0) + \int_{t_0}^{t} dsF(x_s, s) + \int_{t_0}^{t} ds\xi(s)G(x_s, s). \tag{A.1}$$

The term $\xi(s)ds$ is a Gaussian random variable representing the stochastic increment of the Wiener process defined as:

$$dW_t = \xi(t)dt, \tag{A.2}$$

$$P(dW_t) = \frac{1}{\sqrt{2\pi dt}} \exp\left(-\frac{dW_t^2}{2dt}\right), \tag{A.3}$$

$$\langle dW_t \rangle = 0, \quad \langle dW_t^2 \rangle = dt. \tag{A.4}$$

Hence dW_t is a very particular measure, and by integrating over it one is realizing a *stochastic* integral (see Ref. [91] for more details). The Langevin equation can be rewritten as:

$$x(t) = x(t_0) + \int_{t_0}^{t} dsF(x_s, s) + \int_{t_0}^{t} dW_s G(x_s, s) \tag{A.5}$$

The mathematical problems come when one has to define the stochastic integral in terms of discrete sums, as done for the Riemann integral. Considering a general function of time A(t) one can defines the stochastic integral as:

$$\int_{t_0}^{t} dW_s A(s) = lim_{n\to\infty} \sum_{i=0}^{n} A(s_i)(W_i - W_{i-1}), \tag{A.6}$$

where the time interval has been partitioned in n sub-intervals with extremes $t_0 \leq t_1 \leq t_2 \leq .. \leq t_n = t$, while s_i is a general midpoint in each sub-interval $t_{i-1} \leq s_i \leq t_i$. Choosing the value of s_i has very non-trivial consequences on the properties of the stochastic integral and hence on the underlying physics. There are two important ways of doing it. The Itô's stochastic integral is defined by choosing the intermediate point to coincide with the

minimum of the interval $s_i t_{i-1}$

$$\int_{t_0}^{t} dW_s A(s) = lim_{n\to\infty} \sum_{i=0}^{n} A(t_{i-1})(W_i - W_{i-1}); \tag{A.7}$$

while the Stratonovich integral is obtained by choosing the intermediate point as the interval midpoint:

$$\int_{t_0}^{t} dW_s A(s) = lim_{n\to\infty} \sum_{i=0}^{n} A\left(\frac{t_{i-1}+t_i}{2}\right)(W_i - W_{i-1}); \tag{A.8}$$

Let us report on the main differences in these two approaches. In Itô's approach the discretized version of the Langevin reads:

$$x_{i+1} = x_i + F(x_i, t_i)\Delta t + G(x_i, t_i)\Delta W_i \tag{A.9}$$

where $x_i = x(t_i)$ and $\Delta W_i = W_{i+1} - W_i$. This choice makes the Wiener increment $\Delta W_i = W_{i+1} - W_i$ generally statistically independent of x_i. Furthermore, it has the non-trivial consequence of modifying the chain rule. Considere any function of x_t $B(x_t)$ and Taylor expand its infinitesimal change in a time interval up to order dt:

$$d_t B(x_t) = B(x_t + dx_t) - B(x_t) = B'(x_t)dx_t + \frac{1}{2}B''(x_t)dx_t^2 + .. = \tag{A.10}$$

$$= B'(x_t)[F(x_t, t)dx_t + G(x_t, t)dW_t] + \frac{B''(x_t G^2)dW_t^2}{2} \tag{A.11}$$

where we have used Eq.(A.9)); by using that $dW_t^2 = dt$ finally one arrives to celebrated Itô's formula:

$$d_t B(x_t) = [F(x_t, t)B' + \frac{1}{2}G^2 B'']dt + G(x_t, t)B'(x_t)dW_t. \tag{A.12}$$

Hence, when performing any change of variable in an Itô's SDE, Eq. (A.12) needs to be used. As a consequence, the Fokker-Planck equation in the Itô discretization reads:

$$\partial_t P = -\partial_x \left[FP - \frac{1}{2}\partial_x (DP)\right] \tag{A.13}$$

where $D = G^2$.
On the other hand, in the Stratonovich discretization, the Langevin reads:

$$x_{i+1} = x_i + F(x_i, t_i)\Delta t + G\left(\frac{x_i + x_{i+1}}{2}, t_i\right)\Delta W_i \tag{A.14}$$

and hence G and the increment ΔW_i are statistically dependent. On the contrary than the Itô case, there is no change in the chain rule:

$$d_t B(x_t) = F(x_t, t)B'dt + G(x_t, t)B'(x_t)dW_t. \tag{A.15}$$

Finally, in this case the Fokker-Planck reads:

$$\partial_t P = -\partial_x \left[FP - \frac{D}{2}\partial_x P\right]. \tag{A.16}$$

The main differences between the two discretizations manifest when the noise is multiplicative. Indeed the Itô scheme, with respect to the Stratonovich one, adds a term to the deterministic-drift in the Fokker-Planck equation:

$$\partial_t P = -\partial_x \left[\left(F - \frac{1}{2}\partial_x D \right) FP - \frac{D}{2}\partial_x P \right]. \tag{A.17}$$

and hence the deterministic fixed point x^*, $F(x^*) = 0$, and the mean x^* of the stationary distribution P^* are different. Furthermore, The Itô scheme is necessary to correctly simulate and study systems with absorbing states. On the other hand, the Stratonovich one is necessary to study the effects of the time reversal transformation in the dynamics ($\dot{x} \to -\dot{x}$) and hence, for example, to formulate the celebrated fluctuations relations [178].

A.2 Brownian motion

As an example, let us consider Brownian motion, i.e. the random motion of a particle suspended in a medium, like a fluid or a gas, in this case driven by an external force [93, 543, 89]. In the case of large friction, it can be modeled by a Langevin equation for the particle position:

$$\dot{x} = \frac{F(x)}{m\gamma} + \xi_t \tag{A.18}$$

$$\langle \xi \rangle = 0, \quad \langle \xi(t)\xi(t') \rangle = \delta(t - t') k_B T/\gamma. \tag{A.19}$$

where m mass of the particle, γ the friction, F the external force and ξ_i the stochastic noise emerging by the interaction with the medium at a certain temperature T.
Let us give an example of correspondence between a Langevin and Fokker-Planck equations. The one dimensional Brownian particle described in Eq.(A.18) is easily be translated to a Fokker-Planck equation, this case called Smoluchowski equation:

$$\partial_x P = \frac{1}{m\gamma}\partial_x \left[-F + k_B T \partial_x P \right]. \tag{A.20}$$

The current in the Fokker-Planck equation can be decomposed into two contributions:

$$J = J_F + J_D, \tag{A.21}$$

as consequence of the external force, the particle is be driven by a current J_F:

$$J_F = P\frac{F}{\gamma m}, \tag{A.22}$$

but at the same time experience a diffusion current:

$$J_D = -\frac{k_B T}{\gamma m}\partial_x P. \tag{A.23}$$

When the force derives from a potential U, one can use the detailed balance condition to balance the two currents and obtain the equilibrium Boltzmann distribution [543]:

$$J^* = 0 \quad \rightarrow \quad J_F = J_D \tag{A.24}$$

$$P^* \sim \exp\left(-\frac{U(x)}{k_B T}\right). \tag{A.25}$$

To see transient non-equilibrium effects it is necessary to study the time-dependent solution of Eq.(A.18) [93]. On the other hand, a simple example of solvable NESS is the Ornstein-Uhelenbeck process [93], that will be studied in a concrete example of chapter

A.3 Stationary solution of the SLM

Let us consider the SLM for one species

$$\dot{x} = \frac{1}{\tau}\left(1 - \frac{x}{K}\right) + \sqrt{\frac{\sigma}{\tau}}\xi x \tag{A.26}$$

where ξ is a Gaussian white noise:

$$\langle \xi(t) \rangle = 0, \quad \langle \xi(t)\xi(t') \rangle = \delta(t - t'); \tag{A.27}$$

In the following we will derive the stationary solution in general discretizaion by using the parameter α, that for $\alpha = 0$ gives the Stratonovich case while for $\alpha = 1$ the Itô's ones. We start by performing the following change of variables:

$$u = log(x) \tag{A.28}$$

by which we obtain a additive noise Langevin equation (using the Itô's formula for the change of varibale, Eq.A.12):

$$\dot{u} = \frac{1}{\tau}\left(1 - \frac{e^u}{K} - \frac{\alpha\sigma}{2}\right) + \sqrt{\frac{\sigma}{\tau}}\xi. \tag{A.29}$$

In one dimension it is, as always, an equilibrium problem, and hence we can define the potential

$$V(u) = -\frac{2 - \alpha\sigma}{2\tau}u + \frac{e^u}{K\tau}. \tag{A.30}$$

The associated Fokker Planck equation

$$\partial_t P(u,t) = +\partial_u(V'(u)P(u,t)) + \frac{\sigma}{2\tau}\partial_u^2 P(u,t). \tag{A.31}$$

It converges to the stationary distribution

$$P^*(u) = \frac{1}{Z}e^{-\frac{2V(u)\tau}{\sigma}}, \tag{A.32}$$

that can be transformed back to the original variable by including the Jacobian of the transformation:

$$P^*(x) = P^*(u)\frac{du}{dx} = \frac{1}{\Gamma(\frac{2}{\sigma} - \alpha)}\left(\frac{2}{K\sigma}\right)^{\frac{2}{\sigma}-\alpha} x^{\frac{2}{\sigma}-\alpha-1}e^{-\frac{2x}{\sigma K}} = \Gamma\left(x|\frac{2}{\sigma} - \alpha, \frac{\sigma K}{2}\right). \tag{A.33}$$

The result is Gamma distribution with shape and scale parameter $\gamma = \left(\frac{2}{\sigma_i} - 1\right)$, $\theta = \frac{K_i \sigma_i}{2}$ from which the mean and the variance can be easily derived.

$$\langle x \rangle = \gamma\theta = \left(\frac{2}{\sigma} - \alpha\right)\frac{K\sigma}{2} = K - \alpha\frac{K\sigma}{2} \tag{A.34}$$

$$Var^* = \gamma\theta^2 = \left(\frac{2}{\sigma} - \alpha\right)\frac{K^2\sigma^2}{4} \tag{A.35}$$

Appendix B

Data analysis

B.1 Datasets

All the datasets analyzed in this work were obtained from EBI Metagenomics (now Magnify) and have been previously published [544]. Raw data were processed under different versions of EBI Metagenomics pipelines [544] . The consistency of results across studies and pipelines strongly support the robustness and generality of the conclusions here. Supplementary Table S1 reports the references to the original works, description of the Magnify pipeline, and other relevant informations about each dataset. Note that the pipeline version 4.1 uses the algorithm SILVA [545] to assign an OTU classification. Observe also that here we use the term "species" to refer to OTUs, defined accordingly to the methods referred above. Datasets were selected to represent a wide set of diverse biomes: "gut" (human gut), "oral" (human mouth), "lake" and "river" (aquatic ecosystems), "activated sludge", "soil" and "glacier". We considered only datasets with at least 50 samples with more than 10^4 reads. No dataset was excluded a-posteriori.

Biome	Type	EBI ID	Magnify ID	Pipeline Version	NCBI ID	Reference	# Samples	[Range Tot, # Reads N_s]
Gut	c	SRP056641	MGYS00001056	2.0	PRJNA275349	[546]	66	[13842, 102971]
Oral	c	SRP056641	MGYS00001056	2.0	PRJNA275349	[546]	62	[10006, 138172]
River	c	ERP012927	MGYS00001669	3.0	PRJEB11530	[547]	188	[76042, 352675]
Lake	c	ERP012927	MGYS00001669	3.0	PRJEB11530	[547]	198	[57408, 350877]
Soil	c	SRP052295	MGYS00000905	2.0	PRJNA272333	-	112	[11352, 58219]
Sludge	c	ERP009143	MGYS00001064	2.0	PRJEB8105	[547]	575	[22255, 912713]
Glacier	c	ERP017997	MGYS00001292	3.0	PRJEB16145	[546]	30	[79765, 1104214]
Gut F4	l	ERP021896	MGYS00002184	4.1	PRJEB19825	[548]	131	[21008, 51986]
Gut M3	l	ERP021896	MGYS00002184	4.1	PRJEB19825	[548]	334	[15047, 58463]
Oral F4	l	ERP021896	MGYS00002184	4.1	PRJEB19825	[548]	135	[5683, 12651]
Oral M3	l	ERP021896	MGYS00002184	4.1	PRJEB19825	[548]	331	[1052, 2'3567]
Skin L-palm F4	l	ERP021896	MGYS00002184	4.1	PRJEB19825	[548]	134	[12298, 34607]
Skin L-palm M3	l	ERP021896	MGYS00002184	4.1	PRJEB19825	[548]	365	[144, 48475]
Skin R-palm M3	l	ERP021896	MGYS00002184	4.1	PRJEB19825	[548]	358	[135, 91953]

TABLE B.1: Description and references for the datasets used in this work. In column '*Type*', c refers to cross- sectional (across communities) and l to longitudinal (across time).

B.2 Phylogenetic Analysis

All the statistical analyses have been carried out with the tools of the *phyloseq* R library [549]. The phylogenetic tree of each community is obtained by removing the absent species from the three with all possible species. Then, the distance $d_{G,ij}$ between species i and j is calculated as the cophenetic distance [549]. Furthermore, each distance is categorized in one of $n_b = 17$, possible logarithmic bins, where the bin b to which a diven distance is assigned is given by

$$b = int\left(\frac{\log(d) - \min(\log(d))}{[\max(\log(d)) - \min(\log(d))]n_b}\right). \tag{B.1}$$

Note that the min and max are calculated for each community/sample independently. In each bin, the mean distance is calculated by averaging over all the species pairs with a distance within such a bin:

$$d_G(b) = \langle d_{G,ij} \rangle_b = \sum_{i,j \in b}^{N_b} \frac{d_{G,ij}}{N_b}, \tag{B.2}$$

where N_b is the number of species couples within bin b. Figure B.1 shows the histograms of the phylogenetic average distance d_G for the set of considered biomes. Even if the distributions show quantitatively different patterns for each biome, they share the fact that they exhibit a maximum at large distance (slightly below 1) and a monotonous decay to zero with a long tail.

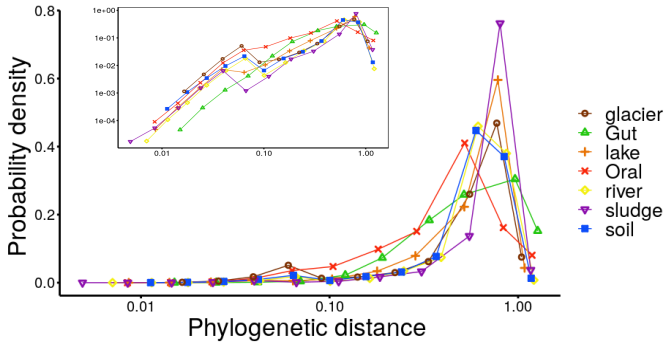


FIGURE B.1: **Phylogenetic distance distribution (log scale)** Histogram of the average phylogenetic distances $d_G$for different biomes (colors). The frequency in each bin (points) is calculated using the discretization given by Eq.(B.1). We only considered bins with at least 10^3 pairs of species. Interestingly, for all the considered biomes there is a peak around 1 and slowly-decaying left tail. The inset shows the same data but in log-log scale.

B.3 Correlation analysis

B.3.1 Cross-sectional data

In each community a, with $a = 1, ..., M$, the count of species i, with $i = 1, ..., N$, is represented by n_i^a; only sufficiently abundant samples are considered, i.e. the total reads for a community a must be larger than 10^4: $N^a = \sum_{i=1}^{N} n_i^a \geq 10^4$.
The relative abundance of species i in community a is calculated as:

$$x_i^a = \frac{n_i^a}{N^a}, \tag{B.3}$$

and the average over communities is defined as:

$$\langle .. \rangle = \frac{1}{M} \sum_{a=1}^{M} (..), \tag{B.4}$$

such that one can calculate the mean and the variance of a species relative abundance:

$$\langle x_i \rangle = \sum_{a=1}^{M} \frac{x_i^a}{M}, \tag{B.5}$$

$$Var_i = \langle x_i^2 \rangle - \langle x_i \rangle^2. \tag{B.6}$$

Another important observable is the rank of species i in community a, r_i^a: the most abundant species has rank $r_i^a = 1$, the second most abundant $r_j^a = 2$, and etc.
Using these ingredients one can construct a set of different observables quantifying species-abundance fluctuations:

$$q_{1i}^a = \frac{x_i^a - \langle x_i \rangle}{\langle x_i \rangle}, \tag{B.7}$$

$$q_{2i}^a = \frac{n_i^a - N^a \langle x_i \rangle}{N^a \langle x_i \rangle}, \tag{B.8}$$

$$q_{3i}^a = \frac{x_i^a - \langle x_i \rangle}{\sqrt{Var_i}}, \tag{B.9}$$

$$q_{4i}^a = \frac{\log x_i^a - \langle \log x_i \rangle}{\sqrt{Var(\log x_i)}}, \tag{B.10}$$

$$q_{5i}^a = 2r_i^a - 1. \tag{B.11}$$

By multiplying a couple corresponding to different species i and j, of the same-type observables, one can estimate pairwise correlations of species-abundance fluctuations:

$$\eta_{k_{ij}} = \langle q_{ki}^a q_{kj}^a \rangle_a = \sum_{a=1}^{N} \frac{q_{ki}^a q_{kj}^a}{N}, \tag{B.12}$$

with $k = 1, 2, 3, 4$ or 5. Finally, by averaging over all possible pairs of species with a mutual distance within a certain bin b, one can compute the averaged correlation in abundance fluctuations as a function of phylogenetic distance:

$$\eta_k(d_G) = \langle \eta_{k,ij} \rangle_{d_{G,ij} \in b} = \sum_{i,j=1, d_{P,ij} \in b}^{N} \frac{\eta_{k,ij}}{N_b}, \tag{B.13}$$

where N_b is the number of species couples in bin b.
To disentangle actual effects of phylogeny from possible spurious ones, we compared the measured correlations with those emerging from two alternative null models.
A) The first null model consists in calculating η_k on a randomized tree, preserving its high-order structure, meaning preserving the transitivity relation, i.e. if $|a - b| < d$ and $|b - c| < d$ also $|a - c| < d$.
B) In the second null model one calculates correlations using a randomization of phylogenetic distances between species, i .e. the distance between a couple is exchanged randomly with another one. These two alternative null models are equivalent for abundant samples. Figures B.2-B.3 show that the decay of abundance-fluctuation correlations with phylogenetic distance is qualitatively independent of the chosen observable, eq.(B.7), and that the null models show constant, almost vanishing, correlations. These results strongly support that the observed decay results from the actual structure of the phylogenetic tree.

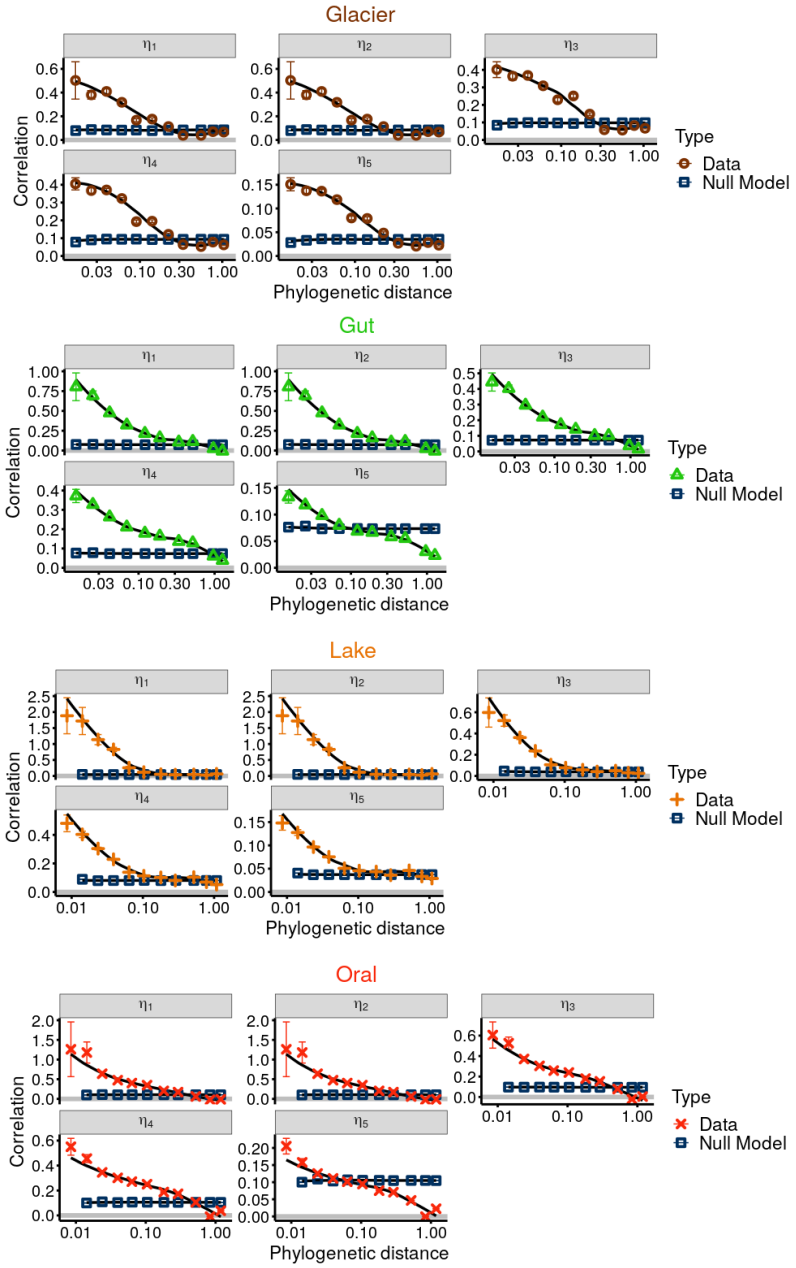


FIGURE B.2: **Correlation decay as a function of phylogenetic distance for different observables (η's) in four different biomes, namely: glacier, gut, lake and oral.** Colored points stand for empirical correlation data, blue squares for the null model and the black line for the average of points weighted by the number of couples in each bin. Bars represent standard errors in each bin.

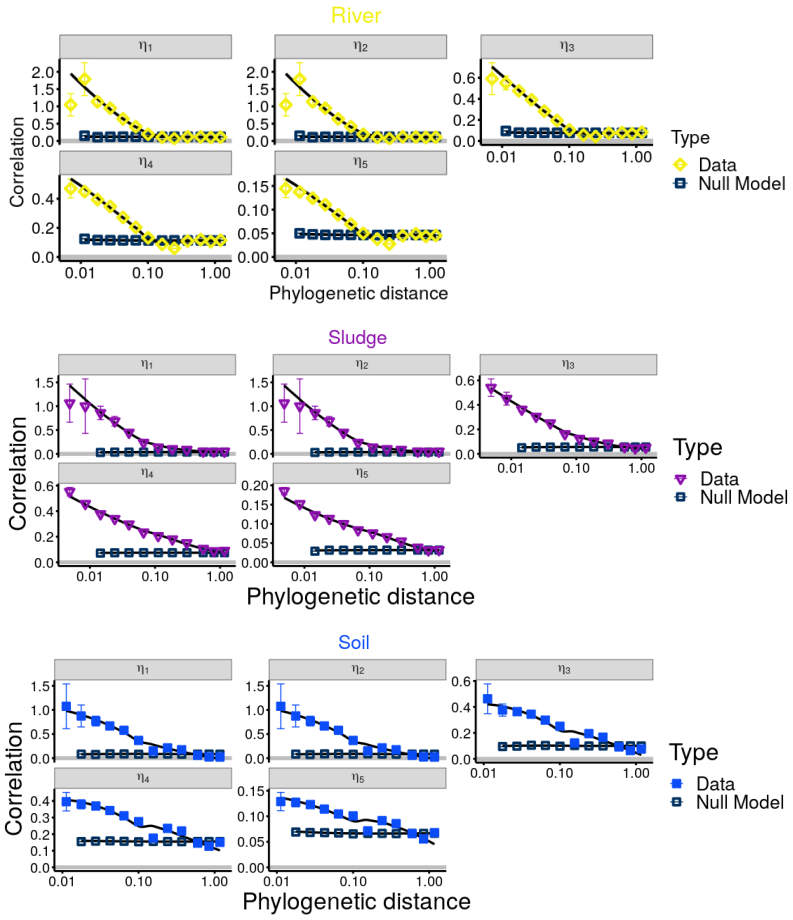


FIGURE B.3: **Correlation decay as a function of phylogenetic distance for different observables in three different biomes: river, sludge and soil**. As in the previous figure but for a different set of biomes.

Biome	χ	λ	R^2
Gut	0.34 ± 0.02	3.42 ± 0.16	0.97
Oral	0.41 ± 0.03	3.9 ± 0.40	0.94
River	0.35 ± 0.05	3.52 ± 0.47	0.83
Lake	0.40 ± 0.05	4.62 ± 0.62	0.88
Soil	0.26 ± 0.02	2.6 ± 0.14	0.94
Sludge	0.29 ± 0.02	3.44 ± 0.19	0.96
Glacier	0.31 ± 0.03	3.00 ± 0.22	0.91
Total	0.33 ± 0.01	3.41 ± 0.13	0.86

TABLE B.2: Stretched-exponential fit parameters for each biome.

B.3.2 Fit to a stretched-exponential function

Here, we present the fitting procedure used to obtain the stretched exponential decay. For each biome, we fitted the η_3 estimator with a stretched exponential function of the form:

$$\eta = e^{-\lambda d_G^{\chi}}, \tag{B.14}$$

by conducting a linear fit between $log(-log(\eta))$ and $\log(d_G)$, i.e. :

$$\log\left(-\log\eta\right) = \log\lambda + \chi\log\left(d_G\right). \tag{B.15}$$

The different points are weighted by the number of couples within the bin. In Fig. B.4 A and in Table A.2 we present the best estimation for λ and χ, including errors, as well as the R^2 coefficient for cross-sectional data and present the resulting fitted curves together with data (plots). As it can be seen, the vast majority of biomes follow a stretched exponential with excellent approximation, i.e. $R^2 \geq 0.9$, except for the lake and river ones where R^2 is smaller, but still larger than 0.8. Furthermore, there is a consistent variability of both the exponent χ and the scale parameter λ. Moreover, by considering all the points together independently of the biome, we still obtain a good fit to the stretched exponential curve ($R^2 = 0.86$), with the best fit parameters reported in chapter 2 $\chi = 0.33 \sim 1/3, \lambda = 3.5$, see Fig. 2.1

In Fig. B.4 B) we present a scatter plot of λ vs χ. Interestingly, we find that the two parameters are not independent there seems to be a linear correlation between them. This results suggests that the macroecological law could be simplified further as function of just one parameter. We will study this dependency in details in a future work.

Finally, in table A.3 and A.4 we report for comparison the result of fitting the correlation patter with a exponential decay, $\eta = \exp\left(-\chi d - \lambda\right)$, and a power-law $\eta = \lambda d^{-\chi}$. Clearly, the quality of the exponential fit is not good, because it cannot capture the scale invariant properties of the phylogenetic structure ($R^2 < 0.8$ a part for gut and oral where $R^2 \sim 0.9$). On the other hand, the power law fit is reasonably good, but in general with smaller R^2 than the stretched exponential one. Nevertheless, in some biomes it is almost as good as it (Oral,Lake, River $R^2 > 0.9$). In particular, the Lake dataset is fitted almost perfectly by the power law function, suggesting a that this biome might have specific features. Finally, the river biome is not well approximated by neither of the three fit, given that $R^2 < 0.9$

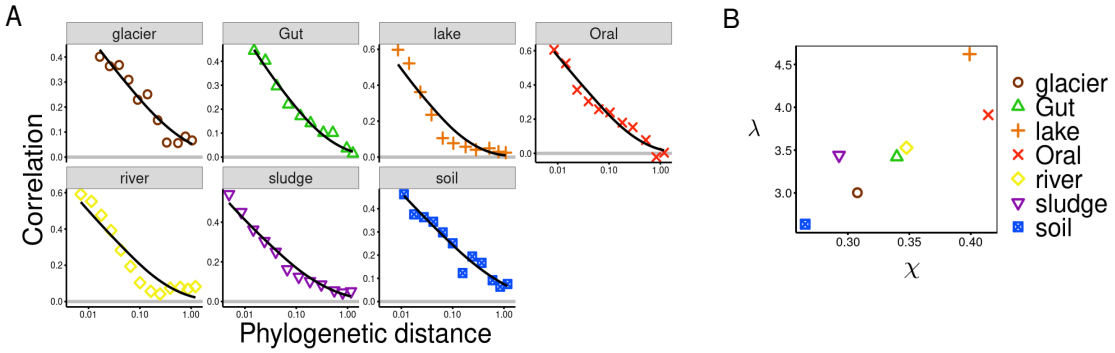


FIGURE B.4: **Stretched exponential fit for cross-sectional data.** A) Plots of the stretched exponential fit together with data. All the biomes follows with very good approximation a stretched exponential decay eq.(B.14), with a poorer fit in the lake and river. Variability of both χ and λ can be appreciated both in Table A.2 and in the plot. Nevertheless, the overall dataset still follows a stretched-exponential decay with $\lambda = 3.41 sim 3.5$ and $\chi = 0.33 \sim 1/3$ with good approximation. B) Scatter plot of λ vs χ obtained from a linear fit for different biomes (see Table A.2). Remarkably, a linear correlation seems to exist between the two parameters, suggesting further redundancies in the macroecological law.

Biome	χ	λ	R^2
Gut	2.53 ± 0.24	1.04 ± 0.12	0.91
Oral	3.98 ± 0.23	1.78 ± 0.09	0.94
River	1.77 ± 0.63	1.20 ± 0.37	0.37
Lake	3.02 ± 0.75	1.17 ± 0.30	0.59
Soil	1.70 ± 0.25	1.06 ± 0.11	0.78
Sludge	2.32 ± 0.42	1.07 ± 0.16	0.67
Glacier	2.15 ± 0.39	1.05 ± 0.15	0.70
Total	2.15 ± 0.39	1.05 ± 0.15	0.70

TABLE B.3: Exponential fit parameters for each biome.

B.3.3 Variability

The results exposed in the preceding sections deal with the decay of the mean Pearson correlation (i.e. mediated in each bin) with phylogenetic distance. On the other hand, here we report on the properties of the full correlation distribution within each bin. For the sake of simplicity, we consider the η_3 estimator only, but the results are similar for other quantifiers. Fig. B.5 shows that the correlation variance exhibits a tendency to decay with the phylogenetic distance as a power law $d^{-\gamma}$, with the exponent $\gamma \in [1/6, 1/2]$ different for each biome. The power-law decay is not perfect as it shows some "oscillations", similar to what happens discrete fractals and could reveal a discrete scale invariance in the underlying tree. We leave for future research a systematic study of this pattern. For the sake of completeness, in Fig. B.6 and Fig. B.7, we plot the η_3 correlation-coefficient distribution within each logarithmic bin, including at least 10^3 couples, for all the considered biomes.

B.3.4 Taxonomic Analysis

To verify that the behavior of correlations is uniform across the phylogenetic tree, i.e. that particularly abundant phylum is not determining the decay, we study the correlation pattern at the larger taxonomic scale of phyla. First, we report on the correlations

Biome	χ	λ	R^2
Gut	0.63 ± 0.06	0.042 ± 0.001	0.88
Oral	0.77 ± 0.14	0.028 ± 0.013	0.71
River	0.50 ± 0.06	0.048 ± 0.009	0.84
Lake	0.68 ± 0.03	0.022 ± 0.002	0.97
Soil	0.36 ± 0.04	0.087 ± 0.009	0.89
Sludge	0.37 ± 0.03	0.005 ± 0.007	0.91
Glacier	0.43 ± 0.05	0.067 ± 0.011	0.85
Total	0.44 ± 0.39	0.070 ± 0.010	0.84

TABLE B.4: Power law fit parameters for each biome.

between species of the same phylum (*intra*-phylum correlation). By averaging separately the correlation between species of the same phylum, we obtain how the correlation of abundances fluctuations decays with the phylogenetic distance for each single phylum. In Fig. B.8 we plotted the intra-phylum correlation for each biome and verify that majority of them decay coherently with general pattern, while a few of them, e.g. actinobacteria, exhibit some deviations. Such taxa-dependent deviations could be responsible for deviations from the typical stretched exponential pattern for specific biomes. To test this claim, we consider as an example the soil biome, where a small non-monotonicity is present in the decay, as illustrated by a characteristic bump around $d = 0.1$ (cfr. with Fig. B.3). In Fig. B.9) (Top), we show that such a deviation is probably due to the behavior of actinobacteria, that present small negative correlation at intermediate distances. Furthermore, in the bottom of Fig. B.9), we show that, by zooming into the actinobacteria phylum and considering correlations at the finer level of orders, the negative correlations around distance $d_G = 0.1$ is caused by the actinomycetales and gaiellales. The other orders contribute mainly with positive correlation. These results seem to suggest that deviation from the typical pattern are driven by few bacterial orders. This analysis could be carried forward by going at even finer taxonomic resolution to identify the drivers of this deviation, with the goal of understanding which are the corresponding ecological traits which could produce such decreasing correlation. All these analyses open exciting routes for future research.

To complement the precedent analysis, in Fig. B.10 we plotted the intra-phylum correlation (same data as before) for each phylum, representing with colors the different biomes. The decaying pattern is found consistently in the most abundant phylum, like acidobacteria, bacteriotedes, proteobacteria and firmicutes. On the other hand, in less abundant phyla correlations are still positive but not showing a universal behavior. We leave for a future work the study of how different order deviate from the general pattern and why.

Finally, in order to understand how the universal pattern changes at the larger scale of phylum and to discover how it emerges from the different phylum and what changes by coarse-graining in Fig. B.11 we report both intra-phylum correlations (red points) and inter-phylum ones (i.e. between species of different phylum) separately. The correlation pattern for these taxonomic relation show two totally different behaviors. Inter-correlations concentrate symmetrically around zero and correspond to very large phylogenetic distances. On the other hand, as showed in the preceding figures, intra-taxonomic correlations decay from positive value to zero with with distance. By averaging over different phyla the stretched-exponential pattern (black line) emerges both in each single biome and in the total case.

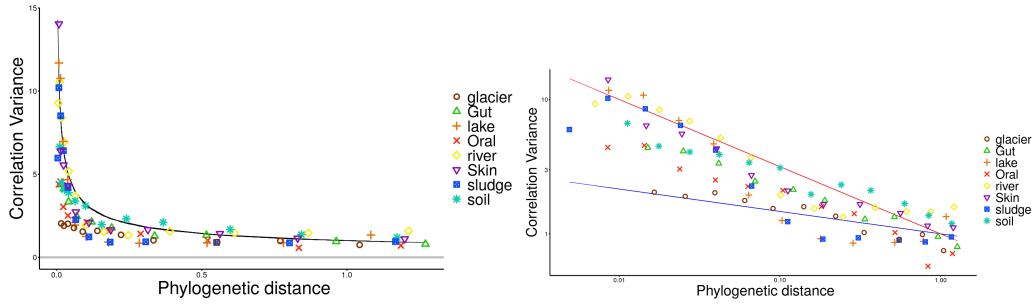


FIGURE B.5: **Quasi-universal pattern of correlation variance.** Points stand for the variance of the correlation distribution within each distance bin. Left: Correlation variance versus phylogenetic distance in natural scale, where the black line is a power-law fit $d^{-\gamma}$, with $\gamma = 1/2$. Right: the same data but represented in a log-log scale. There seems to be a tendency to behave like a power-law but each biome shows a different exponent γ_i (red and blue lines stand for power-law fits with maximum and minimum possible exponents $\gamma_{max} = 1/2$, $\gamma_{min} = 1/6$) and there seem to be oscillations.

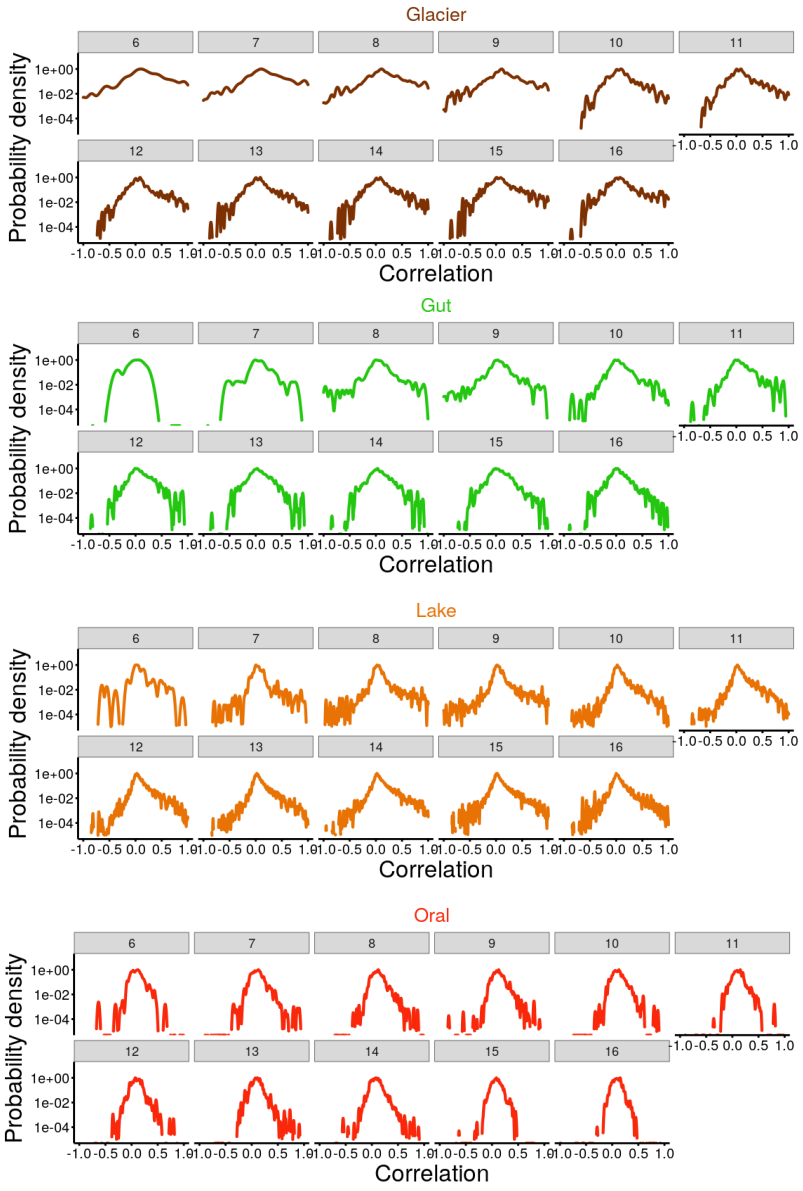


FIGURE B.6: η_3 correlation histogram in each bin (colored lines) in log-scale. Numbers on top of plots indicate the corresponding bin, from 1 to 17. Only bins with at least 10^3 couples are considered.

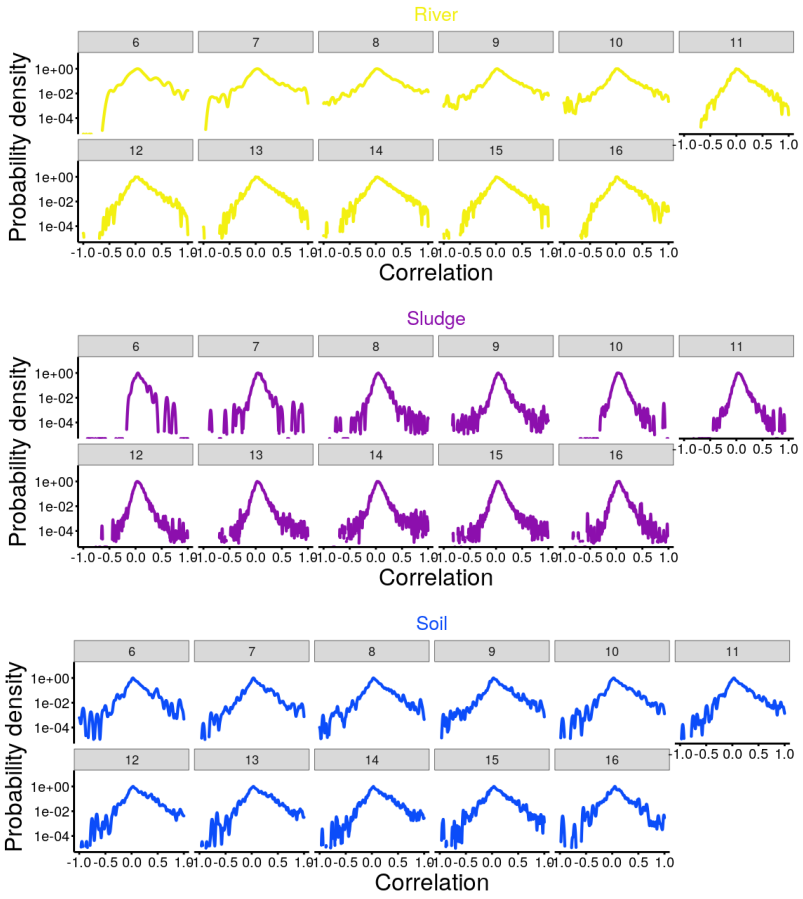


FIGURE B.7: η_3 correlation histogram in each bin (colored lines) in log-scale. Numbers on top of plots indicate the corresponding bin, from 1 to 17. Only bins with at least 10^3 couples are considered.

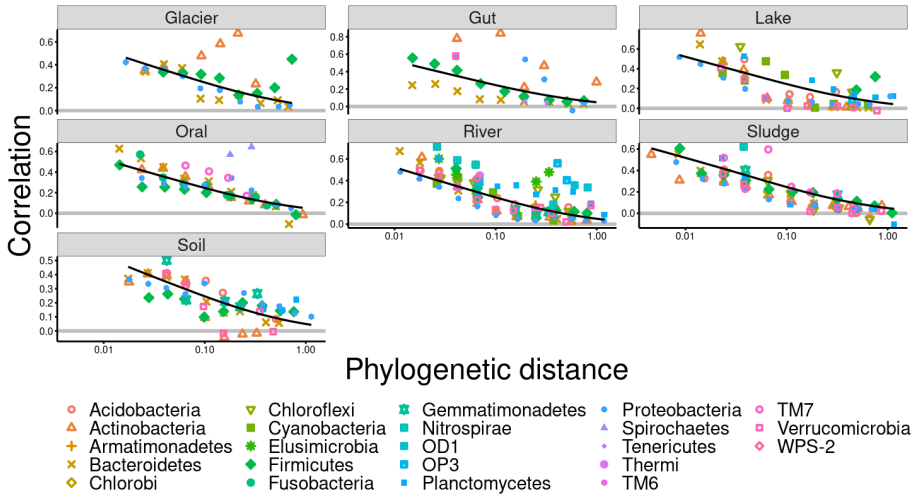


FIGURE B.8: Taxonomic Analysis. Intra-phylum correlations versus phylogenetic distance for each biome (see color code and point shapes in legend). Black lines represent the fit with the average parameters (see Fig. B.4 and Table A.2) while the symbol sizes represent the log of the number of couples within the bin. The colored points are obtained by averaging only species of the same phylum.

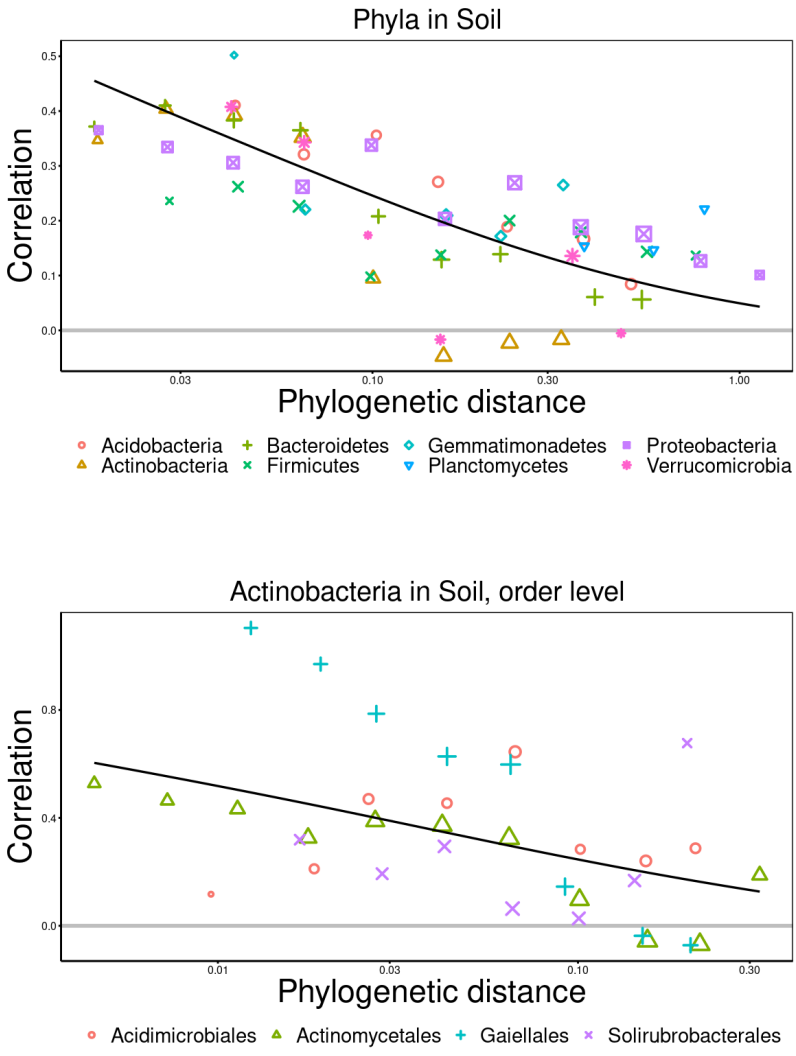


FIGURE B.9: **Taxonomic analysis in the soil biome.** Top: intra-phylum correlation pattern in soil; symbol colors and shapes stand for each different phylum (see legend), while symbol sizes stand for the logarithm of the number of couples within each bin. The phylum actinobacteria is an example of a taxa with large deviations from the main pattern, in particular they show negative correlations around phylogenetic distance of 0.1. Also verrucomicrobia exhibit some negative correlation, but are not less abundant than actinobacteria. Bottom: intra-order correlation pattern in the phylum actibacteria in soil biome. By considering the correlation between species of the order inside the actinobacteria phylum we show that the observed deviation is mostly due to actinomycetales and gaiellales.

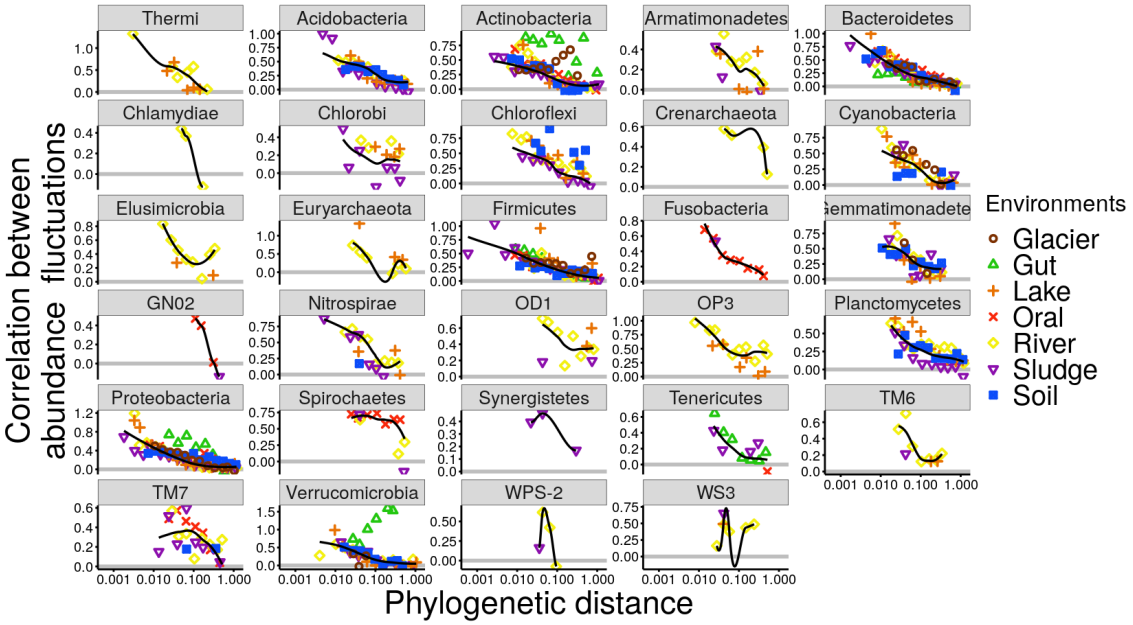


FIGURE B.10: **Correlation versus phylogenetic distance in each phylum for different biomes.** Colored points indicate the corresponding biomes and the black line the averaged behavior over biomes weighted by abundances. Correlation of phyla that are resent in many biomes — such as acidobacteria, bacteriotedes, proteobacteria and firmicutes— tend to follow a positive to null decay. On the other hand, less abundant phyla present large deviations for such an overall trend.

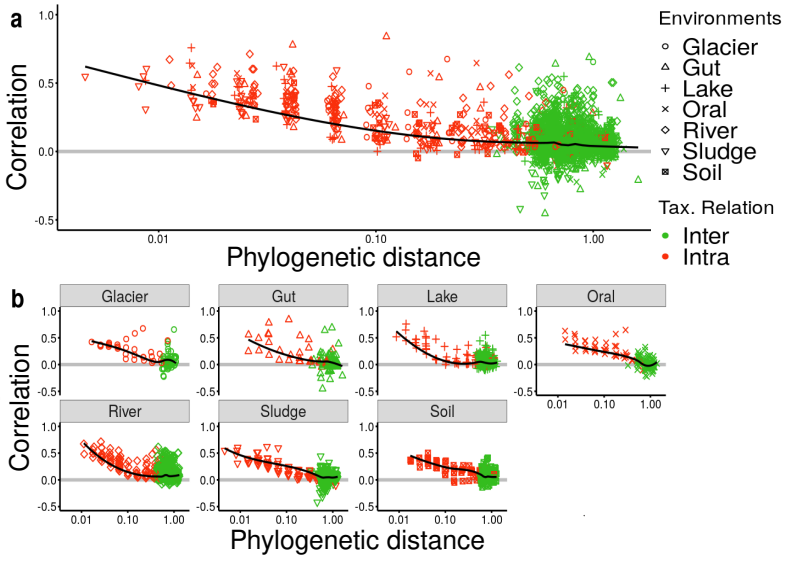


FIGURE B.11: **Taxonomic Analyses.** a: Correlation between abundance fluctuations versus phylogenetic distance for intra-phylum (red points) and inter (green points) phyla. In total, 29 phyla are considered and each point represents the correlation of one of them, within a certain phylogenetic distance, in a particular biome (shapes). Black lines are averages over both taxas and biomes, weighted by abundances in each considered bin. (b) Same data as in (a) above but plotted separately for each biome.

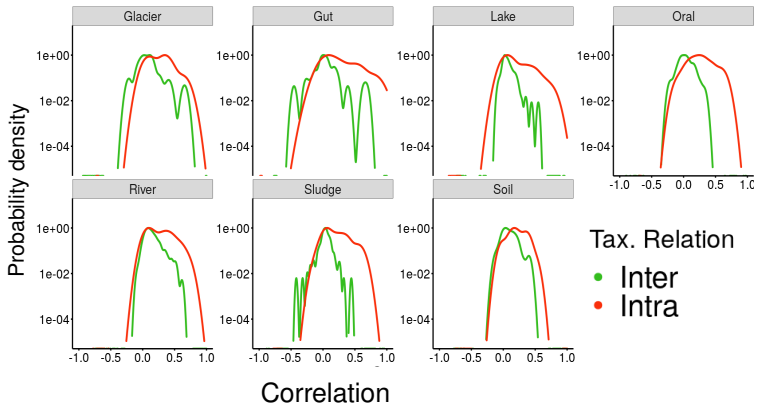


FIGURE B.12: **Histogram of correlations for both intra and inter phyla**. As clearly illustrated, inter-taxonomic correlations have a symmetric distribution centered around zero, while intra ones are skewed to positive values.

B.3.5 Time-series

The analysis of temporal data is analogous to that in Section (B.3.1), but instead of studying fluctuations and correlations between different communities, one considers one single community across time. For each host, $h = 1, .., H$, one has different samples from different times (days) $t = 1, .., T$. All the observables are defined as in Sec.B.3.1 but replacing the community average by a time average $\langle .. \rangle_t = \frac{1}{T} \sum_{t=1}^{T} (..)$. In particular, the equal time correlation between two species abundance fluctuations (i, j) is defined by:

$$\eta_{kij} = \langle q(t)_{k_i} q(t)_{k_j} \rangle_t = \sum_{t=1}^{T} \frac{q(t)_{k_i} q(t)_{k_j}}{T}; \tag{B.16}$$

and the Δt delayed correlation as:

$$\eta_{kij}(\Delta t) = \langle q(t + \Delta t)_{k_i} q(t)_{k_j} \rangle_t = \sum_{t=1}^{T-\Delta t} \frac{q(t + \Delta t)_{k_i} q(t)_{k_j}}{T}. \tag{B.17}$$

In Fig. B.13 A) and in Table A.5, we report the results of fitting the correlations with a stretched-exponential curve as done in Sec.B.3.2. One can see that the correlations follow a stretched exponential of the form of Eq.(B.14) with very good approximation, $R^2 \geq 0.9$ in all biomes and hosts, except for the oral biome of the F4 host. Interestingly, there is some variation in the values of λ and χ both across hosts and biomes. In addition, the gut and oral biomes show different but similar parameters in the cross-sectional and longitudinal datasets. Furthermore, the skin biome seems to deviate from the stretched-exponential fit at very small distances coherently across hosts. Nevertheless, observe that the fit of the total pattern to a single stretched exponential is still good, $R^2 > 0.8$, but that the skin biome deviate in totally different way than the oral and gut (see Fig. B.14). In Fig. B.13 B) we present also a scatter plot of λ vs χ, showing a linear dependency between the parameters, as found in cross-sectional data (see FigB.4). Finally, in tables A.6-A.7 we report for comparison also the results of fitting the equal-time correlations with an exponential decay $\eta = \exp(-\chi x - \lambda)$ and a power-law $\eta = \lambda d^{-\chi}$. Similarly, to what happens in cross-sectional data, the exponential fit is worse than the stretched exponential in all biomes. Also, the power law fit is worse than the stretched one in the gut and oral biomes, but slightly better in the skin, in particular in the right palm. This suggest that this biome has some peculiar characteristics, as already noted in a precedent work by one of us [123]. We leave for future work the investigation of the factors driving such a variability in χ and λ, that appears to be larger than in cross-sectional datasets.

Moreover, in Figures B.15-B.18, we report separately for each biome the behavior of the correlation, for each host and averaging over them, and show that also the delayed correlations can be well fitted by following modification of stretched-exponential function see Sec.2.8:

$$\eta(\Delta t, d_G) = \exp\left(-\frac{\Delta t}{\tau} - \lambda d_G^{1/3}\right), \tag{B.18}$$

where λ and χ are fixed from the equal-time correlations,see Table A.3, and τ is fixed just once for each biome, $\tau = 1$ for Gut and oral and $\tau = 0.5$ for left and right-palm skin .

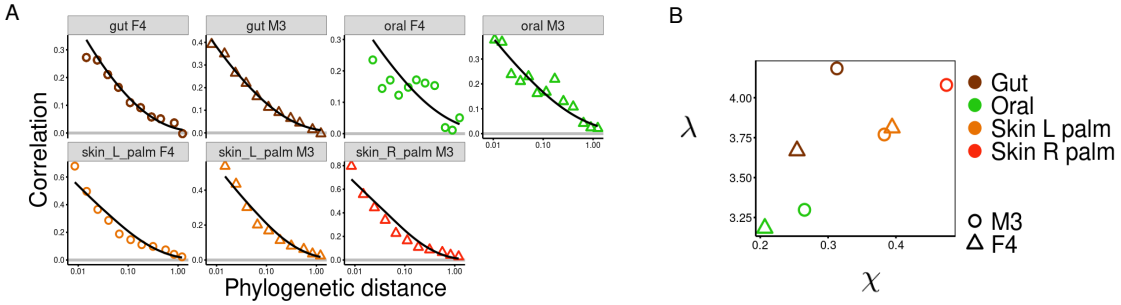


FIGURE B.13: **Stretched exponential fit for longitudinal data.** A) Plots of stretched-exponential fits for different biomes and hosts together with data. Almost all the biomes follow with very good approximation a stretched-exponential decay Eq.(B.14), with a the poorest fit in the F4 oral case. Such large deviations are probably due to some unknown conditions of the host. Furthermore, the skin biome consistently deviates from the stretched exponential at small distances. B)Scatter plot ofλ vs χ obtained from linear fit for different biomes (colors) and hosts (shapes),see Table A.3. Interestingly, a direct linear proportionality emerges between the two parameters, suggesting further redundancies in the macroecological law.

Biome	Host Id	χ	λ	R^2
Gut	M3	0.31 ± 0.01	4.18 ± 0.11	0.98
Gut	F4	0.25 ± 0.01	3.76 ± 0.04	0.99
Oral	M3	0.26 ± 0.01	3.30 ± 0.15	0.90
Oral	F4	0.20 ± 0.04	3.18 ± 0.26	0.54
Skin Left Palm	M3	0.38 ± 0.03	3.77 ± 0.24	0.96
Skin Left Palm	F4	0.39 ± 0.03	3.81 ± 0.39	0.92
Skin Right Palm	M3	0.47 ± 0.05	4.08 ± 0.58	0.89
Total	Total	0.30 ± 0.01	3.73 ± 0.12	0.82

TABLE B.5: Stretched-exponential fit parameters for each biome for *equal-time* temporal data.

Biome	Host Id	χ	λ	R^2
Gut	M3	3.65 ± 0.41	1.38 ± 0.13	0.90
Gut	F4	2.43 ± 0.15	1.65 ± 0.05	0.78
Oral	M3	2.32 ± 0.13	1.33 ± 0.07	0.90
Oral	F4	1.87 ± 0.35	1.70 ± 0.18	0.52
Skin Left Palm	M3	2.53 ± 0.35	1.26 ± 0.21	0.82
Skin Left Palm	F4	2.4 ± 0.36	1.17 ± 0.17	0.84
Skin Right Palm	M3	2.47 ± 0.42	1.08 ± 0.18	0.78

TABLE B.6: Exponential fit parameters for each biome for *equal-time* temporal data.

Biome	Host Id	χ	λ	R^2
Gut	M3	0.64 ± 0.05	0.024 ± 0.004	0.93
Gut	F4	0.53 ± 0.01	0.033 ± 0.001	0.98
Oral	M3	0.55 ± 0.05	0.039 ± 0.005	0.83
Oral	F4	0.54 ± 0.11	0.024 ± 0.004	0.53
Skin Left Palm	M3	0.68 ± 0.03	0.033 ± 0.002	0.98
Skin Left Palm	F4	0.61 ± 0.03	0.037 ± 0.003	0.97
Skin Right Palm	M3	0.65 ± 0.02	0.038 ± 0.001	0.99

TABLE B.7: Power-law fit parameters for each biome for *equal-time* temporal data.

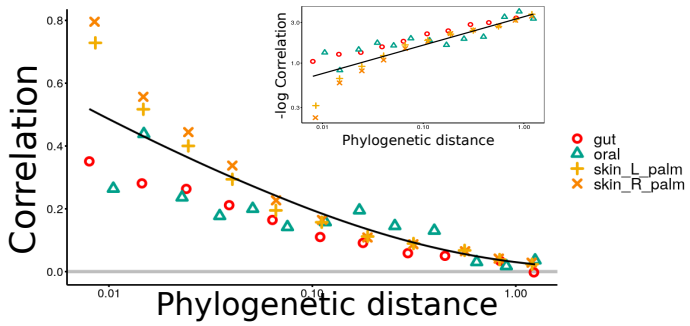


FIGURE B.14: **Macroecological law for temporal data: average over hosts and biomes** Correlations as function of phylogenetic distance averaged over hosts for each biomes. Inset: $-log(Corr)$ as function of distance averaged over hosts. Points represent correlations for each biomes averaged over hosts. The black line is the stretched-exponential fit corresponding to the last line reported in the Table A.3. Observe that, at small distances, gut and oral biomes have a different behavior than the skin one.

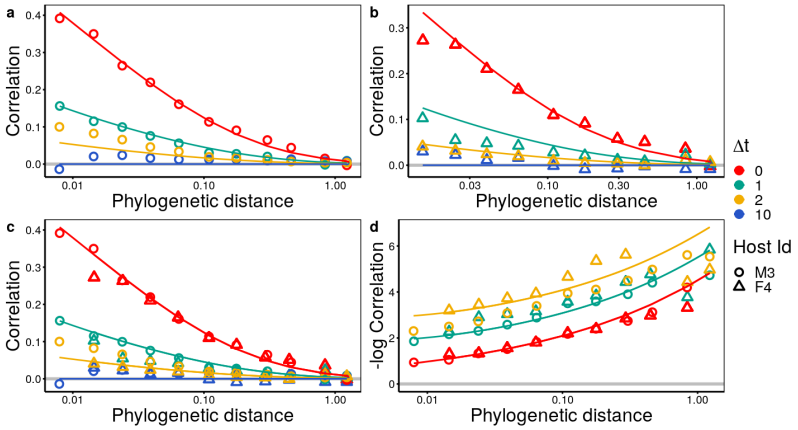


FIGURE B.15: **Macroecological law for temporal data: Gut biome.** Correlations as function of phylogenetic distance for each host (a-b), averaged over hosts (c), $-log(Corr)$ as function of distance averaged over hosts (d). Colored points represent correlations at equal time (red) or one, two and ten day delay(green, yellow, blue). Shapes stand for different hosts. Solid lines are the stretched exponential decay from eq.(B.18). Notably, the delayed correlations can be predicted using eq.(B.18),with the same λ and χ of equal time correlation, just by fixing growth timescale $\tau = 1$. for all the plots. The other parameters for the solid lines are: $(a)\chi = 0.31, \lambda = 4.18; (b)\chi = 0.25, \lambda = 3.76; (c,d)\chi = 0.3, \lambda = 4$.

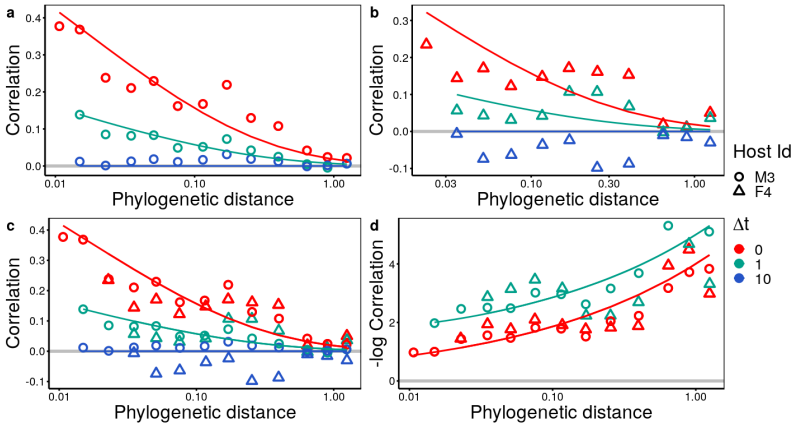


FIGURE B.16: **Macroecological law for temporal data: Oral biome.** Symbols, lines and sub-figure legend as in Fig. B.15. Observe that the F4 host presents major deviations, that are probably caused by some special health conditions. Nevertheless, the correlations are still decaying from positive to null values, and the delayed pattern follows the same, predictable, tendency of the other biomes. Solid lines are the stretched-exponential decay from Eq.(B.18), with $\tau = 1$ in all the plots. The other parameters for the solid lines are: $(a, c, d)\chi = 0.26, \lambda = 3.3; (b)\chi = 0.2, \lambda = 3.18$.

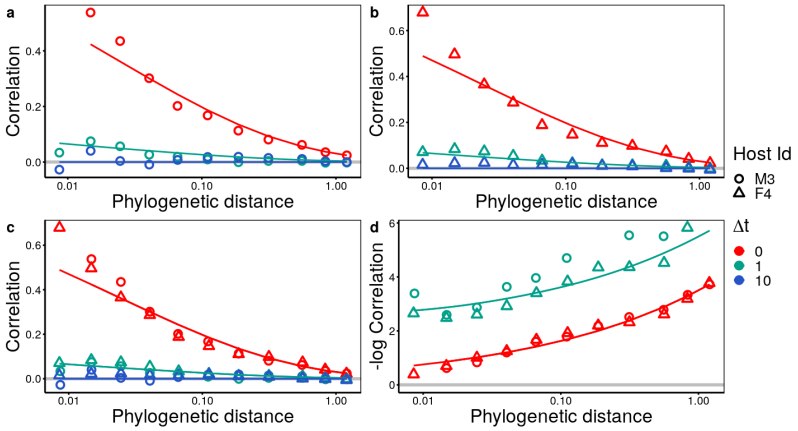


FIGURE B.17: **Macroecological law for temporal data: Skin, left- palm biome.** Symbols, lines and sub-figure legend as in Fig. B.15. Observe that deviations from the stretched-exponential fit are significant at small phylogenetic distance. Notably, the delayed correlations decay to zero way faster than the other biomes. Both observations are coherent with recent macroecological studies on human microbiome that identified a very rapid dynamics in this skin dataset [123]. Solid lines are the stretched exponential decay from eq.(B.18), with $\tau = 0.5$ for all the plots. The other parameters for the solid lines for all plots are: $\chi = 0.38, \lambda = 3.8$.

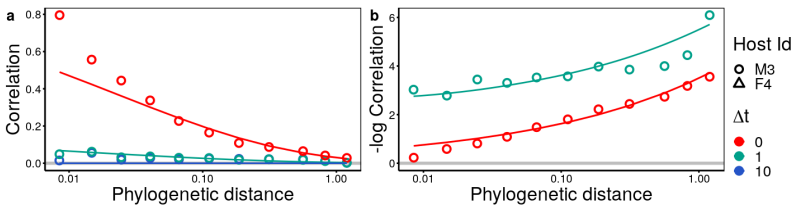


FIGURE B.18: **Macroecological law for temporal data: Skin, right- palm biome.** Correlations (a) and $-log(Corr)$versus phylogenetic distance for host M3. Observe that deviations from the stretched exponential and the rapid decay of the delayed correlation are coherent with the left-palm biome, Fig. B.17. Solid lines are stretched exponential decay from eq.(B.18), with $\tau = 0.5$ The other parameters for the solid lines are: $\chi = 0.47, \lambda = 4$..

Appendix C

Statistical mechanics of phenotypic eco-evolution: mathematical details

C.1 From Microsopic to Macroscopic Process

In this section we present a microscopic description of evolutionary dynamics, and derive the mean field equations. The first part it's devoted to the simple case of a generalized Moran process with constant population size. In the limit of infinite population size a mean-field theory is derived. In the second part we generalize the theory to a general birth-death process with varying population size.

C.1.1 Generalized Moran process

The evolutionary model is interpreted as a many-particle Markov jump process. Let's consider N *indistinguishable* individuals, or particles, that can have a continuum position $x_{\mathbf{i}}$ the phenotypic space, $\mathcal{P}$. The state of the entire system is represented by the the vector

$$\boldsymbol{x}_N = (x_1, x_2, .x_i, .., x_j, ..x_N), \tag{C.1}$$

$$dim(\boldsymbol{x}_N) = N; \tag{C.2}$$

that collects the individual traits. x_i. Any vector different from $\boldsymbol{x}$ for just one individual trait, say individual j, will be indicated as:

$$\widetilde{\boldsymbol{x}}_N^j = (x_1, x_2, ..., x_i, \widetilde{x}_j, ...x_N), \tag{C.3}$$

$$dim(\widetilde{\boldsymbol{x}}_N^j) = N. \tag{C.4}$$

Finally, when always with respect to $\boldsymbol{x}_N$ the j individual has a different trait and, say the 1 individual is absent, we will use the following notation for the population vector:

$$\widetilde{\boldsymbol{x}}_N^{j,\hat{1}} = (x_2, ..., x_i, \widetilde{x}_j, ...x_N), \tag{C.5}$$

$$dim(\widetilde{\boldsymbol{x}}_N^{j,\hat{1}}) = N - 1. \tag{C.6}$$

and also for the differential:

$$d\widetilde{x}_N^{j,\hat{1}} = dx_2 dx_3 .. d\widetilde{x}_j .. dx_N. \tag{C.7}$$

In the following paragraphs, we will study the evolution of probability of a certain configuration $\boldsymbol{x}_N$, $P(\boldsymbol{x}_N)$. The probability, and hence its evolution, satisfies the symmetry of individual indistinguishability that we define as the fact that the probability is unaffected by any permutation of the elements of the configuration vector $\boldsymbol{x}_N$.

The population evolves by implementing the following stochastic process, that we call gen. Moran Process:

1 A first individual with trait x_i is select randomly to reproduce proportionally to its fitness rate $f(x_i, \widetilde{\boldsymbol{x}}^j)$. We restrict our selves to cases where the fitness is composed by a growth rate, $K(x_i)$ and a pairwise interaction $I(x_i, x_k)$:

$$f(x_i, \widetilde{\boldsymbol{x}}^j) \equiv K(x_i) + \frac{1}{N-1} \sum_{k=1, i\neq k}^{N} I(x_i, x_k). \tag{C.8}$$

2 A second individual with trait $\widetilde{x}_j$ is selected randomly to die, i.e. with death rate $d(\tilde{x}_j) = \frac{1}{N-1}$.

3 The second individual dies while the first one reproduces asexually by producing an offspring whose trait will be the parent trait plus a variation, or mutation, δ_j:$x_j = x_i + \delta_j$. The variation is sampled from a probability distribution $\beta(\delta_j)$ that can in principle depend on the parent trait. Hence the total number of individuals is conserved

To translate such process as a stochastic Markov-jump process the step 3 needs to be rewritten as a jump. In detail, two jumps are happening: first, the i individual is jumps from position x_i to x_j, and the magnitude of the jump is ruled by $\beta(\delta_j)$. At the same time, the "dying" individual has to jump from $\widetilde{x}_j$ to x_i, effectively replacing the first individual. This last jump is "non-local" because does not depend on the initial position $\tilde{x}_j$. Such jump is possible for individual indistinguishability. Furthermore, given that the population size is constant we will drop the N pedix from the state notation.
We can write down a general master equation for the probabilistic dynamics of the the population vector $\boldsymbol{x}$ for the gen. Moran process:

$$\partial_t P(\boldsymbol{x}, t) = \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \int_{\mathcal{P}} d\widetilde{x}_j \Big[W_i(\boldsymbol{x}, \widetilde{\boldsymbol{x}}^j) P(\widetilde{\boldsymbol{x}}^j, t) - W_i(\widetilde{\boldsymbol{x}}^j, \boldsymbol{x}) P(\boldsymbol{x}, t) \Big] \tag{C.9}$$

where

$$\begin{aligned} W_i(\boldsymbol{x}, \widetilde{\boldsymbol{x}}^j) &= f(x_i, \widetilde{\boldsymbol{x}}^j)\ \beta(x_j - x_i)\ d(\widetilde{x}_j) \\ &= f(x_i, \widetilde{\boldsymbol{x}}^j)\ \beta(x_j - x_i)/(N-1) \end{aligned} \tag{C.10}$$

is the transition rate to go from a state $\widetilde{\boldsymbol{x}}^j = (x_1, .., x_i, .., \widetilde{x}_j, .., x_N)$ (with just one coordinate differing from those of $\boldsymbol{x}$) to $\boldsymbol{x} = (x_1, .., x_i, .., x_j, ..x_n)$, and $W(\widetilde{\boldsymbol{x}}^j, \boldsymbol{x})$ is the rate for the reverse process.

Marginalization

Now we would like to derive a macroscopic mean field equation for the probability of a position in the phenotypic space. To do that we formulate a coarse-graining procedure. We introduce the one particle *density* of individual with phenotype x at time t can be simply expressed as

$$\rho(x, t) = \sum_{i=1}^{N} \frac{\delta(x - x_i)}{N}. \tag{C.11}$$

Averaging over all possible microscopic configurations one obtains the one-particle probability density

$$\begin{aligned}\phi(x,t) &\equiv \Big\langle \sum_{i=1}^{N} \frac{\delta(x-x_i)}{N} \Big\rangle \\ &= \frac{1}{N}\sum_{i=1}^{N}\int_{\mathcal{P}^N} d\mathbf{x}\delta(x-x_i)P(\mathbf{x},t) \\ &= \int_{\mathcal{P}^{N-1}} dx_2 dx_3..dx_N P(x,x_2,x_3,..,x_N), \end{aligned} \tag{C.12}$$

where in the second to last passage we used the fact that the particles are indistinguishable(the probability function is symmetric for the exchange of two particles), and in the last one the definition of marginal probability. The same can be written for the n-partcle density:

$$\phi^{(n)}(x_1,x_2,..,x_n;t) = \frac{N-n}{N!}\langle \Pi_{j=1}^{n}\sum_{i_j=1}^{N}\delta(x_i-x_{i_j})\rangle \tag{C.13}$$

$$= \int_{\mathcal{P}^{N-n}} dx_{n+1}dx_{n+2}..dx_N P(\boldsymbol{x},t) = P^{(n)}(x_1,x_2,..x_n;t) \tag{C.14}$$

For now let's consider just the one particle density(C.12) and let's derive it to respect to time use the Master eq.(3.3)

$$\begin{aligned}\partial_t\phi^{(1)}(x,t) &= \int_{\mathcal{P}^{N-1}} dx_2..dx_N\partial_t P(\boldsymbol{x},t) \\ &= \int_{\mathcal{P}^{N-1}} d\hat{x}_N^1 \sum_{i=1}^{N}\sum_{i\neq j,1}^{N}\int_{\mathcal{P}} d\widetilde{x}_j P(\widetilde{\boldsymbol{x}}^j,t)\frac{f(x_i,\widetilde{\boldsymbol{x}}^j,t)}{N-1}\beta(x_j-x_i) \\ &- \int_{\mathcal{P}^{N-1}} d\hat{x}_N^1 P(\boldsymbol{x},t)\sum_{i=1}^{N}\sum_{i\neq j,1}^{N}\frac{f(x_i,\boldsymbol{x},t)}{N-1} W_{gain} - W_{loss} \end{aligned} \tag{C.15}$$

Now let's consider the loss contribution (we drop the time dependence):

$$W_{loss} = \int_{\mathcal{P}^{N-1}} d\hat{x}_N^1 P(\boldsymbol{x},t)\sum_{i=1}^{N}\sum_{i\neq j,1}^{N}\frac{f(x_i,\boldsymbol{x})}{N-1}. \tag{C.16}$$

We split the sums in the following way:

$$\sum_{i\geq 1}\sum_{j\geq 1, i\neq j} = \sum_{j>1}^{N}\delta_{i1} + \sum_{i>1}^{N}\sum_{j\geq 1, i\neq j}^{N}.$$

Doing so:

$$\begin{aligned} W_{loss} &= \int_{\mathcal{P}^{N-1}} d\hat{x}_N^1 P(\boldsymbol{x},t) f(x_1,\boldsymbol{x}) + \int_{\mathcal{P}^{N-1}} d\hat{x}_N^1 P(\boldsymbol{x},t)\sum_{i>1} f(x_i,\boldsymbol{x}) \\ &= \int_{\mathcal{P}^{N-1}} d\hat{x}_N^1 [f(x_1,\boldsymbol{x}) + (N-1)f(x_2,\boldsymbol{x})]P(\boldsymbol{x},t) \end{aligned} \tag{C.17}$$

where, once again, we have used the indistinguishability of individuals eq.(C.12).
The gain term is:

$$W_{gain} = \int_{\mathcal{P}^{N-1}} dx_N^{\hat{1}} \sum_{i=1}^{N} \sum_{i\neq j,1}^{N} \int_{\mathcal{P}} d\widetilde{x}_j P(\boldsymbol{x}^j, t) \frac{f(x_i, \widetilde{\boldsymbol{x}}^j)}{N-1} \beta(x_j - x_i). \tag{C.18}$$

Let's split the sums in the following way:

$$\sum_{i\geq 1} \sum_{j\geq 1, i\neq j} = \sum_{j>1}^{N} \delta_{i1} + \sum_{i>1}^{N} \delta_{j1} + \sum_{i>1}^{N} \sum_{j>1, i\neq j}^{N}$$

This is equivalent to say that in the first sum we are considering the case when the particle 1 is reproducing while we are summing over all the possible compatible dying individuals j. In the second term the situation is reversed: individual 1 is dying(i.e. becoming an offspring of j) and we are summing on all the possible j. Finally, the last term considers when individual 1 is not involved in the jump process. Starting with the first term, we integrate over $\widetilde{x}_1$:

$$\begin{aligned} & \sum_{j>1}^{N} \delta_{i1} \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\widetilde{x}_j P(\widetilde{\boldsymbol{x}}^j, t) \frac{f(x_i, \widetilde{\boldsymbol{x}}^j)}{N-1} \beta(x_j - x_1) \\ = & \sum_{j>1} \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\widetilde{x}_j P(\widetilde{\boldsymbol{x}}^j) \frac{f(x_1, \widetilde{\boldsymbol{x}}^j)}{N-1} \beta(x_j - x_1). \end{aligned}$$

Now, thanks to individuals indistinguishability each term of the sum gives the same contribution, such that we can fix $j = 2$ without loss of generality:

$$\begin{aligned} & (N-1) \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\widetilde{x}_2 P(\widetilde{\boldsymbol{x}}^2) \frac{f(x_1, \widetilde{\boldsymbol{x}}^2)}{N-1} \beta(x_2 - x_1) \\ = & \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\widetilde{x}_2 P(\widetilde{\boldsymbol{x}}^2) f(x_1, \widetilde{\boldsymbol{x}}^2) \beta(x_2 - x_1). \end{aligned}$$

Next, we can integrate the mutation function using the fact that it is the only term depending on x_2 and $\int dx_2 \beta(x_2 - x_1) = 1$, leading to:

$$\int_{\mathcal{P}^N} d\tilde{x}_N^{\hat{1},2} P(\widetilde{\boldsymbol{x}}^2) f(x_1, \widetilde{\boldsymbol{x}}^2) \int dx_2 \beta(x_2 - x_1) = \int_{\mathcal{P}^N} d\tilde{x}_N^{\hat{1},2} P(\widetilde{\boldsymbol{x}}^2) f(x_1, \widetilde{\boldsymbol{x}}^2).$$

By considering the second term and applying the same procedure of above by fix $i = 2$ one obtains:

$$\sum_{i>1}^{N} \delta_{j1} \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\widetilde{x}_j P(\widetilde{\boldsymbol{x}}^j) \frac{f(x_i, \widetilde{\boldsymbol{x}}^j)}{N-1} \beta(x_j - x_i) = \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\widetilde{x}_1 P(\widetilde{\boldsymbol{x}}^1) f(x_2, \widetilde{\boldsymbol{x}}^1) \beta(x_1 - x_2).$$

Finally consider the last term. Thanks to indistinguishability we can fix $i = 2, j = 3$ by multiplying for a factor $(N-1)(N-2)$:

$$\begin{aligned} & \sum_{i>1}^{N} \sum_{j>1, i\neq j}^{N} \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\widetilde{x}_j P(\widetilde{\boldsymbol{x}}^j, t) \frac{f(x_i, \widetilde{\boldsymbol{x}}^j)}{N-1} \beta(x_j - x_i) \\ = & (N-2) \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\widetilde{x}_j P(\widetilde{\boldsymbol{x}}^3, t)(x_2, \widetilde{\boldsymbol{x}}^3) \beta(x_3 - x_2). \end{aligned} \tag{C.19}$$

Next, we can integrate the mutation function using the fact that it is the only term depending on x_3 and $\int dx_3\beta(x_3 - x_2) = 1$

$$(N-2)\int_{\mathcal{P}^N} d\tilde{x}_N^{\hat{1},3} d\tilde{x}_3 P(\tilde{\boldsymbol{x}}^3) f(x_2, \tilde{\boldsymbol{x}}^3) \int dx_3 \beta(x_3 - x_2) = (N-2)\int_{\mathcal{P}^N} d\tilde{x}_N^{\hat{1},3} P(\tilde{\boldsymbol{x}}^3) f(x_2, \tilde{\boldsymbol{x}}^3)$$

Putting the three terms together:

$$\begin{aligned} W_{gain} &= \int_{\mathcal{P}^N} d\tilde{x}_N^{\hat{1},2} P(\tilde{\boldsymbol{x}}^2) f(x_1, \tilde{\boldsymbol{x}}^2) + \int_{\mathcal{P}^N} dx_N^{\hat{1}} d\tilde{x}_1 P(\tilde{\boldsymbol{x}}^1) f(x_2, \tilde{\boldsymbol{x}}^1) \beta(x_1 - x_2) \\ &+ (N-2)\int_{\mathcal{P}^N} d\tilde{x}_N^{\hat{1},3} P(\tilde{\boldsymbol{x}}^3) f(x_2, \tilde{\boldsymbol{x}}^3) \end{aligned} \tag{C.20}$$

As a second step, we can consider the explicit form of the fitness function and use the indistinguishability of individuals to simplify more the expression. Namely, the fitness function can be written as the sum over the pairwise fitness over the population:

$$f(x_i, \boldsymbol{x}) = \sum_{k=1, k\neq i}^{N} \frac{f(x_i, x_k)}{N-1} \tag{C.21}$$

with:

$$f(x_i, x_k) = K(x_i) + I(x_i, x_k). \tag{C.22}$$

Let us insert explicitly this expression in the loss term and use individual indistinguishability

$$\begin{aligned} W_{loss} &= \int_{\mathcal{P}^{N-1}} dx_N^{\hat{1}} [f(x_1, \boldsymbol{x}) + (N-1) f(x_2, \boldsymbol{x})] P(\boldsymbol{x}, t) \\ &= \frac{1}{N-1} \int_{\mathcal{P}^{N-1}} dx_N^{\hat{1}} [\sum_{k=2}^{N} f(x_1, x_k) + (N-1) \sum_{k=1,\neq 2}^{N} f(x_2, x_k)] P(\boldsymbol{x}, t) \\ &= \int_{\mathcal{P}} dx_2 P(x_1, x_2) f(x_1, x_2) + (N-1)\int_{\mathcal{P}^2} dx_2 dx_2 P(x_1, x_2, x_3) f(x_2, x_3), \end{aligned} \tag{C.23}$$

where we have fixed $k = 2$ in the first term and $k = 3$ in the second. Similarly, one can reduce the gain term to:

$$\begin{aligned} W_{gain} &= \int_{\mathcal{P}} dx_2 P(x_1, x_2) f(x_1, x_2) \\ &+ \int_{\mathcal{P}^2} dx_2 dx_3 \big[P(x_2, x_3) f(x_2, x_3) \beta(x_1 - x_2) + (N-2) f(x_2, x_3) P(x_1, x_2, x_3) \big]. \end{aligned} \tag{C.24}$$

By combining them we obtain the marginalized Master equation:

$$\begin{aligned} &\partial_t \phi(x, t) = W_{gain} - W_{loss} \\ &= \int_{\mathcal{P}^2} dx_2 dx_3 \big[P(x_2, x_3) f(x_2, x_3) \beta(x_1 - x_2) - f(x_2, x_3) P(x_1, x_2, x_3) \big]. \end{aligned} \tag{C.25}$$

Mean Field Approximation

To go further with the analytic derivation it is necessary to perform a mean field approximation for the probability:

$$P(\boldsymbol{x}, t) = \Pi_{i=1}^{N} \phi(x_i, t), \tag{C.26}$$

that is expected to be true in the thermodynamic limit $N \to \infty$. By applying such ansatz to the marginalized Master eq.(C.25)

$$\begin{aligned} \partial_t \phi(x,t) &= \int_{\mathcal{P}} dx_2 dx_3 \phi(x_2)\beta(x_1 - x_2) \int_{\mathcal{P}} dx_3 \phi(x_3) f(x_2, x_3) \\ &- \phi(x,t) \int_{\mathcal{P}} dx_2 \phi(x_2) \int_{\mathcal{P}} dx_3 \phi(x_3) f(x_2, x_3) \end{aligned} \tag{C.27}$$

and by defining the "marginal fitness" as:

$$f(x,t) = \int_{\mathcal{P}} dx d\widetilde{x} f(x,\widetilde{x})\phi(\widetilde{x},t), \tag{C.28}$$

one finally obtain the Mean-field Master eq:

$$\partial_t \phi(x,t) = \int_{\mathcal{P}} \tilde{x} f(\tilde{x},t)\beta(x-\tilde{x})\phi(x,t) - \phi(x,t)\bar{f}(t) \tag{C.29}$$

with the average fitness defined by

$$\bar{f}(t) = \int_{\mathcal{P}} dx f(x,t)\phi(x,t) = \int_{\mathcal{P}^2} dx d\tilde{x} f(x,\tilde{x})\phi(x,t)\phi(\tilde{x},t). \tag{C.30}$$

C.1.2 Kramers-Moyal expansion

From the macroscopic one-body master equation we are interested in applying a small mutation approximation, i.e. considering that the offspring trait x is a small deviation from the ancestor, $x = \widetilde{x} + \delta$, $\delta \ll 1$. We consider a general mutation kernel $\beta(x - \widetilde{x}; \widetilde{x})$ where we leave the freedom of the dependence on the ancestor trait $\widetilde{x}$. This must be normalized in the jump amplitude $\delta = x - \widetilde{x}$such that the new trait is still inside the phenotypic space $\mathcal{P}$, that for simplicity we consider as an interval $[p_1, p_2]$

$$\int_{p_1-\widetilde{x}}^{p_2-\widetilde{x}} d(x-\widetilde{x})\beta(x-\widetilde{x};\widetilde{x}) = \int_{p_1-\widetilde{x}}^{p_2-\widetilde{x}} d\delta \beta(\delta;\widetilde{x}) = 1 \tag{C.31}$$

Furthermore, we consider it symmetric in this first variable:

$$\beta(\delta, x) = \beta(-\delta, x) \tag{C.32}$$

and with finite first and second moment:

$$\theta(x) = \int_{p_1-\widetilde{x}}^{p_2-\widetilde{x}} d\delta\delta\beta(\delta;x) \tag{C.33}$$

$$\sigma^2(x) = \int_{p_1-\widetilde{x}}^{p_2-\widetilde{x}} d\delta\delta^2\beta(\delta;x) \tag{C.34}$$

To this aim we rewrite the equation separating the part involving mutations from the rest:

$$\dot{\phi}(x,t) = \Delta_1\phi(x,t) + \Delta_2\phi(x,t) \tag{C.35}$$

where

$$\Delta_1\phi(x,t) = \int d\widetilde{x}[f(\widetilde{x})\beta(x-\widetilde{x};\widetilde{x})\phi(\widetilde{x},t) - f(x)\beta(x-\widetilde{x};x)\phi(x,t)] \tag{C.36}$$

and

$$\Delta_2\phi(x) = (f(x) - \bar{f})\phi(x,t) \tag{C.37}$$

If one consider the first term it is easy to see that it does conserve the probability and has the typical form of a Master equation:

$$\Delta_1\phi(x,t) = \int d\widetilde{x}\Big[W(\widetilde{x},\delta)\phi(\widetilde{x},t) - W(x,\delta)\phi(x,t)\Big] \tag{C.38}$$

with transition rate

$$W(\widetilde{x},\delta x) = f(\widetilde{x})\beta(\delta;\widetilde{x}). \tag{C.39}$$

Consider that now this rate can be written as function of the final trait and of the jump amplitude $\widetilde{x} = x - \delta$:

$$W(\widetilde{x},x) = f(\widetilde{x})\beta(x-\widetilde{x};\widetilde{x}) = f(x-\delta)\beta(\delta;x-\delta) := W(t-\delta;\delta) \tag{C.40}$$

In the same way perform the change of variable in the equation:

$$\Delta_1\phi(x,t) \quad = \quad \int_{x-p_2}^{x-p_1} d\delta[W(x-\delta;\delta)\phi(x-\delta) - W(x;\delta)\phi(x)] \tag{C.41}$$

where the change of integration variable is:

$$\int_{p_1}^{p_2} d\widetilde{x}W(\widetilde{x},x)\phi(\widetilde{x},t) \quad = \quad -\int_{x-p_1}^{x-p_2} d\delta W(x-\delta;\delta)\phi(x-\delta) = \int_{x-p_2}^{x-p1} d\delta W(x-\delta;\delta)$$

Now we can perform the classical taylor expansion in the *first* variable of the rate, assuming δ small

$$W(x-\delta;\delta) = W(x;\delta) - \delta\partial_x W(x;\delta) + \frac{\delta^2}{2}\partial_x^2 W(x;\delta) + O(\delta^3) \tag{C.42}$$

and insert it back in the equation (thanks to the Pawula theorem it sufficient to calculate the first two moments, if they are finite):

$$\Delta_1\phi(x) = -\partial_x m_1(x)\phi(x,t) + \frac{1}{2}\partial_x^2 m_2\phi(x,t) \tag{C.43}$$

where

$$m_1(x) \quad = \quad \int_{x-p_2}^{x-p_1} d\delta x\delta f(x)\beta(\delta;x) = 2f(x)\theta(x) \tag{C.44}$$

$$m_2(x) \quad = \quad \int_{x-p_2}^{x-p_1} d\delta x\delta^2 f(x)\beta(\delta;x) = \sigma^2(x)f(x) \tag{C.45}$$

Going back to the full equation we get :

$$\phi(x,t) \quad = \quad (f(x) - \bar{f})\phi(x,t) - \partial_x\theta(x)f(x)\phi(x,t) + \frac{1}{2}\partial_x^2\sigma^2(x)f(x)\phi(x), \tag{C.46}$$

that is a generalization the celebrated Crow-Kimura equation.

C.2 Landau-like Theory

C.2.1 4th order expansion

In this section we analyze the fitness function following the parsimonious principle of Landau[304, 9]. In this approach it is fundamental to understand the most important physical properties of the phenomenon in order to retain them and neglect the rest. This is achieved by expanding the function representing the properties of the system (the Hamiltonian in statistical physics while the fitness here), and individuate the fundamental terms based on symmetry. Hence, as a first step we Taylor expand the fitness function around the average value $\bar{x}$ till fourth order to grasp which terms are essential ones. We start from the basic interaction fitness and expand it in both variables:

$$\begin{aligned} f(x,y) &\approx f(\bar{x},\bar{x}) + f_1^x(\bar{x},\bar{x})(x-\bar{x}) + f_1^y(\bar{x},\bar{x})(y-\bar{x}) + \frac{1}{2}f_2^x(\bar{x},\bar{x})(x-\bar{x})^2 \\ &+ \frac{1}{2}f_2^y(\bar{x},\bar{x})(y-\bar{x})^2 + f_{11}^{xy}(\bar{x},\bar{x})(x-\bar{x})(y-\bar{x}) + \frac{1}{3!}f_3^x(\bar{x},\bar{x})(x-\bar{x})^3 \\ &+ \frac{1}{3!}f_3^y(\bar{x},\bar{x})(y-\bar{x})^3 + \frac{1}{2}f_{21}^{xy}(\bar{x},\bar{x})(x-\bar{x})^2(y-\bar{y}) \\ &+ \frac{1}{2}f_{12}^{xy}(\bar{x},\bar{x})(x-\bar{x})(y-\bar{y})^2 \frac{1}{4!}f_4^x(\bar{x},\bar{x})(x-\bar{x})^4 \\ &+ \frac{1}{4!}f_4^y(\bar{x},\bar{x})(y-\bar{x})^4 \frac{1}{6}f_{31}^{xy}(\bar{x},\bar{x})(x-\bar{x})^3(y-\bar{x}) \\ &+ \frac{1}{6}f_{13}^{xy}(\bar{x},\bar{x})(x-\bar{x})(y-\bar{x})^3 + \frac{1}{4}f_{22}^{xy}(\bar{x},\bar{x})(x-\bar{x})^2(y-\bar{y})^2; \end{aligned} \tag{C.47}$$

where the top indices indicate the variable in which respect the derivative is taken and the bottom one the order of the derivatives, i.e. $f_{22}^{xy}(\bar{x},\bar{x}) = \partial_x^2\partial_y^2 f(x,y)|_{x=y=\bar{x}}$. Once that the interaction fitness expansion has been carried out, by integrating over the trait distribution we obtain the "marginalized" fitness, i.e. the function that appears in the GCK eq:

$$f(x) = \int_{\mathcal{P}} dy f(x,y)\phi(y,t) \tag{C.48}$$

$$\begin{aligned} &\approx f(\bar{x},\bar{x}) + f_1^x(\bar{x},\bar{x})(x-\bar{x}) + \frac{f_2^x(\bar{x},\bar{x})}{2}(x-\bar{x})^2 + \frac{f_2^y(\bar{x},\bar{x})}{2}\Sigma(t) + \frac{1}{3!}f_3^x(\bar{x},\bar{x})(x-\bar{x})^3 \\ &+ \frac{1}{3!}f_3^y(\bar{x},\bar{x})\mu_3(t) + \frac{1}{2}f_{12}^{xy}(\bar{x},\bar{x})(x-\bar{x})\Sigma(t) + \frac{1}{4!}f_4^x(\bar{x},\bar{x})(x-\bar{x})^4 + \frac{1}{4!}f_4^y(\bar{x},\bar{x})\mu_4(t) \\ &+ \frac{1}{6}f_{13}^{xy}(\bar{x},\bar{x})(x-\bar{x})\mu_3(t) + \frac{1}{4}f_{22}^{xy}(\bar{x},\bar{x})(x-\bar{x})^2\Sigma(t); \end{aligned} \tag{C.49}$$

note that the terms proportional to $(y-\bar{x})$, i.e. $f_1^y, f_{11}^{xy}, f_{21}^{xy}$ and f_{31}^{xy} once integrated are zero because $\int_{\mathcal{P}} dy(y-\bar{x})\phi(y) = (\bar{x}-\bar{x}) = 0$. Finally, by averaging the marginal fitness one obtains the mean one:

$$\begin{aligned} \bar{f} &= \int_{\mathcal{P}} dx f(x)\phi(x,t) = f(\bar{x},\bar{x}) + \frac{f_2^x(\bar{x},\bar{x})}{2}\Sigma(t) + \frac{f_2^y(\bar{x},\bar{x})}{2}\Sigma(t) + \frac{1}{3!}f_3^x(\bar{x},\bar{x})\mu_3(t) \\ &+ \frac{1}{3!}f_3^y(\bar{x},\bar{x})\mu_3(t) + \frac{1}{4!}f_4^x(\bar{x},\bar{x})\mu_4(t) + \frac{1}{4!}f_4^y(\bar{x},\bar{x})\mu_4(t) + \frac{1}{4}f_{22}^{xy}\Sigma(t)^2. \end{aligned} \tag{C.50}$$

Now, we calculate the relative fitness, i.e. the selection coefficient of the traits. Such quantity is fundamental because it determines the fixed point of the GCK eq.:

$$\begin{aligned} f(x)-\bar{f} &= f_1^x(\bar{x},\bar{x})(x-\bar{x})+\frac{f_2^x(\bar{x},\bar{x})}{2}[(x-\bar{x})^2-\Sigma(t)]+\frac{1}{3!}f_3^x(\bar{x},\bar{x})[(x-\bar{x})^3-\mu_3] \\ &+ \frac{1}{2}f_{12}^{xy}(x-\bar{x})\Sigma(t)+\frac{1}{4!}f_4^x(\bar{x},\bar{x})[(x-\bar{x})^4-\mu_4]+\frac{1}{6}f_{13}^{xy}(x-\bar{x})\mu_3(t) \\ &+ \frac{1}{4}f_{22}^{xy}\Sigma(t)((x-\bar{x})^2-\Sigma). \end{aligned} \tag{C.51}$$

Here it is fundamental to note that all the fitness term *not* depending on the trait x are canceled out by the mean fitness, that we resume in the symbol$\widetilde{f}$, while the relevant terms are indicated as $F(x)$. Hence, to determine the fixed point of the dynamics it is not necessary to study the full fitness function, but just an effective one $F(x)$ that contains the relevant terms and gives the same relative fitness:

$$f(x)=F(x)+\widetilde{f},\quad f(x)-\bar{f} = F(x)-\bar{F} \tag{C.52}$$

$$\begin{aligned} F(x) &= f_1^x(\bar{x},\bar{x})(x-\bar{x})+\frac{f_2^x(\bar{x},\bar{x})}{2}(x-\bar{x})^2+\frac{1}{3!}f_3^x(\bar{x},\bar{x})(x-\bar{x})^3+\frac{1}{2}f_{12}^{xy}(x-\bar{x})\Sigma(t) \\ &+ \frac{1}{6}f_{13}^{xy}(x-\bar{x})\mu_3(t)+\frac{1}{4}f_{22}^{xy}(x-\bar{x})^2\Sigma(t)+\frac{1}{4!}f_4^x(\bar{x},\bar{x})(x-\bar{x})^4 \end{aligned} \tag{C.53}$$

$$\widetilde{f} = f(\bar{x},\bar{x})+\frac{f_2^y(\bar{x},\bar{x})}{2}\Sigma(t)+\frac{1}{3!}f_3^y(\bar{x},\bar{x})\mu_3(t)+\frac{1}{4!}f_4^y(\bar{x},\bar{x})\mu_4(t). \tag{C.54}$$

Finally, we can write down the effective fitness function in powers of $(x-\bar{x})$ like the Hamiltonian in statistical physics:

$$F(x) = g_1(t)(x-\bar{x})+\frac{g_2(t)}{2}(x-\bar{x})^2+\frac{g_3(t)}{3!}(x-\bar{x})^3+\frac{g_4(t)}{4!}(x-\bar{x})^4 \tag{C.55}$$

$$\begin{aligned} g_1(t) &= f_1^x(\bar{x},\bar{x})+\frac{1}{2}f_{12}^{xy}\Sigma(t)+\frac{1}{6}f_{13}^{xy}\mu_3(t),\quad g_2(t)=f_2^x+\frac{1}{2}f_{22}^{xy}\Sigma(t) \\ g_3(t) &= f_3^x(\bar{x},\bar{x}),\quad g_4(t)=f_4^x(\bar{x},\bar{x}). \end{aligned} \tag{C.56}$$

First of all, let us note that in principle all the coefficients depend on time, given that the average trait evolves. Second, the first and second order coefficients have some correction terms due to higher order terms which depend on the trait distribution moments. To better understand this mathematical structure, we follow once more the approach of Landau and search for a symmetry condition valid at *all* time. For example, the trait distribution $\phi(x,t)$ can be assumed to be always symmetric around its mean, imposing$\mu_3(t)=0$, together with all odd central moments. For simplicity one can also assume the interaction term appearing in the fitness function to be symmetric, leading to vanish all the cross derivative terms of odd order, such as $f_{12}^{xy}(\bar{x})$. In this way, Eq.(3.35) takes the simpler form:

$$\begin{aligned} F(x) &= f_1^x(\bar{x})(x-\bar{x})+\left(\frac{f_2^x(\bar{x})}{2}+\frac{f_{22}^{xy}(\bar{x})}{4}\Sigma(t)\right)(x-\bar{x})^2 \\ &+ \frac{f_3^x(\bar{x})}{3!}(x-\bar{x})^3+\frac{f_4^x(\bar{x})}{4!}(x-\bar{x})^4. \end{aligned} \tag{C.57}$$

Furthermore, at stationarity the symmetry condition has further consequences. First, consider the dynamics of the mean trait:

$$d_t\bar{x} = Cov((x-\bar{x}), f) = f_1^x(\bar{x})\Sigma(t) + \frac{f_3^x(\bar{x})}{3!}\mu_4(t), \tag{C.58}$$

given that the distribution needs to be symmetric around the mean, at the stationary state of the mean $\bar{x}^*$ the third order fitness needs to be zero $f_3^x(\bar{x}^*) = 0$. Hence, this term contributes to the dynamics but vanishes at the stationary state conserving the fact that the stationary mean needs to be an extreme point of the fitness $f_1^x(\bar{x}^*) = 0$. Hence, we can write down an even simpler effective fitness with just two terms that determines the stationary state of the system:

$$F(x) = \frac{g_2}{2}(x-\bar{x}^*)^2 + \frac{g_4}{4!}(x-\bar{x}^*)^4, \tag{C.59}$$

$$g_2 = f_2^x(\bar{x}^*) + \frac{1}{2}f_{22}^{xy}(\bar{x}^*)\Sigma, \tag{C.60}$$

$$g_4 = f_4^x(\bar{x}^*) \tag{C.61}$$

For sake of simplicity we set $\bar{x}^*$ without loss of generality. To study the branching phase we use as a first try a vanishing mutation approximation, assuming that the stationary distribution is a sum of two deltas:

$$\phi^*(x) = \frac{\delta(x-x^*)}{2} + \frac{\delta(x+x^*)}{2}. \tag{C.62}$$

Thanks to this approximation we can determine the variance $\Sigma^* = x^{*2}$ and search for extreme point of the fitness function:

$$\partial_x F(x) = 0, \rightarrow x\left(f_2^x + \frac{f_{22}^{xy}}{2}x^{*2} + \frac{f_4^x}{3!}x^2\right) = 0 \tag{C.63}$$

The possible solutions are:

$$x = 0, \quad x = x^* = \pm\sqrt{-6\frac{f_2^x}{f_4^x + 3f_{22}^{xy}}}. \tag{C.64}$$

The existence condition for the second and third points are:

$$f_2^x > 0, \wedge f_4^x + 3f_{22}^{xy} < 0 \tag{C.65}$$

or

$$f_2^x < 0, \wedge f_4^x + 3f_{22}^{xy} > 0 \tag{C.66}$$

The first point need to be a minimum, hence leading to:

$$\partial_x^2 F|_{x=0} > 0, \rightarrow g_2 = f_2^x + \tfrac{1}{2}f_{22}^{xy}x^{*2} > 0 \tag{C.67}$$

By using Eq.(C.64) the condition turns to be:

$$f_2^x\left(1 - 6\frac{f_{22}^{xy}}{f_4^x + 3f_{22}^{xy}}\right) > 0 \tag{C.68}$$

leading to :

$$f_2^x > 0 \wedge f_4^x < 3f_{22}^{xy} \tag{C.69}$$

Then, $\pm x^*$ need to be maxima, translating into the following conditions

$$\partial_x^2 F|_{x=\pm x^*} < 0, \rightarrow f_4^x < 0 \tag{C.70}$$

that is always trivially true. Summing up, the necessary condition to have branching are:

$$f_2^x > 0, \quad f_4^x < 0, \quad f_4^x < 3f_{22}^{xy}. \tag{C.71}$$

C.2.2 Bimodal trait distribution

Here we derive the equation determining the trait distribution in the branching phase using the 4th order theory and a bimodal ansatz for the stationary distribution:

$$\phi^*(x) = \frac{1}{2}N(-\mu,\omega) + \frac{1}{2}N(\mu,\omega), \tag{C.72}$$

where $N(\mu,\omega)$ is a normal distribution with mean μ and std ω. Its central moments are:

$$\begin{aligned} \bar{x} &= \mu_1 = 0, \quad \mu_2 = Var = \omega^2 + \mu^2, \quad \mu_3 = 0 \\ \mu_4 &= 3\omega^4 + 6\mu^2\omega^2 + \mu^4, \quad \mu_5 = 0, \quad \mu_6 = \mu^6 + 14\mu^4\omega^2 + 45\mu^2\omega^4 + 15\omega^6. \end{aligned} \tag{C.73}$$

From the 4th order theory the necessary conditions are: $f_2^x(0) > 0$ and $f_4^x(0) < 0$. Note that, in this simplified case, we need to find just two paramters, μ and ω.

First, by using the extremal condition for $x = \mu$ we obtain a relation between them:

$$\partial_x F(x)|_{x=\mu} = 0 \quad \rightarrow \mu^2 = -6\left(\frac{f_2^x + f_{22}^{xy}\omega^2}{f_4^x + 6f_{22}^{xy}}\right). \tag{C.74}$$

Second, we consider the stationary equation for the variance Σ:

$$\begin{aligned} d_t\Sigma^* &= \frac{1}{2}\left(f_2^x + \frac{f_4^{xy}}{2}(\mu_2)\right)\left(\mu_4 - \mu_2^2\right) + \frac{f_4^x}{4!}\left(\mu_6 - \mu_4\mu_2\right) \\ &+ \sigma^2\left(f(0) + \frac{f_2^x + f_2^y}{2}\mu_2 + \frac{f_{22}^{xy}}{4}\mu_2^2 + \frac{f_4^x + f_4^y}{4!}\mu_4\right) = 0; \end{aligned}$$

by inserting the explicit expression of the moments, eq.(C.73) it does convert to:

$$\begin{aligned} &\left(f_2^x + \frac{f_{22}^{xy}}{2}(\omega^2+\mu^2)\right)\omega^2(\omega^2+\mu^2) + \frac{f_4^x}{4!}\left(12\omega^6 + 36\omega^4\mu^2 + 7\mu^4\omega^2\right) \\ &+ \sigma^2\left(f(0) + \frac{f_2^x + f_2^y}{2}(\omega^2+\mu^2) + f_{22}^{xy}\left(\omega^2+\mu^2\right)^2 + \frac{f_4^x + f_4^y}{4!}\left(3\omega^4 + 6\mu^2\omega^2 + \mu^4\right)\right) = 0. \end{aligned} \tag{C.75}$$

By inserting Eq.(C.74) in Eq.(C.75) the problem reduces to solving cubic equation. Sadly, it is not feasible analytically and one needs to rely on numerical methods.

C.3 Finite size fluctuations

C.3.1 Derivation from microscopic process

To account for finite size fluctuations we proceed to derive an equation for $\rho(x,t)$, Eq.(C.12), that is a random variable describing the density or frequency of individuals with a certain trait x . Note that the average value of the phenotypic density gives the one-particle probability density, and hence recover all previous results. To derive a stochastic equation for ρ we follow the typical procedure used in statistical physics. First, we write down a Master equation in a discrete setting by "coarse-graining" the phenotypic space. Then, by using a Kramers-Moyal expansion [309] in the space of densities under a "local-noise" approximation we derive a Langevin equation for ρ. Finally, the continuum limit is performed. We start from the set of N individuals($\boldsymbol{x}_N$) and move to a coarse grained scale where we consider M "species", each one with phenotype y_i and density $\rho_i = n_i/N$ that sum to 1. To go from one scale to another we consider the following clustering, or binning, procedure:

1. divide the(one-dimensional) phenotypic space $\mathcal{P}$ in M intervals $I_i =]y_i - \delta_i, y_i + \delta_i]$
2. Individuals with phenotype in the same interval pertain to the same species.
3. integrate over the individual phenotype and collapse them on the species phenotype. All individuals of species have the same fitness, probability, and mutation function. The species fitness is the sum of individual fitness functions.
4. Both the fitness and the mutation function are now a discrete functions, i.e. elements of a matrix. The mutation function from a species i to j depend on the phenotypic distance of the two.

Hence, starting from the individual probability $P(\boldsymbol{x})$ and its Master eq.(C.9), first we write down the probability of the species abundances :

$$\begin{aligned} P(\boldsymbol{n},t) &= P(n_1,n_2,..,n_M,t) = \int_{\mathcal{P}^N} d\boldsymbol{x}\delta(x_1-y_{z_1})\delta(x_2-y_{z_2})..\delta(x_N-y_{z_M})P(\boldsymbol{x},t) \\ &= P(\underbrace{y_1,y_1,..,y_1}_{n_1 \text{ times}},\underbrace{y_2,y_2,..,y_2}_{n_2 \text{ times}},..,\underbrace{y_M,y_M,..,y_M}_{n_M \text{ times}},t). \end{aligned} \tag{C.76}$$

where the y_{z_j} are the species trait to which they pertain. By taking the time derivative and using the Master equation (3.3)we get to

$$\begin{aligned} d_tP(\boldsymbol{n}) &= \int_{\mathcal{P}^N} d\boldsymbol{x}_N\delta(x_1-y_{z_1})..\delta(x_N-y_{z_N}) \\ &\times \sum_{i,j}[\int_{\mathcal{P}} d\tilde{x}_j f(x_i,\tilde{\boldsymbol{x}}^j)\ \beta(x_j-x_i)d(\tilde{x}_j)P(\tilde{\boldsymbol{x}}^j) - P(\boldsymbol{x})f(x_i,\boldsymbol{x})d(x_j)] \end{aligned} \tag{C.77}$$

Now we apply the coarse-graining instructions

- *Coarse-craining the traits* One of the delta function, namely the one of x_i acts on the mutation functions:

$$\begin{aligned} &\int dx_i\delta(x_i-y_{z_i})f(x_i,\tilde{\boldsymbol{x}}^j)\ \beta(x_j-x_i)d(\tilde{x}_j)P(\tilde{\boldsymbol{x}}^j) \\ &= \beta(y_{z_i}-x_j)f(y_{z_i},\tilde{\boldsymbol{x}}^j)d(\tilde{x}_j)P(\tilde{\boldsymbol{x}}^j) \end{aligned} \tag{C.78}$$

  The other $N-1$ deltas act on the fitness and probability giving $f(y_{z_i},\boldsymbol{y}^{z_j,z_{\tilde{j}}})P(\boldsymbol{y}^{z_j,z_{\tilde{j}}},t)$ where we have assumed that the new offspring named x_j, and the dying individual,

$\tilde{x}_j$, pertain respectively to the species z_j and $z_{\tilde{j}}$,and hence the origin vector reads: where

$$\boldsymbol{y}^{z_j,z_{\tilde{j}}} = (\underbrace{y_1,..,y_1}_{n_1 \text{ times}},..,\underbrace{y_{z_i},..,y_{z_i}}_{n_{z_i} \text{ times}},..,\underbrace{y_{z_j},..,y_{z_j}}_{n_{z_j}-1 \text{ times}},..,\underbrace{y_M,..,y_M}_{n_M \text{ times}}) \tag{C.79}$$

- *From continuous to discrete space* Second, we rename the indices as $i = z_i, j = z_j$ and $k = z_{\tilde{j}}$, and sum over the possible transitions from state:

$$\widetilde{\boldsymbol{n}}^{j,k} = (n_1,..,n_i,..,n_j-1,..,n_k+1,..,n_M) \tag{C.80}$$

to state:

$$\boldsymbol{n} = (n_1,..,n_i,..,n_j,..,n_z,..,n_M) \tag{C.81}$$

where a individual of the the species k dies, one of$i$ reproduces by asexually, and the offspring mutates to species j. Note that i, j and k can be the same species. Naturally one defines the species fitness and death rate as the sum of all individuals in pertaining to it:

$$f(n_i,\boldsymbol{n}) = \sum_{j\in i} f(x_j,\boldsymbol{x}) = \sum_{j\in i} f(y_i,\boldsymbol{y}) = f(y_i,\boldsymbol{n})n_i, \qquad d(n_i,\boldsymbol{n}) = d(y_i,\boldsymbol{n})n_i. \tag{C.82}$$

Also, the βs now are discrete rates $\beta(y_i - y_j) = \beta_{i,j}$.

Summing all the terms we find:

$$d_t P(\boldsymbol{n},t) = \sum_{i,j,k}[T_i(\boldsymbol{n},\tilde{\boldsymbol{n}}^{j,k})P(\tilde{\boldsymbol{n}}^{j,k},t) - P(\boldsymbol{n},t)T_i(\tilde{\boldsymbol{n}}^{j,k},\boldsymbol{n})] \tag{C.83}$$

where

$$T_i(\boldsymbol{n},\tilde{\boldsymbol{n}}^{j,k}) = f(y_i,\tilde{\boldsymbol{n}}^{j,k})n_i d(y_k)(n_k+1)\beta_{i,j}. \tag{C.84}$$

Now we are interested in the diffusive approximation in the $N \gg 1$ limit, where we can derive a Langevin equation for the density of i species $\rho_i = \frac{n_i}{N}$, see [309] for a reference. Consider the following notation for the rates $T_i(\boldsymbol{n},\tilde{\boldsymbol{n}}^{j,k}) = T_i^{j,k}$; to calculate the moments of the expansion let us not that effectively there is just a flux of probability from the k species, who loses an individual, to the j one who obtains an offspring. The probability of the i species does not change and hence we can sum over it. Hence, the first moment of the expansion corresponds to(remember that $d_i = 1/(N-1)$):

$$A_j = \frac{1}{N}\sum_{i,k}[T_i^{j,k} - T_i^{k,j}] = \frac{1}{N}\sum_{ik}^{M} f_i b_{ij} n_i \frac{n_k}{N-1} - \frac{1}{N}\sum_{ik}^{M} f_i \beta_{ik} n_i \beta_{\frac{n_j}{N-1}}; \tag{C.85}$$

assuming that $\sum_k n_k = N \approx N-1$, we obtain the deterministic part of the equation:

$$A_j = \sum_{i=1}^{M} \beta_{ij} f_i \rho_i - \bar{f}\rho_j, \tag{C.86}$$

that is the discretized version of the mean-field equation. In the same way we can calculate the diffusion matrix :

$$B_{jk} = \frac{1}{N^2}\sum_i^M [T_i^{j,k} + T_i^{k,j}] = \frac{1}{N}[f_i m_i \beta_{ij} + f_j m_j \beta_{ji} + \sum_k^M f_k m_k (\beta_{ki} m_j + \beta_{kj} m_i)] \quad \text{(C.87)}$$

$$B_{jj} = \frac{1}{N^2}\sum_{i,k,k\neq j}^M [T_i^{j,k} + T_i^{k,j}] = \frac{1}{N}\sum_i f_i b_{ij}\rho_i + \frac{1}{N}\sum_i f_i \rho_i \rho_j, \quad \text{(C.88)}$$

where we used tha $\sum_k \beta_{ik} = 1$. Now, to write down a Langevin equation for ρ_i, it would be necessary to diagonalize the matrix B and find the matrix C that follows the relation $C^T C = B$ to obtain:

$$\dot{\rho}_j = A_j + \sum_{k=1}^M C_{ik}\xi_k. \quad \text{(C.89)}$$

As clearly expressed in [309] this can be done numerically but not analytically in full generality. Given that our aim here is to derive an analytical expression for such finite size fluctuations in the limit of *infinite* species, it is natural to assume a "local noise approximation", i.e. $B_{jk} = 0$ for $j \neq k$. In this case the Langevin equation is trivially determined as:

$$\dot{\rho}_j = \sum_{i=1}^M \beta_{ij} f_i \rho_i - \bar{f}\rho_j + \sqrt{\frac{\sum_i f_i b_{ij}\rho_i + \bar{f}\rho_j}{N}}\xi_j, \quad \text{(C.90)}$$

$$\langle \xi_i(t)\rangle = 0 \quad \text{(C.91)}$$

$$\langle \xi_i(t)\xi_j(t')\rangle = \delta(t-t')\delta_{ij} \quad \text{(C.92)}$$

By taking the continuum limit, i.e. $M \to \infty$, $\delta \to 0$ we finally obtain the stochastic mutation-selection eq.:

$$\dot{\rho}(x,t) = \int_{\mathcal{P}} d\tilde{x} f(\tilde{x})\rho(\tilde{x})\beta(x-\tilde{x}) - \bar{f}\rho(x) + \sqrt{\frac{\int_{\mathcal{P}} d\tilde{x} f(\tilde{x})\rho(\tilde{x})\beta(x-\tilde{x}) + \bar{f}\rho(x)}{N}}\xi(x,t) \quad \text{(C.93)}$$

Finally, by considering a small mutation approximation both in the deterministic and stochastic part we end up with the stochastic generalized Crow-Kimura eq:

$$\begin{aligned}\dot{\rho}(x,t) &= \left(f(x)-\bar{f}\right)\rho(x) - \partial_x \theta f(x)\rho(x) + \frac{1}{2}\partial_x^2 \sigma^2 f(x)\rho(x) \\ &+ \sqrt{\frac{\left(f(x)+\bar{f}\right)\rho(x) - \partial_x\theta f(x)\rho(x) + \frac{1}{2}\partial_x^2\sigma^2 f(x)\rho(x)}{N}}\xi(x,t) \\ \langle \xi(x,t)\rangle &= 0 \\ \langle \xi(x,t)\xi(y,t')\rangle &= \delta(t-t')\delta(x-y) \end{aligned} \quad \text{(C.94)}$$

In the case of trait-independent and undirected mutaions we obtain the following simplified stochastic GCK eq.:

$$\begin{aligned}\partial_t \rho(x) &= (f(x)-\bar{f})\rho(x) + \frac{\sigma^2}{2}\partial_x^2 f(x)\rho(x) + \sqrt{\frac{(f(x)+\bar{f})\rho(x) + \frac{\sigma^2}{2}\partial_x^2 f(x)\rho(x)}{N}}\xi(x,t) \\ \langle \xi(x,t)\rangle &= 0 \\ \langle \xi(x,t)\xi(y,t')\rangle &= \delta(x-y)\delta(t-t') \end{aligned} \quad \text{(C.95)}$$

C.3.2 Langevin eqs. for the moments

In this section we will derive a couple of Langevin equation for the trait mean and variance first in a general fashion and then in the context of Gaussian theory. For sake of simplicity we limit ourselves to trait-independent and consider the mean trait evolution

$$\begin{aligned} d_t\bar{x}(t) &= \int dx(x-\bar{x})\partial_t\rho = f_1\Sigma(t) + \frac{f_2}{2}\mu_3 \\ &+ \int dx(x-\bar{x})\sqrt{\frac{(f+\bar{f})\rho - \partial_x\theta f\rho + 1/2\partial_x^2\sigma^2 f\rho}{N}}\xi(x,t). \end{aligned} \tag{C.96}$$

where we have used Eq(3.51). The stochastic term can be written down as an effective noise in the following form:

$$\eta_{\bar{x}}(t) = \int dx(x-\bar{x})\sqrt{\frac{(f+\bar{f})\rho - \partial_x\theta f\rho + 1/2\partial_x^2\sigma^2 f\rho}{N}}\xi(x,t); \tag{C.97}$$

$$\langle\eta_{\bar{x}}(t)\rangle = 0 \tag{C.98}$$

$$\begin{aligned} \langle\eta_{\bar{x}}(t)\eta_{\bar{x}}(t')\rangle &= \frac{\delta(t-t')}{N}\int dx(x-\bar{x})[(f+\bar{f})\rho - \partial_x\theta f\rho + 1/2\partial_x^2\sigma^2 f\rho] \\ &= \frac{\delta(t-t')}{N}[\overline{(x-\bar{x})^2 f} + \Sigma\bar{f} + 2\overline{(x-\bar{x})\theta f} + \overline{\sigma^2 f}] \end{aligned} \tag{C.99}$$

The same can be done for the trait variance Σ (where we leave implicit the deterministic part as G_Σ)

$$\begin{aligned} d_t\Sigma &= G_\Sigma + \int dx(x-\bar{x})^2\sqrt{\frac{(f+\bar{f})\rho - \partial_x\theta f\rho + 1/2\partial_x^2\sigma^2 f\rho}{N}}\xi(x,t) \\ \eta_\Sigma(t) &= \int dx(x-\bar{x})\sqrt{\frac{(f+\bar{f})\rho - \partial_x\theta f\rho + 1/2\partial_x^2\sigma^2 f\rho}{N}}\xi(x,t) \end{aligned} \tag{C.100}$$

$$\langle\eta_\Sigma(t)\rangle = 0 \tag{C.101}$$

$$\langle\eta_\Sigma(t)\eta_\Sigma(t')\rangle = \frac{\delta(t-t')}{N}\left(\overline{(x-\bar{x})^4 f} + \mu_4\bar{f} + 4\overline{(x-\bar{x})^3\theta f} + 6\overline{(x-\bar{x})^2\sigma^2 f}\right) \tag{C.102}$$

The two effective noises are correlated as follows

$$\langle\eta_{\bar{x}}(t)\eta_\Sigma(t')\rangle = \frac{\delta(t-t')}{N}\left(\overline{(x-\bar{x})^3 f} + \mu_3\bar{f} + 3\overline{(x-\bar{x})^2\theta f} + 3\overline{(x-\bar{x})\sigma^2 f}\right). \tag{C.103}$$

Now by considering undirected mutations together with the Gaussian approximation, the noise moments reduce to a simpler form:

$$\langle\eta_{\bar{x}}(t)\eta_{\bar{x}}(t')\rangle = \frac{\delta(t-t')}{N}\left(f(\bar{x})(2\Sigma+\sigma) + f_2(2\Sigma+\sigma/2)\right) \tag{C.104}$$

$$\langle\eta_\Sigma(t)\eta_\Sigma(t')\rangle = \frac{3\delta(t-t')}{N}\left(2f(\bar{x}) + 3f_2\Sigma\right)\left(\Sigma + \sigma^2\Sigma\right) \tag{C.105}$$

$$\langle\eta_{\bar{x}}(t)\eta_\Sigma(t')\rangle = \frac{3\delta(t-t')}{N}f_1^x(\Sigma + \sigma\Sigma). \tag{C.106}$$

Finally, we get to the couple of simplified Langevin eqs. that determine the trait distribution in Gaussian theory:

$$d_t\bar{x} = f_1^x\Sigma + \sqrt{\frac{f(\bar{x})(2\Sigma+\sigma) + f_2(2\Sigma+\sigma/2)}{N}}\eta_{\bar{x}}(t) \tag{C.107}$$

$$\begin{aligned} d_t\Sigma &= f_2^x(\bar{x})\Sigma^2 + \sigma^2 f(\bar{x}) + \frac{\sigma^2}{2}f_2(\bar{x})\Sigma + \sqrt{3\frac{(2f(\bar{x})+3f_2\Sigma)(\Sigma+\sigma^2\Sigma)}{N}}\eta_\Sigma(t) \\ \langle\eta_{\bar{x}}(t)\rangle &= 0, \quad \langle\eta_\Sigma(t)\rangle = 0, \quad \langle\eta_{\bar{x}}(t)\eta_{\bar{x}}(t')\rangle = \delta(t-t'), \quad \langle\eta_\Sigma(t)\eta_\Sigma(t')\rangle = \delta(t-t') \\ \langle\eta_{\bar{x}}(t)\eta_\Sigma(t')\rangle &= \frac{3\delta(t-t')f_1^x(\Sigma+\sigma\Sigma)}{\sqrt{3\left(f(\bar{x})(2\Sigma+\sigma)+f_2(2\Sigma+\sigma/2)\right)(2f(\bar{x})+3f_2\Sigma)(\Sigma+\sigma^2\Sigma)}} \end{aligned} \tag{C.108}$$

C.3.3 Variance analysis

To study the stationary behavior of the trait variance, we assume that the mean trait has already converged, enabling us to study the variance independently. Furthermore, for sake of simplicity and intuition we consider just the first fitness term $f(\bar{x})$ in the coefficient of the stochastic noise, obtaining the following simplified equation:

$$d_t\Sigma = f_2^x(\bar{x})\Sigma^2 + \sigma^2 f(\bar{x}) + \frac{1}{2}\sigma^2 f_2(\bar{x})\Sigma + \sqrt{\frac{3f(\bar{x})(\Sigma+\sigma^2\Sigma)}{N}}\eta_\Sigma(t). \tag{C.109}$$

Equivalently to Eq.(C.109), one can consider the following Fokker-Planck equation for the variance probability distribution $P(\Sigma)$, in the Ito discretization scheme:

$$\partial_t P(\Sigma,t) = -\partial_\Sigma[A(\Sigma)P(\Sigma,t) - \partial_\Sigma\frac{B(\Sigma)}{2N}P(\Sigma,t)], \tag{C.110}$$

with

$$A(\Sigma) = f_2^x(\bar{x})\Sigma^2 + \sigma^2 f(\bar{x}) + \frac{1}{2}\sigma^2 f_2(\bar{x})\Sigma, \tag{C.111}$$

$$B(\Sigma) = 3f(\bar{x})\left(\Sigma+\sigma^2\Sigma\right). \tag{C.112}$$

Then, we search for a stationary solution $P^*(\Sigma)$:

$$\partial_t P^*(\Sigma) = 0, \qquad A(\Sigma)P^*(\Sigma) - \partial_\Sigma\frac{B(\Sigma)}{2}P^*(\Sigma) = 0 \tag{C.113}$$

leading to:

$$P^* = \frac{1}{Z}\exp\left(-V_N(\Sigma)\right) \tag{C.114}$$

with the potential reading:

$$\begin{aligned} V_N(\Sigma) &= -2N\int d\Sigma\frac{A(\Sigma)}{B(\Sigma)} + \log B(\Sigma) \\ &= \frac{2N}{f(\bar{x}^*)}\left(\left(f_2 - \frac{f_2^x}{2}\right)\sigma^2 + f(\bar{x}^*)\right)\log\left(\Sigma+\sigma^2\right) \\ &- 2N\log\Sigma - \frac{2Nf_2^x}{f(\bar{x}^*)}\Sigma - \log\left(f(\bar{x})\left(\Sigma^2+\Sigma\sigma^2\right)\right), \end{aligned} \tag{C.115}$$

where in chapter 3 for simplicity we have assumed $\left(f_2 - \frac{f_2^x}{2}\right) \approx f_2^x/2$ in the first term. Now, we search for the extreme point of the potential, $\partial_\Sigma V_N(\Sigma^*) = 0$, leading to the following solution:

$$\Sigma^*_{N1,2} = \frac{\sigma^2 b_N}{2f_2^x(x^*)a_N}\Big(\pm \sqrt{\sigma^2 f_2^2(x^*) - 4f(x^*)f_2^x(x^*)a_N c_N/b_N} - f_2(x^*)\Big) \tag{C.116}$$

with the size-dependent coefficients

$$a_N = \left(1 - \frac{27 f_2}{N f_2^x}\right), \quad b_N = \left(1 - \frac{3}{N f_2 \sigma^2}\left(2f(\bar{x}) - 3f_2\sigma^2\right)\right), \quad c_N = \left(2 - \frac{3}{N}\right). \tag{C.117}$$

Depending on the determinant sign Δ_N, Eq.(C.116) has zero $\Delta_N < 0$, one Δ_N or two $\Delta_N > 0$ reals solutions, that correspond to the three regimes reported in chapter 3. The condition of $\Delta_N > 0$ can be translated into a condition for the size N to be smaller than a critical one N^*:

$$N < N^* \frac{f(\bar{x}^*)}{\sigma}\left(\frac{8\sqrt{f(\bar{x}^*)} + 2\sigma\sqrt{f_2^x}}{\sqrt{f_2^x}(16 f(\bar{x}^*) - f_2^x\sigma^2)}\right) \approx \frac{f(\bar{x}^*)}{\sigma} \tag{C.118}$$

C.4 From microscopic to macroscopic entropy production

C.4.1 Detailed balance

The "Detailed-balance condition" is a fundamental sufficient condition for Master Equations to converge to an equilibrium stationary state [89]. Physically it is illuminating, cause it detects the presence of currents capable of driving the system away from equilibrium. Consider a couple of possible microstates:

$$W_i(\mathbf{x}, \widetilde{\mathbf{x}}^j)P^*(\widetilde{\mathbf{x}}^j) = W_i(\widetilde{\mathbf{x}}^j, \mathbf{x})P^*(\mathbf{x}) \tag{C.119}$$

$$f_i(\widetilde{\mathbf{x}}^j)\beta(x_j - x_i)P^*(\widetilde{\mathbf{x}}^j) = f_i(\mathbf{x})\beta(\widetilde{x}_j - x_i), \tag{C.120}$$

where $P^*(\mathbf{x})$ is the stationary probability of being in the configuration $\mathbf{x}$ and $W_i(\widetilde{\mathbf{x}}^j, \mathbf{x})$ the rate of jumping from configuration $\mathbf{x}$ to $\widetilde{\mathbf{x}}^j$. If we translate the condition in words this means that the probability of a transition from a microstate to an other by a reaction is as probable as the inverse. Now we will apply the mean-field condition to the probability: procedure to the detailed-balance condition to derive the equilibrium condition for the GCK Eq. Now, we apply the mean field conditions

$$P(x_1, x_2, x_3, ..., x_N) = \Pi_{i=1}^n \phi(x_i) \tag{C.121}$$

$$f_i(\widetilde{\mathbf{x}}^j)\phi^*(\widetilde{x}_j)\beta(x_j - x_i) = f_i(\mathbf{x})\phi^*(x_j)\beta(\widetilde{x}_j - x_i) \tag{C.122}$$

If furthermore we consider $N \lll 1$ then $f_i(\widetilde{\mathbf{x}}^j) \approx f_i(\mathbf{x})$ the detailed balance condition does not involve the fitness :

$$\phi^*(\widetilde{x}_j)\beta(x_j - x_i) = \phi^*(x_j)\beta(\widetilde{x}_j - x_i) \tag{C.123}$$

The mean-field mutation-selection equation can be rewritten as a non-linear master eq

$$W_y(\widetilde{x} \to x) = f(y)\beta(x - \widetilde{x})\phi(y) \tag{C.124}$$

leading to the (marginalized) master equation:

$$\phi(\dot{x},t) = \int d\tilde{x}dy W_y(\widetilde{x}\to x)\phi(\tilde{x}_1) - \int d\tilde{x}_2 dy W_y(x\to\widetilde{x})\phi(x) \tag{C.125}$$

So, the detailed balance condition for the mean-field mutation-selection Eq. is:

$$W_y(\widetilde{x}\to x)\phi(\tilde{x}) = W_y(x\to\widetilde{x})\phi(x) \tag{C.126}$$
$$f(y)\phi(y)\phi(\tilde{x})\beta(x-y) = f(y)\phi(x)\phi(y)\beta(\tilde{x}-y) \tag{C.127}$$
$$\phi(\tilde{x})\beta(x-y) = \phi(x)\beta(\tilde{x}-y) \tag{C.128}$$

that indeed coincides with Eq.(C.123).

C.4.2 Microscopic Entropy production

Consider the Gibbs-Shannon entropy divided by the number of individuals:

$$S^{micro}(t) = -\frac{1}{N}\int d\mathbf{x}P(\mathbf{x},t)\log P(\mathbf{x},t), \tag{C.129}$$

we are interested in its time evolution. Using the Master Eq.(3.3)

$$\begin{aligned}\dot{S}^{micro}(t) &= -\frac{1}{N}\int d\mathbf{x}\frac{dP(\mathbf{x_N},t)}{dt}\log P(\mathbf{x},t)\\ &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\int d\mathbf{x}d\tilde{x}_j[W_i(\mathbf{x^j},\mathbf{x})P(\mathbf{x^j}t) - W_i(\mathbf{x},\mathbf{x^j})P(\mathbf{x},t)]\log P(\mathbf{x},t)\end{aligned} \tag{C.130}$$

To gain physical insights we follow Lebowitz and Spohn [550] and divide the quantity in two terms using simple manipulations. As first step, using the integration variable symmetries we insert the probability of the other microstates in the log:

$$\dot{S}^{micro}(t) = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\int d\mathbf{x}d\tilde{x}_j[W_i(\mathbf{x^j},\mathbf{x})P(\mathbf{x^j}t) - W_i(\mathbf{x},\mathbf{x^j})P(\mathbf{x},t)]\log\frac{P(\mathbf{x^j}t)}{P(\mathbf{x},t)}. \tag{C.131}$$

Then, by multiplying and dividing for the rates ratio, one obtains two different contributions:

$$\dot{S}^{micro}(t) = \dot{S}^{micro}_{prod} - \dot{S}^{micro}_{flux}, \tag{C.132}$$

where the first is the non-negative Schnakenberg entropy production:

$$\dot{S}^{micro}_{prod} = \frac{1}{2N}\sum_{i,j=1,j\neq i}^{N}\int d\mathbf{x}\widetilde{x}_j[W_i(\mathbf{x^j},\mathbf{x})P(\mathbf{x^j}t) - W_i(\mathbf{x},\mathbf{x^j})P(\mathbf{x},t)]\log\frac{W_i(\mathbf{x^j},\mathbf{x})P(\mathbf{x^j}t)}{W_i(\mathbf{x},\mathbf{x^j})P(\mathbf{x},t)} \geq 0 \tag{C.133}$$

and the second is the entropy flux:

$$S^{micro}_{flux}(t) = \frac{1}{2N}\sum_{i,j=1,j\neq i}^{N}\int d\mathbf{x}d\tilde{x}_j[W_i(\mathbf{x^j},\mathbf{x})P(\mathbf{x^j}t) - W_i(\mathbf{x},\mathbf{x^j})P(\mathbf{x},t)]\log\frac{W_i(\mathbf{x^j},\mathbf{x})}{W_i(\mathbf{x},\mathbf{x^j})} \quad \text{(C.134)}$$

To simplify the notation, in the following section we will remove the t dependency in the probability distribution. Hence, the entropy production is :

$$\begin{aligned}\dot{S}^{micro}_{prod} &= \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\int d\mathbf{x}d\tilde{x}_j[f(x_i,\mathbf{x^j})\beta(x_j-x_i)d_jP(\mathbf{x^j}) + \\ &- f(x_i,\mathbf{x})\beta(\widetilde{x}_j-x_i)d_jP(\mathbf{x}))]\log\frac{f(x_i,\mathbf{x^j})\beta(x_j-x_i)\mu_jP(\mathbf{x^j})}{f(x_i,\mathbf{x})\beta(\widetilde{x}_j-x_i)d_jP(\mathbf{x}))} \quad \text{(C.135)}\end{aligned}$$

We can simplify the expression by choosing $i = 1$ and $j = 2$. Furthermore, by noting that all the variables are integrated the negative term in the current can be eliminated by multiplying the total expression by two:

$$\dot{S}^{micro}_{prod} = \int d\mathbf{x}d\widetilde{x}_2 f(x_1,\mathbf{x^2})\beta(x_2-x_1)P(\mathbf{x^2})\log\frac{f(x_1,\mathbf{x^2})\beta(x_2-x_1)P(\mathbf{x^2})}{f(x_1,\mathbf{x})\beta(\widetilde{x}_2-x_1)P(\mathbf{x})}, \quad \text{(C.136)}$$

and

$$\dot{S}^{micro}_{flux} = \int d\mathbf{x}d\widetilde{x}_2 f(x_1,\mathbf{x^2}\beta(x_2-x_1)P(\mathbf{x^2},t)\log\frac{f(x_1,\mathbf{x^2})\beta(x_1-\widetilde{x}_1)\beta(x_2-x_1)}{f(x_1,\mathbf{x})\beta(\widetilde{x}_1-x_1)\beta(\widetilde{x}_2-x_1)} \quad \text{(C.137)}$$

For notation sake we introduce the currents between microstates:

$$J_i(\mathbf{x^j x}) = W_i(\mathbf{x^j},\mathbf{x})P(\mathbf{x^j}t) - W_i(\mathbf{x},\mathbf{x^j})P(\mathbf{x},t), \quad \text{(C.138)}$$

and the analogue of thermodynamics forces, that we call *evolutionary* forces[319]:

$$R_i(\mathbf{x^j x}) = \log\frac{W_i(\mathbf{x^j},\mathbf{x})}{W_i(\mathbf{x},\mathbf{x^j})}. \quad \text{(C.139)}$$

As a consequence, the entropy flux is the product of current and forces

$$\dot{S}^{micro}_{flux} = \int d\mathbf{x}d\widetilde{x}_2 J_1(\mathbf{x^2 x})R_1(\mathbf{x^2 x}) \quad \text{(C.140)}$$

Furthermore, note that we can decompose the evolutionary forces in two contributions:

$$R_1(\mathbf{x^2 x}) = \log\frac{f(x_1,\mathbf{x^2})}{f(x_1,\mathbf{x})} + \log\frac{\beta(x_2-x_1)}{\beta(\widetilde{x}_2-x_1)}, \quad \text{(C.141)}$$

where the first involves just the fitness while the second only mutations. This relation leads to decompose the entropy flux as:

$$\begin{aligned}\dot{S}^{micro}_{flux} &= \int d\mathbf{x}d\widetilde{x}_2 f(x_1,\mathbf{x^2})\beta(x_2-x_1)P(\mathbf{x^2})\log\frac{f(x_1,\mathbf{x^2})}{f(x_1,\mathbf{x})}\\ &+ \int d\mathbf{x}d\widetilde{x}_2 f(x_1,\mathbf{x^2})\beta(x_2-x_1)P(\mathbf{x^2})\log\frac{\beta(x_2-x_1)}{\beta(\widetilde{x}_2-x_1)}\\ &= \varphi_f(t)+\nu_\beta(t),\end{aligned} \tag{C.142}$$

and similarly the entropy production

$$\begin{aligned}\dot{S}^{micro}_{prod} &= \int d\mathbf{x}d\widetilde{x}_2 f(x_1,\mathbf{x^2})\beta(x_2-x_1)P(\mathbf{x^2})\log\frac{f(x_1,\mathbf{x^2})}{f(x_1,\mathbf{x})}\\ &+ \int d\mathbf{x}d\widetilde{x}_2 f(x_1,\mathbf{x^2})\beta(x_2-x_1)P(\mathbf{x^2})\log\frac{\beta(x_2-x_1)}{\beta(\widetilde{x}_2-x_1)}\\ &+ \int d\mathbf{x}d\widetilde{x}_2 f(x_1,\mathbf{x^2})\beta(x_2-x_1)P(\mathbf{x^2})\log\frac{P(\mathbf{x^2})}{P(\mathbf{x})}\\ &= \varphi_f+\nu_\beta+\dot{S}^{micro}=\dot{S}^{micro}_{flux}+\dot{S}^{micro}.\end{aligned} \tag{C.143}$$

The first contribution φ_f is proportional to fitness difference of the reproducing individual in the present and past state, while the second one ν_β takes into consideration the different probabilities of mutation between individuals.

C.4.3 Mean-field limit approximation

In the following section, we derive the macroscopic entropy production by applyng a mean-field limit the to the microscopic one. The calculation is quite cumbersome, so we carry it out in the simplest scenario. Let us start directly with the approximation ansatz, in this section we use the following approximations:

- Large population size $N \ggg 1$
- Mean field approximation for the probability distribution $P(\mathbf{x}_N) = \phi(x_1)..\phi(x_N)$
- Fitness from two body interaction $f(x_1,\mathbf{x}) = \sum_k^N \frac{1}{N-1} f(x_1)$.
- Gaussian Kernels with zero mean and small variance(in comparison with the phenotypic space), such that also in a finite dominium of $[a,b]$the truncation terms are negligible:

$$\begin{aligned}\beta_x(\delta) &= \frac{e^{-\frac{\delta^2}{2\sigma^2}}}{\sqrt{\frac{\sigma^2\pi}{2}}\Big(Erf((b-x)/\sqrt{2}\sigma)-Erf((a-x)/\sqrt{2}\sigma)\Big)}\\ &\approx \frac{e^{-\frac{\delta^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}=\beta_x(-\delta)\end{aligned} \tag{C.144}$$

First, let us comment that in this limit we can consider as negligible the contribution φ_f. Indeed, given that the fitness is weighted sum of many two-body interaction terms, if just one term is modified the change in the individual fitness is negligible:

$$f(x_1,\mathbf{x}^2)\approx f(x_1,\mathbf{x})\to\varphi_f\approx 0 \tag{C.145}$$

Second, we apply the mean-field ansatz to the net change in entropy :

$$\dot{S}_{mf} = \int d\mathbf{x} d\tilde{x}_2 f(x_1, \mathbf{x^2}) \beta(x_2 - x_1) P(\mathbf{x^2}) \log \frac{\phi(\tilde{x}_2)}{\phi(\tilde{x}_2)}. \tag{C.146}$$

Now, by using the fact that the fitness is weighted sum of two-body interactions we get to:

$$\begin{aligned} \dot{S}_{mf} &= \frac{N-2}{N-1} \int dx_1 dx_2 d\tilde{x}_2 f(x_1) \beta(x_2 - x_1) \phi(x_1) \phi(\tilde{x}_2) \log \frac{\phi(\tilde{x}_2)}{\phi(\tilde{x}_2)} \\ &+ \frac{1}{N-1} \int dx_1 dx_2 d\tilde{x}_2 f(x_1, \tilde{x}_2) \beta(x_2 - x_1) \phi(x_1) \phi(\tilde{x}_2) \log \frac{\phi(\tilde{x}_2)}{\phi(\tilde{x}_2)}, \end{aligned} \tag{C.147}$$

where in the first term the fitness has been marginalized :

$$f(x_1) = \int dx_k f(x, x_k) \phi(x_k). \tag{C.148}$$

Given that $N \ggg 1$, the second contribution is sub-leading and we get to:

$$\dot{S}_{mf} = \int dx_1 dx_2 d\tilde{x}_2 f(x_1) \beta(x_2 - x_1) \phi(x_1) \phi(\tilde{x}_2) \log \frac{\phi(\tilde{x}_2)}{\phi(\tilde{x}_2)}. \tag{C.149}$$

Now, we will like to perform a small mutation approximation, and to do that we use the change of variable $x_2 = \delta + x_1$:

$$\dot{S}_{mf} = \int dx_1 d\tilde{x}_2 d\delta f(x_1) \beta(\delta) \phi(x_1) \phi(\tilde{x}_2) \log \frac{\phi(\tilde{x}_2)}{\phi(\delta + x_1)}. \tag{C.150}$$

By performing the expansion around $\delta = 0$ up to second order, we obtain two terms:

$$\begin{aligned} \dot{S}_{mf} &= \dot{S}^0_{mf} + \dot{S}^2_{mf} \\ &= \int dx_1 d\tilde{x}_2 f(x_1) \phi(x_1) \phi(\tilde{x}_2) \log \frac{\phi(\tilde{x}_2)}{\phi(x_1)} \\ &- \frac{\sigma^2}{2} \int dx_1 d\tilde{x}_2 f(x_1) \phi(x_1) \phi(\tilde{x}_2) \partial^2_{x_1} \log \phi(x_1). \end{aligned} \tag{C.151}$$

The zero order term can be easily rewritten as:

$$\begin{aligned} \dot{S}^0_{mf} &= -\int dx_1 dx_2 f(x_1) \phi(x_1) \log \phi(x_1) + \int dx_1 d\tilde{x}_2 f(x_1) \phi(x_1) \phi(\tilde{x}_2) \log \phi(\tilde{x}_2) \\ &= \Sigma[f, I]. \end{aligned} \tag{C.152}$$

On the other hand, in the second term we can first we can integrate by part in x_1, and obtain:

$$\dot{S}^2_{mf} = \frac{\sigma^2}{2} \int dx_1 d\tilde{x}_2 \partial_{x_1} (f(x_1) \phi(x_1)) \partial_{x_1} \log \phi(x_1), \tag{C.153}$$

where we have assumed the surface term to be zero. Then, we can use the definition of mutation current $j = -(\sigma^2/2) \partial_x f \phi$ and following relation:

$$\frac{\partial_x \phi(x)}{\phi(x)} = -\frac{j(x)}{\sigma^2 f(x) \phi(x)} - \frac{\partial_x f(x)}{f(x)}, \tag{C.154}$$

to rewrite the second term as:

$$\dot{S}^2_{mf} = \frac{1}{2}\int dx \frac{j(x)^2}{\sigma^2 f(x)\phi(x)} + \frac{1}{2}\int j\frac{\partial_x f(x)}{f(x)} \tag{C.155}$$

$$= \frac{1}{2}\overline{\frac{v^2}{\sigma^2 f}} + \frac{1}{2}\overline{v\partial_x \log f}. \tag{C.156}$$

Hence, we obtain the same entropy change derived from the GCK Eq.6.3.
Now, let us consider the contribution to the entropy production ν:

$$\nu_{mf}(t) = \int dx_1 dx_2 d\delta \beta(\delta) f(x_1)\phi(x_1)\phi(x_2)\log\frac{\beta(x_2-x_1)}{\beta(\delta)}, \tag{C.157}$$

where we have approximated the marginalized fitness as before, and renamed $x_2 = \tilde{x}_2$. ν can be decomposed in two contributions:

$$\nu_{mf}(t) = \nu_A(t) + \nu_A(t) \tag{C.158}$$

$$\nu_A(t) = \int dx_1 dx_2 d\delta \beta(\delta) f(x_1)\phi(x_1)\phi(x_2)\log\beta(\delta) = -\bar{f}S[\beta] \tag{C.159}$$

$$\nu_B(t) = \int dx_1 dx_2) f(x_1)\phi(x_1)\phi(x_2)\log\beta(x_2-x_1). \tag{C.160}$$

While the first contribution has a very intuitive form, the second one needs to be worked out with the Gaussian ansatz for the mutational kernel :

$$\begin{aligned}\nu_B &= -\int dx_1 dx_2 d\delta \beta(\delta) f(x_1)\phi(x_1)\phi(x_2)\log\beta(x_2-x_1)\\ &= \frac{1}{2}\int dx_1 dx_2 d\delta \beta(\delta) f(x_1)\phi(x_1)\phi(x_2)\left(\frac{(x_1-x_2)^2}{\sigma^2} + \log(2\pi\sigma^2)\right)\\ &= \frac{1}{2\sigma^2}\left(\overline{x^2 f} + (\overline{x}^2)\overline{f} - 2(\overline{x})\overline{xf}\right) - \frac{1}{2}\log(2\pi\sigma^2)\bar{f}.\end{aligned} \tag{C.161}$$

By using the covariance between the second moment and fitness we obtain that:

$$\Sigma((x-\bar{x})^2, f) = \overline{x^2 f} + (\overline{x}^2)\overline{f} - 2(\overline{x})\overline{xf} - \Sigma\overline{f}, \tag{C.162}$$

it can be further rewritten as:

$$\begin{aligned}\nu_B &= \frac{1}{2\sigma^2}\left(\Sigma((x-\bar{x})^2, f) + \Sigma\bar{f}\right) + S[\beta]\overline{f}\\ &= \frac{1}{2\sigma^2}\left(d_t\Sigma + (\Sigma-\sigma^2)\bar{f}\right) + S[\beta]\overline{f}.\end{aligned} \tag{C.163}$$

Hence, by summing the two terms, ν_A and ν_B we get to:

$$\nu_{mf}(t) = \frac{1}{2\sigma^2}\left(d_t\Sigma + (\Sigma-\sigma^2)\bar{f}\right). \tag{C.164}$$

In resume, we have calculated the entropy production in adaptive evolution using a mean-field limit and a Gaussian mutation approximation, and it reads:

$$\dot{S}^{mf}_{prod} = \dot{S}^{mf}_{flux} + \dot{S}^{mf} \tag{C.165}$$

$$\dot{S}^{mf}_{flux} = \frac{1}{2\sigma^2}\left(d_t\Sigma + (\Sigma - \sigma^2)\bar{f}\right) \tag{C.166}$$

$$\dot{S}^{mf} = \Sigma[f, I] + \frac{1}{2}\frac{\overline{v^2}}{\sigma^2 f} + \frac{1}{2}\overline{v\partial_x \log f} \tag{C.167}$$

C.5 Simulation algorithm

To simulate exactly and efficiently the microscopic Master Eq.(C.9) for the Growth-competition model, we used the following version of the Gillespie algorithm. Consider N individuals at time t_0:

- The fitness function of all individuals is calculated as the weighted sum over the interaction with all the population:

$$f_i(\boldsymbol{x}) = \sum_{j=1, j\neq i}^{N} \frac{f(x_i, x_j)}{N-1} \tag{C.168}$$

  for all i.

- A random uniform distributed number $\chi \in [0,1]$ is generated to calculate the time interval for the next event:

$$\delta t = -\log\chi/(\sum_{i=1}^{N} f_i(\boldsymbol{x})) \tag{C.169}$$

- Consider the representation of the reproduction probabilities of all individuals $q_i = f_i(\boldsymbol{x})/(\sum_{i=1}^{N} f_i(\boldsymbol{x}))$. A random uniform distributed number $\eta \in [0,1]$ is generated. Using a standard search algorithm the value of i such that

$$q_i < \eta <= q_{i+1} \tag{C.170}$$

  is found. The i individual will reproduce.

- The j individual is sampled randomly and will die. Hence, the j individual is replaced by the offspring of the i individual. Its new trait $\tilde{x}_j$ is the mother one plus a variation δ_j sampled form the distribution $\beta\delta$:

$$\tilde{x}_j = x_i + \delta_j \tag{C.171}$$

- time is updated to $t_0 = t + \delta t$. Iterate.

Appendix D

Evolution of tolerance supplementary materials

D.1 The Microscopic process

Consider the dynamics of a population of N individual bacteria cyclically exposed to antibiotics, as described in chapter 4. Each bacteria is characterized by its intrinsic lag time $\tau_i \in \mathcal{P} = \mathbb{R}^+$, where $\mathcal{P}$ denotes the continuous one−dimensional phenotype space of possible lag time values. In addition, each individual can be in two different states, as indicated by the variable y_i: if $y_i = 1$ the individual is in the awake state, such that it can reproduce, while if $y_i = -1$ it is in dormant state, where no activity happens. Finally, the environment, represented by the parameter η, varies between two states characterized by the presence ($\eta = -1$) or absence ($\eta = 1$) of antibiotics. The environmental dynamics is considered to be cyclical with period T_{max}, $\eta(t) = \eta(t + T_{max})$, and within each cycle:

$$\eta = \begin{cases} -1 & \text{if} \quad 0 \leq T \leq T_a \\ +1 & \text{if} \quad T_a < T \leq T_{max} \end{cases} \tag{D.1}$$

where $T \in [0, T_{max}]$ is the time within a cycle and T_a is the duration of the antibiotic phase.

The whole population state is thus represented by the vectors $\boldsymbol{\tau}_N = (\tau_1, \tau_2, .., \tau_i, .., \tau_N)$ and $\boldsymbol{y}_N = (y_1, y_2, .., y_i, .., y_N)$. The vector subindex indicates the dimensionality, i.e., the number of bacteria: $dim(\boldsymbol{\tau}_N) = N, dim(\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1}) = N - 1$. To further proceed, let us consider that all individuals are indistinguishable, i.e., that the system state is invariant under individual permutations; for example: $\boldsymbol{\tau}_N = (\tau_1, \tau_2, .., \tau_i, .., \tau_N) = \boldsymbol{\tau}'_{\boldsymbol{N}} = (\tau_i, \tau_1, \tau_N, \ldots, \tau_2, \ldots, \tau_3)$. The following notation is implemented in order to account for changes in the population; consider a reference state $\boldsymbol{\tau}_N = (\tau_1, \tau_2, \ldots, \tau_i, \ldots, \tau_j, \ldots, \tau_N)$; then a "modified" state with respect to $\boldsymbol{\tau}_N$ is defined as $\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1} = (\tau_1, \tau_2, \ldots, \tilde{\tau}_i, \ldots, \tau_{j-1}, \tau_{j+1}, .., \tau_N) = (\tau_1, \tau_2, \ldots, \tilde{\tau}_i, \ldots, \tau_{N-1})$, where the i individual has a different trait, $\tilde{\tau}_i \neq \tau_i$, and the j one is absent, while the rest of individuals have the same traits in both states. Note that the property of indistinguishability has been used to derive the last equality.

Having defined the states, one can define the dynamics. In particular, each individual cell can perform diverse stochastic "reactions" (using the jargon of stochastic processes), that are illustrated in Fig 1 of chapter 4.

Any individual i in the growing state($y_i = 1$) with trait $\tilde{\tau}_i$ is exposed to a demographic-adaptive process (Δ_D):

- It can try to reproduce asexually at rate $g(\tau_i, \boldsymbol{\tau}_N) = b \in \mathbb{R}^+$. On the one hand, if the antibiotics are present ($\eta = -1$), the reproduction does not happen and the

individual is killed. On the other hand, in fresh environment, ($\eta = 1$), the reproduction is successful and the individual is replaced by two offspring with phenotypes $\tau_i = \tilde{\tau}_i + \xi_i$ and $\tau_j = \tilde{\tau}_i + \xi_j$, where ξ_i, ξ_j are the variations, which are sampled from probability distribution $\beta(\xi_i = \tau_i - \tilde{\tau}_i; \tilde{\tau}_i)$ and $\beta(\xi_j = \tau_j - \tilde{\tau}_i; \tilde{\tau}_i)$ respectively. If the population size N has reached the carrying capacity K this reproduction event must be accompanied by the death of individual j with trait τ_j, chosen at random (i.e. with rate $\mu = \frac{1}{K-1}$), much as in the Moran process. In principle, β is an arbitrary probability distribution normalized in the jump amplitude (the first argument): $\int_{-\tilde{\tau}}^{\infty} \beta(\tau - \tilde{\tau}; \tilde{\tau}) d(\tau - \tilde{\tau}) = 1$, but it can depend on the initial phenotype, that is indicated with the second argument. In the following, for the sake of notation, the second argument will be omitted when not necessary.

- If in the awake state, it can die by natural causes at rate $\mu(\tau_i, \boldsymbol{\tau}_N) = d \in \mathbb{R}^+$.

Furthermore, any individual with trait τ_i can switch on and off from dormancy with the following processes (Δ_{GD}):

- If dormant ($y_i = 1$), it can wake up at rate $A(\tau_i, \boldsymbol{\tau}_N) = 1/\tau_i$.
- If awake ($y_i = 1$) and in the presence of antibiotics ($\eta = -1$) it can sense their presence and switch to dormancy at rate $S(\tau_i, \boldsymbol{\tau}_N) = s \in \mathbb{R}^+$.

D.1.1 Master equation

The general trait-dependent expression of the rates will be used for the sake of theoretical comprehensiveness, and it will be reduced to the particular one used in chapter 4 when necessary. From these processes it is possible to write down a master equation for probability of the system configurations $P(\boldsymbol{\tau}_N, \mathbf{y}_N, t)$:

$$\frac{dP(\boldsymbol{\tau}_N, \mathbf{y}_N, t)}{dt} = \Delta_D P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) + \Delta_{GD} P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t), \tag{D.2}$$

where Δ_{BD} stands for the time change with respect to the demographic part of the dynamics, and the Δ_{GD} for the *growing-dormant* state transitions. In addition, it is useful to separate the demographic change into two contributions, the *birth-death* part, i.e. the set of processes that change the population size N, and the *Moran* processes part that drives the dynamics when the carrying capacity K is reached and does not change N:

$$\frac{dP(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{dt} = (1 - \delta_{N,K})\Delta_{BD} P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) + \delta_{N,K}\Delta_M P(\boldsymbol{\tau}_N, \mathbf{y}_N, t) + \Delta_{GD} P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t). \tag{D.3}$$

Birth-death contribution To take into account the change in population size, it is assumed that the probability of a state $(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)$ can be written as the ratio of some population state function $\psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)$, $\psi : \mathcal{P} \to \mathcal{R}^{0,+}$, and its normalization:

$$P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) \equiv \frac{\psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y} \in \{\pm 1\}^N} \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}. \tag{D.4}$$

ψ is non-negative, and its integral is heuristically assumed not to diverge, but it is not a probability function since it is not normalized. It can be thought as a scalar field generalizing, to population configurations, the concept of the number of individuals of a particular species in a discrete set of them (e.g. in evolutionary game theory, where the frequency of species i is written as $p_i = \frac{n_i}{N}$, where n_i is the number of individual of species

i and N the population size). The time evolution of the probability density is specified by:

$$\begin{aligned}
&\Delta_{BD} P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) = \\
&\frac{\dot{\psi}(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm 1\}^N} \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)} + \\
- &\frac{\psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm 1\}^N} \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)} \frac{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm 1\}^N} \dot{\psi}(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm 1\}^N} \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)} = \\
= &\frac{\dot{\psi}(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} \sum_{\mathbf{y}\in\{\pm 1\}^N} d\boldsymbol{\tau}_N \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)} - P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) \frac{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm 1\}^N} \dot{\psi}(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm 1\}^N} \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)} = \\
= &\frac{\dot{\psi}(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} \sum_{\mathbf{y}\in\{\pm 1\}^N} d\boldsymbol{\tau}_N \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)} - \Lambda_N(t) P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t), \qquad \text{(D.5)}
\end{aligned}$$

where we have defined the normalizing factor:

$$\Lambda_N(t) \equiv \frac{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm 1\}^N} \dot{\psi}(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm 1\}^N} \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}. \qquad \text{(D.6)}$$

Note that in (D.5) the probability is conserved, so once the equation of motion for ψ is written down, one has an equation for P. Thus, for the Δ_{BD} term it is mandatory to take into account this normalization, whereas it is not necessary for the Δ_{GD} and Δ_M terms, which do not change N.

Let us consider the various processes that change the population size:

- Birth, that increases the population's size of one, e.g. from a state $\tilde{\tau}^{i,\hat{j}}_{N-1}$ to τ_N, with rate:

$$\mathcal{B}(\boldsymbol{\tau}_N, \boldsymbol{y}_N | \tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1} \tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1}) = \beta(\tau_i - \tilde{\tau}_i)\beta(\tau_j - \tilde{\tau}_i) g(\tilde{\tau}_i, \tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1}) \delta_{\eta,1} \delta_{\tilde{y}_i,1} \delta_{y_i,1} \delta_{y_j,1}. \qquad \text{(D.7)}$$

  where the i individual with trait $\tilde{\tau}_i$ produces two offspring with traits τ_i, τ_j. Naturally, to reproduce individual i must be in the growing state and antibiotics must be absent.

- Death of the i individual by the effect of antibiotics, which reduces the population size from N to $N-1$ at rate:

$$\mathcal{Q}(\boldsymbol{\tau}^{\hat{i}}_{N-1}, \boldsymbol{y}^{\hat{i}}_{N-1} | \boldsymbol{\tau}_N, \boldsymbol{y}_N) = g(\tau_i, \boldsymbol{\tau}_N) \delta_{\eta,-1} \delta_{y_i,1}, \qquad \text{(D.8)}$$

  where the index $\hat{i}$ represent the absence of the i individual with respect to $\boldsymbol{\tau}_N$.

- Natural death of individual i at rate:

$$\mathcal{T}(\boldsymbol{\tau}^{\hat{i}}_{N-1}, \boldsymbol{y}^{\hat{i}}_{N-1} | \boldsymbol{\tau}_N, \boldsymbol{y}_N) = \mu(\tau_i, \boldsymbol{\tau_N}) \delta_{y_i,1}. \qquad \text{(D.9)}$$

  Thus, by adding these two rates (eq.D.8-D.9) one gets the full death rate:

$$\mathcal{D}(\boldsymbol{\tau}^{\hat{i}}_{N-1}, \boldsymbol{y}^{\hat{i}}_{N-1} | \boldsymbol{\tau}_N, \boldsymbol{y}_N) \equiv \mathcal{Q}(\boldsymbol{\tau}^{\hat{i}}_{N-1}, \boldsymbol{y}^{\hat{i}}_{N-1} | \boldsymbol{\tau}_N, \tilde{\boldsymbol{y}}_N) + \mathcal{T}(\boldsymbol{\tau}^{\hat{i}}_{N-1}, \boldsymbol{y}^{\hat{i}}_{N-1} | \boldsymbol{\tau}_N, \boldsymbol{y}_N). \qquad \text{(D.10)}$$

Putting together all these contributions, Δ_{BD} can be written as:

$$\Delta_{BD}P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t) = \tag{D.11}$$

$$\begin{aligned}
&= \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{\tilde{y}_i\in\{\pm1\}}\int_{\mathcal{P}} d\tilde{\tau}_i \frac{1}{N-1}\mathcal{B}(\boldsymbol{\tau}_N,\boldsymbol{y}_N|\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1},\tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1})\frac{\psi(\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1},\tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1},t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm1\}^N}\psi(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)}\\
&+ \sum_{i=1}^{N+1}\sum_{y_i\in\{\pm1\}}\int_{\mathcal{P}} d\tau_i \frac{1}{N+1}\mathcal{D}(\boldsymbol{\tau}^{\hat{i}}_N,\boldsymbol{y}^{\hat{i}}_N|\boldsymbol{\tau}_{N+1},\boldsymbol{y}_{N+1})\frac{\psi(\boldsymbol{\tau}_{N+1},\boldsymbol{y}_{N+1},t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm1\}^N}\psi(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)}\\
&- \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N+1}\sum_{\tilde{y}_i\in\{\pm1\}}\int_{\mathcal{P}^2} d\tilde{\tau}_i d\tilde{\tau}_j \frac{1}{N}\mathcal{B}(\tilde{\boldsymbol{\tau}}^{i,j}_{N+1},\tilde{\boldsymbol{y}}^{i,j}_{N+1}|\boldsymbol{\tau}_N,\boldsymbol{y}_N)\frac{\psi(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm1\}^N}\psi(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)}\\
&- \sum_{i=1}^{N}\sum_{y_i\in\{\pm1\}}\mathcal{D}(\boldsymbol{\tau}^{\hat{i}}_{N-1},\boldsymbol{y}^{\hat{i}}_{N-1}|\boldsymbol{\tau}_N,\boldsymbol{y}_N)\frac{\psi(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm1\}^N}\psi(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)}
\end{aligned}$$

$$- \quad P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,\boldsymbol{y}_N,t)\Lambda_N(t). \tag{D.12}$$

Note that in the first three lines it is necessary to normalise the sums by the initial population size to respect individuals' indistinguishability. Consider for example the first line: it quantifies the probabilities that the state $\boldsymbol{\tau}_N$ is the result of a birth event in the state $\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1}$ where the i individual has reproduced. One of the offspring is indicated as x_i but there are $N-1$ possibilities of assigning the index j to the second one. Due to individual indistinguishability all these possibilities are equivalent, and therefore the sum must be normalized by $N-1$.

To rewrite this expression in terms of probabilities, we normalize it in the following way:

$$\begin{aligned}
\frac{\psi(\boldsymbol{\tau}_N)}{\int_{\mathcal{P}^M} d\boldsymbol{\tau}_M\psi(\boldsymbol{\tau}_M,t)} &= \frac{\psi(\boldsymbol{\tau}_N)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N\psi(\boldsymbol{\tau}_N,t)}\frac{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N\psi(\boldsymbol{\tau}_N,t)}{\int_{\mathcal{P}^M} d\boldsymbol{\tau}_M\psi(\boldsymbol{\tau}_M,t)}\\
&= P(\boldsymbol{\tau}_N)\frac{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N\psi(\boldsymbol{\tau}_N,t)}{\int_{\mathcal{P}^M} d\boldsymbol{\tau}_M\psi(\boldsymbol{\tau}_M,t)} \equiv P(\boldsymbol{\tau}_N)a_{N,M}
\end{aligned} \tag{D.13}$$

where the re-scaling factor is defined as:

$$a_{N,M} \equiv \frac{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N\psi(\boldsymbol{\tau}_N,t)}{\int_{\mathcal{P}^M} d\boldsymbol{\tau}_M\psi(\boldsymbol{\tau}_M,t)}. \tag{D.14}$$

Thanks to this definition, Δ_{BD} reads:

$$\begin{aligned}
\Delta_{BD}P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t) &= a_{N-1,N}\sum_{i=1}^{N}\sum_{j=1,i\neq j}^{N}\sum_{\tilde{y}_i\in\{\pm1\}}\int_{\mathcal{P}} d\tilde{\tau}_i \frac{1}{N-1}\mathcal{B}(\boldsymbol{\tau}_N,\boldsymbol{y}_N|\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1},\tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1})P(\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1},\tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1})\\
&+ a_{N+1,N}\sum_{i=1}^{N+1}\sum_{y_i\in\{\pm1\}}\int_{\mathcal{P}} d\tau_i \frac{1}{N+1}\mathcal{D}(\boldsymbol{\tau}^{\hat{i}}_N,\boldsymbol{y}^{\hat{i}}_N|\boldsymbol{\tau}_{N+1},\boldsymbol{y}_{N+1})P(\boldsymbol{\tau}_{N+1},\boldsymbol{y}_{N+1},t)\\
&- \sum_{i=1}^{N}\sum_{j=1\neq i}^{N+1}\sum_{\tilde{y}_i\in\{\pm1\}}\int_{\mathcal{P}^2} d\tilde{\tau}_i d\tilde{\tau}_j \frac{1}{N}\mathcal{B}(\tilde{\boldsymbol{\tau}}^{i,j}_{N+1},\tilde{\boldsymbol{y}}^{i,j}_{N+1}|\boldsymbol{\tau}_N,\boldsymbol{y}_N)P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)\\
&- \sum_{i=1}^{N}\sum_{y_i\in\{\pm1\}}\mathcal{D}(\boldsymbol{\tau}^{\hat{i}}_{N-1},\boldsymbol{y}^{\hat{i}}_{N-1}|\boldsymbol{\tau}_N,\boldsymbol{y}_N)P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)\\
&- \Lambda_N(t)P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,\boldsymbol{y}_N,t).
\end{aligned} \tag{D.15}$$

Let us now calculate explicitly the normalization factor $\Lambda_N(t)$:

$$\begin{aligned}
&\Lambda_N(t) = \\
&\frac{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm1\}} \dot{\psi}(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\mathbf{y}\in\{\pm1\}} \psi(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)} = \\
&= a_{N-1,N} \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N-1} \sum_{j=1, i\neq j}^{N} \sum_{\tilde{y}_i\in\{\pm1\}} \int_{\mathcal{P}} d\tilde{\tau}_i \frac{1}{N-1} \mathcal{B}(\boldsymbol{\tau}_N, \boldsymbol{y}_N | \tilde{\boldsymbol{\tau}}_{N-1}^{i,\hat{j}} \tilde{\boldsymbol{y}}_{N-1}^{i,\hat{j}}) P(\tilde{\boldsymbol{\tau}}_{N-1}^{i,\hat{j}}, \tilde{\boldsymbol{y}}_{N-1}^{i,\hat{j}}) \\
&+ a_{N+1,N} \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N+1} \sum_{y_i\in\{\pm1\}} \int_{\mathcal{P}} d\tau_i \frac{1}{N+1} \mathcal{D}(\boldsymbol{\tau}_N^{\hat{i}}, \boldsymbol{y}_N^{\hat{i}} | \boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1}) P(\boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1}, t) \\
&- \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N} \sum_{j=1\neq i}^{N+1} \sum_{\tilde{y}_i=\pm1} \int_{\mathcal{P}^2} d\tilde{\tau}_i d\tilde{\tau}_j \frac{1}{N} \mathcal{B}(\tilde{\boldsymbol{\tau}}_{N+1}^{i,j}, \tilde{\boldsymbol{y}}_{N+1}^{i,j} | \boldsymbol{\tau}_N, \boldsymbol{y}_N) P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) \\
&- \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N} \sum_{\tilde{y}_i\in\{\pm1\}} \mathcal{D}(\boldsymbol{\tau}_{N-1}^{\hat{i}}, \boldsymbol{y}_{N-1}^{\hat{i}} | \boldsymbol{\tau}_N, \boldsymbol{y}_N) P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t).
\end{aligned} \tag{D.16}$$

The expression can be split in four terms:

$$\Lambda_N(t) = \Lambda_1 + \Lambda_2 - \Lambda_3 - \Lambda_4, \tag{D.17}$$

and each of them can be calculated separately. The first term is:

$$\begin{aligned}
\Lambda_1 &= a_{N-1,N} \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N} \sum_{j=1, i\neq i}^{N} \sum_{\tilde{y}_i\in\{\pm1\}} \int_{\mathcal{P}} d\tilde{\tau}_i \frac{1}{N-1} \mathcal{B}(\boldsymbol{\tau}_N, \boldsymbol{y}_N | \tilde{\boldsymbol{\tau}}_{N-1}^{i,\hat{j}} \tilde{\boldsymbol{y}}_{N-1}^{i,\hat{j}}) P(\tilde{\boldsymbol{\tau}}_{N-1}^{i,\hat{j}}, \tilde{\boldsymbol{y}}_{N-1}^{i,\hat{j}}) \\
&= a_{N-1,N} \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \sum_{\tilde{y}_i=\pm1} \\
&\times \int_{\mathcal{P}} d\tilde{\tau}_i \frac{1}{N-1} \beta(\tau_i - \tilde{\tau}_i) \beta(\tau_j - \tilde{\tau}_i) g(\tilde{\tau}_i, \tilde{\boldsymbol{\tau}}_{N-1}^{i,\hat{j}}) \delta_{\eta,1} \delta_{\tilde{y}_i,1} \delta_{y_i,1} \delta_{y_j,1} P(\tilde{\boldsymbol{\tau}}_{N-1}^{i,\hat{j}}, \tilde{\boldsymbol{y}}_{N-1}^{i,\hat{j}}).
\end{aligned} \tag{D.18}$$

Using the fact that individuals are indistinguishable the sums $\sum_{i=1}^{N}$, $\sum_{j=1,i\neq j}^{N}$ can be replaced with a factor $N \cdot (N-1)$ by fixing i and j, e.g. $i = 1, j = 2$:

$$\begin{aligned}
\Lambda_1 &= N \cdot a_{N-1,N} \sum_{\boldsymbol{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{\tilde{y}_1\in\{\pm1\}} \\
&\times \int_{\mathcal{P}} d\tilde{\tau}_1 \beta(\tau_1 - \tilde{\tau}_1) \beta(\tau_2 - \tilde{\tau}_1) g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}}) \delta_{\eta,1} \delta_{\tilde{y}_1,1} \delta_{y_1,1} \delta_{y_2,1} P(\tilde{\boldsymbol{\tau}}_N^{1,\hat{2}}, \tilde{\boldsymbol{y}}_N^{1,\hat{2}}) \\
&= N \cdot a_{N-1,N} \sum_{\boldsymbol{y}\in\{\pm1\}} \int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}} \sum_{\tilde{y}_1=\pm1} \int_{\mathcal{P}} d\tau_1 \beta(\tau_1 - \tilde{\tau}_1) \\
&\times \int_{\mathcal{P}} d\tau_2 \beta(\tau_2 - \tilde{\tau}_1) g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}}) \delta_{\eta,1} \delta_{\tilde{y}_1,1} \delta_{y_1,1} \delta_{y_2,1} P(\tilde{\boldsymbol{\tau}}_N^{1,\hat{2}}, \tilde{\boldsymbol{y}}_N^{1,\hat{2}}).
\end{aligned} \tag{D.19}$$

By using the normalization of the β's in the variable $\delta = \tau - \tilde{\tau}$, $\int_0^\infty d\tau_1 \beta(\tau_1 - \tilde{\tau}_1) = \int_{-\tilde{\tau}_1}^\infty d\delta \beta(\delta) = \int_0^\infty d\tau_2 \beta(\tau_2 - \tilde{\tau}_1) = 1$, one obtains:

$$\Lambda_1 = N \cdot a_{N-1,N} \sum_{\boldsymbol{y}\in\{\pm1\}} \int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}} \sum_{\tilde{y}_1\in\{\pm1\}} g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}}) \delta_{\eta,1} \delta_{\tilde{y}_1,1} \delta_{y_1,1} \delta_{y_2,1} P(\tilde{\boldsymbol{\tau}}_N^{1,\hat{2}}, \tilde{\boldsymbol{y}}_N^{1,\hat{2}}). \tag{D.20}$$

By performing a marginalization $\sum_{y\in\{\pm 1\}} P(\tilde{\boldsymbol{\tau}}_N^{1,\hat{2}}, \tilde{\boldsymbol{y}}_N^{1,\hat{2}}) = P(\tilde{\boldsymbol{\tau}}_N^{1,\hat{2}}, \tilde{y}_1)$ and defining the average birth and death rates:

$$\bar{g}_N = \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N g(\tau_i, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, y_i, t)\delta_{y_i,1}, \tag{D.21}$$

$$\bar{\mu}_N = \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \mu(\tau_i, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, y_i, t)\delta_{y_i,1}, \tag{D.22}$$

one readily obtains:

$$\begin{aligned}\Lambda_1 &= N\cdot a_{N-1,N}\sum_{\mathbf{y}\in\{\pm1\}}\int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}} \sum_{\tilde{y}_1\in\{\pm1\}} g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}})\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1}\delta_{y_2,1}P(\tilde{\boldsymbol{\tau}}_N^{1,\hat{2}}, \tilde{\boldsymbol{y}}_N^{1,\hat{2}}) \\ &= a_{N-1,N}N\bar{g}_{N-1}\delta_{\eta,1}.\end{aligned} \tag{D.23}$$

Similarly, one can work out the second term:

$$\begin{aligned}\Lambda_2 &= a_{N+1,N}\sum_{\mathbf{y}\in\{\pm1\}}\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N+1}\sum_{y_i\in\{\pm1\}}\int_{\mathcal{P}} d\tau_i \frac{1}{N+1}\mathcal{D}(\boldsymbol{\tau}_N^{\hat{i}}, \boldsymbol{y}_N^{\hat{i}}|\boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1})P(\boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1}, t) \\ &= a_{N+1,N}\sum_{\mathbf{y}\in\{\pm1\}}\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{y_{N+1}\in\{\pm1\}} \\ &\times \int_{\mathcal{P}} d\tau_{N+1}P(\boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1}, t)\left[\mu(\tau_{N+1}, \boldsymbol{\tau_{N+1}})\delta_{y_{N+1},1} + g(\tau_{N+1}, \boldsymbol{\tau}_{N+1})\delta_{\eta,-1}\delta_{y_{N+1},1}\right] = \\ &= a_{N+1,N}(\bar{\mu}_{N+1} + \bar{g}_{N+1}\delta_{\eta,-1}),\end{aligned} \tag{D.24}$$

where i has been fixed to $N+1$. The third one reads:

$$\begin{aligned}&\Lambda_3 = \\ &= \sum_{\mathbf{y}\in\{\pm1\}}\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N}\sum_{j=1\neq i}^{N+1}\sum_{\tilde{y}_i\in\{\pm1\}}\int_{\mathcal{P}^2} d\tilde{\tau}_i d\tilde{\tau}_j \frac{1}{N}\mathcal{B}(\tilde{\boldsymbol{\tau}}_{N+1}^{i,j}, \tilde{\boldsymbol{y}}_{N+1}^{i,j}|\boldsymbol{\tau}_N, \boldsymbol{y}_N)P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) \\ &= \sum_{\mathbf{y}\in\{\pm1\}}\int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N}\sum_{\tilde{y}_i\in\{\pm1\}} \\ &\quad \int_{\mathcal{P}} d\tilde{\tau}_i\beta(\tau_i - \tilde{\tau}_{N+1})\beta(\tau_{N+1} - \tilde{\tau}_i)g(\tilde{\tau}_i, \tilde{\boldsymbol{\tau}}_N^{i,\hat{j}})\delta_{\eta,1}\delta_{\tilde{y}_i,1}\delta_{y_i,1}\delta_{y_{N+1},1}P(\tilde{\boldsymbol{\tau}}_N^{i,\widehat{N+1}}, \tilde{\boldsymbol{y}}_N^{i,\widehat{N+1}}) = \\ &= N\cdot\int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}_N^1 \sum_{\tilde{y}_1\in\{\pm1\}} \\ &\times \int_{\mathcal{P}} d\tau_1\beta(\tau_1 - \tilde{\tau}_1)\int_P d\tau_{N+1}\beta(\tau_{N+1} - \tilde{\tau}_1)g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_N^1)P(\tilde{\boldsymbol{\tau}}_N^1, \tilde{\boldsymbol{y}}_N^1)\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1} = \\ &= N\cdot\int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}_N^1 \sum_{\tilde{y}_1\in\{\pm1\}} g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_N^1)P(\tilde{\boldsymbol{\tau}}_N^1, \tilde{\boldsymbol{y}}_N^1)\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1} = N\bar{g}_N\delta_{\eta,1},\end{aligned} \tag{D.25, D.26}$$

where the indices has been fixed as $i = 1, j = N+1$. Finally, the fourth one term can be written as:

$$\begin{aligned}
\Lambda_4 &= \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N} \sum_{y_i\in\{\pm1\}} \mathcal{D}(\boldsymbol{\tau}^{\hat{i}}_{N-1}, \boldsymbol{y}^{\hat{i}}_{N-1}|\boldsymbol{\tau}_N, \boldsymbol{y}_N) P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) \\
&= \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N} \sum_{y_i\in\{\pm1\}} P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)\big[\mu(\tau_i, \boldsymbol{\tau_N})\delta_{y_i,1} + g(\tau_i, \boldsymbol{\tau}_N)\delta_{\eta,-1}\delta_{y_i,1}\big] = \\
&= N \sum_{\mathbf{y}\in\{\pm1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{y_1\in\{\pm1\}} P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)\big[\mu(\tau_1, \boldsymbol{\tau_N})\delta_{y_1,1} + g(\tau_1, \boldsymbol{\tau}_N)\delta_{\eta,-1}\delta_{y_1,1}\big] \\
&= N(\bar{\mu}_N + \bar{g}_N\delta_{\eta,-1}).
\end{aligned} \tag{D.27}$$

Thus, by collecting the four terms (D.23, D.24, D.25, D.27) one finally obtains an explicit expression for the normalization factor:

$$\begin{aligned}
&\Lambda_N(t) = \\
&= N\left[a_{N-1,N}\cdot\bar{g}_{N-1} - \bar{g}_N\right]\delta_{\eta,1} + \left(a_{N+1,N}\cdot\bar{\mu}_{N+1} - N\cdot\mu_N\right) + \left[a_{N+1,N}\bar{g}_{N+1} - N\cdot\bar{g}_N\right]\delta_{\eta,-1}
\end{aligned} \tag{D.28}$$

Moran process contribution. When the systems reaches the carrying capacity K each birth must be compensated by a random death. This part of the dynamics follows a master equation that conserves the system size $N = K$:

$$\begin{aligned}
\Delta_M P(\boldsymbol{\tau}_n, \boldsymbol{y}_N, t) &= \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \sum_{\tilde{y}_i\in\{\pm1\}} \sum_{\tilde{y}_j\in\{\pm1\}} \int_{\mathcal{P}^2} d\tilde{\tau}_i d\tilde{\tau}_j \Big[W_M(\boldsymbol{\tau}_N, \boldsymbol{y}_N|\tilde{\boldsymbol{\tau}}^{ij}_N, \tilde{\boldsymbol{y}}^{ij}_N) P(\tilde{\boldsymbol{\tau}}^{ij}_N, \tilde{\boldsymbol{y}}^{ij}_N, t) \\
&- W_M(\tilde{\boldsymbol{\tau}}^{ij}_N, \tilde{\boldsymbol{y}}^{ij}_N|\boldsymbol{\tau}_N, \boldsymbol{y}_N) P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)\Big],
\end{aligned} \tag{D.29}$$

where the transition rate is given by:

$$W_M(\boldsymbol{\tau}_N, \boldsymbol{y}_N|\tilde{\boldsymbol{\tau}}^{ij}_N, \tilde{\boldsymbol{y}}^{ij}_N) = \frac{1}{K-1}\beta(\tau_i - \tilde{\tau}_i)\beta(\tau_j - \tilde{\tau}_i) g(\tilde{\tau}_i, \tilde{\boldsymbol{\tau}}^{i,j})\delta_{\eta,1}\delta_{\tilde{y}_i,1}\delta_{y_i,1}\delta_{y_j,1}. \tag{D.30}$$

Growing and Dormant state contribution The part representing the transitions between growing and dormant states is relatively easy to define given that N is conserved:

$$\begin{aligned}
&\Delta_{GD} P(\boldsymbol{\tau}_N, \mathbf{y}_N, t) = \\
&= \sum_{i=1}^{N} \sum_{\tilde{y}_i\in\{\pm1\}} \Big[W_{GD}(\boldsymbol{\tau}_N, \mathbf{y}_N|, \boldsymbol{\tau}_N, \tilde{\mathbf{y}}^i_N) P(\boldsymbol{\tau}_N, \tilde{\mathbf{y}}^i_N, t) - W_{GD}(\boldsymbol{\tau}_N, \tilde{\mathbf{y}}^i_N|, \boldsymbol{\tau}_N, \mathbf{y}_N) P(\boldsymbol{\tau}_N, \mathbf{y}_N, t)\Big],
\end{aligned} \tag{D.31}$$

where the transition rate $\tilde{y}_i \to y_i$ is given by:

$$W_{GD}(\boldsymbol{\tau}_N, \mathbf{y}_N|\boldsymbol{\tau}_N, \tilde{\mathbf{y}}^i_N) = \delta_{\tilde{y}_i,-1}\delta_{y_i,1} A(\tau_i, \boldsymbol{\tau}_N) + \delta_{\tilde{y}_i,1}\delta_{y_i,-1}\delta_{\eta,-1} S(\tau_i, \boldsymbol{\tau}_N), \tag{D.32}$$

the δ stand for Kronecker deltas introduced because each term only contributes in a specific medium or state.

Thus, summarizing, the master equation defining our proposed model reads:

$$\frac{dP(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)}{dt} = (1-\delta_{N,K})\Delta_{BD}P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) + \delta_{N,K}\Delta_M P(\boldsymbol{\tau}_N, \mathbf{y}_N, t) + \Delta_{GD}P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) \quad \text{(D.33)}$$

where the birth-death contribution Δ_{BD} is given by eqs.(D.11,D.28,D.7,D.10), the Moran process contribution Δ_M by eqs.(D.29,D.30) and the growing-dormant transition $\Delta_{GD}$is specified by eqs. (D.31,D.32).

D.1.2 Gillespie algorithm

To integrate the master equation (D.2) numerically we use the Gillespie algorithm. The Gillespie algorithm is a kinetic Monte Carlo Method [551] initially proposed in the field of chemistry to simulate individual trajectories of a set of chemical reactions [400]. However, since its initial formulation, the Gillespie algorithm has gained great importance e.g. statistical physics and biophysics [552]. In general each individual simulation of the Gillespie algorithm represents an exact trajectory of the probability density function that is ruled by the master eq. (D.2) [551].

At each time step one possible event is randomly chosen with a probability proportional to its rate. In our case there are four different events in the antibiotic-exposure phase and five events in the fresh medium phase. Moreover, each bacteria can be in two different states: growing state and dormant state.

Antibiotic-exposure phase: 1) A bacteria in the growing state can be randomly chosen to die with rate $d = const$. Overall, bacteria death at rate $\varphi_{death} = N_G \cdot d$, where N_G is the number of growing individuals. 2) A bacteria in growing state can be randomly chosen to die while trying to reproduce with rate $b = const$. Overall, bacteria are killed by antibiotics with rate $\varphi_{killed\ by\ a.} = b \cdot N_G$. 3) A bacteria in the growing state can sense the environment and enter in dormant state at rate $s = const$. Overall, bacteria enter in dormant sate at rate $\varphi_{enter\ d.s.} = s \cdot N_G$. 4) A bacteria in the dormant state can randomly chosen to exit dormant state at rate $\varphi^i_{exit\ d.s.} = \frac{1}{\tau^i_{lag}}$. Overall, bacteria exit from dormant state at rate: $\sum_i \varphi^i_{exit\ d.s.}$.

Fresh medium phase: 1) A bacteria in the growing state can be randomly chosen for die at a constant rate d. Overall, bacteria death at rate φ_{death}. 2) A bacteria in the growing state can be randomly chosen to reproduce at constant rate b. Overall, bacteria reproduce at rate $\varphi_{birth} = b \cdot N_G$. Bacteria reproduce asexually by duplication, so both offspring will inherit the same phenotype τ^i_{lag} plus a variation. Here we explore two possibilities for the variation $\sigma(\tau)$, additive amplitude $\sigma_A(\tau) \sim \alpha_A = const$ and multiplicative amplitude $\sigma_m(\tau) \sim \alpha_M \cdot \tau$.. 3)A bacteria in dormant state can randomly chosen to exit dormant state with rate $\varphi^i_{exit\ d.s.} = \frac{1}{\tau^i_{lag}}$. Overall, bacteria exit from dormant state at rate: $\sum_i \varphi^i_{exit\ d.s.}$. Importantly, there is also a maximum carrying capacity K, so once it is reached for each reproduction event , a randomly chosen member of the population is deleted (following a Moran process).

Since the rate $\varphi^i_{exit\ d.s.}$ is specific to each bacteria i, the search of the randomly chosen reaction could be computationally expensive, thus a binary search is implemented. In the worst case, binary search needs $O(Ln(N))$ comparisons, whereas direct search needs $O(N)$ comparisons.

D.2 From Microscopic to Macroscopic Process

D.2.1 Marginalization

The master equation (D.2,D.15, D.29, D.31) rules the system's evolution at the microscopic level (individual-based). Our aim now is to derive the "one-cell" probability density , see appendix B.3.5, $\phi^{(1)}(\tau, y, t)$: given a certain pair (τ, y), $\phi^{(1)}$ gives the probability of finding an individual in that state at a time t, i.e. it is the one-particle probability density in the jargon of statistical physics. $\phi^{(1)}$ is an experimentally measurable magnitude that provides a macroscopic description (population-based) of the system.
To obtain the one-cell density one needs to marginalize $P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t)$, i.e. to remove the dependence of $\boldsymbol{P}$ in all individuals but one [77]. Mathematically, the one-cell probability density is defined as:

$$\phi^{(1)}(\tau, y, t) \equiv \frac{1}{N}\Big\langle \sum_{i=1}^{N} \sum_{\boldsymbol{y}_i \in \{\pm 1\}} \delta(\tau - \tau_i)\delta_{y,y_i} \Big\rangle, \tag{D.34}$$

and, by writing down explicitly the average operator $\langle \dots \rangle$, it does coincide with the marginalized phenotypic distribution for the first individual:

$$\begin{aligned}\phi^{(1)}(\tau, y, t) &\equiv \frac{1}{N}\langle \sum_{i=1}^{N} \sum_{\boldsymbol{y}_i} \delta(\tau - \tau_i)\delta_{y,y_i} \rangle = \frac{1}{N} \sum_{\boldsymbol{y} \in \{\pm 1\}} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N \sum_{i=1}^{N} \delta(\tau - \tau_i)\delta_{y,y_i} P(\boldsymbol{\tau}, \boldsymbol{y}, t) = \\ &= \sum_{y_j \in \{\pm 1\}, j \in [2,N]} \int_{\mathcal{P}^{N-1}} d\tau_2 \dots d\tau_N P(\tau, \tau_2, \dots, \tau_N, y, y_2, y_3, .., y_N) = P^{(1)}(\tau, y, t).\end{aligned}$$

In the same way the $n-$cells probability density can be defined:

$$\begin{aligned}\phi^{(n)}(\boldsymbol{\tau}_n, \boldsymbol{y}_n, t) &= \frac{N-n}{N!} \Big\langle \prod_{j=1}^{n} \sum_{i_j=1}^{N} \sum_{\boldsymbol{y}_N} \delta(\tau_j - \tau_{i_j})\delta_{y_j,y_i} \Big\rangle \\ &= P^n(\tau_1, \tau_2, .., \tau_n, y_1, y_2, .., y_n, t).\end{aligned} \tag{D.35}$$

Taking the time derivative of $\phi^{(1)}$, by using the master eq.(D.2) one obtains:

$$\begin{aligned}\partial_t \phi^{(1)}(\tau, y, t) &= \sum_{y_j \in \{\pm 1\}, j \in [2,N]} \int_{\mathcal{P}^{N-1}} d\tau_2 d\tau_3 .. d\tau_N \partial_t P(\boldsymbol{\tau}_N, \mathbf{y}_N, t) = \\ &= \sum_{y_j \in \{\pm 1\}, j \in [2,N]} \int_{\mathcal{P}^{N-1}} d\tau_2 d\tau_3 .. d\tau_N \Big[\Delta_D P(\boldsymbol{\tau}_N, \mathbf{y}_N, t) + \Delta_{GD} P(\boldsymbol{\tau}_N, \mathbf{y}_N, t)\Big] = \\ &= \Delta_{BD}\phi^{(1)}(\tau, y, t) + \Delta_M \phi^{(1)}(\tau, y, t) + \Delta_{GD}\phi^{(1)}(\tau, y, t).\end{aligned} \tag{D.36}$$

Here on, we consider separately the birth-death (Δ_{BD}, $D.15$), Moran (Δ_M, $D.29$) and growing-dormant (Δ_{GD}, $D.32$) contributions.

Birth and death contribution. The Birth-Death is defined as:

$$\Delta_{BD}\phi^{(1)}(\tau, y, t) = \sum_{y_k \in \{\pm 1\}, k \in [2,N]} \int_{\mathcal{P}^{N-1}} d\tau_2 \dots d\tau_N \Delta_{BD} P(\boldsymbol{\tau}_N, \mathbf{y}_N, t). \tag{D.37}$$

The birth-death contribution of the master eq.(D.15) is composed by five terms: the four positive/negative contribution of birth and death, plus the normalizing factor. This last one does not depend on the variables, and thus it is not relevant for the marginalization.

Therefore, we focus here on the first four terms one at the time:

$$\Delta_{BD}\phi^{(1)}(\tau,y,t) = B_1 + B_2 - B_3 - B_4 - \Lambda_N\phi^{(1)}(\tau,y,t) \tag{D.38}$$

The first one takes into account the positive contribution of reproduction:

$$\begin{aligned} B_1 &= a_{N-1,N} \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^N} d\boldsymbol{\tau}^{\hat{1}}_{N-1} \sum_{i=1}^{N}\sum_{j=1,i\neq j}^{N}\sum_{\tilde{y}_i\in\{\pm1\}} \\ &\times \int_{\mathcal{P}} d\tilde{\tau}_i \frac{1}{N-1}\mathcal{B}(\boldsymbol{\tau}_N,\boldsymbol{y}_N|\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1}\tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1})P(\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1},\tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1},t). \end{aligned} \tag{D.39}$$

Let us split the i,j sums in three contributions: $\sum_{i=1}^{N}\sum_{j=1,i\neq j}^{N} = \delta_{i,1}\sum_{j=2}^{N} + \delta_{j,1}\sum_{i=2}^{N} + \sum_{i=2}^{N}\sum_{j=2\neq i}^{N}$, to decompose the expression in three contributions:

$$B_1 = \left(B_1^a + B_1^b + B_1^c\right) a_{N-1,N} \tag{D.40}$$

The first two terms, B_1^a y B_1^b give a similar contribution: in the first sum, the individual 1 duplicates and gives birth to the new individual 1 and the j one, while the second term is just the reversed process. Indeed, there are $(N-1)$ ways of doing so, because when the first individual is chosen there are other $N-1$ that could be its offspring. Hence, the first term is then computed as follows:

$$\begin{aligned} B_1^a &= \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}^{\hat{1}}_{N-1}\delta_{i,1}\sum_{j=2}^{N}\sum_{\tilde{y}_i\in\{\pm1\}} \\ &\times \int_{P} d\tilde{\tau}_i \frac{P(\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1},\tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1},t)}{N-1}\beta(\tau_i-\tilde{\tau}_i)\beta(\tau_j-\tilde{\tau}_i)g(\tilde{\tau}_i,\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1})\delta_{\eta,1}\delta_{\tilde{y}_i,1}\delta_{y_i,1}\delta_{y_j,1}, \end{aligned} \tag{D.41}$$

and using the fact that individuals are indistinguishable, one can rewrite the $\sum_{j=2}^{N}$ by fixing $j=2$ and pulling out a $N-1$ factor:

$$\begin{aligned} B_1^a &= \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^N} d\tilde{\boldsymbol{\tau}}^{1}_{N}\sum_{\tilde{y}_1\in\{\pm1\}} P(\tilde{\boldsymbol{\tau}}^{1,\hat{2}}_{N-1},\tilde{\boldsymbol{y}}^{1,\hat{2}}_{N-1},t)\beta(\tau_1-\tilde{\tau}_1)\beta(\tau_2-\tilde{\tau}_1)g(\tilde{\tau}_1,\tilde{\boldsymbol{\tau}}^{1,\hat{2}}_{N-1}) \\ &\times \delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1}\delta_{y_2,1} \\ &= \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}^{1,\hat{2}}_{N-1}\sum_{\tilde{y}_1\in\{\pm1\}} P(\tilde{\boldsymbol{\tau}}^{1,\hat{2}}_{N-1},\tilde{\boldsymbol{y}}^{1,\hat{2}}_{N-1},t)g(\tilde{\tau}_1,\tilde{\boldsymbol{\tau}}^{1,\hat{2}}_{N-1})\beta(\tau_1-\tilde{\tau}_1)\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1}\delta_{y_2,1}. \end{aligned} \tag{D.42}$$

The second contribution of the sum has a very similar form:

$$\begin{aligned} B_1^b &= \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}^{\hat{1}}_{N-1}\delta_{j,1}\sum_{i=2}^{N-1}\sum_{\tilde{y}_i\in\{\pm1\}} \\ &\times \int_{P} d\tilde{\tau}_i \frac{P(\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1},\tilde{\boldsymbol{y}}^{i,\hat{j}}_{N-1})}{N-1}\beta(\tau_i-\tilde{\tau}_i)\beta(\tau_j-\tilde{\tau}_i)g(\tilde{\tau}_i,\tilde{\boldsymbol{\tau}}^{i,\hat{j}}_{N-1})\delta_{\eta,1}\delta_{\tilde{y}_i,1}\delta_{y_i,1}\delta_{y_j,1} \\ &= \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}^{\hat{1},2}_{N-1}\sum_{\tilde{y}_1\in\{\pm1\}} P(\tilde{\boldsymbol{\tau}}^{2,\hat{1}}_{N-1},\tilde{\boldsymbol{y}}^{2,\hat{1}}_{N-1})\beta(\tau_1-\tilde{\tau}_2)g(\tilde{\tau}_2,\tilde{\boldsymbol{\tau}}^{2,\hat{1}}_{N-1})\delta_{\eta,1}\delta_{\tilde{y}_2,1}\delta_{y_2,1}\delta_{y_1,1} \end{aligned} \tag{D.43}$$

where the $\sum_{i=2}^{N}$ pulls out a $(N-1)$ factor by fixing $i=2$. By renaming $\tilde{\tau}_2 \to \tilde{\tau}_1$ one obtains:

$$\begin{aligned} B_1^b &= \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}} \sum_{\tilde{y}_1\in\{\pm1\}} P(\tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}}, \tilde{\boldsymbol{y}}_{N-1}^{1,\hat{2}}, t) g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}}) \\ &\times \beta(\tau_1-\tilde{\tau}_1)\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1}\delta_{y_2,1}. \end{aligned} \tag{D.44}$$

Finally, the third contribution of the sum, $\sum_{i=2}^{N}\sum_{j=2\neq i}^{N}$ reads:

$$\begin{aligned} B_1^c &= \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_{N-1}^{\hat{1}} \sum_{i=2}^{N}\sum_{j=2\neq i}^{N} \sum_{\tilde{y}_i=\in\{\pm1\}} \\ &\times \int_{\mathcal{P}} d\tilde{\tau}_i \frac{P(\tilde{\boldsymbol{\tau}}_{N-1}^{i,\hat{j}}, \tilde{\boldsymbol{y}}_{N-1}^{i,\hat{j}})}{N-1} \beta(\tau_i-\tilde{\tau}_i)\beta(\tau_j-\tilde{\tau}_i) g(\tilde{\tau}_i, \tilde{\boldsymbol{\tau}}_{N-1}^{i,\hat{j}})\delta_{\eta,1}\delta_{\tilde{y}_i,1}\delta_{y_i,1}\delta_{y_j,1} \\ &= (N-2) \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_{N-1}^{\hat{1}} \sum_{\tilde{y}_2\in\{\pm1\}} \\ &\times \int_{\mathcal{P}} d\tilde{\tau}_2 P(\tilde{\boldsymbol{\tau}}_{N-1}^{2,\hat{3}}, \tilde{\boldsymbol{y}}_{N-1}^{2,\hat{3}}) \beta(\tau_2-\tilde{\tau}_2)\beta(\tau_3-\tilde{\tau}_2) g(\tilde{\tau}_2, \tilde{\boldsymbol{\tau}}_{N-1}^{2,\hat{3}})\delta_{\eta,1}\delta_{\tilde{y}_2,1}\delta_{y_2,1}\delta_{y_3,1}. \end{aligned} \tag{D.45}$$

where first sum, $\sum_{i=2}^{N}$, pulls a $N-1$ factor by fixing $i=2$, whereas the second sum, $\sum_{j=2\neq i}^{N}$, pulls a $N-1$ factor by fixing $j=3$. By splitting the following factor $d\boldsymbol{\tau}_{N-1}^{\hat{1}}$ as $d\boldsymbol{\tau}_{N-3}^{\hat{1},\hat{2}} d\tau_2 d\tau_3$ and performing the integrals $\int_P d\tau_2 \beta(\tau_2-\tilde{\tau}_2) = \int_P d\tau_3 \beta(\tau_3-\tilde{\tau}_2) = 1$ one gets to:

$$B_1^c = (N-2) \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-2}} d\tilde{\boldsymbol{\tau}}_{N-2}^{\hat{1},2,\hat{3}} \sum_{\tilde{y}_2\in\{\pm1\}} P(\tilde{\boldsymbol{\tau}}_{N-1}^{2,\hat{3}}, \tilde{\boldsymbol{y}}_{N-1}^{2,\hat{3}}) g(\tilde{\tau}_2, \tilde{\boldsymbol{\tau}}_{N-1}^{2,\hat{3}})\delta_{\eta,1}\delta_{\tilde{y}_2,1}\delta_{y_2,1}\delta_{y_3,1}. \tag{D.46}$$

By summing all the three contributions ($D.42$, $D.44$, $D.46$) one obtains the marginalization of the positive birth contribution:

$$\begin{aligned} B_1 &= 2a_{N-1,N} \sum_{y_k=\pm1,k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}} \sum_{\tilde{y}_1\in\{\pm1\}} P(\tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}}, \tilde{\boldsymbol{y}}_{N-1}^{1,\hat{2}}, t) g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_{N-1}^{1,\hat{2}}) \\ &\times \beta(\tau_1-\tilde{\tau}_1)\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1}\delta_{y_2,1} + \\ &+ (N-2)a_{N-1,N} \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-2}} d\tilde{\boldsymbol{\tau}}_{N-2}^{\hat{1},2,\hat{3}} \sum_{\tilde{y}_2\in\{\pm1\}} P(\tilde{\boldsymbol{\tau}}_{N-1}^{2,\hat{3}}, \tilde{\boldsymbol{y}}_{N-1}^{2,\hat{3}}) g(\tilde{\tau}_2, \tilde{\boldsymbol{\tau}}_{N-1}^{2}) \delta_{\eta,1}\delta_{\tilde{y}_2,1}\delta_{y_2,1}. \end{aligned} \tag{D.47}$$

The marginalization of second term of the birth-death process, i.e. positive contribution of death, is straightforward:

$$\begin{aligned} B_2 &= a_{N+1,N} \sum_{i=1}^{N+1} \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_{N-1}^{\hat{1}} \sum_{i=1}^{N+1} \sum_{y_{N+1}\in\{\pm1\}} \\ &\times \int_{\mathcal{P}} d\tau_i \frac{1}{N+1} \mathcal{D}(\boldsymbol{\tau}_N^{\hat{i}}, \boldsymbol{y}_N^{\hat{i}} | \boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1}) P(\boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1}, t) = \\ &= a_{N+1,N} \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_{N-1}^{\hat{1}} \sum_{y_i\in\{\pm1\}} \int_{\mathcal{P}} d\tau_{N+1} [\mu(\tau_{N+1}, \boldsymbol{\tau_{N+1}})\delta_{y_{N+1},1} P(\boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1}, t) \\ &+ g(\tau_{N+1}, \boldsymbol{\tau}_{N+1})\delta_{\eta,-1}\delta_{y_{N+1},1}] P(\boldsymbol{\tau}_{N+1}, \boldsymbol{y}_{N+1}, t)], \end{aligned} \tag{D.48}$$

where the index i has been fixed to $i = N+1$. The third term of the master equation quantifies the negative contribution of reproduction

$$\begin{aligned} & B_3 = \\ = & \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \sum_{i=1}^{N} \sum_{j=1\neq i}^{N+1} \sum_{\tilde{y}_i\in\{\pm1\}} \int_{\mathcal{P}^2} d\tilde{\tau}_i d\tilde{\tau}_j \frac{1}{N} \mathcal{B}(\tilde{\boldsymbol{\tau}}^{i,j}_{N+1}, \tilde{\boldsymbol{y}}^{i,j}_{N+1} | \boldsymbol{\tau}_N, \boldsymbol{y}_N) P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) \\ = & \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \sum_{i=1}^{N} \sum_{\tilde{y}_i\in\{\pm1\}} \int_{\mathcal{P}^2} d\tilde{\tau}_i d\tau_{N+1} \mathcal{B}(\tilde{\boldsymbol{\tau}}^{i,N+1}_{N+1}, \tilde{\boldsymbol{y}}^{i,N+1}_{N+1} | \boldsymbol{\tau}_N \boldsymbol{y}_N) P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t). \end{aligned}$$

where the index j has been fixed to $N+1$. By splitting the sum $\sum_{i=1}^{N} = \delta_{i,1} + \sum_{i=2}^{N}$, and integrating the variations kernels one obtains:

$$\begin{aligned} B_3 = & \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \sum_{\tilde{y}_1\in\{\pm1\}} P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) g(\tau_1, \boldsymbol{\tau}_N) \delta_{\eta,1} \delta_{\tilde{y}_1,1} \delta_{y_1,1} \\ + & (N-1) \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \sum_{\tilde{y}_2\in\{\pm1\}} P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t) g(\tau_2, \boldsymbol{\tau}_N) \delta_{\eta,+1} \delta_{\tilde{y}_2,1} \delta_{y_2,1}. \end{aligned} \tag{D.49}$$

where in the second term we have fixed $i = 2$. Finally, the fourth term, i.e. the negative contribution of death, is very similar to the second one:

$$B_4 = \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \sum_{i=1}^{N} \sum_{y_i\in\{\pm1\}} \mathcal{D}(\hat{\boldsymbol{\tau}^i}_{N-1}, \hat{\boldsymbol{y}^i}_{N-1} | \boldsymbol{\tau}_N, \boldsymbol{y}_N) P(\boldsymbol{\tau}_N, \boldsymbol{y}_N, t), \tag{D.50}$$

by splitting the sum as $\sum_{i=1}^{N} = \delta_{i,1} + \sum_{i=2}^{N}$, one obtains:

$$\begin{aligned} & B_4 = \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \sum_{y_1\in\{\pm1\}} \left[\mu(\tau_1, \boldsymbol{\tau_N}) \delta_{y_1,1} + g(\tau_1, \boldsymbol{\tau}_N) \delta_{\eta,-1} \delta_{y_1,1}\right] P(\boldsymbol{\tau}_N, \boldsymbol{y}_N) \\ + & (N-1) \sum_{y_k\in\{\pm1\},k\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \sum_{y_2\in\{\pm1\}} \left[\mu(\tau_2, \boldsymbol{\tau_N}) \delta_{y_2,1} + g(\tau_2, \boldsymbol{\tau}_N) \delta_{\eta,-1} \delta_{y_2,1}\right] P(\boldsymbol{\tau}_N, \boldsymbol{y}_N). \end{aligned} \tag{D.51}$$

By putting all terms together (D.38, D.47,D.48, D.49, D.51) and marginalizing also on the y_1 variable one finally obtains the marginalized birth-death contribution:

$$\begin{aligned} & \Delta_{BD}\phi^{(1)}(\tau,t) = \sum_{y_k\in\{\pm1\},k\in[1,N]} \int_{\mathcal{P}^{N-1}} d\tau_2 \dots d\tau_N \Delta_{BD} P(\boldsymbol{\tau}_N, \mathbf{y}_N, t) \\ = & 2a_{N-1,N} \int_{\mathcal{P}^{N-1}} d\tilde{\boldsymbol{\tau}}^{1,\hat{2}}_{N-1} P(\tilde{\boldsymbol{\tau}}^{1,\hat{2}}_{N-1}, \tilde{y}_1 = 1, t) g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}^{1,\hat{2}}_{N-1}) \beta(\tau_1 - \tilde{\tau}_1) \delta_{\eta,1} + \\ + & a_{N-1,N}(N-2) \int_{\mathcal{P}^{N-1}} d\tilde{\tau}_2 d\boldsymbol{\tau}^{\hat{1},\hat{3}}_{N-2} P(\tilde{\boldsymbol{\tau}}^2_{N-1}, \tilde{y}_2 = 1, t) g(\tilde{\tau}_2, \tilde{\boldsymbol{\tau}}^2_{N-1}) \delta_{\eta,1} \\ + & a_{N+1,N} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \int_{\mathcal{P}} d\tau_{N+1} \left[\mu(\tau_{N+1}, \boldsymbol{\tau_{N+1}}) + g(\tau_{N+1}, \boldsymbol{\tau}_{N+1}) \delta_{\eta,-1}\right] P(\boldsymbol{\tau}_{N+1}, y_{N+1} = 1, t) \\ - & \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \left[g(\tau_1, \boldsymbol{\tau}_N) \delta_{\eta,1} P(\boldsymbol{\tau}_N, \tilde{y}_1 = 1, t) + \left(\mu(\tau_1, \boldsymbol{\tau_N}) + g(\tau_1, \boldsymbol{\tau}_N) \delta_{\eta,-1}\right) P(\boldsymbol{\tau}_N, y_1 = 1, t)\right] \\ - & (N-1) \cdot \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \\ \times & \left[g(\tau_2, \boldsymbol{\tau}_N) \delta_{\eta,1} P(\boldsymbol{\tau}_N, \tilde{y}_2 = 1, t) + (\mu(\tau_2, \boldsymbol{\tau_N}) + g(\tau_2, \boldsymbol{\tau}_N) \delta_{\eta,-1}) P(\boldsymbol{\tau}_N, y_2 = 1, t)\right] \\ - & \Lambda_N(t) \phi^{(1)}(\tau,t). \end{aligned} \tag{D.52}$$

Moran Process Contribution The marginalization of the Moran process contribution (when $N = K$)

$$\Delta_M \phi^{(1)}(\tau, y, t) = \sum_{y_k \in \{\pm 1\}, k \in [2,N]} \int_{\mathcal{P}^{N-1}} d\tau_2 \dots d\tau_N \Delta_M P(\boldsymbol{\tau}_N, \mathbf{y}_N, t). \tag{D.53}$$

is analogous to the birth-death one. It is necessary to consider separately the positive and negative contributions,

$$\Delta_M \phi^{(1)}(\tau, y, t) = M_1 - M_2. \tag{D.54}$$

The first one reads:

$$\begin{aligned} M_1 &= \sum_{y_k \in \{\pm 1\}, k \in [2,N]} \int_{\mathcal{P}^{N-1}} d\tau_2 \dots d\tau_N \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \sum_{\tilde{y}_i \in \{\pm 1\}} \sum_{\tilde{y}_j \in \{\pm 1\}} \\ &\times \int d\tilde{\tau}_i d\tilde{\tau}_j W_M(\boldsymbol{\tau}_N, \boldsymbol{y}_N | \tilde{\boldsymbol{\tau}}_N^{ij}, \tilde{\boldsymbol{y}}_N^{ij}) P(\tilde{\boldsymbol{\tau}}_N^{ij}, \tilde{\boldsymbol{y}}_N^{ij}, t). \end{aligned} \tag{D.55}$$

As done before, one can split the sum as $\sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} = \delta_{i,1} \sum_{j=2}^{N} + \delta_{j,1} \sum_{i=2}^{N} + \sum_{i=2}^{N} \sum_{j=2, j\neq i}^{N}$ to respectively separate three contributions:

$$M_1 = M_1^a + M_1^b + M_1^c. \tag{D.56}$$

The calculation is identical to the first term of the birth-death part (D.47). The three contributions read:

$$\begin{aligned} M_1^a &= M_1^b = \sum_{y_k \in \{\pm 1\}, k \in [2,N]} \int_{\mathcal{P}^N} d\tilde{\boldsymbol{\tau}}_N^{1,2} \sum_{\tilde{y}_1 \in \{\pm 1\}} P(\tilde{\boldsymbol{\tau}}_N^{1,2}, \tilde{\boldsymbol{y}}_N^{1,2}, t) g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_N^{1,2}) \beta(\tau_1 - \tilde{\tau}_1) \delta_{\eta,1} \delta_{\tilde{y}_1,1} \delta_{y_1,1} \delta_{y_2,1}, \\ M_1^c &= (N-2) \sum_{y_k \in \{\pm 1\}, k \in [2,N]} \int_{\mathcal{P}^N} d\hat{\boldsymbol{\tau}}_N^{\hat{1}} \sum_{\tilde{y}_2 \in \{\pm 1\}} \\ &\times \int_{\mathcal{P}} d\tilde{\tau}_2 P(\tilde{\boldsymbol{\tau}}_N^{2,3}, \tilde{\boldsymbol{y}}_N^{2,3}) \beta(\tau_2 - \tilde{\tau}_2) \beta(\tau_3 - \tilde{\tau}_2) g(\tilde{\tau}_2, \tilde{\boldsymbol{\tau}}_{N-1}^{2,3}) \delta_{\eta,1} \delta_{\tilde{y}_2,1} \delta_{y_2,1} \delta_{y_3,1}. \end{aligned} \tag{D.57}$$

By summing them all, one obtains:

$$\begin{aligned} M_1 &= \sum_{y_k \in \{\pm 1\}, k \in [2,N]} \int_{\mathcal{P}^N} d\tilde{\boldsymbol{\tau}}_N^{1,2} \sum_{\tilde{y}_1 \in \{\pm 1\}} P(\tilde{\boldsymbol{\tau}}_N^{1,2}, \tilde{\boldsymbol{y}}_N^{1,2}, t) g(\tilde{\tau}_1, \tilde{\boldsymbol{\tau}}_N^{1,2}) \beta(\tau_1 - \tilde{\tau}_1) \delta_{\eta,1} \delta_{\tilde{y}_1,1} \delta_{y_1,1} \delta_{y_2,1} \\ &+ (N-2) \sum_{y_k \in \{\pm 1\}, k \in [2,N]} \int_{\mathcal{P}^N} d\hat{\boldsymbol{\tau}}_N^{\hat{1}} \sum_{\tilde{y}_2 \in \{\pm 1\}} \\ &\times \int_{\mathcal{P}} d\tilde{\tau}_2 P(\tilde{\boldsymbol{\tau}}_N^{2,3}, \tilde{\boldsymbol{y}}_N^{2,3}) \beta(\tau_2 - \tilde{\tau}_2) \beta(\tau_3 - \tilde{\tau}_2) g(\tilde{\tau}_2, \tilde{\boldsymbol{\tau}}_{N-1}^{2,3}) \delta_{\eta,1} \delta_{\tilde{y}_2,1} \delta_{y_2,1} \delta_{y_3,1}. \end{aligned} \tag{D.58}$$

The marginalization of the negative term is analogous to the birth-death one (D.49), leading to:

$$\begin{aligned} M_2 &= \sum_{y_k\epsilon\{\pm1\},k\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_N^{\hat{1}} \sum_{\tilde{y}_1\epsilon\{\pm1\}} P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t) g(\tau_1,\boldsymbol{\tau}_N)\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1} \\ &+ (N-1) \sum_{y_k\epsilon\{\pm1\},k\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_{N-1}^{\hat{1}} \sum_{\tilde{y}_2\epsilon\{\pm1\}1} P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t) g(\tau_2,\boldsymbol{\tau}_N)\delta_{\eta,1}\delta_{\tilde{y}_2,1}\delta_{y_2,1}. \end{aligned}$$

By summing the two terms and summing over $y_1$one obtains the marginalized contribution of the Moran process:

$$\begin{aligned} &\Delta_M\phi^{(1)}(\tau,t) = \\ =& \sum_{y_k\epsilon\{=\pm1\},k\epsilon[2,N]} \int_{\mathcal{P}^N} d\tilde{\boldsymbol{\tau}}_N^{1,2} \sum_{\tilde{y}_1\epsilon\{\pm1\}} P(\tilde{\boldsymbol{\tau}}_N^{1,2},\tilde{\boldsymbol{y}}_N^{1,2},t) g(\tilde{\tau}_1,\tilde{\boldsymbol{\tau}}_N^{1,2})\beta(\tau_1-\tilde{\tau}_1)\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1}\delta_{y_2,1} \\ &+ (N-2) \sum_{y_k in\{\pm1\},k\epsilon[2,N]} \int_{\mathcal{P}^N} d\boldsymbol{\tau}_N^{\hat{1}} \sum_{\tilde{y}_2 in\{\pm1\}} \\ &\times \int_{\mathcal{P}} d\tilde{\tau}_2 P(\tilde{\boldsymbol{\tau}}_N^{2,3},\tilde{\boldsymbol{y}}_N^{2,3})\beta(\tau_2-\tilde{\tau}_2)\beta(\tau_3-\tilde{\tau}_2)g(\tilde{\tau}_2,\tilde{\boldsymbol{\tau}}_{N-1}^{2,3})\delta_{\eta,1}\delta_{\tilde{y}_2,1}\delta_{y_2,1}\delta_{y_3,1} \\ &- \sum_{y_k in\{\pm1\},k\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_N^{\hat{1}} \sum_{\tilde{y}_1=\pm1} P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t) g(\tau_1,\boldsymbol{\tau}_N)\delta_{\eta,1}\delta_{\tilde{y}_1,1}\delta_{y_1,1} \\ &- (N-1) \sum_{y_k in\{\pm1\},k\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_{N-1}^{\hat{1}} \sum_{\tilde{y}_2 in\{\pm1\}} P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t) g(\tau_2,\boldsymbol{\tau}_N)\delta_{\eta,2}\delta_{\tilde{y}_2,1}\delta_{y_2,1}\delta_{y_3,1} \end{aligned} \tag{D.59}$$

Growing and Dormant state contribution Using the definition of this part of the process from the master eq.(D.31, D.32):

$$\begin{aligned} &\Delta_{GD}\phi^{(1)}(\tau,y,t) = \sum_{y_j in\{\pm1\},j\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\tau_2\ldots d\tau_N \Delta_{GD} P(\boldsymbol{\tau}_N,\mathbf{y}_N,t) = \\ =& \sum_{y_j in\{\pm1\},j\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_N^{\hat{1}} \sum_{i=1}^{N} \sum_{\tilde{y}_i in\{\pm1\}} \\ &\times \Big[W_{GD}(\mathbf{y}_N|\tilde{\mathbf{y}}_N^i,\boldsymbol{\tau}_N)P(\boldsymbol{\tau}_N,\tilde{\mathbf{y}}_N^i,t) - W_{GD}(\tilde{\mathbf{y}}_N^i|\mathbf{y}_N,\boldsymbol{\tau}_N)P(\boldsymbol{\tau}_N,\mathbf{y}_N,t)\Big] = \\ =& \sum_{y_j in\{\pm1\},j\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\boldsymbol{\tau}_{N-1}^{\hat{1}} \sum_{i=1}^{N} \sum_{\tilde{y}_i in\{\pm1\}} \\ &\times \Big[\big\{\delta_{\tilde{y}_i,-1}\delta_{y_i,1}A(\tau_i,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_i,1}\delta_{y_i,-1}\delta_{\eta,-1}S(\tau_i,\boldsymbol{\tau}_N)\big\}P(\boldsymbol{\tau}_N,\tilde{\mathbf{y}}_N^i,t) \\ &- \big\{\delta_{\tilde{y}_i,1}\delta_{y_i,-1}A(\tau_i,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_i,-1}\delta_{y_i,1}\delta_{\eta,-1}S(\tau_i,\boldsymbol{\tau}_N)\big\}P(\boldsymbol{\tau}_N,\mathbf{y}_N,t)\Big]., \end{aligned} \tag{D.60}$$

and splitting the sum in i as $\sum_{i=1}^{N} = \sum_{i=1}^{N}\delta_{i,1} + \sum_{i=2}^{N}$, the expression is composed by two terms:

$$\Delta_{GD}\phi^{(1)}(\tau,y,t) = G_1 + G_2, \tag{D.61}$$

as defined now. Let us consider first the second term, which takes into account the transitions in which individual 1 is not involved:

$$\begin{aligned} G_2 &= \sum_{y_j in\{\pm1\}, j\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{1}_{N-1} \sum_{i=2}^{N} \sum_{\tilde{y}_i in\{\pm1\}} \\ &\times \left\{\delta_{\tilde{y}_i,-1}\delta_{y_i,1}A(\tau_i,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_i,1}\delta_{y_i,-1}\delta_{\eta,-1}S(\tau_i,\boldsymbol{\tau}_N)\right\}P(\boldsymbol{\tau}_N,\tilde{\mathbf{y}}^{i}_N,t) \\ &- \left\{\delta_{\tilde{y}_i,1}\delta_{y_i,-1}A(\tau_i,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_i,-1}\delta_{y_i,1}\delta_{\eta,-1}S(\tau_i,\boldsymbol{\tau}_N)\right\}P(\boldsymbol{\tau}_N,\mathbf{y}_N,t)\Big]. \end{aligned} \tag{D.62}$$

Using individual indistinguishability, the index i is fixed to $i = 2$ and removing the sum $\sum_{i=2}^{N}$ by pulling out a factor $(N-1)$:

$$\begin{aligned} G_2 &= (N-1) \sum_{y_j in\{\pm1\}, j\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{1}_{N-1} \sum_{\tilde{y}_2 in\{\pm1\}} \\ &\times \Big[\left\{\delta_{\tilde{y}_i,-1}\delta_{y_2,1}A(\tau_2,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_2,1}\delta_{y_2,-1}\delta_{\eta,-1}S(\tau_2,\boldsymbol{\tau}_N)\right\}P(\boldsymbol{\tau}_N,\tilde{\mathbf{y}}^{2}_N,t) \\ &- \left\{\delta_{\tilde{y}_2,1}\delta_{y_2,-1}A(\tau_2,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_2,-1}\delta_{y_2,1}\delta_{\eta,-1}S(\tau_2,\boldsymbol{\tau}_N)\right\}P(\boldsymbol{\tau}_N,\mathbf{y}_N,t)\Big], \end{aligned} \tag{D.63}$$

and by marginalizing this expression in $y_j, j = 3,\ldots,N$ one obtains:

$$\begin{aligned} G_2 &= (N-1) \sum_{y_2 in\{\pm1\}} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{1}_{N-1} \sum_{\tilde{y}_2 in\{\pm1\}} \\ &\times \Big[\left\{\delta_{\tilde{y}_2,-1}\delta_{y_2,1}A(\tau_2,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_2,1}\delta_{y_2,-1}\delta_{\eta,-1}S(\tau_2,\boldsymbol{\tau}_N)\right\}P(\boldsymbol{\tau}_N,y_1,\tilde{y}_2,t) \\ &- \left\{\delta_{\tilde{y}_2,1}\delta_{y_2,-1}A(\tau_2,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_2,-1}\delta_{y_2,1}\delta_{\eta,-1}S(\tau_2,\boldsymbol{\tau}_N)\right\}P(\boldsymbol{\tau}_N,y_1,y_2,t)\Big]. \end{aligned} \tag{D.64}$$

Obviously, if the individual before the transition was in a state, $\tilde{y}_2 in\{\pm1\}$, after it the state is opposite one $y_2 in\{\pm1\}$, implying that the sum of the contribution is zero:

$$\begin{aligned} G_2 &= (N-1) \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{1}_{N-1} \\ &\times \Big[\left\{A(\tau_2,\boldsymbol{\tau}_N)P(\boldsymbol{\tau}_N,y_1,\tilde{y}_2=-1,t) - S(\tau_2,\boldsymbol{\tau}_N)P(\boldsymbol{\tau}_N,y_1,y_2=+1,t)\right\}_{y_2=+1,\tilde{y}_2=-1} \\ &+ \left\{S(\tau_2,\boldsymbol{\tau}_N)P(\boldsymbol{\tau}_N,y_1,\tilde{y}_2=+1,t) - A(\tau_2,\boldsymbol{\tau}_N)P(\boldsymbol{\tau}_N,y_1,y_2=-1,t)\right\}_{y_2=-1,\tilde{y}_2=+1}\Big] = 0. \end{aligned} \tag{D.65}$$

Thus, the first term $\sum_{i=1}^{N}\delta_{i,1}$, which takes in count the transitions in which individual 1 is involved, is the only net contribution to the growing-dormant part:

$$\begin{aligned} \Delta_{GD}\phi^{(1)}(\tau,y,t) = G_1 &= \sum_{y_j in\{\pm1\}, j\epsilon[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{1}_{N-1} \sum_{\tilde{y}_i in\{\pm1\}} \Big[\left\{\delta_{\tilde{y}_1,-1}\delta_{y_1,1}A(\tau_1,\boldsymbol{\tau}_N) + \right. \\ &\left. +\delta_{\tilde{y}_1,1}\delta_{y_1,-1}\delta_{\eta,-1}S(\tau_1,\boldsymbol{\tau}_N)\right\}P(\boldsymbol{\tau}_N,\tilde{\mathbf{y}}^{1}_N,t) \\ &- \left\{\delta_{\tilde{y}_1,1}\delta_{y_1,-1}A(\tau_1,\boldsymbol{\tau}_N) + \delta_{\tilde{y}_1,-1}\delta_{y_1,1}\delta_{\eta,-1}S(\tau_1,\boldsymbol{\tau}_N)\right\}P(\boldsymbol{\tau}_N,\mathbf{y}_N,t)\Big]. \end{aligned}$$

This is because index 1 is the only one that remains when function $P(\boldsymbol{\tau}_N,\boldsymbol{y}_N,t)$ is marginalized, i.e., $\tau = \tau_1$. It is possible to separate the previous equation in two contributions:

$$\Delta_{GD}\phi^{(1)}(\tau,y,t) = \Delta_{GD}\phi^{(1)}(\tau,y=+1,t) + \Delta_{GD}\phi^{(1)}(\tau,y=+1,t), \tag{D.66}$$

where

$$\Delta_{GD}\phi(\tau, y=+1, t) = \sum_{y_j in \{\pm 1\}, j\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \times \Big[A(\tau_1, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, y_1=+1, t) - S(\tau_1, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, y_1=-1, t)\Big], \tag{D.67}$$

and

$$\Delta_{GD}\phi(\tau, y=-1, t) = \sum_{y_j in \{\pm 1\}, j\in[2,N]} \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1} \times \Big[S(\tau_1, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, \tilde{y}_1=+1, t) - A(\tau_1, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, y_1=-1, t)\Big]. \tag{D.68}$$

Given that the rates do not depend on the variable y, one can marginalize in the variables using the sum $\sum_{y_j in \{\pm 1\}, j\in[2,N]}$ and obtain:

$$\Delta_{GD}\phi^{(1)}(\tau, t) = \Delta_{GD}\phi^{(1)}_G(\tau, t) + \Delta_{GD}\phi^{(1)}_D(\tau, t), \tag{D.69}$$

where the growing and dormant densities have been defined as:

$$\phi^{(1)}_G(\tau, t) \equiv \phi^{(1)}(\tau, y=+1, t), \tag{D.70}$$

$$\phi^{(1)}_D(\tau, t) \equiv \phi^{(1)}(\tau, y=-1, t); \tag{D.71}$$

while their evolution is given by:

$$\Delta_{GD}\phi_G(\tau, t) \equiv \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1}\Big[A(\tau_1, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, \tilde{y}_1=-1, t) - S(\tau_1, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, y_1=+1, t)\Big] \tag{D.72}$$

$$\Delta_{GD}\phi_D(\tau, t) \equiv \int_{\mathcal{P}^{N-1}} d\hat{\boldsymbol{\tau}}^{\hat{1}}_{N-1}\Big[S(\tau_1, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, y_1=+1, t) - A(\tau_1, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, \tilde{y}_1=-1, t)\Big]. \tag{D.73}$$

D.2.2 Mean Field Approximation

To proceed further, one can first apply the individual rate approximation, i.e the rates depend just on the single individual traits:

$$g(\tau_i, \boldsymbol{\tau}_N) = g(\tau_i), \qquad \mu(\tau_i, \boldsymbol{\tau}_N) = \mu(\tau_i), \qquad A(\tau_i, \boldsymbol{\tau}_N) = A(\tau_i), \qquad S(\tau_i, \boldsymbol{\tau}_N) = S(\tau_i) \tag{D.74}$$

Thanks to this approximation it is possible to integrate over most of individuals traits, but it leaves the dependence on the two-cells density as a kind of "correlation kernel" in the birth-death contribution eq.(D.52):

$$\begin{aligned}\Delta_{BD}\phi^{(1)}(\tau_1, t) &= 2a_{N-1,N}\delta_{\eta,1}\int_{\mathcal{P}} \phi^{(1)}_G(\tilde{\tau}_1, t) g(\tilde{\tau}_1)\beta(\tau_1 - \tilde{\tau}_1) d\tilde{\tau}_1 \\ &+ a_{N-1,N}(N-2)\int_{\mathcal{P}} d\tau_2 \phi^{(2)}(\tau_1, \tau_2, y_2=1, t) g(\tau_2)\delta_{\eta,1} \\ &+ a_{N+1,N}[\bar{\mu}_{N+1} + \delta_{\eta,-1}\bar{g}_{N+1}]\phi^{(1)}(\tau_1, t) - [\mu(\tau_1) + g(\tau_1)]\,\phi^{(1)}_G(\tau_1, t) \\ &- (N-1)\int_{\mathcal{P}} d\tau_2\,[g(\tau_2) + \mu(\tau_2)]\,\phi^{(2)}(\tau_1, \tau_2, y_2=1, t) - \Lambda_N(t)\phi^{(1)}(\tau, t);\end{aligned} \tag{D.75}$$

where we have used the definition of growing/dormant densities (D.70), the 2 – *cells* density function (D.35) and the fact that $1 = \delta_{\eta,1} + \delta_{\eta,-1}$.
Similarly for the Moran process contribution (D.59) reads:

$$\begin{aligned}\Delta_M \phi^{(1)}(\tau_1, t) &= 2 \int_{\mathcal{P}} d\tilde{\tau}_1 \phi^{(1)}(\tilde{\tau}_1) g(\tilde{\tau}_1) \beta(\tau_1 - \tilde{\tau}_1) - g(\tau_1)\phi_G^{(1)}(\tau_1, t) \\ + \quad &\int_{\mathcal{P}^2} d\tilde{\tau}_2 d\tau_2 \phi^{(2)}(\tau_1, \tilde{\tau}_2, \tilde{y}_2 = 1, t) g(\tilde{\tau}_2) \left[(N-1)\beta(\tau_2 - \tilde{\tau}_2) - (N-1)\delta(\tau_2 - \tilde{\tau}_2) \right] \end{aligned} \tag{D.76}$$

On the contrary, in the marginalized growing-dormant part, eq.(D.72-D.73) , the rate approximation (D.74) is enough to reach the final expression:

$$\Delta_{GD} \phi_G^{(1)}(\tau, t) = A(\tau)\phi_D^{(1)}(\tau, t) - \delta_{\eta,-1} S(\tau)\phi_G^{(1)}(\tau, t), \tag{D.77}$$

$$\Delta_{GD} \phi_D^{(1)}(\tau, t) = \delta_{\eta,-1} S(\tau)\phi_G^{(1)}(\tau, t) - A(\tau)\phi^{(1)}(\tau, t). \tag{D.78}$$

To simplify further the birth-death (D.75) and Moran (D.76) part one proceeds with a second mean field approximation. Specifically, we assume that the correlation between individuals is negligible. This is expected to be true in the thermodynamic limit $N \to \infty$, see appendix B.3.5 In such a case one can factorize the probability distribution:

$$P(\tau_1, \ldots, \tau_N) = \phi^{(N)}(\tau_1, \ldots, \tau_N) = \prod_{i=1}^{N} \phi(\tau_i), \tag{D.79}$$

and the 1-cell density can be replaced by its mean-field approximation (removing also the individual-subindexτ_1):

$$\phi^{(1)}(\tau, t) \equiv \phi(\tau, t), \tag{D.80}$$

$$\phi^{(2)}(\tau_1, \tau_2, t) \equiv \phi(\tau_1, t)\phi(\tau_2, t). \tag{D.81}$$

An important consequence of this simplification is that the normalization coefficients $a_{N,M}$ fulfills the relationship:

$$a_{N,M} = \frac{\int_{P^N} d\boldsymbol{\tau}_N \psi(\boldsymbol{\tau}_N, t)}{\int_{P^M} d\boldsymbol{\tau}_M \psi(\boldsymbol{\tau}_M, t)} = \frac{\prod_{i=1}^{N} \int_P d\tau_i \psi(\tau_i, t)}{\prod_{j=1}^{M} \int_{P^M} d\tau_j \psi(\tau_j, t)} = a^{N-M}, \tag{D.82}$$

and, moreover, the different marginal averages no longer dependent on the system size:

$$\begin{aligned}\bar{g}_N &= \int_{P^N} d\boldsymbol{\tau}_N g(\tau_k, \boldsymbol{\tau}_N) P(\boldsymbol{\tau}_N, y_i, t) \delta_{y_k,1} \\ &= \prod_{i=1}^{N} \int_P d\tau_i g(\tau_k) P(\boldsymbol{\tau}_N, y_i, t) \delta_{y_k,1} = \int_P d\tau_k g(\tau_k) \phi(\tau_k, t) \delta_{y_k,1} \equiv \bar{g}. \end{aligned} \tag{D.83}$$

Hence, the birth-death term (D.75) becomes:

$$\begin{aligned}
\Delta_{BD}\phi(\tau,t) &= \frac{2\delta_{\eta,1}}{a}\int_{\mathcal{P}} d\tilde{\tau}_1\phi(\tilde{\tau},t)g(\tilde{\tau}_1)\beta(\tau_1-\tilde{\tau}_1)+\frac{(N-2)}{a}\phi(\tau_1,t)\bar{g}\delta_{\eta,1}+\\
&+ a\left[\bar{\mu}+\bar{g}\delta_{\eta,-1}\right]\phi(\tau,t)-(g(\tau)+\mu(\tau))\phi_G(\tau,t)-(N-1)(\bar{g}+\bar{\mu})\phi(\tau,t)\\
&- \left\{N\Big(\frac{1}{a}-1\Big)\bar{g}\delta_{\eta,1}+\bar{\mu}(a-N)-(N-a)\bar{g}\delta_{\eta,-1}\right\}\cdot\phi(\tau,t)=\\
&= \frac{2\delta_{\eta,1}}{a}\left(\int_{\mathcal{P}} d\tilde{\tau}_1\phi(\tilde{\tau},t)g(\tilde{\tau}_1)\beta(\tau_1-\tilde{\tau}_1)-\bar{g}\phi(\tau,t)\right)\\
&- (\delta_{\eta,1}+\delta_{\eta,-1})\left(g(\tau)\phi_G(\tau,t)-\bar{g}\phi(\tau,t)\right)+\\
&- (\mu(\tau))\phi_G(\tau,t)-\bar{\mu}\phi(\tau,t)). \qquad \text{(D.84)}
\end{aligned}$$

In practical terms $\phi(\tau,t)$ is a frequency, and, for this reason, one can evaluate the normalization constant a as $\frac{1}{a}=\frac{N-1}{N}$. Furthermore, by implementing the mean-field approximation it has been assumed the thermodynamic limit $N\to\infty$, i.e. infinite population size, hence the the normalization constant can be approximated as: $\frac{N-1}{N}\approx 1$. Thanks to all of these, equation (D.84) simplifies to:

$$\begin{aligned}
\Delta_{BD}\phi(\tau,t) &= \delta_{\eta,1}\left(2\int_{\mathcal{P}} d\tilde{\tau}_1\phi(\tilde{\tau},t)g(\tilde{\tau}_1)\beta(\tau_1-\tilde{\tau}_1)-g(\tau)\phi_G(\tau,t)-\bar{g}\phi(\tau,t)\right)\\
&- \delta_{\eta,-1}(g(\tau)\phi_G(\tau,t)-\bar{g}\phi(\tau,t))-(\mu(\tau)\phi_G(\tau,t)-\bar{\mu}\phi(\tau,t)). \qquad \text{(D.85)}
\end{aligned}$$

Importantly, if one applies the mean-field approximation (D.79) to the Moran process contribution (D.76) it coincides with the contribution of the birth-death process in the fresh medium (see [77]):

$$\Delta_M\phi(\tau,t) = \delta_{\eta,1}\left(2\int_{\mathcal{P}} d\tilde{\tau}_1\phi(\tilde{\tau},t)g(\tilde{\tau}_1)\beta(\tau_1-\tilde{\tau}_1)-g(\tau)\phi_G(\tau,t)-\bar{g}\phi(\tau,t)\right). \qquad \text{(D.86)}$$

Thus, in the mean-field description, i.e. in the thermodynamics limit, it is not necessary to take into account if the population size is constant or varying: $\Delta_D\phi(\tau,t)=\Delta_{BD}\phi(\tau,t)$. By using the decomposition in growing and dormant densities (D.70) in the birth-death part, it is possible to separate such contribution in two equations:

$$\begin{aligned}
\Delta_{BD}\phi_G(\tau,t) &= \delta_{\eta,1}\left(2\int_{\mathcal{P}} d\tilde{\tau}_1\phi(\tilde{\tau},t)g(\tilde{\tau}_1)\beta(\tau_1-\tilde{\tau}_1)-(g(\tau)+\bar{g})\phi_G(\tau,t)\right)\\
&- \delta_{\eta,-1}(g(\tau)-\bar{g})\phi_G(\tau,t)-(\mu(\tau)-\bar{\mu})\phi_G(\tau,t), \qquad \text{(D.87)}\\
\Delta_{BD}\phi_D(\tau,t) &= -(\bar{g}[\delta_{\eta,1}-\delta_{\eta,-1}]-\bar{\mu})\phi_D(\tau,t). \qquad \text{(D.88)}
\end{aligned}$$

By summing the birth-death (D.87)and the growing-dormant (D.77) contributions, and by rewriting the deltas as

$$\delta_{\eta,1}=\frac{\eta+1}{2}, \qquad \delta_{\eta,-1}=\frac{1-\eta}{2} \qquad \text{(D.89)}$$

one finally obtains the general mean-field equations :

$$\begin{aligned}\dot{\phi}_G(\tau,t) &= \frac{1+\eta}{2}\Big[-g(\tau)\phi_G(\tau,t)+2\int_0^\infty d\tilde{\tau}g(\tilde{\tau})\beta(\tau-\tilde{\tau})\phi_G(\tilde{\tau},t)+\\ &- \phi_G(\tau,t)\int_0^\infty d\tilde{\tau}\phi_G g(\tilde{\tau})(\tilde{\tau},t)\Big]\\ &- \frac{1-\eta}{2}\Big[\Big(g(\tau)-\int_0^\infty d\tilde{\tau}g(\tilde{\tau})\phi_G(\tilde{\tau},t)\Big)+S(\tau)\Big]\phi_G(\tau,t)\\ &+ A(\tau)\phi_D(\tau,t)-\mu(\tau)\phi_G(\tau,t)+\bar{\mu}\phi(\tau,t) \end{aligned} \tag{D.90}$$

$$\dot{\phi}_D(\tau,t) = -\Big(\eta\int_0^\infty d\tilde{\tau}g(\tilde{\tau})\phi_G(\tilde{\tau},t)+A(\tau)\Big)\phi_D(\tau,t)\ +S(\tau)\frac{1-\eta}{2}\phi_G(\tau,t)+\bar{\mu}\phi_D(\tau,t). \tag{D.91}$$

By adding these two eqs. one can easily compute the total marginalized distribution :

$$\begin{aligned}\dot{\phi}(\tau,t) &= \dot{\phi}_G(\tau,t)+\dot{\phi}_D(\tau,t)\\ &= \frac{1+\eta}{2}\Big(2\int_0^\infty d\tilde{\tau}g(\tilde{\tau})\beta(\tau-\tilde{\tau})\phi_G(\tilde{\tau},t)-g(\tau)\phi_G(\tau,t)\\ &- \Big[\int_0^\infty d\tilde{\tau}g(\tilde{\tau})\phi_G(\tilde{\tau},t)\Big]\phi(\tau,t)\Big)\\ &- \frac{1-\eta}{2}\Big(g(\tau)\phi_G(\tau,t)-\bar{g}\phi(\tau,t)\Big)-\mu(\tau)\phi_G(\tau,t)+\bar{\mu}\phi(\tau,t). \end{aligned} \tag{D.92}$$

Finally, fixing the rates as

$$g(\tau_i)=b=\text{const},\qquad \mu(\tau_i)=0,\qquad A(\tau_i)=1/\tau_i\qquad ,S(\tau_i)=s=\text{const}; \tag{D.93}$$

and by inserting them in the mean-field equation (D.90) one finds the equations in chapter 4.

D.3 Small Variation Approximation

When variations are small, it is possible to perform a diffusive approximation to equation (D.92). Specifically, we consider the term proportional to the variation kernel, $\beta(\tau-\tilde{\tau};\tilde{\tau})$, which regulates the probability of changing from trait $\tilde{\tau}$ to τ during reproduction, and expand it around $\tilde{\tau}\sim\tau$. Note that β depends on the magnitude of variation, $\delta=\tau-\tilde{\tau}$, but can also depend explicitly on the value of the starting trait, $\tilde{\tau}$, (i.e. state dependent, or multiplicative variation). Given that β is a probability distribution of variation, hence: $\int_{-\tilde{\tau}}^\infty d\delta\beta(\delta;\tilde{\tau})=1$. Furthermore, one can assume it to be symmetric in δ: $\beta(\delta;\tilde{\tau})=\beta(-\delta;\tilde{\tau})$; and with finite first and second moments:

$$\mu(\tilde{\tau})=\int_{-\tilde{\tau}}^\infty d\delta\delta\beta(\delta;\tilde{\tau}),\qquad \sigma(\tilde{\tau})=\int_{-\tilde{\tau}}^\infty d\delta\delta^2\beta(\delta;\tilde{\tau}). \tag{D.94}$$

Let us separate the mean-field equation for the birth-death change of the growing individual as(D.87) in two parts:

$$\Delta_{BD}\phi_G(\tau,t)=\Delta_1\phi_G(\tau,t)+\Delta_2\phi_G(\tau,t), \tag{D.95}$$

where the first part involves the local variation events:

$$\Delta_1\phi_G(\tau,t) = (1+\eta)\left(\int_0^\infty d\tilde{\tau}\phi_G(\tilde{\tau},t)g(\tilde{\tau})\beta(\tau-\tilde{\tau};\tilde{\tau}) - g(\tau)\phi_G(\tau,t)\right), \quad \text{(D.96)}$$

and the second involves the remaining rates:

$$\begin{aligned}\Delta_2\phi_G(\tau,t) &= \eta(g(\tau)-\bar{g})\phi_G(\tau,t) - (\mu(\tau)-\bar{\mu})\phi_G = \\ &= \left[(g(\tau)-\mu(\tau)) - (\bar{g}-\bar{\mu})\right]\phi_G(\tau,t), \end{aligned} \quad \text{(D.97)}$$

where the deltas have been replaced as in eq.(D.89). Considering now Δ_1(D.96), which does conserve the probability and has the typical structure of a master equation [92]):

$$\Delta_1\phi_G(\tau,t) = (\eta+1)\int_0^\infty d\tilde{\tau}\Big[W(\tilde{\tau},\tau)\phi_G(\tilde{\tau},t) - W(\tau,\tilde{\tau})\phi(\tau,t)\Big], \quad \text{(D.98)}$$

with the "effective" stochastic transition rate from phenotype $\tilde{\tau}$ to τ:

$$W(\tilde{\tau},\tau) = g(\tilde{\tau})\beta(\tau-\tilde{\tau};\tilde{\tau}). \quad \text{(D.99)}$$

Thus it is possible to perform a Kramers-Moyal expansion of eq.(D.98) around $\tau = \tilde{\tau}$. The final trait can be written as the initial one plus the "jump" amplitude δ: $\tau = \tilde{\tau} + \delta$, allowing one to rewrite the transition rate as function just of the initial trait and of the jump amplitude:

$$W(\tilde{\tau},\tau) = W(\tau-\delta,\tau) := W(\tau-\delta;\delta) = g(\tau-\delta)\beta(\delta;\tau-\delta). \quad \text{(D.100)}$$

Hence, equation(D.96) can be rewritten as:

$$\Delta_1\phi_G(\tau,t) = (\eta(t)+1)\int_{-\infty}^{\tau} d\delta[W(\tau-\delta;\delta)\phi(\tau-\delta,t) - W(\tau;-\delta)\phi(\tau,t)], \quad \text{(D.101)}$$

where we have performed the following change of variable:

$$\int_0^\infty d\tilde{\tau}W(\tilde{\tau},\tau)\phi_G(\tilde{\tau},t) = -\int_\tau^{\tau-\infty} d\delta W(\tau-\delta;\delta)\phi(\tau-\delta,t) = \int_{-\infty}^{\tau} d\delta W(\tau-\delta;\delta)\phi(\tau-\delta,t).$$

One can consider δ as a small change in the *first* variable of the rate, and Taylor expand uo to second order:

$$\begin{aligned}W(\tau-\delta;\delta)\phi(\tau-\delta,t) &= W(\tau;\delta)\phi(\tau,t) - \delta\partial_\tau W(\tau;\delta)\phi(\tau,t) + \\ &+ \frac{\delta^2}{2}\partial_\tau^2 W(\tau;\delta)\phi(\tau,t) + O(\delta^3). \end{aligned} \quad \text{(D.102)}$$

Inserting it back in the equation and noting that $W(\tau;\delta) = W(\tau;-\delta)$, one obtains:

$$\Delta_1\phi_G(\tau,t) = (\eta+1)\left(-\partial_\tau m_1(\tau)\phi_G(\tau,t) + \partial_\tau^2 m_2(\tau)\phi_G(\tau,t)\right), \quad \text{(D.103)}$$

where

$$m_1(\tau) = 2\int_{-\infty}^{\tau} d\delta\delta\beta(\delta;\tau)g(\tau)\phi_G(\tau,t) = \theta(\tau)g(\tau)\phi_G(\tau,t), \quad \text{(D.104)}$$

$$m_2(\tau) = \int_{-\infty}^{\tau} d\delta\delta^2\beta(\delta;\tau)g(\tau)\phi_G(\tau,t) = \frac{\sigma^2(\tau)}{2}g(\tau)\phi_G(\tau,t). \quad \text{(D.105)}$$

Plugging the eq. for Δ_1 (D.103) into eq.(D.95), and then inserting it back in the mean-field equation for the growing population (D.92), one obtains:

$$\begin{aligned}\partial_t\phi_G(\tau,t) &= \left(\left\{g(\tau)\eta-\mu(\tau)\right\}-\left\{\bar{g}\eta-\bar{\mu}\right\}\right)\phi_G(\tau,t)+\\ &- (1+\eta)\partial_\tau\theta(\tau)g(\tau)\phi_G(\tau,t)+\frac{1+\eta}{2}\partial_\tau^2\sigma^2(\tau)g(\tau)\phi_G(\tau,t)\\ &+ A(\tau)\phi_D(\tau,t)-\frac{1-\eta}{2}S(\tau)\phi_G(\tau,t). \end{aligned} \tag{D.106}$$

By adding $\partial_t\phi_G$ (D.106) and $\partial_t\phi_D$ (D.91) one obtains the following expression for $\partial_t\phi(\tau,t)$ within the small-variation approximation:

$$\begin{aligned}\partial_t\phi(\tau,t) &= \left\{g(\tau)\eta-\mu(\tau)\right\}\phi_G(\tau,t)\\ &- \left\{\bar{g}\eta-\bar{\mu}\right\}\phi(\tau,t)+\\ &- (1+\eta)\partial_\tau\theta(\tau)g(\tau)\phi_G(\tau,t)+\frac{1+\eta}{2}\partial_\tau^2\sigma^2(\tau)g(\tau)\phi_G(\tau,t). \end{aligned} \tag{D.107}$$

By choosing the rates as in main text(D.93), one obtains the equations for the growing and dormant densities:

$$\begin{aligned}\dot{\phi}_G(\tau,t) &= b\left(1-\int_0^\infty d\tilde{\tau}\phi_G(\tilde{\tau},t)\right)\phi_G(\tau,t)-s\frac{(1-\eta)}{2}\phi_G(\tau,t)+\frac{1}{\tau}\phi_D(\tau,t)+\\ &- b(1+\eta(t))\partial_\tau\theta(\tau)\phi_G(\tau,t)+b\left(\frac{1+\eta}{2}\right)\partial_\tau^2\sigma^2(\tau)\phi_G(\tau,t), \end{aligned} \tag{D.108}$$

$$\dot{\phi}_D(\tau,t) = -\frac{1}{\tau}\phi_D(\tau,t)+s\frac{(1-\eta)}{2}\phi_G(\tau,t)-b\phi_D(\tau,t)\cdot\int_0^\infty d\tilde{\tau}\phi_G(\tilde{\tau},t). \tag{D.109}$$

In addition, one can define an effective fitness function $f(\tau,t)\equiv b\frac{\phi_G(\tau,t)}{\phi(\tau,t)}$, leading to the following expression of eq.(D.107) :

$$\begin{aligned}\partial_t\phi(\tau,t) &= \eta(t)\left(f(\tau,t)-\bar{f}\right)\phi(\tau,t)-(1+\eta(t))\partial_\tau\theta(\tau)f(\tau)\phi(\tau,t)+\\ &+ \frac{1+\eta(t)}{2}\partial_\tau^2\sigma^2(\tau)f(\tau)\phi(\tau,t). \end{aligned} \tag{D.110}$$

This expression for the marginal probability distribution is a version of the celebrated Crow-Kimura selection-mutation equation for "infinite alleles" in population genetics [404, 77]. From here on, we refer eq. (D.110) as the *generalized Crow-Kimura* equation (GCK). The first term of this equation, $\phi(\tau,t)\left(f(\tau,t)-\bar{f}\right)$, is a replicator-like term for continuous phenotypes. The replicator dynamics captures the essence of selection. If a phenotype τ has fitness which is below the average fitness $\bar{f}$ it will tend to disappear, whereas if it is above $\bar{f}$ it will tend to expand through the population. The second term, that resembles the divergence of a current from a Fokker-Planck eq.([92]), captures the effects of mutation/variation:

$$\partial_t\phi(\tau,t) = \eta(t)\left(f(\tau,t)-\bar{f}\right)\phi(\tau,t)-\partial_\tau J(\tau,t) \tag{D.111}$$

$$J(\tau,t) = (1+\eta(t))\theta(\tau)g(\tau)\phi_G(\tau,t)-\left(\frac{1+\eta(t)}{2}\right)\partial_\tau\sigma^2(\tau)g(\tau)\phi_G(\tau,t). \tag{D.112}$$

Note that this term has the typical structure of a conserved local probability current of a Fokker-Planck equation, modeling variations as a reaction-diffusion dynamics [77] and

then allowing $\phi(\tau, t)$ to move across the phenotype space. Let us note that the main structural difference with the classic Crow-Kimura eq. is the presence of the fitness function in the variation terms. This dependence is typical of phenotypic evolution and couple together ecological and evolutionary timescales.

Consider the two types of variation kernel presented in chapter 4. The additive case has no direct dependence on the initial phenotype; so we choose :

$$\beta_A(\tau - \tilde{\tau}) = \frac{e^{-\frac{(\tau-\tilde{\tau})^2}{2\alpha_A^2}}}{Z_A}, \qquad Z_A = \sqrt{2}\alpha_A \int_{-\frac{\tilde{\tau}}{\sqrt{2}\alpha_A}}^{\infty} e^{-y^2/2} dy = \alpha_A \frac{\sqrt{\pi}}{\sqrt{2}} Erfc\left(-\frac{\tilde{\tau}}{\sqrt{2}\alpha_A}\right), \quad \text{(D.113)}$$

where, due to the truncation in the domain $[-\tilde{\tau}, \infty]$, the normalization function Z_A presents a dependence on the phenotypic state and $Erfc$ stands for the complementary error function. Its first two cumulants are:

$$\theta_A(\tau) = \frac{\int_{-\tau}^{\infty} \delta e^{-\frac{\delta^2}{2\alpha_A^2}}}{Z_A} d\delta = 2\alpha_A e^{-\frac{\tau^2}{2\alpha_A^2}} Erfc\Big(-\frac{\tau}{\sqrt{2}\alpha_A}\Big)^{-1}, \quad \text{(D.114)}$$

$$\sigma_A^2(\tau) = \alpha_A^2 \left[1 - Erfc\left(\frac{\tau}{\sqrt{2}\alpha_A}\right)^{-1} e^{-\frac{\tau^2}{2\alpha_A^2}} \left(\frac{\tau}{\sqrt{2}\alpha_A} - e^{-\frac{\tau^2}{2\alpha^2}} Erfc\left(\frac{\tau}{\sqrt{2}\alpha_A}\right)^{-1}\right)\right]. \quad \text{(D.115)}$$

For sufficiently small α it is possible to ignore the $Erfc$ function, such that:

$$\theta_A(\tau) \approx 0 \qquad \sigma_A^2(\tau) \approx \alpha_A^2. \quad \text{(D.116)}$$

On the other hand, is the multiplicative case we consider a Gaussian with the variance proportional to the phenotype:

$$\beta_M(\tau - \tilde{\tau}; \tilde{\tau}) = \frac{e^{-\frac{(\tau-\tilde{\tau})^2}{2\alpha_M^2\tilde{\tau}^2}}}{Z_M}, \quad \text{(D.117)}$$

$$Z_M = \int_{-\tilde{\tau}}^{\infty} d\delta e^{-\frac{\delta^2}{2\alpha_M^2\tilde{\tau}^2}} = \sqrt{2}\alpha_M\tilde{\tau} \int_{-\frac{1}{\sqrt{2}\alpha_M}}^{\infty} e^{-y^2} dy = \alpha_M\tilde{\tau}\sqrt{\frac{\pi}{2}} Erfc\left(-\frac{1}{\sqrt{2}\alpha_M}\right); \quad \text{(D.118)}$$

whose first two moments are:

$$\theta_M(\tau) = \sqrt{\frac{2}{\pi}}\tau\alpha_M e^{-\frac{1}{2\alpha_M^2}} Erfc\left(-\frac{1}{\sqrt{2}\alpha_M}\right)^{-1},$$

$$\sigma_M^2(\tau) = \alpha_M^2\tau^2 \left[1 - Erfc\left(-\frac{1}{\sqrt{2}\alpha_M}\right)^{-1} e^{-\frac{1}{2\alpha_M^2}} \left(\frac{1}{\sqrt{2}\alpha_M} - e^{-\frac{1}{2\alpha_M^2}} Erfc\left(-\frac{1}{\sqrt{2}\alpha_M}\right)^{-1}\right)\right]. \quad \text{(D.119)}$$

For sufficiently small α it is possible to approximate them to:

$$\theta_M(\tau) \approx 0, \qquad \sigma_M^2(\tau) \approx \tau^2\alpha_M^2. \quad \text{(D.120)}$$

D.4 Deviation between theory and simulations

Systematic deviations between theory and simulations have been reported in chapter 4. Hence, one needs to check if the performed approximations are valid for the range of parameters used in simulations.

D.4.1 Validity of the small variation approximation and border effects

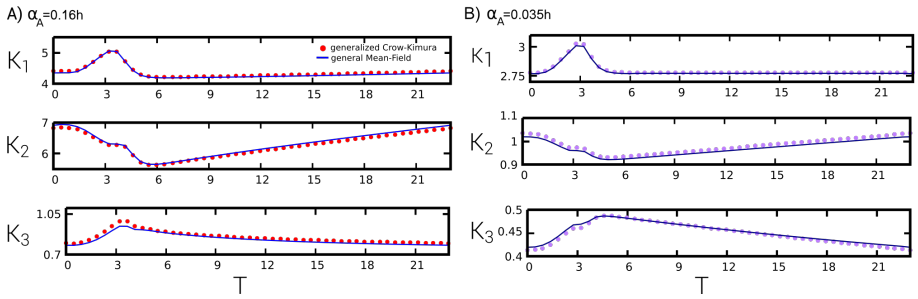


FIGURE D.1: **Validity of the small variation approximation in the additive case.** Evolution of the first three cumulants (K_1, K_2, K_3) in a asymptotic cycle calculated by the numerical integration of both the general mean-field equation (D.92)(solid line) and the generalized Crow-Kimura equation (D.110)(points) for two different values of α_A, in **(A)** $\alpha_A = 0.16h$ and in **(B)** $\alpha_A = 0.035h$. The first variation value is the best fit the experimental results (main text), while the second causes negligible border effects. In both cases the small-variation approximation is a good approximation. Parameter values: $T_a = 3h$, the rest of the parameters, as well as the initial conditions, are kept fixed as specified in chapter 4).

In this section we explore the validity of two different approximations used in the previous section: (i) the small-variation approximation, that leads from the general mean field equation (D.92) to the GCK equation (D.110), and (ii) the neglection of the border effects in the truncated domain $\delta \in]-\tilde{\tau}, \infty[$, (D.116-D.120) in equation (D.110).

Validity of the small variation approximation. The generalized Crow-Kimura equation (D.110) is derived through a diffusion approximation of the full mean-field equation (D.92). This approximation is valid only in the limit of small variation, thus here we verify if the optimal values of α_A and α_M, i.e. the ones that best fits to experimental results, meet this condition. At a practical level, the objective of this section is to compare the results given by the general mean-field eq. (D.92) and the generalized Crow-Kimura eq. (D.107) for the full density both in the additive (D.114) and multiplicative(D.119) case. In order to quantitatively compare these two equations one can define the parameter δ:

$$\delta^{as.}(\alpha) \equiv \frac{1}{M}\sum_{j=0}^{M}\delta_j^{as.} = \frac{1}{M}\sum_{j=0}^{M}|\langle\tau_j^{G.M.F.}\rangle - \langle\tau_j^{G.C.K.}\rangle| \tag{D.121}$$

where $M+1 = \frac{T_{max}}{\Delta t} + 1$ is the number of measures taken through a whole cycle (note that +1 is added since measures starts at $t = 0$), $T_{max} = 23h$ is the total duration of one single experimental cycle, $\Delta t = 0.5h$ is the interval between measures, $\langle\tau_j^{G.M.F.}\rangle$ and $\langle\tau_j^{G.C.K.}\rangle$ are the mean of the general mean field and the generalized Crow-Kimura equation respectively at time $j\cdot\Delta t$, and super-index $as.$ in $\delta^{as.}(\alpha)$ points out that comparison is performed in the asymptotic state (see Fig (D.5) for a visual definition of $\delta^{as.}$).

On the one hand, the numerical integration in the additive amplitude case (D.113) is straightforward. Fig D.1 shows the time evolution of the first three cumulants (K_1,K_2 and K_3) of $\phi(t,\tau)$ along a cycle in the asymptotic state both for the general mean field and the GCK equation. Moreover, in Fig D.2 the parameter δ is plotted for different variation

amplitudes α_A, showing the range of applicability of the small-variation approximation. On the other hand, the multiplicative case presents some subtleties since the multiplica-

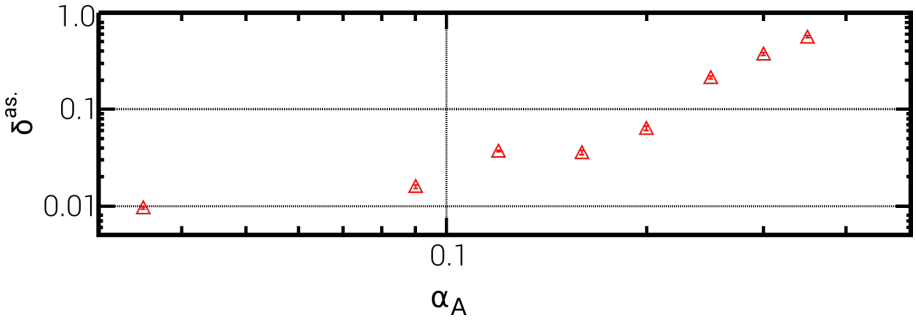


FIGURE D.2: **Range of validity of the small variation approximation in the additive scenario**. Systematic comparison of equations (D.110) and (D.92) via parameter (D.121) for different α_A values, both axis in log−scale. Note that the deviation monotonically increases with α_A. In particular for $\alpha_A = 0.16h$, the one used in chapter 4 results, the deviation $\delta_{st.} = (3.5 \pm 0.2) \cdot 10^{-2}h$, (D.121) is small enough to use equation (D.110). Parameter values: $T_a = 3h$, the other as in chapter 4.

tive variation kernel (D.117) has a singularity in $\tau = 0$. In addition, since its variance is proportional to $\alpha \cdot \tilde{\tau}$, one has to numerically impose that:

$$\lim_{\tilde{\tau}\to 0} \beta(\tau - \tilde{\tau}) = \delta(\tau - \tilde{\tau}). \tag{D.122}$$

Individuals in the vicinity of $\tilde{\tau} = 0$ experience either negligible or null variations after reproduction, thus one can safely decompose the variation contribution of the general mean field equation (D.92) in two terms:

$$\begin{aligned} & 2\delta_{\eta,1} b \int_0^\infty d\tilde{\tau}\beta(\tau - \tilde{\tau})\phi_G(\tilde{\tau}, t) = \\ = \; & 2\delta_{\eta,1} b \left(\int_0^\epsilon d\tilde{\tau}\delta(\tau - \tilde{\tau})\phi_G(\tilde{\tau}, t) + \int_\epsilon^\infty d\tilde{\tau}\beta(\tau - \tilde{\tau})\phi_G(\tilde{\tau}, t) \right) \end{aligned} \tag{D.123}$$

where ϵ is a small parameter that delimits the vicinity of $\tilde{\tau} = 0$, such that variations of individuals with $\tilde{\tau} < \epsilon$ can be considered as null. The initial condition $\phi(t = 0, \tau) = \phi_G(t = 0, \tau)$ is defined by cases: it is a truncated Gaussian (see methods in chapter 4), with the peak $\mu\epsilon$, and standard deviation is $\sigma = 1h$ 16min for $\tau \geq \epsilon$, and it is set to zero for $0 < \tau < \epsilon$. Fig (D.3) shows the integration of the mean-field equation with this numerical prescription and it is compared with the GCK eq. In conclusion, the validity of the small-variation approximation both in the additive and multiplicative case is confirmed.

Border effects. We also study the importance of border effects, i.e the dependence on $\tilde{\tau}$ (initial phenotype) in the normalization/moments of β given by the truncation of the probability density in $]-\tilde{\tau}, \infty[$. Equivalently, we check if their negligence, eqs.(D.116,D.120), is a good approximation in the range of variation amplitudes α_A and α_M that fits the experimental results. In Fig (D.4) the exact moments, (D.114-D.119), are compared with the approximated ones,(D.116,D.120), both in the additive and multiplicative case. Clearly, the results indicate the border effects are considerable in the additive case, but negligible in the multiplicative one.

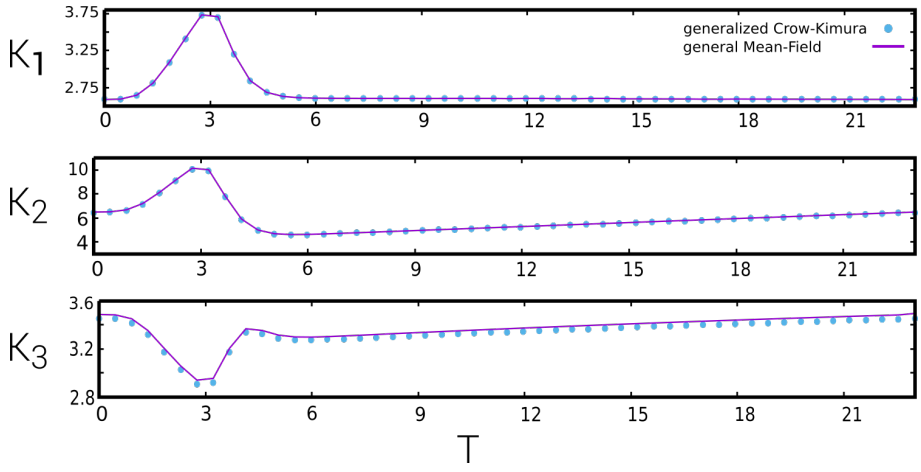


FIGURE D.3: **Validity of the small variation approximation in the multiplicative scenario**. Evolution of the first three cumulants (K_1, K_2, K_3) in a asymptotic cycle calculated by the numerical integration of both the general mean-field equation (D.92)(solid line) and the generalized Crow-Kimura equation (D.110)(points). The generalized Crow-Kimura equation is a good approximation of the general mean-field equation for the α_M-value that best fits the experimental results, e.g.$\delta_{st.} = (1.5 \pm 0.3) \cdot 10^{-3} h$. Parameter values: $T_a = 3h, \alpha_M = 0.048$, the remaining are specified in chapter 4.

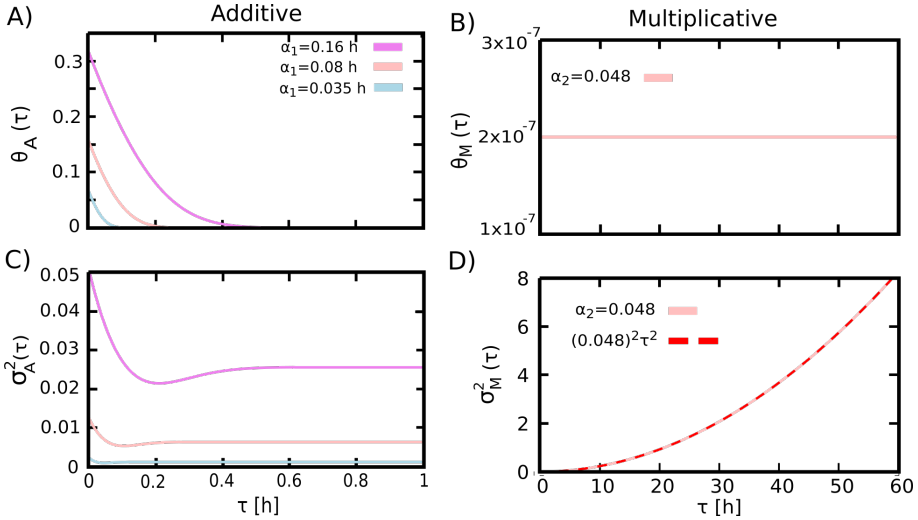


FIGURE D.4: **Border effects in the variation functions**. First (**A**-**B**) and second (**C**-**D**) moment of the variation functions as function of the trait τ in additive (left) and multiplicative (right) variation cases. In **(A)** the additive case, $\theta_A(\tau)$ is positive in $\tau = 0$ but decays rapidly to zero as τ increases, such that it is sufficient to restrict τ axis between 0 and 1. In **(C)** the same is shown for the second moment $\sigma_A^2(\tau)$. Note that the magnitude of the dependence decreases with α_A). Nevertheless, the value used in chapter 4, $\alpha_A = 0.16h$, is too large to neglect (D.116) this effect in the generalized Crow-Kimura eq. On the other hand, in the multiplicative case chapter 4 value, $\alpha_M = 0.048$, is small enough to avoid the border effects in the GCK eq. In **(B)** one can observe that $\theta_A(\tau)$ almost vanishes, and in **(C)** that the exact moment $\sigma_M^2(\tau)$ coincides quite well with the approximation $\alpha_M \tau^2$(D.120).

D.4.2 Finite-size effects

Here we complement the analysis of the finite-size effects of chapter 4 with the number of growing and dormant individuals as a function of time during a asymptotic cycle (Fig. D.9). In addition, the definition of the parameter $\delta^{st.}(T_a)$ (see Fig D.5), which os analogous to eq.(D.121), to measure the difference between the macroscopic theory and the simulations is:

$$\delta^{as.}(T_a) \equiv \frac{1}{M}\sum_{j=0}^{M}\delta_j^{as.} = \frac{1}{M}\sum_{j=0}^{M}|\langle\tau_j^{sim.}\rangle - \langle\tau_j^{theo.}\rangle|. \tag{D.124}$$

where $M+1 = \frac{T_{max}}{\Delta t}+1$ is the number of measures taken through a whole cycle (note that +1 is added since measures starts at $t = 0$), $T_{max} = 23h$ is the total duration of one single experimental cycle, $\Delta t = 0.5h$ is the waiting interval between measures, $\langle\tau_j^{sim.}\rangle$ and $\langle\tau_j^{theo.}\rangle$ are the mean lag time calculated from simulations and the generalized Crow-Kimura equation at time $j\cdot\Delta t$, respectively and the super-index *as* indicates that it is calculated in the asymptotic state.

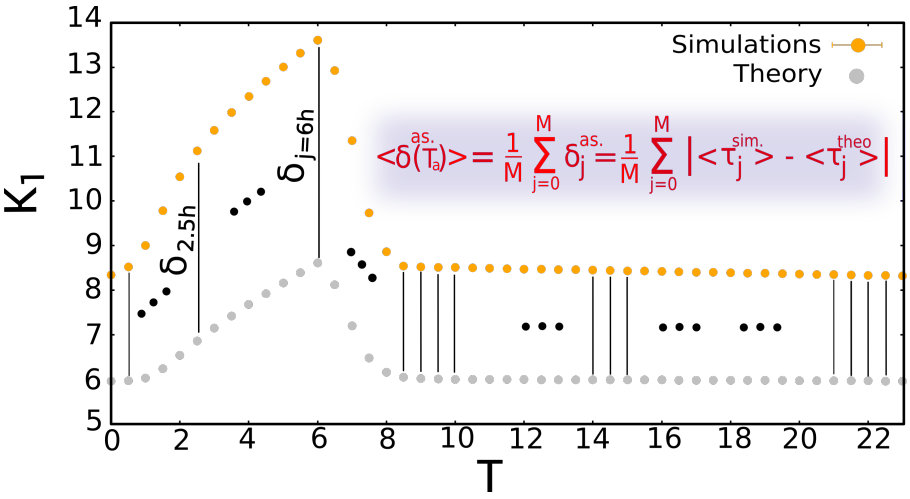


FIGURE D.5: **Schematic definition of the Parameter** δ (illustrated in the multiplicative amplitude scenario). $\alpha_M = 0.048$, $T_a = 6h$

D.5 Spontaneous shifting to the dormant state

In this section we investigate how the main results are changed when there is a non-vanishing spontaneous rate to enter the dormant state in the absence of antibiotics, e.g. when starving. Let us call such a generalized rate $S(\tau_i, \boldsymbol{\tau}_N)$; in chapter 4 we use

$$S(\tau_i, \boldsymbol{\tau}_N) = \begin{cases} s_k = s & \text{if} \quad 0 \le T \le T_a \\ s_f = 0 & \text{if} \quad T_a < T \le T_{max} \end{cases} \tag{D.125}$$

where the sub-indexes k and f denote "killing medium" and "fresh medium", respectively.

Given that bacteria can enter in dormant state when they sense lack of resources, i.e. when they approach starvation (i.e. the system carrying capacity), we consider a small, but non-zero, value for the rate to enter the dormant state in the fresh medium. In particular, we keep s_k constant as in chapter 4 but consider two possible scenarios for s_f:

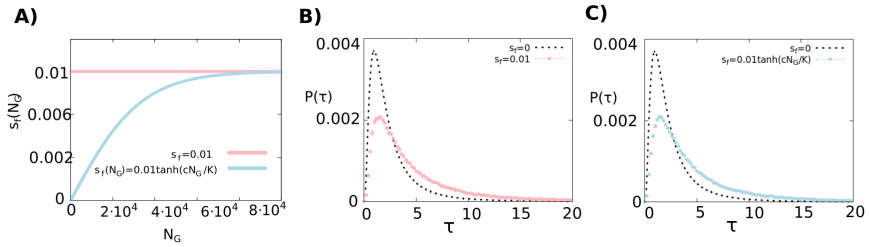


FIGURE D.6: **A)** Functional dependence of the rate to enter the dormant state in fresh medium s_f and the number of particles in growing state N_G. **B)** and **C)** Lag-time probability distribution function, $P(\tau)$, at the end of the tenth cycle (for the multiplicative case) for a constant s_f and $s_f = s_k\,(\tanh[-cN_G/K])$, respectively. The results in chapter 4, for $s_f = 0$, are represented by a dashed line. Observe that in both cases the value of K_1, K_2, K_3 increase, but the qualitative form of the distributions remains unchanged. Parameter values: $h_\tau = 0.01$, $\alpha_M = 0.048$, $T_a = 3h$, $c = 3$, $T = 23h$, the rest of parameters are fixed as in chapter 4.

1. a constant rate $s_f = \tilde{s} = 0.01$ all across the fresh-medium phase, or

2. a state-dependent rate, increasing as the carrying capacity is approached; in particular we consider a sigmoidal-like function $s_f = \tilde{s}\,(\tanh[-cN_G/K])$, $(\tilde{s} = 0.01)$.

Computational results for the variants of the model obtained by implementing these types of rate are displayed in D.6 Figs, corresponding to the end of the $10th$ exposure cycle (results are presented only for the multiplicative version of the model, but similar curves and conclusions can be obtained for the additive version).

Observe that, on the one hand, a non-vanishing s_f increases the values of K_1, K_2, K_3, but maintains the functional form of the lag-time distributions, see Fig D.6B and D.6C. Thus, one can recover the results of chapter 4 by just tuning the mutational amplitude parameter, α_M. In conclusion, considering $s_f = 0$ —though not a completely realistic assumption— is a reasonable approximation which eases the tractability of the model by reducing the computational times without significantly modifying the qualitative results.

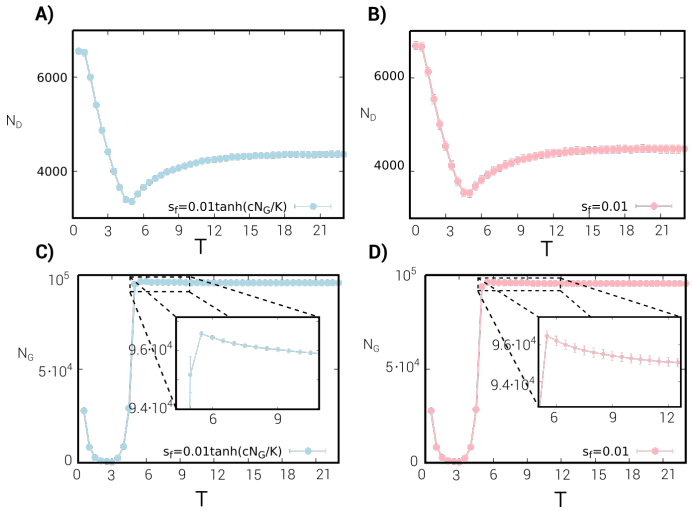


FIGURE D.7: **Dynamics of the averaged population structure within one cycle**. Abundances N_G and N_D along the 10th cycle for s_f constant —**A)** and **C)**— and $s_f = \tilde{s}\,(\tanh[-cN_G/K])$ —**B)** and **D)**—. The two scenarios are very similar to each other. At the beginning of the cycle N_D is non-vanishing, since bacteria entered this state during the fresh phase of the previous cycle. During the antibiotic exposure phase both, N_D and N_G, decrease to a minimum as the bacteria die. When the antibiotic is removed and a fresh medium is added, N_G grows towards the system's carrying capacity. Observe that the bacteria still enter the dormant state, but the reproduction rate is much higher, in such a way that an overall reduction of N_G is only observed as the system approaches the carrying capacity.

D.6 Additional Figures

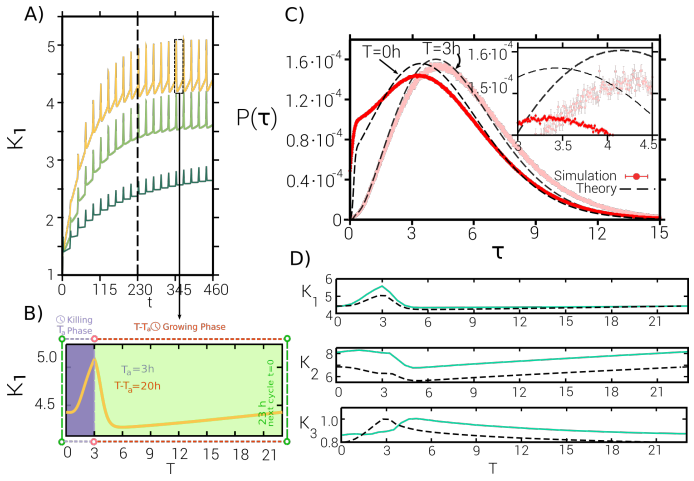


FIGURE D.8: **Characterization of the asymptotic state in the additive version of the model. (A)** Relaxation of the mean lag-time to its asymptotic state (curves obtained from the integration of Equation (D.110) with $T_a = 3h$ and additive model. The different curves correspond to three different values of the variation amplitude, α_A; from the lowest to the highest: $\alpha_A = 0.035h$, $\alpha_A = 0.1h$ and $\alpha_A = 0.16h$. **(B)** Zoom of the curve $\alpha_A = 0.16h$ for one single cycle. In particular, the mean lag time, K_1, is shown along a cycle in the asymptotic regime. At $T = 0$ the antibiotic is added and the system enters in the "killing phase" ($T \in [0, T_a]$). When the antibiotic is present, the system experiences a selection pressure towards longer lag times, in consequence, K_1 increases. At $T = T_a$ the antibiotic is washed and the fresh medium is added ($T \in [T_a, T_{max}]$). In this regime, the selection pressure is towards shorter lag times and K_1 relaxes back to the initial value. **(C)** Lag-time probability distribution at $T = 0$ (leftmost curve) and $T = T_a = 3h$ (rightmost curve) as derived theoretically (Eq.(D.110), dashed lines) and computationally (dots). In the asymptotic state the system oscillates between these two limiting probability distributions, both of them exhibiting weak tails. **(D)** Evolution of the first three cumulants, K_1, K_2 and K_3, (mean, variance , and skewness respectively) along a cycle in the asymptotic state (both theoretical and computational results are displayed). Observe that in C/D the theory correctly predicts the properties of the distribution, however there are deviations due to finite size effects.

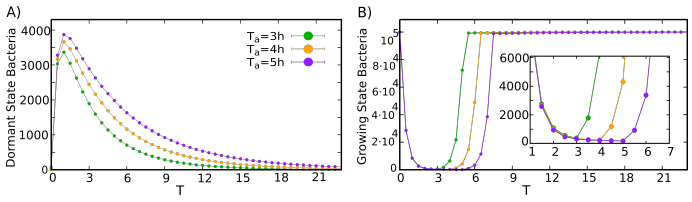


FIGURE D.9: **Simulated number of cells via the Gillespie algorithm** —additive-amplitude scenario, linear scale—. **a)** Number of cell in dormant state during a whole cycle in the asymptotic state for different exposition times T_a. **b)** Same as a) but for growing state bacteria. Obviously, the minimum number of bacteria in growing state $N_{G,min}$ is reached at $t = T_a$, since, during the antibiotic exposure phase, that number can only decrease. Parameters: $\alpha_A = 0.16h$, the rest are fixed as in chapter 4.

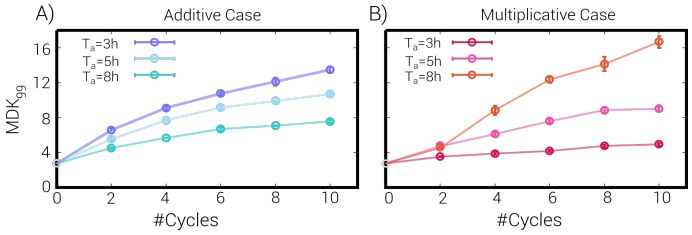


FIGURE D.10: **Evolution of the MDK_{99} over 10 exposure cycles**. **a)** Additive case. **b)** Multiplicative case. In our simulations, the MDK_{99} is calculated analogously to the experimental procedure. After a certain number cycles (i.e. # in the figure) of antibiotics-fresh environment, the evolved population is posed back in the antibiotic phase for a long time. The maximum number of cycles is 10 as in the experiments. The MDK_{99} is estimated by the time necessary to kill to the 99% of the population. Both in the additive and multiplicative case, the MDK_{99} increases with the # and T_a. Interestingly, in the multiplicative case by increasing T_a the change in MDK_{99} is bigger than the additive one.

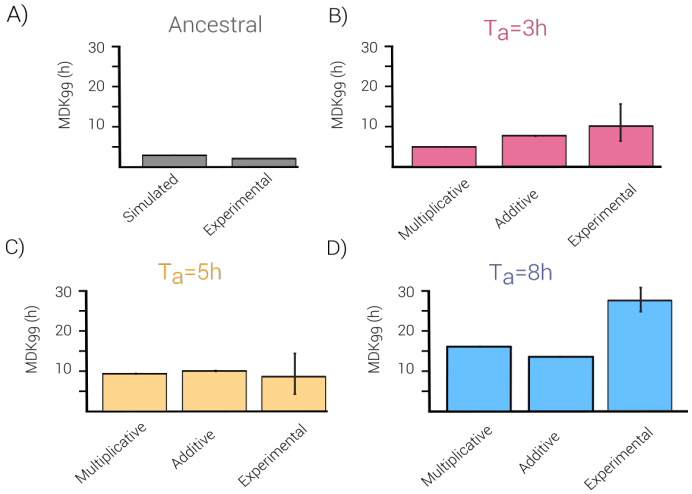


FIGURE D.11: **Comparison of experimental and simulated MDK_{99} of the evolved population after 10 cycles of exposure**. For both the additive and multiplicative cases, we observe that the simulated MDK_{99} falls within —or it is very close— the experimental values (i.e. the mean plus error) for $T_a = 3h$ and $5h$, but outside for the case of $T_a = 8h$. This result is to be expected given the higher noise of the experimental measurements. In particular the experimental mean is $\langle\tau\rangle^{exp.}_{T_a=8h} = 10 \pm 1h$ higher than the theoretical prediction of $8h$.

D.7 Movies

We provide three videos to help visualize the system dynamics:

- (V1) https://github.com/MCMateu/Phenotypic-dependent-variability/blob/main/AsymptoticAdditive.mp4
- (V2) https://github.com/MCMateu/Phenotypic-dependent-variability/blob/main/AsymptoticMultiplicative.mp4
- (V3) https://github.com/MCMateu/Phenotypic-dependent-variability/blob/main/TransitoryAdditive.mp4

The first two videos show the evolution of $\phi(\tau, t)$ along an antibiotic exposure cycle in the asymptotic state, calculated both by the simulation via Gillespie algorithm and by numerical integration of generalized Crow-Kimura equation (D.107) in the additive (V1) and multiplicative (V2) case respectively. Note that, as in Fig 4 of chapter 4, there exists a small deviation between theory and simulation, due to finite-size effects. The duration of the killing phase is $T_a = 3h$, and the duration of the whole cycle is $T = 23h$. The full length of the video is 23 s, thus each second approximately matches $1h$ of the cycle. During the first three seconds the distributions move towards longer lag times (to the right side), since individuals with longer lag times are the one that survives, until the mean lag time reaches a maximum value when antibiotics are removed. Then, the distribution moves back towards the initial configuration. Numerical value the of parameters: $\alpha_A = 0.16h$, $\alpha_M = 0.048$, $b = 2.4h^{-1}$, $s = 0.12h^{-1}$, $d = 3.6 \cdot 10^{-5}h^{-1}$.

On the other hand, the third video (V3) shows the transient dynamics of $\phi(\tau, t)$ as calculated by the numerical integration of the general mean-field equation (D.92) and the generalized Crow-Kimura (D.107) in the additive case. The full length of the video is $50s$ at a speed of $25fps$, the full video counts with 1250 frames. Each cycle has 50 frames, so the total duration of each cycle is approximately $2s$. The duration of the killing phase is $T_a = 5h$. The oscillations observed in the video correspond with the killing phase, when the density moves to the right, and with the growing phase, when it moves to the left. Note that the two distributions matches almost perfectly. Numerical value the of parameters: $\alpha_A = 0.035h$, $b = 2.4h^{-1}$, $s = 0.12h^{-1}$, $d = 3.6 \cdot 10^{-5}h^{-1}$.

Appendix E

Resumen y conclusiones en castellano

Resumen

La física estadística estudia las propiedades macroscópicas de agregados de muchas moléculas o componentes, desde gases hasta sistemas ferromagnéticos. En lugar de seguir la trayectoria individual de cada componente, la física estadística adopta un enfoque de conjunto y estudia la probabilidad de una cierta configuración colectiva. Al pasar al nivel macroscópico, se revelan nuevos fenómenos colectivos y cooperativos, como diferentes "fases" y transiciones particulares entre ellas [9]. Al aplicar este paradigma a sistemas no físicos, como los biológicos o sociales, los físicos han contribuido a crear la disciplina de los "sistemas complejos", cuyo objetivo es estudiar los fenómenos colectivos naturales. Los ecosistemas son un ejemplo paradigmático de sistemas complejos, ya que están compuestos por un gran número de especies cuyas interacciones generan propiedades emergentes colectivas, como la diversidad, la estabilidad y las funciones. Sin embargo, las comunidades ecológicas son mucho más complejas que los gases, ya que en las primeras las diferencias individuales tienen un efecto, mientras que en los últimos se promedian debido al enorme tamaño de 10^{23} moléculas por mol. Además, mientras que en la física estadística el sistema está en equilibrio termodinámico, las comunidades ecológicas se encuentran lejos del equilibrio debido a las interacciones no triviales entre especies y las constantes fluctuaciones del entorno. Por último, las especies están siempre en un proceso de "evolución" que, aunque en escalas de tiempo largas, provoca una transformación continua de los componentes e interacciones del sistema. Por lo tanto, construir una física estadística de los sistemas ecológicos es un objetivo mucho más difícil que el enfrentado por Boltzmann y Gibbs.

Para facilitar la tarea, intentaremos construir una teoría tan ambiciosa en uno de los sistemas más accesibles disponibles: las poblaciones bacterianas. De hecho, gracias a los avances tecnológicos recientes, las comunidades bacterianas (o más en general, microbianas), como el microbioma intestinal humano, pueden ser muestreadas, analizadas y secuenciadas fácilmente, obteniendo las abundancias de típicamente 10^3 especies para un total del orden de 10^{13} individuos. A primera vista, este número no es tan grande como el número de Avogadro, pero es más que el número total de árboles en el planeta. Por lo tanto, las comunidades microbianas son el sistema ecológico ideal para estudiar fenómenos emergentes.

Efectivamente, las características de las poblaciones bacterianas, como su corto tiempo de división y su alta tasa de mutación, las convierten en sistemas excelentes para estudiar fenómenos evolutivos. La aparición de tolerancia a los antibióticos y el surgimiento de diferentes especies son solo algunos ejemplos de los complejos procesos evolutivos que se pueden observar en las poblaciones bacterianas [10, 11]. Estos procesos tienen implicaciones significativas en medicina, especialmente en la comprensión del papel de la

microbiota intestinal en enfermedades crónicas y el alarmante aumento de la resistencia a los antibióticos. Al desarrollar una física predictiva de las comunidades microbianas utilizando la física estadística de no equilibrio, podemos obtener una comprensión más profunda de los mecanismos subyacentes que impulsan estos fenómenos biológicos. La física estadística de no equilibrio proporciona herramientas y conceptos poderosos para estudiar sistemas complejos, incluyendo la dinámica de las poblaciones, la aparición de comportamiento colectivo y el papel de las fluctuaciones y la irreversibilidad.
La tesis se divide en tres partes, correspondientes a los argumentos investigados: ecología, evolución y física estadística fuera del equilibrio. En el Capítulo 1, ofrecemos una visión general de las motivaciones y el contenido de esta tesis, junto con una introducción general a los sistemas complejos y la ecología y evolución de las bacterias (considerar la siguiente sección para un resumen en español).
En la Parte I, nos enfocamos en la macroecología microbiana y exploramos las interacciones dentro de los ecosistemas bacterianos. A través del análisis de datos, descubrimos una ley macroecológica universal que relaciona la correlación entre especies con su distancia filogenética. Mediante el desarrollo de un modelo estocástico basado en la física estadística, atribuimos este patrón a fluctuaciones ambientales acopladas, conocidas como filtro ambiental.
La parte II se adentra en la eco-evolución bacteriana. En el Capítulo 3 establecemos un nuevo marco teórico utilizando herramientas de física estadística para estudiar la distribución de fenotipos y diversos fenómenos evolutivos, como la especiación simpátrica. Además, en el Capítulo 4 empleamos este marco para investigar la evolución de la tolerancia a los antibióticos en bacterias mediante la adaptación del tiempo de lag. Presentamos un modelo estocástico que reproduce los resultados experimentales y obtenemos predicciones analíticas utilizando nuestro marco teórico.
En la Parte III, examinamos el concepto de irreversibilidad en la física estadística de sistemas fuera de equilibrio. El Capítulo 5 se centra en las propiedades geométricas de las corrientes en termodinámica estocástica y sus implicaciones. Analizamos la relación entre irreversibilidad, disipación y rupturas de simetría de las corrientes en estados estacionarios fuera de equilibrio. En particular, conseguimos generalizar el principio de Prigogine utilizando el exceso de entropía.
En el Capítulo 6, exploramos las propiedades irreversibles de la evolución Darwiniana utilizando el marco teórico general presentado anteriormente. Descubrimos que la evolución se mantiene constantemente fuera de equilibrio debido a la presencia de selección y mutaciones, y estudiamos la irreversibilidad en ejemplos como la especiación.
Finalmente, en el Capítulo 7, presentamos conclusiones generales y sugerimos posibles direcciones para futuras investigaciones. Al emplear la física estadística, esta tesis contribuye a la comprensión de los sistemas complejos, proporcionando conocimientos sobre el comportamiento colectivo, la ecología y la evolución de las comunidades bacterianas.
En el siguiente capítulo proponemos una concisa introducción al mundo de los microbios.

Un rápido viaje en el mundo microbiano

Los microbios son organismos microscópicos, tanto unicelulares como pluricelulares, como bacterias, arqueas, protistas y algunas especies de algas y hongos. Aunque los microbios han sido utilizados por los seres humanos en la preparación de alimentos desde la antigüedad, fueron descubiertos por primera vez por el biólogo neerlandés Leeuwenhoek en 1677 [33].
Las bacterias son los organismos más abundantes en la Tierra (aproximadamente 10^{30} células [34], una biomasa total de $70Gt$ de carbono) y se encuentran en toda la biosfera,

desde los intestinos humanos hasta los glaciares, desde el suelo hasta los lodos activados. Las bacterias son organismos unicelulares (procariotas), similares a un sistema "abierto", capaces de aprovechar la energía del entorno y convertirla en biomasa y residuos. Las bacterias se "reproducen" mediante división asexual; el tiempo de duplicación puede variar desde 20 minutos para la E. coli en condiciones de laboratorio hasta dos años para los microbios que viven en sedimentos del fondo marino. Su metabolismo es altamente diverso, ya que pueden realizar la respiración molecular, la fermentación e incluso algunas especies son capaces de la fotosíntesis.
La ecología microbiana estudia cómo crecen, interactúan y se mantiene la diversidad de especies en las comunidades microbianas. Esta cuestión es especialmente relevante para las bacterias, debido a la gran diversidad observada a nivel taxonómico de cepas [35]. Por ejemplo, en las poblaciones marinas de la cianobacteria Prochlorococcus (género), se observan cientos de cepas diferentes [31]. ¿Cómo pueden coexistir tantas cepas?
Muchos procesos ecológicos posibles, o "fuerzas" ecológicas, pueden surgir en comunidades diversas, pero en general son consecuencia de las propiedades metabólicas de los individuos.
La fuerza ecológica que generalmente se considera dominante es la "competencia". Puede surgir entre dos individuos cuando comparten un recurso común de suministro limitado [36]. Cuando hay competencia, el crecimiento de una especie implica una reducción en el número de la especie competidora, lo que posiblemente puede llevar a extinciones.
En términos generales, la competencia se asocia con el "Principio de Exclusión Competitiva", que establece que "dos especies o poblaciones no pueden habitar el mismo nicho: una consistentemente superará a la otra" [37]. El nicho de una especie es un concepto ecológico abstracto que consiste en sus interacciones con su comunidad y las condiciones ambientales necesarias para su supervivencia. Por lo tanto, si la exclusión competitiva está en acción, solo puede haber un número de especies igual o menor al número de nichos disponibles.
También pueden surgir interacciones mutualistas entre especies bacterianas gracias a los subproductos del metabolismo celular. De hecho, a menudo ocurre que dos especies competidoras pueden coexistir si una es capaz de metabolizar el subproducto de la otra. Este fenómeno se llama "cross-feeding" y se encuentra en muchas comunidades bacterianas [38, 39, 40]. Por ejemplo, un par de cepas de las familias Pseudomonadaceae y Enterobacteriaceae pueden sobrevivir juntas porque la primera puede crecer utilizando acetato, que es un subproducto de la segunda [39].
Los subproductos metabólicos de una especie también pueden inhibir la presencia de otra, es decir, un comportamiento llamado "antagonismo". Por ejemplo, la bacteria Streptomyces coelicolor es capaz de producir antibióticos, que luego se utilizan para neutralizar a otras especies como E. coli [41].
La depredación también está presente en los procariotas, aunque generalmente se considera rara (solo se han identificado quince especies depredadoras) [42, 43]. La bacteria altamente móvil B. bacteriovorus ha sido reportada como capaz de ingresar al espacio periplasmático de otras bacterias y consumir su contenido citoplasmático [44].

Las interacciones ecológicas también pueden producir efectos colectivos, como las funciones comunitarias. De hecho, las funciones ecológicas no son solo propiedades individuales, sino más bien propiedades colectivas de las comunidades que emergen de una compleja red de interacciones moleculares, fisiológicas y organismales [45], como el ciclo del carbono en ecosistemas cerrados [46], el ciclo de nutrientes en el suelo [47] y el ciclo del nitrógeno en el océano [48]. Por último, los ecosistemas microbianos muestran redundancia funcional: aunque la presencia y abundancia de especies varía entre comunidades, se conserva el número y tipo de funciones realizadas [38, 49]. Aunque las

comunidades bacterianas exhiben una diversidad taxonómica astronómica, muchas especies comparten las mismas funciones metabólicas, lo que sugiere que la selección no actúa solo a nivel de especies.
Un efecto colectivo adicional que surge en las comunidades bacterianas son los *biofilms*, es decir, comunidades de bacterias que se adhieren colectivamente a superficies, creando un medio físico adaptado para interactuar a través de señalización e intercambio de nutrientes.
Finalmente, las fuerzas ecológicas no son solo interacciones "internas" entre especies, sino también factores "externos" como fluctuaciones ambientales y demográficas, migraciones, estructura espacial, etc. Por ejemplo, se sabe que el caos espacial en modelos de meta-poblaciones amplía enormemente el número de especies coexistentes en comparación con la exclusión competitiva [50, 51, 52].
Por lo tanto, uno podría preguntarse: ¿Es posible desentrañar los efectos de estas diferentes fuerzas? ¿Cuál es la principal responsable de la existencia de la biodiversidad? Existe un debate interminable sobre la importancia relativa de las fuerzas deterministas, como la exclusión competitiva, o las estocásticas, como las fluctuaciones ambientales y demográficas, en la determinación del mantenimiento de la biodiversidad en la comunidad. En comunidades muy diversas, como el prochlorococcus o el plancton marino [53], la existencia de un número astronómico de nichos es hipotético y no completamente razonable. Por lo tanto, se puede cuestionar la importancia relativa de la exclusión competitiva.
El cross-feeding parece tener la capacidad de mantener la diversidad, ya que las comunidades que crecen en una sola fuente de carbono pueden soportar hasta 40 taxa gracias al reciclaje de metabolitos.
Además, las comunidades bacterianas tienen diferentes "diversidades" dependiendo del nivel taxonómico en el que se consideren. ¿Cómo se puede elegir la escala taxonómica correcta? Aunque la mayoría de nuestro conocimiento sobre ecología microbiana ha surgido de entornos de laboratorio simples y controlados, las respuestas a estas preguntas fundamentales se pueden obtener interrogando a comunidades naturales. En los últimos años, los proyectos del microbioma humano y programas similares han muestreado y caracterizado un número astronómico de comunidades [54, 55]. Comprender los patrones generales que emergen en estas comunidades es útil tanto a nivel fundamental como con importantes consecuencias para la salud general en el caso del microbioma intestinal humano [56].

Conclusiones

En esta tesis hemos estudiado fenómenos colectivos en ecología y evolución microbiana, utilizando la física estadística (del no equilibrio). En esta sección, derivamos algunas conclusiones generales y comentamos sobre posibles desarrollos futuros. Antes de adentrarnos en perspectivas futuras, resumamos brevemente los resultados reportados en esta tesis:

- En el Capítulo 2, hemos revelado la emergencia de una nueva ley macroecológica en comunidades microbianas. Esta ley empírica establece en términos cuantitativos que el promedio de la función de correlación entre pares de especies disminuye desde valores positivos hasta valores nulos a medida que aumenta la distancia filogenética, aproximadamente siguiendo una función de "stretched exponential".
  Al definir un modelo ecológico estocástico (el CSLM), hemos demostrado que las fluctuaciones ambientales acopladas (*multiplicativas*), también conocidas como filtrado ambiental, son responsables de este patrón universal. Al vincular las preferencias ecológicas con la distancia filogenética de las especies, podemos formular un modelo en el cual los árboles filogenéticos se pueden utilizar directamente. Por último, pero no menos importante, analizamos datos temporales para una comunidad fija, mostrando que la ley macroecológica también se cumple cuantitativamente en este contexto y que las correlaciones temporales retardadas se reproducen naturalmente en nuestro modelo con filtrado ambiental. Por lo tanto, los resultados reportados en este capítulo nos permiten concluir que solo el *filtrado ambiental* (y no, por ejemplo, la competencia entre especies) explica el patrón observado empíricamente de correlaciones decrecientes con la distancia filogenética, y por lo tanto, es probablemente la fuerza ecológica dominante en las comunidades microbianas naturales.

- En el Capítulo 3, introdujimos un marco eco-evolutivo para los rasgos bacterianos. En particular, describimos la evolución adaptativa como la dinámica de la distribución de rasgos de la población en espacios fenotípicos, sintetizada por la llamada ecuación Generalized Crow-Kimura (GCK). Al modelar explícitamente la selección, las mutaciones y la deriva de tamaño finito, generalizamos la celebrada teoría de la dinámica adaptativa a escenarios eco-evolutivos. Con nuestro formalismo, podemos describir exactamente la población y la distribución de rasgos incluso después de una ramificación evolutiva, y en particular, predecir el pico de las dos subpoblaciones. Por último, generalizamos nuestra descripción para incluir efectos estocásticos de tamaño finito, revelando la posible frustración de la ramificación evolutiva. La frustración se debe a las peculiares fluctuaciones multiplicativas de la varianza de rasgos, que pueden crear un estado metastable o un estado absorbente en la dinámica. Por lo tanto, el trabajo presentado en este capítulo contribuye al desarrollo de una teoría eco-evolutiva para comunidades microbianas, lo que permite arrojar más luz sobre la asombrosa diversidad observada empíricamente en los rasgos y las interacciones de las comunidades microbianas. Nuestra esperanza es que este trabajo haga que este tipo de enfoques cuantitativos para comunidades eco-evolutivas complejas sean accesibles a un público más amplio, que incluya físicos, biólogos y ecólogos.

- En el Capítulo 4, aplicamos el marco eco-evolutivo mencionado anteriormente a la aparición de la tolerancia a los antibióticos mediante la adaptación del tiempo de retraso en las bacterias. Experimentos de laboratorio recientes en E. coli han demostrado que las bacterias adaptan su tiempo de retraso para sobrevivir a los

antibióticos en un estado latente [10]. Las distribuciones de la población de tiempos de retraso resultantes muestran una proporcionalidad directa entre el tiempo de retraso promedio y su varianza. Identificamos esta característica como típica de los procesos multiplicativos y diseñamos un modelo evolutivo donde la variación del tiempo de retraso de la descendencia es proporcional al valor de la célula madre. El modelo puede describirse de manera efectiva mediante la ecuación GCK con un nuevo término de "variaciones multiplicativas" y es capaz de reproducir las distribuciones empíricas de retraso. Por lo tanto, el trabajo presentado en este capítulo arroja luz sobre la evolución de la tolerancia a los antibióticos y, además, representa un ejemplo de una clase general de problemas evolutivos donde se puede aplicar nuestro marco matemático.

- La última parte de la tesis se dedicó a la física estadística de no equilibrio y, en particular, en el Capítulo 5 ilustramos las propiedades geométricas y topológicas generales de las corrientes de velocidad en los estados estacionarios de no equilibrio. Estas propiedades geométricas pueden explicarse como una ruptura de simetría inducida por la irreversibilidad y la disipación. Mientras que en dimensiones iguales o mayores que dos, los sistemas estocásticos con ruido aditivo generan velocidades de corriente sin divergencia, la adición de ruido *multiplicativo* induce situaciones geométricas más complicadas. En particular, en este caso, las velocidades tipo gradiente empujan al sistema hacia regiones del espacio de fase con baja fluctuación, como estados absorbentes. Es notable que, gracias a este marco geométrico, podemos dar una interpretación termodinámica clara de la entropía en exceso y relacionarla con el principio de Prigogine, resolviendo un problema debatido con una larga historia.

- Para concluir, en el Capítulo 6 estudiamos la naturaleza irreversible de la evolución utilizando el marco de trabajo introducido en el Capítulo 3, tanto de manera general como en varios escenarios concretos, como la evolución neutral, la selección estabilizadora y la ramificación evolutiva. Descubrimos que la evolución, en general, está fuera del equilibrio debido a la presencia contemporánea de una corriente de mutación y diferencias de selección. Aunque estas fuerzas se equilibran en un estado estacionario, sus efectos diferentes en la población alejan al sistema del equilibrio. Además, encontramos que las diversas contribuciones a la producción de entropía se correlacionan con los regímenes evolutivos y las transiciones evolutivas que experimenta la población. En la discusión, en la sección 3.7, comentamos sobre una posible generalización estocástica del cambio en la entropía de Shannon y, en particular, destacamos su naturaleza *multiplicativa*. Por lo tanto, el trabajo presentado en este capítulo arroja luz sobre el aspecto fundamental del tiempo en la adaptación y revela el papel de las mutaciones y la selección en la creación de irreversibilidad durante la evolución darwiniana.

Epílogo

En el libro "¿Qué es la vida?", Erwin Schrödinger sentó las bases de la biología contemporánea al analizar los fenómenos biológicos desde la perspectiva de la física [533]. Su pregunta central se plantea en la introducción: *"¿Cómo pueden explicarse los eventos en el espacio y el tiempo que tienen lugar dentro de los límites espaciales de un organismo vivo mediante la física y la química?"*. En el libro, Schrödinger explora diferentes conceptos y fenómenos como el caos molecular, la evolución darwiniana, el origen físico de las mutaciones y, finalmente, la naturaleza de la herencia biológica. Al discutir este último argumento,

Schrödinger teorizó que el material genético debe ser una molécula, pero una en particular que él define como un "cristal aperiódico", capaz de codificar una gran cantidad de información. Aunque algunas teorías de la herencia ya estaban en circulación, la intuición de Schrödinger se puede considerar una predicción teórica sobre las propiedades del ADN, y estimuló un vigoroso interés en investigar las bases moleculares de la genética. De hecho, Watson y Crick reconocieron a Schrödinger como fuente de inspiración.
En la última parte del libro titulada "Sobre el determinismo y el libre albedrío", Schrödinger se adentra en reflexiones más metafísicas sobre la naturaleza de la conciencia, influenciado principalmente por Schopenhauer y las Upanishads indias. El inicio del capítulo dice: *"Como recompensa por los esfuerzos que he tenido que asumir para exponer los aspectos puramente científicos de nuestro problema "sine ira et studi, permítanme agregar mi propia visión, necesariamente subjetiva, de sus implicaciones filosóficas."*
Consideramos que el uso de un científico de exponer sus ideas filosóficas sea un ejercicio muy saludable y necesario. De hecho, la mayoría de los grandes físicos del siglo pasado, como Einstein, Bohr, Pauli, Heisenberg y, por supuesto, Schrödinger, escribieron extensamente sobre sus ideas filosóficas. Por lo tanto, con humildad, proponemos aquí algunas reflexiones generales.
Aquí no estamos interesados en la conciencia, pero compartimos con Schrödinger la necesidad de no interpretar al individuo como la base de la naturaleza. De hecho, las relaciones ecológicas colectivas, la transferencia genética lateral (LGT) y la simbiosis en la evolución nos invitan a repensar el concepto de individuo.
En la sección 1.3.1expusimos brevemente que durante la evolución biológica ocurrieron muchos eventos de endosimbiosis. Por ejemplo, las mitocondrias eucariotas son los remanentes de una proteobacteria capturada por, probablemente, un arquea. Los virus también han desempeñado un papel importante en la configuración de la vida tal como la conocemos. En 2001, el Consorcio Internacional de Secuenciación del Genoma Humano estableció que el 8% del genoma de los mamíferos, y por lo tanto de los humanos, consiste en remanentes de antiguos virus, mientras que hasta el 40% está compuesto por repeticiones de secuencias genéticas que también se cree que tienen origen viral [534, 17]. Los virus antiguos eran endógenos, como el VIH moderno, y para reproducirse tenían que insertar su material genético en el genoma de su huésped. Algunos de ellos eran capaces de infectar células germinales, como óvulos o espermatozoides, y transmitieron su genoma a las generaciones futuras [535]. Esta parte híbrida del genoma a menudo es beneficiosa y en ratones se ha descubierto que brinda protección contra virus similares. Sin embargo, el hecho más sorprendente es que este genoma viral es responsable de un proceso celular clave en el desarrollo de la placenta en mamíferos [536]. T. Heidmann, líder del grupo de investigación mencionado sobre la placenta, resumió estos hechos de manera provocativa: "Nuestros genes no solo son *nuestros* genes. También son genes retrovirales" [17].
Las interacciones ecológicas entre especies de diferentes dominios pueden ser tan estrechas como para poner en crisis nuestro concepto de individuo. Los líquenes son organismos compuestos que surgen de algas o cianobacterias que viven entre los filamentos de varias especies de hongos en una relación simbiótica. Los hongos se benefician de los nutrientes producidos por las algas o cianobacterias a través de la fotosíntesis. Por otro lado, las algas o cianobacterias se benefician al estar protegidas del medio ambiente por los filamentos de los hongos [537]. La simbiosis también es extremadamente frecuente en los ambientes marinos, especialmente como simbiosis quimiosintéticas entre bacterias e invertebrados, donde estos últimos son los productores primarios y proporcionan la mayor parte del carbono orgánico necesario para la nutrición del hospedador animal [538]. En un estudio reciente [539], se ha informado de una simbiosis tridireccional en

una planta de hierba tropical. Se pensaba que esta planta en particular crecía en temperaturas del suelo altas gracias a un hongo simbiótico que vive en sus hojas. Sin embargo, se descubrió que la responsabilidad de la tolerancia al calor recae en un virus que ha infectado al hongo.
Los seres humanos también han desarrollado relaciones simbióticas íntimas con microbios intestinales, conocidos como el microbioma, los cuales tienen una influencia significativa en la salud [540, 541, 56]. Además, evidencia reciente sugiere que la composición del microbioma intestinal es heredable, lo que significa que se transmite de nuestros ancestros [542].
A nivel más amplio, todas las plantas y animales viven en simbiosis con sus microbiomas, que consisten en bacterias, arqueas, hongos y protistas. Por lo tanto, no somos simplemente individuos, sino ecosistemas, o más específicamente, holobiontes. Como M. Sheldrake afirma en su libro "Entangled Life" [537], las comunidades microbianas nos hacen comprender que el concepto de "individuo" es simplemente una categoría científica con ciertos límites de aplicabilidad. La biología, que solía ser el estudio de organismos individuales, se ha transformado en ecología, el estudio de las interacciones entre los organismos vivos.

En esta perspectiva, la física estadística y los sistemas complejos brindan conceptos útiles para comprender el nuevo mundo colectivo que emerge de las interacciones de un número astronómico de seres vivos. Al aceptar que un individuo no es una entidad aislada, nos enfrentamos a un nuevo mundo interconectado que se despliega ante nuestros ojos.

Bibliography

[1] Emanuele Coccia. *Metamorphoses*. Polity, 2021.

[2] Alfred J. Lotka. "Elements of Physical Biology." In: *Nature* 116 (1925), pp. 461–461.

[3] M. Loreau. *From Populations to Ecosystems: Theoretical Foundations for a New Ecological Synthesis*. Princeton University Press, 2010.

[4] Yinon M. Bar-On, Rob Phillips, and Ron Milo. "The biomass distribution on Earth". In: *Proceedings of the National Academy of Sciences* 115.25 (2018), pp. 6506–6511.

[5] Pablo A. Marquet et al. "Scaling and power-laws in ecological systems". In: *Journal of Experimental Biology* 208.9 (May 2005), pp. 1749–1769.

[6] Ian A. Hatton et al. "The predator-prey power law: Biomass scaling across terrestrial and aquatic biomes". In: *Science* 349.6252 (2015), aac6284.

[7] Brian Enquist, James Brown, and GB West. "Allometric scaling of plant energetics and population density". In: *Nature* 395 (Jan. 1998), pp. 163–165.

[8] John Damuth. "Cope's rule, the island rule and the scaling of mammalian population density". In: *Nature* 365 (1993), pp. 748–750.

[9] J. Binney et al. Oxford, 1992.

[10] Nathalie Q. Balaban et al. "Bacterial Persistence as a Phenotypic Switch". In: *Science* 305.5690 (2004), pp. 1622–1625.

[11] Benjamin H. Good et al. "The dynamics of molecular evolution over 60,000 generations". In: *Nature* 551 (45 2017), p. 021007.

[12] P.H. Barrett et al. *Charles Darwin's Notebooks, 1836-1844: Geology, Transmutation of Species, Metaphysical Enquiries*. Cambridge University Press, 1988.

[13] Charles Darwin. *On the origin of species, 1859*. Routledge, 2004.

[14] J. R. Brown and W. Ford Doolittle. "Archaea and the prokaryote-to-eukaryote transition." In: *Microbiology and molecular biology reviews : MMBR* 61 4 (1997), pp. 456–502.

[15] Doolittle Wf. "Phylogenetic Classification and the Universal Tree". In: *Science* 284 (1999), pp. 2124–2128.

[16] William Martin. "Symbiogenesis, gradualism, and mitochondrial energy in eukaryote origin". In: *Periodicum Biologorum* 119 (Sept. 2017), pp. 141–158.

[17] David Quammen. *The Tangled Tree: A Radical New History of Life*. Simon Schuster, 2018.

[18] Kevin de Queiroz. "Ernst Mayr and the modern concept of species". In: *Proceedings of the National Academy of Sciences* 102.suppl_1 (2005), pp. 6600–6607.

[19] Ernest Mayr. *Systematics and the Origin of Species*. Columbia Univ. Press, New York, 1942.

[20] Carl R. Woese and George E. Fox. "Phylogenetic structure of the prokaryotic domain: The primary kingdoms". In: *Proceedings of the National Academy of Sciences of the United States of America* 74 (1977), pp. 5088 –5090.

[21] Emile Zuckerkandl and Linus Pauling. "Molecules as documents of evolutionary history". In: *Journal of Theoretical Biology* 8.2 (1965), pp. 357–366.

[22] Carl R. Woese, O Kandler, and Mark L. Wheelis. "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." In: *Proceedings of the National Academy of Sciences of the United States of America* 87 (1990), pp. 4576 –4579.

[23] Fred Griffith. "The Significance of Pneumococcal Types". In: *Journal of Hygiene* 27.2 (1928), 113–159.

[24] Howard Ochman, Jeffrey G. Lawrence, and Eduardo A. Groisman. "Lateral gene transfer and the nature of bacterial innovation". In: *Nature* 405 (2000), pp. 299–304.

[25] Maureen L. Coleman et al. "Genomic Islands and the Ecology and Evolution of Prochlorococcus". In: *Science* 311.5768 (2006), pp. 1768–1770.

[26] L. Margulis and R. Fester. *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*. Mit Press. MIT Press, 1991.

[27] Laura A Hug et al. "A new view of the tree of life". In: *Nature microbiology* 1.5 (2016), pp. 1–6.

[28] Pablo Yarza et al. "Uniting the classification of cultured and uncultured Bacteria and Archaea using 16S rRNA gene sequences". In: *Nature reviews. Microbiology* 12 (Aug. 2014), pp. 635–45.

[29] Sallie W. Chisholm et al. "A novel free-living prochlorophyte abundant in the oceanic euphotic zone". In: *Nature* 334 (1988), pp. 340–343.

[30] Benjamin Callahan, Paul Mcmurdie, and Susan Holmes. "Exact sequence variants should replace operational taxonomic units in marker-gene data analysis". In: *The ISME Journal* 11 (July 2017).

[31] Nadav Kashtan et al. "Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild Prochlorococcus". In: *Science* 344.6182 (2014), pp. 416–420.

[32] Siddhartha Mandal et al. "Analysis of composition of microbiomes: a novel method for studying microbial composition". In: *Microbial ecology in health and disease* 26.1 (2015), p. 27663.

[33] Nick Lane. "The unseen world: reflections on Leeuwenhoek (1677) 'Concerning little animals'". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1666 (2015), p. 20140344.

[34] William B. Whitman, David C. Coleman, and William J. Wiebe. "Prokaryotes: The unseen majority". In: *Proceedings of the National Academy of Sciences* 95.12 (1998), pp. 6578–6583.

[35] Thea Van Rossum et al. "Diversity within species: interpreting strains in microbiomes". In: *Nature Reviews Microbiology* 18 (June 2020), pp. 1–16.

[36] Michael E. Hibbing et al. "Bacterial competition: surviving and thriving in the microbial jungle". In: *Nature Reviews Microbiology* 8 (2010), pp. 15–25.

[37] S.E. Kingsland. *Modeling Nature*. University of Chicago Press, 1995.

[38] Joshua E. Goldford et al. "Emergent simplicity in microbial community assembly". In: *Science* 361.6401 (2018), pp. 469–474.

[39] Sylvie Estrela et al. "Functional attractors in microbial community assembly". In: *Cell Systems* 13.1 (2022), 29–42.e7.

[40] Jared Kehe et al. “Positive interactions are common among culturable bacteria”. In: *Science Advances* 7.45 (2021), eabi7159.

[41] Mervyn J Bibb. “Regulation of secondary metabolism in streptomycetes”. In: *Current Opinion in Microbiology* 8.2 (2005). Cell Regulation, pp. 208–215.

[42] Mark O. Martin. “Predatory prokaryotes: an emerging research opportunity.” In: *Journal of molecular microbiology and biotechnology* 4 5 (2002), pp. 467–77.

[43] Gregory Velicer and Helena Mendes-Soares. “Bacterial predators”. In: *Current biology : CB* 19 (Feb. 2009), R55–6.

[44] Renee Elizabeth Sockett. “Predatory Lifestyle of Bdellovibrio bacteriovorus”. In: *Annual Review of Microbiology* 63.1 (2009), pp. 523–539.

[45] Alvaro Sanchez et al. “The community-function landscape of microbial consortia”. In: *Cell Systems* 14.2 (2023), pp. 122–134.

[46] Luis Miguel de Jesús Astacio et al. “Closed microbial communities self-organize to persistently cycle carbon”. In: *Proceedings of the National Academy of Sciences* 118.45 (2021), e2013564118.

[47] Cameron Wagg et al. “Soil biodiversity and soil community composition determine ecosystem multifunctionality”. In: *Proceedings of the National Academy of Sciences* 111.14 (2014), pp. 5266–5270.

[48] Jonathan P. Zehr and Bess B. Ward. “Nitrogen Cycling in the Ocean: New Perspectives on Processes and Paradigms”. In: *Applied and Environmental Microbiology* 68.3 (2002), pp. 1015–1024.

[49] Stilianos Louca et al. “Function and functional redundancy in microbial systems”. In: *Nature ecology & evolution* 2.6 (2018), pp. 936–943.

[50] Alan Hastings et al. “Chaos in Ecology: Is Mother Nature a Strange Attractor?” In: *Annual Review of Ecology and Systematics* 24 (1993), pp. 1–33.

[51] J. Bascompte and R.V. Sole. *Modeling Spatiotemporal Dynamics in Ecology*. Springer Berlin Heidelberg, 1998.

[52] Aditya Mahadevan, Michael T. Pearce, and Daniel S. Fisher. “Spatiotemporal Ecological Chaos Enables Gradual Evolutionary Diversification Without Niches or Tradeoffs”. In: (2022).

[53] G. E. Hutchinson. “The Paradox of the Plankton”. In: *The American Naturalist* 95.882 (1961), pp. 137–145.

[54] Peter Turnbaugh et al. “The human microbiome project: exploring the microbial part of ourselves in a changing world”. In: *Nature* 449 (Nov. 2007), p. 804.

[55] Luke R. Thompson et al. “A communal catalogue reveals Earth’s multiscale microbial diversity”. English. In: *Nature* 551.7681 (2017), pp. 457–463. ISSN: 0028-0836.

[56] Yong Fan and Oluf Pedersen. “Gut microbiota in human metabolic health and disease”. In: *Nature Reviews Microbiology* 19 (2020), pp. 55 –71.

[57] James H Brown et al. *Macroecology*. University of Chicago Press, 1995.

[58] Brian J. McGill et al. “Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework”. In: *Ecology Letters* 10.10 (2007), pp. 995–1015.

[59] S.P. Hubbell and Princeton University Press. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, 2001.

[60] Jacopo Grilli, Tim Rogers, and Stefano Allesina. "Modularity and stability in ecological communities". In: *Nature communications* 7.1 (2016), p. 12031.

[61] Jacopo Grilli et al. "Feasibility and coexistence of large ecological communities". In: *Nature communications* 8.1 (2017), p. 14389.

[62] Stefano Allesina and Jacopo Grilli. "Models for large ecological communities—a random matrix approach". In: May 2020, pp. 74–92.

[63] Ada Altieri et al. "Properties of Equilibria and Glassy Phases of the Random Lotka-Volterra Model with Demographic Noise". In: *Phys. Rev. Lett.* 126 (25 2021), p. 258301.

[64] M. Kimura. "The neutral theory of molecular evolution." In: *Scientific American* 241 5 (1987), 98–100, 102, 108 passim.

[65] R.H. MacArthur and E.O. Wilson. *The Theory of Island Biogeography*. Princeton University Press, 2001.

[66] Stephen P. Hubbell. "The neutral theory of biodiversity and biogeography and Stephen Jay Gould". In: *Paleobiology* 31.sp5 (2005), pp. 122 –132.

[67] Igor Volkov et al. "Neutral Theory and Relative Species Abundance in Ecology". In: *Nature* 424 (Sept. 2003), pp. 1035–7.

[68] Sandro Azaele et al. "Dynamical evolution of ecosystems". In: *Nature* 444 (2006), pp. 926–928.

[69] Robert M May. "Will a large complex system be stable?" In: *Nature* 238 (1972), pp. 413–414.

[70] Wenping Cui, Robert Marsland, and Pankaj Mehta. "Effect of Resource Dynamics on Species Packing in Diverse Ecosystems". In: *Phys. Rev. Lett.* 125 (4 2020), p. 048101.

[71] Valentina Ros et al. *Quenched complexity of equilibria for asymmetric Generalized Lotka-Volterra equations*. 2023.

[72] Stefano Allesina and Si Tang. "Stability criteria for complex ecosystems". In: *Nature* 483.7388 (2012), pp. 205–208.

[73] Sandro Azaele et al. "Statistical mechanics of ecological systems: Neutral theory and beyond". In: *Rev. Mod. Phys.* 88 (3 2016), p. 035003.

[74] David Alonso, Rampal S. Etienne, and Alan J. McKane. "The merits of neutral theory". In: *Trends in Ecology Evolution* 21.8 (2006), pp. 451–457.

[75] Alan J. McKane, David Alonso, and Ricard V. Solé. "Analytic solution of Hubbell's model of local community dynamics". In: *Theoretical Population Biology* 65.1 (2004), pp. 67–73.

[76] Simone Pigolotti et al. "Stochastic Spatial Models in Ecology: A Statistical Physics Approach". In: *Journal of Statistical Physics* 172 (July 2018).

[77] Matteo Sireci and Miguel A Muñoz. "Statistical physics of phenotypic evolution: Adaptive dynamics and beyond". unpublished.

[78] Charles K. Fisher and Pankaj Mehta. "The transition between the niche and neutral regimes in ecology". In: *Proceedings of the National Academy of Sciences* 111.36 (2014), pp. 13111–13116.

[79] Rafael D'Andrea, Theo Gibbs, and James P. O'Dwyer. "Emergent neutrality in consumer-resource dynamics". In: *PLOS Computational Biology* 16.7 (July 2020), pp. 1–17.

[80] Matteo Sireci, Miguel A. Muñoz, and Jacopo Grilli. "Environmental fluctuations explain the universal decay of species-abundance correlations with phylogenetic distance". In: *bioRxiv* (2023).

[81] D.C. Krakauer. *Worlds Hidden in Plain Sight: The Evolving Idea of Complexity at the Santa Fe Institute, 1984-2019*. SFI Press, 2018.

[82] Philip W. Anderson. "More is different." In: *Science* 177 4047 (1972), pp. 393–6.

[83] Bernard Derrida. "Non-equilibrium steady states: Fluctuations and large deviations of the density and of the current". In: *Journal of Statistical Mechanics* 2007 (Mar. 2007).

[84] Lorenzo Bertini et al. "Macroscopic fluctuation theory". In: *Rev. Mod. Phys.* 87 (2 2015), pp. 593–636.

[85] M. Henkel, H. Hinrichsen, and S. Lübeck. *Non-Equilibrium Phase Transitions: Volume 1: Absorbing Phase Transitions*. Springer Netherlands, 2008.

[86] R. Livi and P. Politi. *Nonequilibrium Statistical Physics: A Modern Perspective*. Cambridge University Press, 2017.

[87] P. Gaspard. *The Statistical Mechanics of Irreversible Phenomena*. Cambridge University Press, 2022.

[88] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, 2001.

[89] L. Peliti and S. Pigolotti. *Stochastic Thermodynamics: An Introduction*. Princeton University Press, 2021.

[90] Giovanni Gallavotti. *Nonequilibrium and Irreversibility*. Springer.

[91] C. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer Series in Synergetics. Springer, 2009. ISBN: 9783540707127.

[92] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*. Vol. 1. Elsevier, 1992.

[93] Hannes Risken. *Fokker-Planck Equation, of Springer Series in Synergetics*. 1996.

[94] T Chou, K Mallick, and R K P Zia. "Non-equilibrium statistical mechanics: from a paradigmatic model to biological transport". In: *Reports on Progress in Physics* 74.11 (2011), p. 116601.

[95] Shiling Liang, Daniel Maria Busiello, and Paolo De Los Rios. "Emergent thermophoretic behavior in chemical reaction systems". In: *New Journal of Physics* (2022).

[96] Antonio Celani et al. "Anomalous Thermodynamics at the Microscale". In: *Phys. Rev. Lett.* 109 (26 2012), p. 260603.

[97] Leroy R. Taylor. "Aggregation, Variance and the Mean". In: *Nature* 189 (1961), pp. 732–735.

[98] Didier Sornette. "Multiplicative processes and power laws". In: *Phys. Rev. E* 57.4 (1998), p. 4811.

[99] Andrea Giometto et al. "Sample and population exponents of generalized Taylor's law". In: *Proceedings of the National Academy of Sciences* 112.25 (2015), pp. 7755–7760.

[100] Jacopo Grilli. "Macroecological laws describe variation and diversity in microbial communities". In: *Nature communications* 11.1 (2020), pp. 1–11.

[101] William R. Shoemaker, Kenneth J Locey, and Jay T. Lennon. "A macroecological theory of microbial biodiversity". In: *Nature Ecology &Evolution* 1 (2017).

[102] Ashish B. George and James O'Dwyer. "Universal abundance fluctuations across microbial communities, tropical forests, and urban populations". In: *bioRxiv* (2022).

[103] Luís Bettencourt and Daniel Zünd. "Demography and the emergence of universal patterns in urban systems". In: *Nature communications* 11 (Sept. 2020), p. 4584.

[104] Luis M. A. Bettencourt. "Urban growth and the emergent statistics of cities". In: *Science Advances* 6.34 (2020), eaat8812.

[105] Alberto Stefano Sassi et al. "Protein Concentration Fluctuations in the High Expression Regime: Taylor's Law and Its Mechanistic Origin". In: *Phys. Rev. X* 12 (1 2022), p. 011051.

[106] Miguel Angel Muñoz. "Nature of different types of absorbing states". In: *Physical Review E* 57 (1997), pp. 1377–1383.

[107] Geoffrey Grinstein and Miguel Angel Muñoz. "The statistical mechanics of absorbing states". In: 1997.

[108] C. Van den Broeck, J. M. R. Parrondo, and R. Toral. "Noise-Induced Nonequilibrium Phase Transition". In: *Phys. Rev. Lett.* 73 (25 1994), pp. 3395–3398.

[109] C. Van den Broeck et al. "Nonequilibrium phase transitions induced by multiplicative noise". In: *Phys. Rev. E* 55 (4 1997), pp. 4084–4094.

[110] MA Muñoz. *Multiplicative noise in non-equilibrium phase transitions: A tutorial; in Advances in Condensed Matter and Statistical Physics, Ed. E. Koroutcheva and R. Cuerno*. 2004.

[111] Walter Genovese, Miguel A. Muñoz, and J. M. Sancho. "Nonequilibrium transitions induced by multiplicative noise". In: *Phys. Rev. E* 57 (3 1998), R2495–R2498.

[112] G Grinstein, Miguel A Muñoz, and Yuhai Tu. "Phase structure of systems with multiplicative noise". In: *Physical Review Letters* 76.23 (1996), p. 4376.

[113] Francesc Sagués, José M. Sancho, and Jordi García-Ojalvo. "Spatiotemporal order out of noise". In: *Rev. Mod. Phys.* 79 (3 2007), pp. 829–882.

[114] Amos Maritan and Jayanth R. Banavar. "Chaos, noise, and synchronization". In: *Phys. Rev. Lett.* 72 (10 1994), pp. 1451–1454.

[115] Miguel A. Muñoz and Romualdo Pastor-Satorras. "Stochastic Theory of Synchronization Transitions in Extended Systems". In: *Phys. Rev. Lett.* 90 (20 2003), p. 204101.

[116] Ronald Dickman et al. "Paths to Self-Organized Criticality". In: *Brazilian Journal of Physics* 30 (Mar. 2000).

[117] M. A. Muñoz et al. "Critical Behavior of Systems with Many Absorbing States". In: *Phys. Rev. Lett.* 76 (3 1996), pp. 451–454.

[118] Alessandro Vespignani et al. "Driving, Conservation, and Absorbing States in Sandpiles". In: *Phys. Rev. Lett.* 81 (25 1998), pp. 5676–5679.

[119] Friedrich Schlögl. "Chemical reaction models for non-equilibrium phase transitions". In: *Zeitschrift für Physik* 253 (1972), pp. 147–161.

[120] J L Cardy and R L Sugar. "Directed percolation and Reggeon field theory". In: *Journal of Physics A: Mathematical and General* 13.12 (1980), p. L423.

[121] J L Cardy and P Grassberger. "Epidemic models and percolation". In: *Journal of Physics A: Mathematical and General* 18.6 (1985), p. L267.

[122] Giorgio Nicoletti et al. "The emergence of scale-free fires in Australia". In: *iScience* 26.3 (2023), p. 106181. ISSN: 2589-0042.

[123] Silvia Zaoli and Jacopo Grilli. "A macroecological description of alternative stable states reproduces intra- and inter-host variability of gut microbiome". In: *Science Advances* 7.43 (2021), eabj2882.

[124] William R. Shoemaker and Jacopo Grilli. "Macroecological patterns in coarse-grained microbial communities". In: *bioRxiv* ().

[125] Karoline Faust et al. "Signatures of ecological processes in microbial community time series". In: *Microbiome* 6.1 (2018), pp. 1–13.

[126] Richard Wolff, William Shoemaker, and Nandita Garud. "Ecological stability emerges at the level of strains in the human gut microbiome". In: *Mbio* (2023), e02502–22.

[127] Stephen Jay Gould. *The Structure of Evolutionary Theory*. Murray, Harvard University Press.

[128] Stuart A Kauffman et al. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA, 1993.

[129] Luca Peliti. "Introduction to the statistical theory of Darwinian evolution". In: *arXiv preprint cond-mat/9712027* (1997).

[130] Barbara Drossel. "Biological evolution and statistical physics". In: *Advances in physics* 50.2 (2001), pp. 209–295.

[131] Michael Lässig and Angelo Valleriani. *Biological evolution and statistical physics*. Vol. 585. Springer, 2008.

[132] Ping Ao. "Laws in Darwinian evolutionary theory". In: *Physics of life Reviews* 2.2 (2005), pp. 117–156.

[133] J Arjan GM De Visser and Joachim Krug. "Empirical fitness landscapes and the predictability of evolution". In: *Nature Reviews Genetics* 15.7 (2014), pp. 480–490.

[134] D. Lack, L.M. Ratcliffe, and P.T. Boag. *Darwin's Finches*. Cambridge University Press.

[135] Éva Kisdi and Stefan A. H. Geritz. "Adaptive dynamics: a framework to model evolution in the ecological theatre". In: *Journal of Mathematical Biology* 61.1 (2010), pp. 165–169.

[136] Stefan A. H. Geritz et al. "Dynamics of Adaptation and Evolutionary Branching". In: *Phys. Rev. Lett.* 78 (10 1997), pp. 2024–2027.

[137] Kisdi E´. Mesze´NA G. Metz J.A.J. Geritz S.A.H. "Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree". In: *Evolutionary Ecology* 12 (1998), pp. 35–57.

[138] Doebeli M. Dieckmann U. "Speciation along environmental gradients." In: *Nature* 421 (2003), pp. 259 –264.

[139] Metz J. Tautz D. (Eds.) Dieckmann U. Doebeli M. *Adaptive Speciation*. Cambridge University Press, 2004.

[140] Fabio Dercole, Fabio Della Rossa, and Pietro Landi. "The transition from evolutionary stability to branching: A catastrophic evolutionary shift". In: *Scientific reports* 6.1 (2016), pp. 1–8.

[141] Florence Debarre and Sarah P. Otto. "Evolutionary dynamics of a quantitative trait in a finite asexual population". In: *Theoretical Population Biology* 108 (2016), pp. 75–88.

[142] U. Dieckmann, B. O'Hara, and W. Weisser. "The evolutionary ecology of dispersal". In: *Trends in Ecology & Evolution* (1999).

[143] U Dieckmann et al. "Adaptive dynamics of infectious diseases". In: *Pursuit of virulence management* (2002), pp. 460–463.

[144] Rodrigo Caetano, Yaroslav Ispolatov, and Michael Doebeli. "Evolution of diversity in metabolic strategies". In: *eLife* 10 (2020).

[145] Maren L. Friesen et al. "Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in Escherichia coli". In: *Evolution* 58.2 (Feb. 2004), pp. 245–260.

[146] Adriano Bonforti and Ricard Solé. "Unicellular-multicellular evolutionary branching driven by resource limitations". In: *Journal of The Royal Society Interface* 19.191 (2022), p. 20220018.

[147] Ulf Dieckmann and Richard Law. "The dynamical theory of coevolution: a derivation from stochastic ecological processes". In: *Journal of Mathematical Biology* 34.5 (1996), pp. 579–612.

[148] Joe Yuichiro Wakano and Yoh Iwasa. "Evolutionary Branching in a Finite Population: Deterministic Branching vs. Stochastic Branching". In: *Genetics* 193.1 (2013), pp. 229–241.

[149] Benjamin Allen, Martin A. Nowak, and Ulf Dieckmann. "Adaptive Dynamics with Interaction Structure." In: *The American Naturalist* 181.6 (2013), E139–E163.

[150] Stefan A. H. Geritz, Johan A. J. Metz, and Claus Rueffler. "Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case". In: *Journal of Mathematical Biology* 72.4 (2016), pp. 1081–1099.

[151] Leimar O. "Multidimensional convergence stability". In: *Evol. Ecol. Res.* 11 (2009), pp. 191–208.

[152] Iaroslav Ispolatov, Vaibhav Madhok, and Michael Doebeli. "Individual-based models for adaptive diversification in high-dimensional phenotype spaces". In: *Journal of Theoretical Biology* 390 (2016), pp. 97 –105.

[153] Hiroshi Ito and Akira Sasaki. "Evolutionary branching under multi-dimensional evolutionary constraints". In: *Journal of Theoretical Biology* 407 (2016), pp. 409–428.

[154] Daniel S. Fisher. "Inevitability of Red Queen evolution driven by organismic complexity and simple feedback via environmental modification". In: *bioRxiv* (2021).

[155] J. Ripa and U. Dieckmann. "Mutant invasions and adaptive dynamics in variable environments". In: *Evolution* 67.5 (2013), pp. 1279–1290.

[156] Doebeli M. Dieckmann U. "On the origin of species by sympatric speciation." In: *Nature* 400 (1999), pp. 354 –357.

[157] Christine C. Spencer et al. "Adaptation increases the likelihood of diversification in an experimental bacterial lineage". In: *Proceedings of the National Academy of Sciences* 105.5 (2008), pp. 1585–1589.

[158] Michael Doebeli, Christoph Hauert, and Timothy Killingback. "The Evolutionary Origin of Cooperators and Defectors". In: *Science* 306.5697 (2004), pp. 859–862.

[159] Karen M Page and Martin A Nowak. "Unifying evolutionary dynamics". In: *Journal of theoretical biology* 219.1 (2002), pp. 93–98.

[160] Matthew J Melissa et al. "Population genetics of polymorphism and divergence in rapidly evolving populations". In: *Genetics* 221.4 (Apr. 2022).

[161] Albert F. Bennett and Richard E. Lenski. "Phenotypic and evolutionary adaptation of a model bacterial system to stressful thermal environments." In: *EXS* 83 (1997), pp. 135–54.

[162] Barry R. Bochner. "Global phenotypic characterization of bacteria". In: *Fems Microbiology Reviews* 33 (2008), pp. 191 –205.

[163] Lorenzo Fant, Iuri Macocco, and Jacopo Grilli. "Eco-evolutionary dynamics lead to functionally robust and redundant communities". In: *bioRxiv* (2021).

[164] Paula Villa Martín et al. "Eco-evolutionary Model of Rapid Phenotypic Diversification in Species-Rich Communities". In: *PLOS Computational Biology* 12.10 (Oct. 2016), pp. 1–17.

[165] Benjamin H. Good and Layton B. Rosenfeld. "Eco-evolutionary feedbacks in the human gut microbiome". In: *bioRxiv* (2022).

[166] Simon van Vliet and Michael Doebeli. "The role of multilevel selection in host microbiome evolution". In: *Proceedings of the National Academy of Sciences* 116 (2019), pp. 20591 –20597.

[167] Marta Ciechonska et al. "Emergent expression of fitness-conferring genes by phenotypic selection". In: *PNAS Nexus* 1.3 (June 2022).

[168] David Berry et al. "Tracking heavy water (D2O) incorporation for identifying and sorting active microbial cells". In: *Proceedings of the National Academy of Sciences* 112 (2014), E194 –E203.

[169] Michael M. Desai, Daniel S. Fisher, and Andrew W. Murray. "The Speed of Evolution and Maintenance of Variation in Asexual Populations". In: *Current Biology* 17.5 (2007), pp. 385–394. ISSN: 0960-9822.

[170] Christina Homberger, Lars Barquist, and Jörg Vogel. "Ushering in a new era of single-cell transcriptomics in bacteria". In: *microLife* 3 (Sept. 2022). uqac020. ISSN: 2633-6693.

[171] Stilianos Louca et al. "Function and functional redundancy in microbial systems". In: *Nature Ecology & Evolution* 2 (2018), pp. 936–943.

[172] Karna Gowda et al. "Genomic structure predicts metabolite dynamics in microbial communities". In: *Cell* 185.3 (2022), 530–546.e25. ISSN: 0092-8674.

[173] Matti Gralka, Shaul Pollak, and Otto X. Cordero. "Fundamental metabolic strategies of heterotrophic bacteria". In: *bioRxiv* (2022).

[174] S.R. De Groot and P. Mazur. *Non-Equilibrium Thermodynamics*. Dover Publications, 2013.

[175] I. Prigogine. *Introduction to Thermodynamics of Irreversible Processes*. Wiley, 1968.

[176] H. Spohn. *Large Scale Dynamics of Interacting Particles*. Springer Berlin Heidelberg, 2012.

[177] Sheldon Goldstein et al. "On the Nonequilibrium Entropy of Large and Small Systems". In: *Stochastic Dynamics Out of Equilibrium* (2017).

[178] Udo Seifert. "Stochastic thermodynamics, fluctuation theorems and molecular machines". In: *Reports on progress in physics* 75.12 (2012), p. 126001.

[179] Lars Onsager. "Reciprocal Relations in Irreversible Processes. I," in: *Phys. Rev.* 37 (4 1931), pp. 405–426.

[180] Matteo Polettini. "Macroscopic constraints for the minimum entropy production principle". In: *Phys. Rev. E* 84 (5 2011), p. 051117.

[181] Daniel M Busiello, Jorge Hidalgo, and Amos Maritan. "Entropy production for coarse-grained dynamics". In: *New Journal of Physics* 21.7 (2019), p. 073004.

[182] Christopher W. Lynn et al. "Broken detailed balance and entropy production in the human brain". In: *Proceedings of the National Academy of Sciences* 118.47 (2021), e2109889118.

[183] Christopher W. Lynn et al. "Decomposing the Local Arrow of Time in Interacting Systems". In: *Phys. Rev. Lett.* 129 (11 2022), p. 118101.

[184] Ville Mustonen and Michael Lässig. "Fitness flux and ubiquity of adaptive evolution". In: *Proceedings of the National Academy of Sciences* 107.9 (2010), pp. 4248–4253. ISSN: 0027-8424.

[185] Benjamin Andrae et al. "Entropy Production of Cyclic Population Dynamics". In: *Phys. Rev. Lett.* 104 (21 2010), p. 218102.

[186] Christopher Battle et al. "Broken detailed balance at mesoscopic scales in active biological systems". In: *Science* 352.6285 (2016), pp. 604–607.

[187] Cesare Nardini et al. "Entropy Production in Field Theories without Time-Reversal Symmetry: Quantifying the Non-Equilibrium Character of Active Matter". In: *Phys. Rev. X* 7 (2 2017), p. 021007.

[188] Qiwei Yu and Yuhai Tu. "Energy Cost for Flocking of Active Spins: The Cusped Dissipation Maximum at the Flocking Transition". In: *Phys. Rev. Lett.* 129 (27 2022), p. 278001.

[189] Federica Ferretti et al. "Signatures of irreversibility in microscopic models of flocking". In: *Phys. Rev. E* 106 (3 2022), p. 034608.

[190] Zak Frentz, Seppe Kuehn, and Stanislas Leibler. "Strongly Deterministic Population Dynamics in Closed Microbial Communities". In: *Phys. Rev. X* 5 (4 2015), p. 041014.

[191] Barrere Julien Ratzke Christoph and Jeff Gore. "Strength of species interactions determines biodiversity and stability in microbial communities". In: *Nature Ecology Evolution* 4 (2020), pp. 376–383.

[192] Matti Gralka et al. "Trophic Interactions and the Drivers of Microbial Community Assembly". In: *Current Biology* 30.19 (2020), R1176–R1188.

[193] Jonathan Friedman, Logan M. Higgins, and Jeff Gore. "Community structure follows simple assembly rules in microbial microcosms". In: *Nature Ecology &Evolution* 1 (2017).

[194] Rachel E. Szabo et al. "Ecological stochasticity and phage induction diversify bacterioplankton communities at the microscale". In: *bioRxiv* (2021).

[195] Jiliang Hu et al. "Emergent phases of ecological diversity and dynamics mapped in microcosms". In: *Science* 378.6615 (2022), pp. 85–89.

[196] Ruth E. Ley et al. "Worlds within worlds: evolution of the vertebrate gut microbiota". In: *Nature Reviews Microbiology* 6 (2008), pp. 776–788.

[197] Catherine A. Lozupone and Rob Knight. "Global patterns in bacterial diversity". In: *Proceedings of the National Academy of Sciences* 104.27 (2007), pp. 11436–11440. ISSN: 0027-8424.

[198] Manimozhiyan Arumugam et al. "Enterotypes of the human gut microbiome: [Plus]Corrigendum [Plus]Addendum". English. In: *Nature* 473.7346 (2011), pp. 174–180. ISSN: 0028-0836.

[199] Laura Grieneisen et al. "Gut microbiome heritability is nearly universal but environmentally contingent". In: *Science* 373.6551 (2021), pp. 181–186.

[200] Jennifer B. H. Martiny et al. "Microbiomes in light of traits: A phylogenetic perspective". In: *Science* 350.6261 (2015), aac9323.

[201] James I. Prosser et al. "The role of ecological theory in microbial ecology". In: *Nature Reviews Microbiology* 5 (2007), pp. 384–392.

[202] Pablo A. Marquet et al. "On Theory in Ecology". In: *BioScience* 64.8 (July 2014), pp. 701–710.

[203] Jack A. Gilbert and Christopher L. Dupont. "Microbial Metagenomics: Beyond the Genome". In: *Annual Review of Marine Science* 3.1 (2011), pp. 347–371.

[204] Jethro Johnson et al. "Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis". In: *Nature Communications* 10 (Nov. 2019).

[205] Ashley Shade et al. "Macroecology to Unite All Life, Large and Small". In: *Trends in Ecology Evolution* 33.10 (2018), pp. 731–744.

[206] Brian W Ji et al. "Macroecological dynamics of gut microbiota". In: *Nature microbiology* 5.5 (2020), pp. 768–775.

[207] Jacopo Grilli. "Macroecological laws describe variation and diversity in microbial communities". In: *Nature communications* 11.1 (2020), pp. 1–11.

[208] Lana Descheemaeker and Sophie de Buyl. "Stochastic logistic models reproduce experimental time series of microbial communities". In: *eLife* 9 (2020). Ed. by Sandeep Krishna and Aleksandra M Walczak, e55650. ISSN: 2050-084X.

[209] Po-Yi Ho, Benjamin H Good, and Kerwyn Casey Huang. "Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas". In: *Elife* 11 (2022), e75168.

[210] James P O'Dwyer and Ryan Chisholm. "A mean field model for competition: from neutral ecology to the Red Queen". In: *Ecology letters* 17.8 (2014), pp. 961–969.

[211] Jacopo Grilli et al. "Higher-order interactions stabilize dynamics in competitive network models". In: *Nature* 548.7666 (2017), pp. 210–213.

[212] J Timothy Wootton. "Indirect effects in complex ecosystems: recent progress and future challenges". In: *Journal of Sea Research* 48.2 (2002), pp. 157–172.

[213] Campbell O Webb et al. "Phylogenies and community ecology". In: *Annual review of ecology and systematics* 33.1 (2002), pp. 475–505.

[214] Janneke HilleRisLambers et al. "Rethinking community assembly through the lens of coexistence theory". In: *Annual review of ecology, evolution, and systematics* 43 (2012), pp. 227–248.

[215] Marc W Cadotte and Caroline M Tucker. "Should environmental filtering be abandoned?" In: *Trends in ecology & evolution* 32.6 (2017), pp. 429–437.

[216] Brent C Emerson and Rosemary G Gillespie. "Phylogenetic analysis of community assembly and structure over space and time". In: *Trends in ecology & evolution* 23.11 (2008), pp. 619–630.

[217] Robert Poulin et al. "Phylogeny determines the role of helminth parasites in intertidal food webs". In: *Journal of Animal Ecology* (2013), pp. 1265–1275.

[218] Boris R Krasnov et al. "Co-occurrence and phylogenetic distance in communities of mammalian ectoparasites: limiting similarity versus environmental filtering". In: *Oikos* 123.1 (2014), pp. 63–70.

[219] Eduardo Pérez-Valera et al. "Fire modifies the phylogenetic structure of soil bacterial co-occurrence networks". In: *Environmental Microbiology* 19.1 (2017), pp. 317–327.

[220] Jeannine Cavender-Bares et al. "The merging of community ecology and phylogenetic biology". In: *Ecology letters* 12.7 (2009), pp. 693–715.

[221] Christopher A Gaulke et al. "Ecophylogenetics clarifies the evolutionary association between mammals and their gut microbiota". In: *MBio* 9.5 (2018), e01348–18.

[222] Patricio Jeraldo et al. "Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes". In: *Proceedings of the National Academy of Sciences* 109.25 (2012), pp. 9692–9698.

[223] James P O'Dwyer, Steven W Kembel, and Jessica L Green. "Phylogenetic diversity theory sheds light on the structure of microbial communities". In: *PLoS computational biology* 8.12 (2012), e1002832.

[224] Alex Mitchell et al. "EBI Metagenomics in 2017: Enriching the analysis of microbial communities, from sequence reads to assemblies". In: *Nucleic acids research* 46 (Oct. 2017).

[225] J. Laherrére and D. Sornette. "Theoretical microbial ecology without species". In: *The European Physical Journal B* 2 (4 1998), 525–539.

[226] Robert MacArthur. "Species packing and competitive equilibrium for many species". In: *Theoretical Population Biology* 1.1 (1970), pp. 1–11.

[227] Anna Posfai, Thibaud Taillefumier, and Ned S. Wingreen. "Metabolic Trade-Offs Promote Diversity in a Model Ecosystem". In: *Phys. Rev. Lett.* 118 (2 2017), p. 028103.

[228] Peter Jung and Peter Hänggi. "Dynamical systems: A unified colored-noise approximation". In: *Phys. Rev. A* 35 (10 1987), pp. 4464–4466.

[229] Paul H Harvey, Mark D Pagel, et al. *The comparative method in evolutionary biology*. Vol. 239. Oxford university press Oxford, 1991.

[230] John J Wiens and Catherine H Graham. "Niche conservatism: integrating evolution, ecology, and conservation biology". In: *Annu. Rev. Ecol. Evol. Syst.* 36 (2005), pp. 519–539.

[231] Leonardo Pacciani-Mori et al. "Dynamic metabolic adaptation can promote species coexistence in competitive microbial communities". In: *PLoS computational biology* 16.5 (2020), e1007896.

[232] Bruno Burlando. "The fractal dimension of taxonomic systems". In: *Journal of Theoretical Biology* 146.1 (1990), pp. 99–114. ISSN: 0022-5193.

[233] Emilio Hernandez-Garcia et al. "Simple models for scaling in phylogenetic trees". In: *International Journal of Bifurcation and Chaos* 10 (Mar. 2010), pp. 805–811.

[234] Chi Xue, Zhiru Liu, and Nigel Goldenfeld. "Scale-invariant topology and bursty branching of evolutionary trees emerge from niche construction". In: *Proceedings of the National Academy of Sciences* 117.14 (2020), pp. 7879–7887. ISSN: 0027-8424.

[235] James P O'Dwyer, Steven W Kembel, and Thomas J Sharpton. "Backbones of evolutionary history test biodiversity theory for microbes". In: *Proceedings of the National Academy of Sciences* 112.27 (2015), pp. 8356–8361.

[236] Marten Scheffer and Egbert H Van Nes. "Self-organized similarity, the evolutionary emergence of groups of similar species". In: *Proceedings of the National Academy of Sciences* 103.16 (2006), pp. 6230–6235.

[237] Francisco Ramos et al. "Crystallization and melting of bacteria colonies and Brownian bugs". In: *Physical Review E* 77.2 (2008), p. 021102.

[238] Akshit Goyal et al. "Interactions between strains govern the eco-evolutionary dynamics of microbial communities". In: *bioRxiv* (2021).

[239] K. G. Wilson. "Problems in Physics with many Scales of Length". In: *Scientific American* 241 (1979), pp. 158–179.

[240] Efi Efrati et al. "Real-space renormalization in statistical mechanics". In: *Reviews of Modern Physics* 86.2 (2014), p. 647.

[241] Mikhail Tikhonov. "Theoretical microbial ecology without species". In: *Phys. Rev. E* 96 (3 2017), p. 032410.

[242] H. Risken and H. Haken. *The Fokker-Planck Equation: Methods of Solution and Applications Second Edition*. Springer, 1989.

[243] Motoo Kimura. "Diffusion models in population genetics". In: *Journal of Applied Probability* 1.2 (1964), pp. 177–232.

[244] J. F. Crow and M. Kimura. *An Introduction to Population Genetics Theory*. Blackburn Press, 2009.

[245] S.H. Rice. *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer, 2004.

[246] Martin A Nowak. *Evolutionary dynamics*. Harvard University Press, 2006.

[247] E Kussell and M Vucelja. "Non-equilibrium physics and evolution—adaptation, extinction, and ecology: a Key Issues review". In: *Reports on Progress in Physics* 77.10 (2014), p. 102602.

[248] George R. Price. "The nature of selection". In: *Journal of Theoretical Biology* 175.3 (1995), pp. 389 –396. ISSN: 0022-5193.

[249] S. A. Frank. "Natural selection. IV. The Price equation". In: *Journal of Evolutionary Biology* 25.6 (2012), pp. 1002–1019.

[250] Steven A. Frank. "The Price Equation Program: Simple Invariances Unify Population Dynamics, Thermodynamics, Probability, Information and Inference". In: *Entropy* 20.12 (2018).

[251] G. R. Price. "Selection and Covariance". In: 227 (Aug. 1970), pp. 520–521.

[252] S. A. Frank. "Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory". In: *Journal of Evolutionary Biology* 25.12 (2012), pp. 2377–2396.

[253] Samir Okasha. *Evolution and the Levels of Selection*. Oxford University Press, 2006.

[254] Jussi Lehtonen, Samir Okasha, and Heikki Helanterae. "Fifty years of the Price equation". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797 (2020), p. 20190350.

[255] Sean H Rice. "A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution". In: *BMC Evolutionary Biology* 8.262 (2008).

[256] Ulf Dieckmann and Johan A.J. Metz. "Surprising evolutionary predictions from enhanced ecological realism". In: *Theoretical Population Biology* 69.3 (2006). ESS Theory Now, pp. 263 –281.

[257] Eva Kisdi et al. "Evolutionary Optimisation Models and Matrix Games in the Unified Perspective of Adaptive Dynamics". In: *Selection* 2 (Aug. 2001).

[258] R.A. Fisher. *The Genetical Theory of Natural Selection — A Complete Variorum Edition*. Oxford University Press, 1990.

[259] S. Wright. "The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution". In: *Proceedings of the XI International Congress of Genetics* 8 (1932), pp. 209–222.

[260] Daniel S Fisher. "Asexual evolution waves: fluctuations and universality". In: *Journal of Statistical Mechanics: Theory and Experiment* 2013.01 (2013), P01011.

[261] Oskar Hallatschek. "The noisy edge of traveling waves". In: *Proceedings of the National Academy of Sciences* 108.5 (2011), pp. 1783–1787. ISSN: 0027-8424. DOI: 10.1073/pnas.1013529108. URL: https://www.pnas.org/content/108/5/1783.

[262] Oskar Hallatschek et al. "Genetic drift at expanding frontiers promotes gene segregation". In: *Proceedings of the National Academy of Sciences* 104.50 (2007), pp. 19926–19930.

[263] John Maynard Smith. "Time in the Evolutionary Process". In: *The Study of Time: Proceedings of the First Conference of the International Society for the Study of Time Oberwolfach (Black Forest) — West Germany*. Ed. by J. T. Fraser, F. C. Haber, and G. H. Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, 1972, pp. 200–205.

[264] Karl Sigmund and Martin A Nowak. "Evolutionary game theory". In: *Current Biology* 9.14 (1999), R503–R505.

[265] Josef Hofbauer and Karl Sigmund. "Evolutionary game dynamics". In: *Bulletin of the American Mathematical Society* 40.4 (2003), pp. 479–519.

[266] Josef Hofbauer, Karl Sigmund, et al. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[267] Erwin Frey. "Evolutionary game theory: Theoretical concepts and applications to microbial communities". In: *Physica A: Statistical Mechanics and its Applications* 389.20 (2010), pp. 4265–4298.

[268] John M McNamara. "Towards a richer evolutionary game theory". In: *Journal of the Royal Society Interface* 10.88 (2013), p. 20130544.

[269] Jörgen W Weibull. *Evolutionary game theory*. MIT press, 1997.

[270] J.A.J. Metz et al. *Adaptive Dynamics: A Geometrical Study of the Consequences of Nearly Faithful Reproduction*. IIASA Working Paper. IIASA, Laxenburg, Austria, 1995.

[271] Odo Diekmann. "Beginners guide to adaptive dynamics". In: *Banach Center Publications* (2004), pp. 47–86.

[272] Fabio Dercole and Sergio Rinaldi. "Analysis of evolutionary processes". In: *Analysis of Evolutionary Processes*. Princeton University Press, 2008.

[273] Åke Brännström, Jacob Johansson, and Niels Von Festenberg. "The hitchhiker's guide to adaptive dynamics". In: *Games* 4.3 (2013), pp. 304–328.

[274] Santiago F Elena and Richard E Lenski. "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation". In: *Nature Reviews Genetics* 4.6 (2003), pp. 457–469.

[275] Benjamin J Callahan, Tadashi Fukami, and Daniel S Fisher. "Rapid evolution of adaptive niche construction in experimental microbial populations". In: *Evolution* 68.11 (2014), pp. 3307–3316.

[276] Rubing Chen and Edward C Holmes. "Avian influenza virus exhibits rapid evolutionary dynamics". In: *Molecular biology and evolution* 23.12 (2006), pp. 2336–2341.

[277] Akshit Goyal et al. “Interactions between strains govern the eco-evolutionary dynamics of microbial communities”. In: *Elife* 11 (2022), e74987.

[278] Antoine Giraud et al. “Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut”. In: *science* 291.5513 (2001), pp. 2606–2608.

[279] John W. Drake and John J. Holland. “Mutation rates among RNA viruses”. In: *Proceedings of the National Academy of Sciences* 96.24 (1999), pp. 13910–13913.

[280] M.Lynch. “Evolution of the mutation rate”. In: *Trends in genetics* 26 (2010).

[281] Ariadna Villanueva et al. “Genotype-to-Protein Map and Collective Adaptation in a Viral Population”. In: *Biophysica* 2.4 (2022), pp. 381–399. ISSN: 2673-4125.

[282] Thomas W. Schoener. “The Newest Synthesis: Understanding the Interplay of Evolutionary and Ecological Dynamics”. In: *Science* 331.6016 (2011), pp. 426–429.

[283] F. Pelletier, D. Garant, and A.P. Hendry. “Eco-evolutionary dynamics”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1523 (2009), pp. 1483–1489.

[284] C. K. Biebricher and M. Eigen. “What Is a Quasispecies?” In: *Quasispecies: Concept and Implications for Virology*. Ed. by Esteban Domingo. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–31.

[285] Matti Gralka, Shaul Pollak, and Otto X. Cordero. “Fundamental metabolic strategies of heterotrophic bacteria”. In: *bioRxiv* (2022).

[286] Michael Doebeli. *Adaptive Diversification*. Princeton University Press, 2011.

[287] Ulf Dieckmann. “Can adaptive dynamics invade?” In: *Trends in Ecology Evolution* 12.4 (1997), pp. 128 –131.

[288] Nicolas Champagnat, Regis Ferriere, and Sylvie Meleard. “Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models”. In: *Theoretical Population Biology* 69.3 (2006), pp. 297 –321.

[289] Gerard Ben Arous, N Champagnat, and D Ferriere. “The canonical equation of adaptive dynamics: a mathematical view”. English (US). In: *Selection* 2 (2001), pp. 71–81.

[290] Nicolas Champagnat, Regis Ferriere, and Sylvie Meleard. “From Individual Stochastic Processes to Macroscopic Models in Adaptive Evolution”. In: *Stochastic Models* 24.sup1 (2008), pp. 2–44.

[291] Nicolas Champagnat and Amaury Lambert. “Evolution of discrete populations and the canonical diffusion of adaptive dynamics”. In: *Annals of Applied Probability* 17.1 (2007), 102 – 155.

[292] M.Bauer et al. “Ecological feedback in quorum-sensing microbial populations can induce heterogeneous production of autoinducers”. In: *elife* 6 (2017).

[293] Erwin Frey, Johannes Knebel, and Peter Pickl. “Mean-field equation for a stochastic many-particle model of quorum-sensing microbial populations”. In: (2018).

[294] Gil Jorge Barros Henriques et al. “On the importance of evolving phenotype distributions on evolutionary diversification”. In: *PLOS Computational Biology* 17.2 (Feb. 2021), pp. 1–21.

[295] P. A. P. Moran. “Random processes in genetics”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 54.1 (1958), 60–71.

[296] "N.G. Van Kampen". *Stochastic processes in physics and chemistry*. Third Edition. Elsevier, 2007.

[297] Jacopo Grilli et al. "Higher-order interactions stabilize dynamics in competitive network models". In: *Nature* 548.7666 (2017), pp. 210–213.

[298] Ross Cressman and Yi Tao. "The replicator equation and other game dynamics". In: *Proceedings of the National Academy of Sciences* 111.supplement_3 (2014), pp. 10810–10817.

[299] Henry P. McKean. "A CLASS OF MARKOV PROCESSES ASSOCIATED WITH NONLINEAR PARABOLIC EQUATIONS". In: *Proceedings of the National Academy of Sciences of the United States of America* 56 (1966), pp. 1907 –1911.

[300] René Carmona, François Delarue, and Aimé Lachapelle. "Control of McKean–Vlasov dynamics versus mean field games". In: *Mathematics and Financial Economics* 7 (2013), pp. 131–166.

[301] Christian Kuehn. "Moment Closure—A Brief Review". In: *Control of Self-Organizing Nonlinear Systems*. Ed. by Eckehard Schöll, Sabine H. L. Klapp, and Philipp Hövel. Cham: Springer International Publishing, 2016, pp. 253–271.

[302] Michael Doebeli and Iaroslav Ispolatov. "Continuously stable strategies as evolutionary branching points". In: *Journal of Theoretical Biology* 266.4 (2010), pp. 529–535.

[303] Simone Pigolotti, Cristóbal López, and Emilio Hernández-García. "Species Clustering in Competitive Lotka-Volterra Models". In: *Phys. Rev. Lett.* 98 (25 2007), p. 258101.

[304] Mehran Kardar. *Statistical Physics of Fields*. Cambridge University Press, 2007.

[305] Avadh B Saxena, Ivan C. Christov, and Avinash Khare. "Higher-Order Field Theories: ϕ^6 , ϕ^8 and Beyond". In: *Nonlinear Systems and Complexity* (2019).

[306] Jonas Cremer, Anna Melbinger, and Erwin Frey. "Evolutionary and population dynamics: A coupled approach". In: *Phys. Rev. E* 84 (5 2011), p. 051921.

[307] Anna Melbinger, Jonas Cremer, and Erwin Frey. "Evolutionary game theory in growing populations". In: *Physical Review Letters* 105.17 (2010), p. 178101.

[308] George W. A. Constable et al. "Demographic noise can reverse the direction of deterministic selection". In: *Proceedings of the National Academy of Sciences* 113.32 (2016), E4745–E4754.

[309] Arne Traulsen, Jens Christian Claussen, and Christoph Hauert. "Stochastic differential equations for evolutionary dynamics with demographic noise and mutations". In: *Phys. Rev. E* 85 (4 2012), p. 041901.

[310] David Claessen et al. "Delayed evolutionary branching in small populations". In: *Evolutionary Ecology Research* 9.1 (2007), 51 – 69.

[311] Kotil S.E. and Vetsigian K. "Emergence of evolutionarily stable communities through eco-evolutionary tunnelling". In: *Nature Ecology and Evolution* (2018).

[312] Emanuele Crosato et al. "Dynamical demographic phases explain how population growth and mutation control the evolutionary impact of bottlenecks". In: *Physical Review Research* 5.1 (2023), p. 013093.

[313] Sidney Redner. "Random multiplicative processes: An elementary tutorial". In: *American Journal of Physics* 58.3 (1990), pp. 267–273.

[314] Jiliang Hu et al. "Emergent phases of ecological diversity and dynamics mapped in microcosms". In: *Science* 378.6615 (2022), pp. 85–89.

[315] Lorenzo Fant et al. *Stable cooperation emerges in stochastic multiplicative growth*. 2022.

[316] Alma Dal Co et al. "Short-range interactions govern the dynamics and functions of microbial communities". In: *Nature Ecology & Evolution* 4 (2020), pp. 366 –375.

[317] Olivia M Ghosh and Benjamin H. Good. "Emergent evolutionary forces in spatial models of luminal growth and their application to the human gut microbiota". In: *Proceedings of the National Academy of Sciences of the United States of America* 119 (2022).

[318] Ilan N. Rubin, Yaroslav Ispolatov, and Michael Doebeli. "Adaptive diversification and niche packing on rugged fitness landscapes". In: *Journal of Theoretical Biology* 562 (2023), p. 111421.

[319] Riccardo Rao and Stanislas Leibler. "Evolutionary dynamics, evolutionary forces, and robustness: A nonequilibrium statistical mechanics perspective". In: *Proceedings of the National Academy of Sciences* 119.13 (2022), e2112083119.

[320] Charles D Kocher and Ken A Dill. "Darwinian evolution as a dynamical principle". In: *Proceedings of the National Academy of Sciences* 120.11 (2023), e2218390120.

[321] Andrew P Hendry. *Eco-evolutionary dynamics*. Princeton university press, 2016.

[322] Thomas W. Schoener. "The Newest Synthesis: Understanding the Interplay of Evolutionary and Ecological Dynamics". In: *Science* 331.6016 (2011), pp. 426–429.

[323] F. Pelletier, D. Garant, and A.P. Hendry. "Eco-evolutionary dynamics". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1523 (2009), pp. 1483–1489.

[324] Jay T Lennon and Stuart E Jones. "Microbial seed banks: the ecological and evolutionary implications of dormancy". In: *Nature reviews microbiology* 9.2 (2011), pp. 119–130.

[325] William R Shoemaker and Jay T Lennon. "Evolution with a seed bank: the population genetic consequences of microbial dormancy". In: *Evolutionary applications* 11.1 (2018), pp. 60–75.

[326] Dawn Fraser and Mads Kaern. "A chance at survival: gene expression noise and phenotypic diversification strategies". In: *Molecular microbiology* 71.6 (2009), pp. 1333–1340.

[327] Michael Guppy and Philip Withers. "Metabolic depression in animals: physiological perspectives and biochemical generalizations". In: *Biological Reviews* 74.1 (1999), pp. 1–40.

[328] Alexander Sturm and Jonathan Dworkin. "Phenotypic diversity as a mechanism to exit cellular dormancy". In: *Current Biology* 25.17 (2015), pp. 2272–2277.

[329] Erik S Barton et al. "Herpesvirus latency confers symbiotic protection from bacterial infection". In: *Nature* 447.7142 (2007), pp. 326–329.

[330] Robert L Bertrand. "Lag phase is a dynamic, organized, adaptive, and evolvable period that prepares bacteria for cell division". In: *Journal of bacteriology* 201.7 (2019), e00697–18.

[331] Simon van Vliet. "Bacterial dormancy: how to decide when to wake up". In: *Current Biology* 25.17 (2015), R753–R755.

[332] Emily SC Rittershaus, Seung-Hun Baek, and Christopher M Sassetti. "The normalcy of dormancy: common themes in microbial quiescence". In: *Cell host & microbe* 13.6 (2013), pp. 643–651.

[333] Dylan Z Childs, CJE Metcalf, and Mark Rees. "Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants". In: *Proceedings of the Royal Society B: Biological Sciences* 277.1697 (2010), pp. 3055–3064.

[334] Jennifer R Gremer and D Lawrence Venable. "Bet hedging in desert winter annual plants: optimal germination strategies in a variable environment". In: *Ecology Letters* 17.3 (2014), pp. 380–387.

[335] K Björn Rechinger et al. "Early protein synthesis of Lactobacillus delbrueckii ssp. bulgaricus in milk revealed by methionine labeling and two-dimensional gel electrophoresis". In: *Electrophoresis: An International Journal* 21.13 (2000), pp. 2660–2669.

[336] Nadja Larsen et al. "Differential expression of proteins and genes in the lag phase of Lactococcus lactis subsp. lactis grown in synthetic medium and reconstituted skim milk". In: *Appl. Environ. Microbiol.* 72.2 (2006), pp. 1173–1179.

[337] Peter Van Bodegom. "Microbial maintenance: a critical review on its quantification". In: *Microbial ecology* 53.4 (2007), pp. 513–523.

[338] Daniel Paredes-Sabja, Peter Setlow, and Mahfuzur R Sarker. "Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved". In: *Trends in microbiology* 19.2 (2011), pp. 85–94.

[339] Erik S. Wright and Kalin H. Vetsigian. "Stochastic exits from dormancy give rise to heavy-tailed distributions of descendants in bacterial populations". In: *Molecular Ecology* 28.17 (2019), pp. 3915–3928.

[340] Slava S Epstein. "Microbial awakenings". In: *Nature* 457.7233 (2009), pp. 1083–1083.

[341] S Buerger et al. "Microbial scout hypothesis, stochastic exit from dormancy, and the nature of slow growers". In: *Appl. Environ. Microbiol.* 78.9 (2012), pp. 3221–3228.

[342] Edo Kussell and Stanislas Leibler. "Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments". In: *Science* 309.5743 (2005), pp. 2075–2078.

[343] Goldberg A. Ronin I. et al. Fridman O. "Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations". In: *Nature* 513 (2014), 418–421.

[344] Irit Levin-Reisman et al. "Antibiotic tolerance facilitates the evolution of resistance". In: *Science* 6327 (), 826–830.

[345] Tom Philippi and Jon Seger. "Hedging one's evolutionary bets, revisited". In: *Trends in Ecology and Evolution* 4.2 (1989), pp. 41 –44.

[346] Jan-Willem Veening, Wiep Klaas Smits, and Oscar P. Kuipers. "Bistability, Epigenetics, and Bet-Hedging in Bacteria". In: *Annual Review of Microbiology* 62.1 (2008), pp. 193–210.

[347] Paula Villa Martín, Miguel A Muñoz, and Simone Pigolotti. "Bet-hedging strategies in expanding populations". In: *PLoS computational biology* 15.4 (2019), e1006529.

[348] Ye Xu and Kalin Vetsigian. "Phenotypic variability and community interactions of germinating Streptomyces spores". In: *Scientific reports* 7.1 (2017), pp. 1–13.

[349] Edo Kussell. "Evolution in Microbes". In: *Annual Review of Biophysics* 42.1 (2013), pp. 493–514.

[350] Benjamin H. Good et al. "The dynamics of molecular evolution over 60,000 generations". English. In: *Nature* 551.7678 (Nov. 2017), pp. 45–50. ISSN: 0028-0836.

[351] Santiago F Elena and Richard E Lenski. "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation". In: *Nature Reviews Genetics* 4.6 (2003), pp. 457–469.

[352] Max Müller. "Ueber den Einfluss von Fiebertemperaturen auf die Wachsthumsgeschwindigkeit und die Virulenz des Typhus". In: *Zeitschrift für Hygiene und Infektionskrankheiten, medizinische Mikrobiologie, Immunologie und Virologie* 20 (1895), p. 245.

[353] Asher Brauner et al. "Distinguishing between resistance, tolerance and persistence to antibiotic treatment". In: *Nature Reviews Microbiology* 14.5 (2016), p. 320.

[354] Michael A Kohanski, Daniel J Dwyer, and James J Collins. "How antibiotics kill bacteria: from targets to networks". In: *Nature Reviews Microbiology* 8.6 (2010), pp. 423–435.

[355] Bing Li et al. "The importance of lag time extension in determining bacterial resistance to antibiotics". In: *Analyst* 141.10 (2016), pp. 3059–3067.

[356] Asher Brauner et al. "An experimental framework for quantifying bacterial tolerance". In: *Biophysical journal* 112.12 (2017), pp. 2664–2671.

[357] Stefan A. H. Geritz et al. "Dynamics of Adaptation and Evolutionary Branching". In: *Phys. Rev. Lett.* 78 (10 1997), pp. 2024–2027.

[358] Stefan A. H. Geritz et al. "Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree". In: *Evolutionary Ecology* 12 (1998), pp. 35–37.

[359] Diekmann O. *Diekmann O. 2004. A beginner's guide to adaptive dynamics school*. Banach Center Publications, 2004.

[360] Ulf Dieckmann and Johan AJ Metz. "Surprising evolutionary predictions from enhanced ecological realism". In: *Theoretical Population Biology* 69.3 (2006), pp. 263–281.

[361] Josef Hofbauer and Karl Sigmund. "Evolutionary game dynamics". In: *Bulletin of the American Mathematical Society* 40.4 (2003), pp. 479–519.

[362] Ulf Dieckmann and Richard Law. "The dynamical theory of coevolution: a derivation from stochastic ecological processes". In: *Journal of mathematical biology* 34.5-6 (1996), pp. 579–612.

[363] Michael Doebeli and Ulf Dieckmann. "Evolutionary branching and sympatric speciation caused by different types of ecological interactions". In: *The american naturalist* 156.S4 (2000), S77–S101.

[364] Doebeli Michael Dieckmann Ulf. "On the origin of species by sympatric speciation". In: *Nature* 400 (1999), 354–357.

[365] Michael Doebeli and Ulf Dieckmann. "Speciation along environmental gradients". In: *Nature* 421.6920 (2003), pp. 259–264.

[366] Ulf Dieckmann et al. *Adaptive Speciation*. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, 2004.

[367] Jeffrey A Fletcher and Michael Doebeli. "A simple and general explanation for the evolution of altruism". In: *Proceedings of the Royal Society B: Biological Sciences* 276.1654 (2009), pp. 13–19.

[368] Michael Doebeli, Christoph Hauert, and Timothy Killingback. "The Evolutionary Origin of Cooperators and Defectors". In: *Science* 306.5697 (2004), pp. 859–862.

[369] Michael Doebeli and Graeme D. Ruxton. "Evolution of disperal rates in metapopulations models: branching and cyclic dynamics in phenotype space". In: *Evolution* 51.6 (1997), pp. 1730–1741.

[370] Gil Jorge Barros Henriques et al. "On the importance of evolving phenotype distributions on evolutionary diversification". In: *PLOS Computational Biology* 17.2 (Feb. 2021), pp. 1–21.

[371] Ilan N. Rubin and Michael Doebeli. "Rethinking the evolution of specialization: A model for the evolution of phenotypic heterogeneity". In: *Journal of Theoretical Biology* 435 (2017), pp. 248–264.

[372] Michael Doebeli and Yaroslav. "Towards a mechanistic foundation of evolutionary theory." In: *Elife* 6.:e23804 (2017).

[373] Kalin Kotil Seyfullah Enes.Vetsigian. "Emergence of evolutionarily stable communities through eco-evolutionary tunnelling". In: *Nature Ecology and Evolution* (2018).

[374] Yoh Iwasa, Franziska Michor, and Martin A. Nowak. "Stochastic Tunnels in Evolutionary Dynamics". In: *Genetics* 166.3 (2004), pp. 1571–1579.

[375] Teppo Hiltunen, Marko Virta, and Anna-Liisa Laine. "Antibiotic resistance in the wild: an eco-evolutionary perspective". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1712 (2017), p. 20160039.

[376] Otto X Cordero and Martin F Polz. "Explaining microbial genomic diversity in light of evolutionary ecology". In: *Nature reviews. Microbiology* 12.4 (2014), 263—273.

[377] Alvaro Sanchez and Jeff Gore. "Feedback between Population and Evolutionary Dynamics Determines the Fate of Social Microbial Populations". In: *PLOS Biology* 11.4 (Apr. 2013), pp. 1–9.

[378] Benjamin J Callahan, Tadashi Fukami, and Daniel S Fisher. "Rapid evolution of adaptive niche construction in experimental microbial populations". In: *Evolution* 68.11 (2014), pp. 3307–3316.

[379] Eitan Rotem et al. "Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence". In: 107.28 (2010), pp. 12541–12546.

[380] Philipp W Messer, Stephen P Ellner, and Nelson G Hairston Jr. "Can population genetics adapt to rapid evolution?" In: *Trends in Genetics* 32.7 (2016), pp. 408–418.

[381] Juan A Bonachela, Meike T Wortel, and Nils Chr Stenseth. "Eco-evolutionary Red Queen dynamics regulate biodiversity in a metabolite-driven microbial system". In: *Scientific reports* 7.1 (2017), pp. 1–9.

[382] Sergio M Vallina et al. "Models in microbial ecology". In: *Encyclopedia of Microbiology*. Elsevier, 2019, pp. 211–246.

[383] Meike T Wortel et al. "Continual evolution through coupled fast and slow feedbacks". In: *Proceedings National Academy of Sciences* 117.8 (2020), pp. 4234–4242.

[384] Irit Levin-Reisman et al. "Automated imaging with ScanLag reveals previously undetectable bacterial growth phenotypes". In: *Nature methods* 7.9 (2010), 737—739. ISSN: 1548-7091.

[385] József Baranyi. "Stochastic modelling of bacterial lag phase". In: *International journal of food microbiology* 73.2-3 (2002), pp. 203–206.

[386] A Métris et al. "Modelling the variability of lag times and the first generation times of single cells of E. coli". In: *International journal of food microbiology* 100.1-3 (2005), pp. 13–19.

[387] Stefany Moreno-Gámez et al. "Wide lag time distributions break a trade-off between reproduction and survival in bacteria". In: *Proceedings of the National Academy of Sciences* 117.31 (2020), pp. 18729–18736.

[388] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*. Vol. 1. Elsevier, 1992.

[389] Thomas M Norman et al. "Stochastic switching of cell fate in microbes". In: *Annual review of microbiology* 69 (2015), pp. 381–403.

[390] Yusuke Himeoka and Namiko Mitarai. "When to wake up? The optimal waking-up strategies for starvation-induced persistence". In: *PLoS computational biology* 17.2 (2021), e1008655.

[391] A.-L. Barabási and R. Albert. "Emergence of scaling in random networks". In: *Science* 286.5439 (1999), pp. 509–512.

[392] M Mitzenmacher. "A Brief History of Generative Models for Power Law and Lognormal Distributions". In: *Internet Mathematics* 1 (2002), pp. 226–251.

[393] M. E. J. Newman. "Power laws, Pareto distributions and Zipf's law". In: *Contemp Phys* 46.5 (May 2005), pp. 323–351. ISSN: 0010-7514. eprint: cond-mat/0412004.

[394] Susanna C Manrubia and Damián H Zanette. "Stochastic multiplicative processes with reset events". In: *Physical Review E* 59.5 (1999), p. 4945.

[395] Yuhai Tu, G Grinstein, and M.A. Muñoz. "Systems with multiplicative noise: critical behavior from KPZ equation and numerics". In: *Physical Review Letters* 78.2 (1997), p. 274.

[396] Miguel A Muñoz, Francesca Colaiori, and Claudio Castellano. "Mean-field limit of systems with multiplicative noise". In: *Physical Review E* 72.5 (2005), p. 056102.

[397] Walter Genovese and Miguel A Muñoz. "Recent results on multiplicative noise". In: *Physical Review E* 60.1 (1999), p. 69.

[398] Erik van Nimwegen. "Influenza Escapes Immunity Along Neutral Networks". In: *Science* 314.5807 (2006), pp. 1884–1886.

[399] Katia Koelle et al. "Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans". In: *Science* 314.5807 (2006), pp. 1898–1903.

[400] Daniel T Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of computational physics* 22.4 (1976), pp. 403–434.

[401] Matthias Bauer et al. "Ecological feedback in quorum-sensing microbial populations can induce heterogeneous production of autoinducers". In: *Elife* 6 (2017), e25773.

[402] Erwin Frey, Johannes Knebel, and Peter Pickl. "Mean-field equation for a stochastic many-particle model of quorum-sensing microbial populations". In: *arXiv preprint arXiv:1802.05307* (2018).

[403] Motoo Kimura. "Diffusion models in population genetics". In: *Journal of Applied Probability* 1.2 (1964), pp. 177–232.

[404] James F Crow, Motoo Kimura, et al. "An introduction to population genetics theory." In: *An introduction to population genetics theory.* (1970).

[405] Josef Hofbauer. "The selection mutation equation". In: *Journal of mathematical biology* 23.1 (1985), pp. 41–53.

[406] Karen M Page and Martin A Nowak. "Unifying evolutionary dynamics". In: *Journal of theoretical biology* 219.1 (2002), pp. 93–98.

[407] Katsuhiko Sato and Kunihiko Kaneko. "On the distribution of state values of reproducing cells". In: *Physical biology* 3.1 (2006), p. 74.

[408] Katsuhiko Sato and Kunihiko Kaneko. "Evolution equation of phenotype distribution: General formulation and application to error catastrophe". In: *Physical Review E* 75.6 (2007), p. 061909.

[409] Thierry Mora and Aleksandra M Walczak. "Effect of phenotypic selection on stochastic gene expression". In: *The Journal of Physical chemistry B* 117.42 (2013), pp. 13194–13205.

[410] Jonas Cremer, Anna Melbinger, and Erwin Frey. "Evolutionary and population dynamics: a coupled approach". In: *Physical Review E* 84.5 (2011), p. 051921.

[411] Jonas Cremer, Anna Melbinger, and Erwin Frey. "Growth dynamics and the evolution of cooperation in microbial populations". In: *Scientific reports* 2.1 (2012), pp. 1–6.

[412] F S Gnesotto et al. "Broken detailed balance and non-equilibrium dynamics in living systems: a review". In: *Reports on Progress in Physics* 81.6 (2018), p. 066601.

[413] Anne C Dalziel, Sean M Rogers, and Patricia M Schulte. "Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology". In: *Molecular ecology* 18.24 (2009), pp. 4997–5017.

[414] Clemente F Arias et al. "toyLIFE: a computational framework to study the multilevel organisation of the genotype-phenotype map". In: *Scientific reports* 4.1 (2014), pp. 1–10.

[415] Susanna Manrubia and José A Cuesta. "Distribution of genotype network sizes in sequence-to-structure genotype–phenotype maps". In: *Journal of The Royal Society Interface* 14.129 (2017), p. 20160976.

[416] Susanna Manrubia et al. "From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics". In: *arXiv preprint arXiv:2002.00363* (2020).

[417] Aaron Novick and Milton Weiner. "Enzyme induction as an all-or-none phenomenon". In: *Proceedings of the National Academy of Sciences (USA)* 43.7 (1957), p. 553.

[418] Josep Casadesús and David Low. "Epigenetic gene regulation in the bacterial world". In: *Microbiology and molecular biology reviews* 70.3 (2006), pp. 830–856.

[419] Josep Casadesús and David A Low. "Programmed heterogeneity: epigenetic mechanisms in bacteria". In: *Journal of Biological Chemistry* 288.20 (2013), pp. 13929–13935.

[420] Irine Ronin et al. "A long-term epigenetic memory switch controls bacterial virulence bimodality". In: *Elife* 6 (2017), e19599.

[421] Alper Mutlu et al. "Phenotypic memory in Bacillus subtilis links dormancy entry and exit by a spore quantity-quality tradeoff". In: *Nature Comm.* 9.1 (2018), pp. 1–12.

[422] Oana Carja and Joshua B Plotkin. "The evolutionary advantage of heritable phenotypic heterogeneity". In: *Scientific reports* 7.1 (2017), pp. 1–12.

[423] Emrah Şimşek and Minsu Kim. "Power-law tail in lag time distribution underlies bacterial persistence". In: *Proceedings of the National Academy of Sciences* 116.36 (2019), pp. 17635–17640.

[424] Jorge Hidalgo, Simone Pigolotti, and Miguel A. Muñoz. "Stochasticity enhances the gaining of bet-hedging strategies in contact-process-like dynamics". In: *Phys. Rev. E* 91 (3 2015), p. 032114.

[425] Paula Villa Martin et al. "Eco-evolutionary model of rapid phenotypic diversification in species-rich communities". In: *PLoS computational biology* 12.10 (2016), e1005139.

[426] Jiajun Zhang and Tianshou Zhou. "Computation of stationary distributions in stochastic models of cellular processes with molecular memory". In: *bioRxiv* (2019), p. 521575.

[427] Carl E Pearson. *Numerical methods in engineering & science*. CRC Press, 1986.

[428] Ken Sekimoto. "Langevin equation and thermodynamics". In: *Progress of Theoretical Physics Supplement* 130 (1998), pp. 17–27.

[429] A. Einstein. *Investigations on the Theory of the Brownian Movement*. Dover Books on Physics Series. Dover Publications, 1956.

[430] Lars Onsager. "Reciprocal Relations in Irreversible Processes. II." In: *Phys. Rev.* 38 (12 1931), pp. 2265–2279.

[431] Giovanni Gallavotti and Ezechiel Godert David Cohen. "Dynamical ensembles in nonequilibrium statistical mechanics". In: *Physical review letters* 74.14 (1995), p. 2694.

[432] Jorge Kurchan. "Fluctuation theorem for stochastic dynamics". In: *Journal of Physics A: Mathematical and General* 31.16 (1998), p. 3719.

[433] Joel L Lebowitz and Herbert Spohn. "A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics". In: *Journal of Statistical Physics* 95.1 (1999), pp. 333–365.

[434] Paul Glansdorff and Ilya Prigogine. "Thermodynamic theory of structure, stability and fluctuations". In: 1971.

[435] Peter G. Bergmann and Joel L. Lebowitz. "New Approach to Nonequilibrium Processes". In: *Phys. Rev.* 99 (2 1955), pp. 578–587.

[436] J. Schnakenberg. "Network theory of microscopic and macroscopic behavior of master equation systems". In: *Rev. Mod. Phys.* 48 (4 1976), pp. 571–585.

[437] Xiaona Fang et al. "Nonequilibrium physics in biology". In: *Rev. Mod. Phys.* 91 (4 2019), p. 045004.

[438] M. Polettini. "Nonequilibrium thermodynamics as a gauge theory". In: *Europhysics Letters* 97.3 (2012), p. 30003.

[439] Rep Kubo. "The fluctuation-dissipation theorem". In: *Reports on progress in physics* 29.1 (1966), p. 255.

[440] Robert Graham. "Covariant formulation of non-equilibrium statistical thermodynamics". In: *Zeitschrift für Physik B Condensed Matter* 26 (1977), pp. 397–405.

[441] R Graham and T Tél. "Existence of a potential for dissipative dynamical systems". In: *Physical review letters* 52.1 (1984), p. 9.

[442] R Graham and T Tél. "On the weak-noise limit of Fokker-Planck models". In: *Journal of statistical physics* 35.5 (1984), pp. 729–748.

[443] R Graham and T Tél. "Weak-noise limit of Fokker-Planck models and nondifferentiable potentials for dissipative dynamical systems". In: *Physical Review A* 31.2 (1985), p. 1109.

[444] P L Garrido. "Quasipotentials in the nonequilibrium stationary states or a method to get explicit solutions of Hamilton–Jacobi equations". In: *Journal of Statistical Mechanics: Theory and Experiment* 2021.11 (2021), p. 113206.

[445] Marc Mendler and Barbara Drossel. "Predicting properties of the stationary probability currents for two-species reaction systems without solving the Fokker-Planck equation". In: *Physical Review E* 102.2 (2020), p. 022208.

[446] L. Bertini et al. "Clausius Inequality and Optimality of Quasistatic Transformations for Nonequilibrium Stationary States". In: *Phys. Rev. Lett.* 110 (2 2013), p. 020601.

[447] Lorenzo Bertini et al. "Minimum Dissipation Principle in Stationary Non-Equilibrium States". In: *Journal of Statistical Physics* 116 (2004), pp. 831–841.

[448] Matteo Polettini. "Of dice and men. Subjective priors, gauge invariance, and nonequilibrium thermodynamics". In: (2013).

[449] Christian Maes and Karel Netočnỳ. "Revisiting the Glansdorff–Prigogine criterion for stability within irreversible thermodynamics". In: *Journal of Statistical Physics* 159.6 (2015), pp. 1286–1299.

[450] Christian Van den Broeck and Massimiliano Esposito. "Three faces of the second law. II. Fokker-Planck formulation". In: *Physical Review E* 82.1 (2010), p. 011144.

[451] Simone Pigolotti et al. "Generic properties of stochastic entropy production". In: *Physical review letters* 119.14 (2017), p. 140604.

[452] Daniel Maria Busiello and Simone Pigolotti. "Hyperaccurate currents in stochastic thermodynamics". In: *Physical Review E* 100.6 (2019), p. 060102.

[453] Takahiro Hatano and Shin ichi Sasa. "Steady-state thermodynamics of Langevin systems". In: *Physical review letters* 86.16 (2001), p. 3463.

[454] Leticia F Cugliandolo and Vivien Lecomte. "Rules of calculus in the path integral representation of white noise Langevin equations: the Onsager–Machlup approach". In: *Journal of Physics A: Mathematical and Theoretical* 50.34 (2017), p. 345001.

[455] Paul Glansdorff and Ilya Prigogine. "On a general evolution criterion in macroscopic physics". In: *Physica* 30.2 (1964), pp. 351–374.

[456] Sosuke Ito. "Information geometry, trade-off relations, and generalized Glansdorff–Prigogine criterion for stability". In: *Journal of Physics A: Mathematical and Theoretical* 55.5 (2022), p. 054001.

[457] Christopher Jarzynski. "Nonequilibrium equality for free energy differences". In: *Physical Review Letters* 78.14 (1997), p. 2690.

[458] Gavin E Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". In: *Physical Review E* 60.3 (1999), p. 2721.

[459] Ilya Prigogine and Grégoire Nicolis. "On symmetry-breaking instabilities in dissipative systems". In: *The Journal of Chemical Physics* 46.9 (1967), pp. 3542–3550.

[460] Philip W. Anderson and Daniel L. Stein. "Broken Symmetry, Emergent Properties, Dissipative Structures, Life". In: 1987.

[461] Philip W. Anderson. "Can Broken Symmetry Occur in Driven Systems ?" In: ed. by G Nicolis; G Dewel; J W Turner. Order and fluctuations in equilibrium and nonequilibrium statistical mechanics:XVllth International So/ray Conference on Physics. 1981.

[462] G. Jona-Lasinio. "Spontaneous Symmetry Breaking: Variations on a Theme". In: *Progress of Theoretical Physics* 124.5 (2010), pp. 731–746.

[463] Yongjoo Baek, Yariv Kafri, and Vivien Lecomte. "Dynamical Symmetry Breaking and Phase Transitions in Driven Diffusive Systems". In: *Phys. Rev. Lett.* 118 (3 2017), p. 030604.

[464] Pablo I Hurtado and Pedro L Garrido. "Spontaneous symmetry breaking at the fluctuating level". In: *Physical review letters* 107.18 (2011), p. 180601.

[465] N. Tizón-Escamilla et al. "Order and Symmetry Breaking in the Fluctuations of Driven Systems". In: *Phys. Rev. Lett.* 119 (9 2017), p. 090602.

[466] Riccardo Rao and Massimiliano Esposito. "Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics". In: *Physical Review X* 6.4 (2016), p. 041064.

[467] Daniel Maria Busiello and Amos Maritan. "Entropy production in master equations and Fokker–Planck equations: facing the coarse-graining and recovering the information loss". In: *Journal of Statistical Mechanics: Theory and Experiment* 2019.10 (2019), p. 104013.

[468] Andre C Barato and Udo Seifert. "Thermodynamic uncertainty relation for biomolecular processes". In: *Physical review letters* 114.15 (2015), p. 158101.

[469] Gianmaria Falasco, Massimiliano Esposito, and Jean-Charles Delvenne. "Unifying thermodynamic uncertainty relations". In: *New Journal of Physics* 22.5 (2020), p. 053046.

[470] R Dean Astumian. "Kinetic asymmetry allows macromolecular catalysts to drive an information ratchet". In: *Nature communications* 10.1 (2019), pp. 1–14.

[471] Daniel Maria Busiello et al. "Dissipation-driven selection of states in non-equilibrium chemical networks". In: *Communications Chemistry* 4.1 (2021), pp. 1–7.

[472] Avinash Vicholous Dass et al. "Equilibrium and non-equilibrium furanose selection in the ribose isomerisation network". In: *Nature communications* 12.1 (2021), pp. 1–10.

[473] Andreas Dechant, Shin-ichi Sasa, and Sosuke Ito. "Geometric decomposition of entropy production in out-of-equilibrium systems". In: *Physical Review Research* 4.1 (2022), p. L012034.

[474] Kohei Yoshimura et al. "Housekeeping and excess entropy production for general nonlinear dynamics". In: *Phys. Rev. Res.* 5 (1 2023), p. 013017.

[475] Christian Maes and Karel Netocny. "Minimum entropy production principle". In: *Scholarpedia* 8(7):9664 (2013).

[476] See supplemental materials for analytical derivations and mathematical details.

[477] Jürgen Schnakenberg. "Network theory of microscopic and macroscopic behavior of master equation systems". In: *Reviews of Modern physics* 48.4 (1976), p. 571.

[478] Benoît Mahault, Evelyn Tang, and Ramin Golestanian. "A topological fluctuation theorem". In: *Nature Communications* 13.1 (2022), pp. 1–9.

[479] Haidong Feng and Jin Wang. "Potential and flux decomposition for dynamical systems and non-equilibrium thermodynamics: Curvature, gauge field, and generalized fluctuation-dissipation theorem". In: *The Journal of Chemical Physics* 135.23 (2011), p. 234511.

[480] Aleksei V Chechkin et al. "Brownian yet non-Gaussian diffusion: from superstatistics to subordination of diffusing diffusivities". In: *Physical Review X* 7.2 (2017), p. 021002.

[481] Giorgio Nicoletti and Daniel Maria Busiello. "Mutual information disentangles interactions from changing environments". In: *Physical review letters* 127.22 (2021), p. 228301.

[482] P L Garrido. "Notes about the macroscopic fluctuating theory". In: *Journal of Statistical Mechanics: Theory and Experiment* 2021.2 (2021), p. 024001.

[483] Michael te Vrugt, Hartmut Löwen, and Raphael Wittkowski. "Classical dynamical density functional theory: From fundamentals to applications". In: *Advances in Physics* 69.2 (2020), pp. 121–247.

[484] Federico Cornalba, Tony Shardlow, and Johannes Zimmer. "A Regularized Dean–Kawasaki Model: Derivation and Analysis". In: *SIAM Journal on Mathematical Analysis* 51.2 (2019), pp. 1137–1187.

[485] R. Kupferman, G. A. Pavliotis, and A. M. Stuart. "Itô versus Stratonovich white-noise limits for systems with inertia and colored multiplicative noise". In: *Phys. Rev. E* 70 (3 2004), p. 036120.

[486] Sara Cerasoli et al. "Asymmetry relations and effective temperatures for biased Brownian gyrators". In: *Physical Review E* 98.4 (2018), p. 042149.

[487] Andreas Dechant and Shin-ichi Sasa. "Entropic bounds on currents in Langevin systems". In: *Phys. Rev. E* 97 (6 2018), p. 062101.

[488] George R. Price. "The nature of selection". In: *Journal of Theoretical Biology* 175.3 (1995), pp. 389–396.

[489] Ricardo Gutiérrez and Carlos Pérez-Espigares. "Dynamical phase transition to localized states in the two-dimensional random walk conditioned on partial currents". In: *Phys. Rev. E* 104 (4 2021), p. 044134.

[490] Manfred Eigen and Peter Schuster. "The hypercycle". In: *Naturwissenschaften* 65.1 (1978), pp. 7–41.

[491] Lennart Dabelow, Stefano Bo, and Ralf Eichhorn. "Irreversibility in active matter systems: Fluctuation theorem and mutual information". In: *Physical Review X* 9.2 (2019), p. 021009.

[492] Arya Datta, Patrick Pietzonka, and Andre C. Barato. "Second Law for Active Heat Engines". In: *Phys. Rev. X* 12 (3 2022), p. 031034.

[493] Eörs Szathmáry. "Toward major evolutionary transitions theory 2.0". In: *Proceedings of the National Academy of Sciences* 112.33 (2015), pp. 10104–10111. ISSN: 0027-8424.

[494] María Rebolleda-Gómez and Michael Travisano. "Adaptation, chance, and history in experimental evolution reversals to unicellularity". In: *Evolution; International Journal of Organic Evolution* 73 (2018), pp. 73 –83.

[495] R.A. Fisher. In: *Ann. Eugen.* 1.1 (1941), pp. 53—63.

[496] Steven A. Frank and Montgomery Slatkin. "Fisher's fundamental theorem of natural selection". In: *Trends in Ecology Evolution* 7.3 (1992), pp. 92 –95.

[497] Yoh Iwasa. "Free fitness that always increases in evolution". In: *Journal of Theoretical Biology* 135.3 (1988), pp. 265–281.

[498] Marc Harper. "Information Geometry and Evolutionary Game Theory". In: (2009).

[499] Michael Doebeli, Yaroslav Ispolatov, and Burt Simon. "Point of View: Towards a mechanistic foundation of evolutionary theory". In: *eLife* 6 (2017). Ed. by Wenying Shou, e23804. ISSN: 2050-084X.

[500] Michael J. Wiser, Noah Ribeck, and Richard E. Lenski. "Long-Term Dynamics of Adaptation in Asexual Populations". In: *Science* 342.6164 (2013), pp. 1364–1367.

[501] Stanislas Leibler and Edo Kussell. "Individual histories and selection in heterogeneous populations". In: *Proceedings of the National Academy of Sciences* 107.29 (2010), pp. 13183–13188.

[502] Michal Hledík, Nick Barton, and Gašper Tkačik. "Accumulation and maintenance of information in evolution". In: *Proceedings of the National Academy of Sciences* 119.36 (2022), e2123152119.

[503] Marc Harper. "The Replicator Equation as an Inference Dynamic". In: (2009).

[504] "A Mathematical Theory of Communication". In: *The Bell System Technical Journal* 27 (1948), pp. 379–423.

[505] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006. ISBN: 0471241954.

[506] Lou Jost. "Entropy and Diversity". In: *Oikos* 113.2 (2006), pp. 363–375.

[507] Juan M. R. Parrondo, Jordan M. Horowitz, and Takahiro Sagawa. "Thermodynamics of information". In: *Nature Physics* 11 (2015), pp. 131 –139.

[508] Jaco Fuchs, Sebastian Goldt, and Udo Seifert. "Stochastic thermodynamics of resetting". In: *EPL (Europhysics Letters)* 113.6 (2016), p. 60009.

[509] C. E. Fernández Noa et al. "Entropy production as a tool for characterizing nonequilibrium phase transitions". In: *Phys. Rev. E* 100 (1 2019), p. 012104.

[510] Tânia Tomé and Mário J. de Oliveira. "Entropy Production in Nonequilibrium Systems at Stationary States". In: *Phys. Rev. Lett.* 108 (2 2012), p. 020601.

[511] Andre C Barato and Udo Seifert. "Thermodynamic uncertainty relation for biomolecular processes." In: *Physical review letters* 114 15 (2015), p. 158101.

[512] Jordan M. Horowitz and Todd R. Gingrich. "Thermodynamic uncertainty relations constrain non-equilibrium fluctuations". In: *Nature Physics* 16 (2020), pp. 15–20.

[513] Schuyler Nicholson et al. "Time–information uncertainty relations in thermodynamics". In: *Nature Physics* (2020), pp. 1–5.

[514] Manfred Eigen. "Selforganization of matter and the evolution of biological macromolecules". In: *Naturwissenschaften* 58 (1971), pp. 465–523.

[515] Susanna Manrubia et al. "From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics". In: *Physics of Life Reviews* 38 (2021), pp. 55–106.

[516] Alexander Blokhuis. "Physical aspects of origins of life scenarios". PhD thesis. Université Paris sciences et lettres, Nov. 2019.

[517] Alex Blokhuis et al. "The generality of transient compartmentalization and its associated error thresholds". In: *Journal of Theoretical Biology* 487 (2020), p. 110110.

[518] Charles D. Kocher and Ken A. Dill. "Darwinian evolution as a dynamical principle". In: *Proceedings of the National Academy of Sciences* 120.11 (2023), e2218390120.

[519] Alan Ianeselli et al. "Physical non-equilibria for prebiotic nucleic acid chemistry". In: *Nature Reviews Physics* 5.3 (Mar. 2023), pp. 185–195.

[520] Jacob Moran and Mikhail Tikhonov. "Defining Coarse-Grainability in a Model of Structured Microbial Ecosystems". In: *Phys. Rev. X* 12 (2 2022), p. 021038.

[521] H. Stanley and Paul Meakin. "Multifractal phenomena in physics and chemistry". In: *Nature* 335 (Sept. 1988), pp. 405–409.

[522] G. Jona-Lasinio. "Renormalization group and probability theory". In: *Physics Reports* 352.4 (2001), pp. 439–458.

[523] Serena Bradde and William Bialek. "Pca meets rg". In: *Journal of statistical physics* 167 (2017), pp. 462–475.

[524] Leenoy Meshulam et al. "Coarse graining, fixed points, and scaling in a large population of neurons". In: *Physical review letters* 123.17 (2019), p. 178103.

[525] Guillermo B. Morales, Serena di Santo, and Miguel A. Muñoz. "Quasiuniversal scaling in mouse-brain neuronal activity stems from edge-of-instability critical dynamics". In: *Proceedings of the National Academy of Sciences* 120.9 (2023), e2208998120.

[526] J. Bouchaud. "Weak ergodicity breaking and aging in disordered systems". In: *Journal de Physique I* 2.9 (1992), pp. 1705–1713.

[527] Massimo Bernaschi et al. "Strong ergodicity breaking in aging of mean-field spin glasses". In: *Proceedings of the National Academy of Sciences* 117.30 (2020), pp. 17522–17527.

[528] M. Mezard, G. Parisi, and M.A. Virasoro. *Spin Glass Theory And Beyond: An Introduction To The Replica Method And Its Applications*. World Scientific Publishing Company, 1987.

[529] Daniel R. Amor, Christoph Ratzke, and Jeff Gore. "Transient invaders can induce shifts between alternative stable states of microbial communities". In: *Science Advances* 6.8 (2020), eaay8676.

[530] Jay T. Lennon et al. "Principles of seed banks and the emergence of complexity from dormancy". In: *Nature Communications* 12 (2021).

[531] Gabin Laurent, David Lacoste, and Pierre Gaspard. "Emergence of homochirality in large molecular systems". In: *Proceedings of the National Academy of Sciences* 118.3 (2021), e2012741118.

[532] Eugene V. Koonin, Mart Krupovic, and Valerian V. Dolja. "The global virome: How much diversity and how many independent origins?" In: *Environmental Microbiology* 25.1 (2023), pp. 40–44.

[533] E. Schrödinger. *What is Life? the Physical Aspect of the Living Cell & Mind and Matter*. University P., 1967.

[534] Eric Lander et al. "Initial Sequencing and Analysis of the Human Genome." In: *Nature* 409 (Feb. 2001).

[535] Welkin Johnson. "Origins and evolutionary consequences of ancient endogenous retroviruses". In: *Nature Reviews Microbiology* 17 (Apr. 2019), p. 1.

[536] Anne Dupressoir et al. "Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene". In: *Proceedings of the National Academy of Sciences* 106.29 (2009), pp. 12127–12132.

[537] M. Sheldrake. *Entangled Life: How Fungi Make Our Worlds, Change Our Minds & Shape Our Futures*. Random House Publishing Group, 2020.

[538] Nicole Dubilier, Claudia Bergin, and Christian Lott. "Dubilier N, Bergin C, Lott C.. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Micro 6: 725-740". In: *Nature reviews. Microbiology* 6 (Nov. 2008), pp. 725–40.

[539] Luis M. Márquez et al. "A Virus in a Fungus in a Plant: Three-Way Symbiosis Required for Thermal Tolerance". In: *Science* 315.5811 (2007), pp. 513–515.

[540] Min Li et al. "Symbiotic gut microbes modulate human metabolic phenotypes". In: *Proceedings of the National Academy of Sciences* 105.6 (2008), pp. 2117–2122.

[541] E. Yong. *I Contain Multitudes: The Microbes Within Us and a Grander View of Life*. HarperCollins, 2016.

[542] Ruth E. Ley et al. "Evolution of Mammals and Their Gut Microbes". In: *Science* 320.5883 (2008), pp. 1647–1651.

[543] L. Peliti. *Statistical Mechanics in a Nutshell*. Princeton University Press, 2011.

[544] Alex L Mitchell et al. "EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies". In: *Nucleic Acids Research* 46.D1 (Oct. 2017), pp. D726–D735.

[545] Christian Quast et al. "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools". In: *Nucleic Acids Research* 41.D1 (Nov. 2012), pp. D590–D596.

[546] Roberto Ambrosini et al. "Diversity and Assembling Processes of Bacterial Communities in Cryoconite Holes of a Karakoram Glacier". In: *Microbial ecology* 73.4 (May 2017), pp. 827–837.

[547] Juan Pablo Niño-García, Clara Ruiz-González, and Paul A. Giorgio. "Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks". In: *The ISME Journal* 10 (2016), pp. 1755–1766.

[548] J. Gregory Caporaso et al. "Moving pictures of the human microbiome". In: *Genome Biology* 12 (2011), R50 –R50.

[549] Paul J. McMurdie and Susan Holmes. "phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data". In: *PLOS ONE* 8.4 (Apr. 2013), pp. 1–11.

[550] Joel L. Lebowitz. "Macroscopic laws, microscopic dynamics, time's arrow and Boltzmann's entropy". In: *Physica A: Statistical Mechanics and its Applications* 194.1 (1993), pp. 1 –27. ISSN: 0378-4371.

[551] Andrew Charteris Tadrowski. "Role of phenotype switching during biological evolution in static environments". In: (2017).

[552] Subhadip Raychaudhuri. "Kinetic Monte Carlo simulation in biophysics and systems biology". In: *Theory and applications of Monte Carlo simulations* (2013).